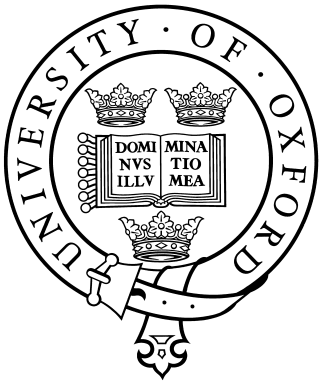
Validation of a climate model for extreme event attribution studies

Neil R. Massey
Wolfson College
University of Oxford

A Thesis submitted to the Faculty of Physical Sciences for the degree of
Doctor of Philosophy at the University of Oxford
Michaelmas Term, 2015



Atmospheric, Oceanic and Planetary Physics
Department of Physics
University of Oxford

Validation of a climate model for extreme event attribution studies

Neil R. Massey, Wolfson College, University of Oxford
Michaelmas term 2015.

Abstract

Probabilistic event attribution (PEA) presents a method of quantifying the change in risk of extreme weather events, both in magnitude and probability of occurrence, due to anthropogenic climate change. Studies have so far covered numerous different extreme events in different regions (Stott et al., 2016). One method of PEA relies on computing large ensembles of climate models for two different climate scenarios, one which represents the observed climate forcings, and another the climate forcings with anthropogenic greenhouse gas contributions removed. Other PEA studies using different methodologies have also been conducted.

Until now there has been no formal validation of the climate models used in PEA studies. This thesis presents two ways of validating the models and applies these methods to a very large ensemble of a climate model simulating the observed climate for the period 1960 to 2010. The premise is, that for an attribution statement to be meaningful, the model should be able to accurately represent relevant weather statistics over a range of climate scenarios. In order to compare to observations, the period 1960–2010 is used as the range of scenarios.

Conceived as an e–Science project, this thesis presents technical development of methodologies in the detection and attribution of extreme weather events to climate change. The first method identifies and tracks storm–like features in meteorological data. A novel set of algorithms transforms the meteorological data to a hierarchical equal–area triangular grid, identifies storm–like feature points in the data and grows the points into objects. These objects are tracked as they evolve over time by a hybrid prediction–optimisation routine, which minimises a cost function to find a locally optimal set of feature tracks. Applying the algorithms to the large ensemble of climate models, and also to the ERA–Interim data, shows that the model can successfully capture the track length, persistence and position of low–pressure systems over Europe. However, the depression depth of the systems is not as well represented.

The second method applies a forecast verification technique to the distributions of climate variables. Comparing the temperature, precipitation, mean sea level pressure (MSLP) and winds between the large ensemble and ERA–40 and ERA–Interim shows that the model can accurately represent temperature, precipitation and windspeed variables over Europe, after a bias correction has been applied. However, for low MSLP, there are irreparable biases in the low tails of the distribution. Overall, it is shown that using a large ensemble of climate models is a valid method of investigating the change in risk of extreme weather events due to climate change, as long as the variable to be attributed is carefully chosen.

Acknowledgements

The work in this Thesis was funded by a NERC e-Science studentship award. e-Science is research that can only be made possible through distributed computing power. My studentship is inter-discplinary, with a large computing component, but still with a focus on atmospheric science. I would like to thank the Research Councils UK for this funding, without which I would have been unable to study for a DPhil.

I would like to thank my supervisor, Prof. Myles Allen for his insight, guidance and patience. Thank you for giving me the opportunity to start, and the time and means to finish, this DPhil.

I would like to thank Tolu Aina, without whose technical ability the volunteer distributed computing aspect of this project would have been a lot more difficult. Thanks also to the climate*prediction*.net volunteers, without whom computing such large ensembles of climate models would be impossible, particularly for graduate students.

Thanks also to David Hassell, Simon Wilson and Richard Jones from the Met Office for their help and assistance with the HadAM3P climate model, ancillary files and forcing scenarios.

Thanks to my parents for recognising the importance of an education, instilling in me the desire to learn and for supporting me throughout my three stints at university.

Finally, I would like to thank my wife, Claire for her continued love and support, and just the right amount of nagging!

Contents

1 Introduction and background **3**

1.1 Attribution of extreme weather events 4
1.2 Definition of an extreme weather event 6
1.3 Methods of event attribution 7
1.3.1 Methods using coupled ocean–atmosphere models 8
1.3.2 Methods using large ensembles of atmosphere only models 9
1.3.3 Analogue based methods 10
1.3.4 Empirical and observation based methods 10
1.3.5 Broad scale methods 11
1.4 Attribution of extreme events over Europe 12
1.5 The mid–latitude atmosphere, extra–tropical cyclones and storm tracks . . . 14
1.5.1 The structure of the atmosphere 15
1.5.2 Storm tracks 16
1.5.3 The life–cycle of the mid–latitude cyclone 17
1.5.4 Baroclinic instability 18
1.6 Identification and tracking of mid–latitude storms 21
1.6.1 Transforming data prior to identification 22
1.6.2 Identification of storm–like features 23
1.6.3 Tracking of storm–like features 25
1.6.4 Track post–processing 28
1.6.5 Track analysis 28
1.6.6 Application of tracking algorithms to climate data 30
1.6.7 Inter–comparison of tracking methods 33
1.7 Definition of sub regions of Europe 34
1.8 Aims and scope of this thesis 36

2 Feature tracking on a hierarchical triangular mesh 41

2.1 Previous work . . . 41
2.2 Generating a hierarchical equal area triangular mesh . . . 42
2.2.1 Initialisation of the triangular mesh . . . 43
2.2.2 Subdividing the mesh . . . 45
2.2.3 Labelling the mesh . . . 46
2.2.4 Calculating the adjacency relationships of the triangles . . . 47
2.2.5 Optimisation of the mesh to equal area triangles . . . 49
2.2.6 Locating points on the mesh . . . 52
2.2.7 Storing data on the mesh . . . 53
2.2.8 Regridding regularly gridded data onto the mesh . . . 55
2.3 Objective feature detection on the mesh . . . 58
2.3.1 Background field removal . . . 60
2.3.2 Feature object detection . . . 61
2.3.3 Deriving the feature point from the object . . . 63
2.4 Feature tracking . . . 64
2.4.1 Geostrophic wind as a steering vector . . . 65
2.4.2 Tracking rules . . . 66
2.4.3 Assigning objects to tracks . . . 73
2.4.4 Track optimisation . . . 74
2.4.5 Post–processing of tracks . . . 78
2.5 Summary and conclusions . . . 78

3 Validation and sensitivity of a storm–tracking algorithm 81

3.1 Application to ERA-Interim data . . . 82
3.1.1 Northern Hemisphere winter . . . 83
3.1.2 Southern Hemisphere winter . . . 85
3.1.3 Summary of validation of algorithm on ERA-Interim data . . . 88
3.2 Application to individual storm test cases . . . 89
3.3 Sensitivity of the tracking algorithm to temporal resolution . . . 95

3.3.1 Degrading 6 hourly data to 12 and 24 hours 96
3.3.2 Comparing tracks in different temporal resolution data 98
3.4 Summary and conclusions 104
3.5 Future work 105

4 Design of a climate model validation experiment using a very large ensemble 107
4.1 The HadAM3P climate model 108
4.2 Forcings at the boundary layer 112
4.2.1 Sea–surface temperatures 113
4.2.2 Sea ice fraction 120
4.2.3 Sea–surface temperatures and sea ice fraction forcings summary . . . 121
4.3 Atmospheric composition 121
4.3.1 Greenhouse gas concentrations 121
4.3.2 Ozone 125
4.3.3 Natural and anthropogenic sulphate concentrations 125
4.3.4 Orography 128
4.4 Solar forcing 128
4.5 Initial conditions for the model 129
4.5.1 Generation of the initial conditions 130
4.5.2 Initial condition perturbation 131
4.6 Computation via Distributed Computing 133
4.7 Summary 136

5 Validation of a very large ensemble of a climate model on seasonal timescales 137
5.1 Choosing a dataset to verify against 138
5.1.1 Near surface air temperature 142
5.1.2 Precipitation at the surface 146
5.1.3 Conclusions and validation method for global data 148
5.2 Validating seasonal means of climate variables 150
5.2.1 Validating seasonal mean temperature 151
5.2.2 Validating seasonal mean precipitation 153

5.2.3 Validating seasonal mean MSLP 157
5.2.4 Validating seasonal mean upper–level wind 160
5.2.5 Summary and conclusions of validating seasonal means. 162
5.3 Validating extremes of seasonal means 163
5.3.1 Validating extremes of seasonal mean temperature 165
5.3.2 Validating extremes of seasonal mean precipitation 167
5.3.3 Validating extremes of seasonal mean MSLP 169
5.4 Variability of seasonal mean temperature 172
5.5 Variability of seasonal mean precipitation 174
5.6 Summary and conclusions 176

6 Validation of variables on a daily timescale from a very large ensemble of a climate model 179
6.1 Methods used to validate daily data 181
6.1.1 Bias correction 182
6.1.2 Non–parametric estimate of confidence intervals 183
6.1.3 Non–parametric goodness of fit test 184
6.1.4 Model verification plots 184
6.1.5 Root mean square deviation of percentiles 186
6.2 Daily data over five European Regions 187
6.2.1 Daily maximum temperature 187
6.2.2 Daily minimum temperature 190
6.2.3 Daily accumulated precipitation 192
6.2.4 Daily mean sea level pressure 197
6.2.5 Daily maximum windspeed 200
6.3 Summary and conclusions 202

7 Validation of weather systems in a very large ensemble of climate models 205
7.1 Applying the algorithms to the Europe–Atlantic region 206
7.2 Counts of tracks over the Europe–Atlantic region 210
7.3 Spatial density of tracks over the Europe–Atlantic region 212

7.4 Track metrics over the Europe–Atlantic region 216
7.4.1 Track lifetime 217
7.4.2 Track length 219
7.4.3 Track intensity 220
7.4.4 Track propagation speed 222
7.5 Distribution of variables associated with cyclones 224
7.5.1 Mean sea level pressure 225
7.5.2 Precipitation 226
7.5.3 Windspeed at 10m 229
7.6 Summary and conclusions 231

8 Conclusions and future work **235**
8.1 Summary, conclusions and discussion 235
8.2 Future work : the Weather-@-Home experiments 242
8.2.1 Improvements to the model 242
8.2.2 Improvements to the experimental design 243
8.2.3 Future attribution studies and extreme event sets 245

1 Introduction and background

Extreme weather events occur, by their definition, infrequently and there is not an adequate sample size in the observations to detect changes in the frequency of occurrence or magnitude of the extreme events that may be due to anthropogenic climate change. However, seemingly every time an extreme weather event occurs, such as flooding, a windstorm or a heatwave, the question is asked within the scientific community and media broadcasts: "was this extreme event caused by climate change?" It is the emergent science of probabilistic event attribution (PEA) which is concerned with answering this question.

PEA cannot definitively answer whether climate change has caused an event, but instead tries to determine whether there is an increased likelihood of the event occurring under an anthropogenically altered climate, and the magnitude of this increased risk. In order to determine whether there is an increased risk of extreme weather events, PEA, in effect, conducts two experiments. One experiment calculates the likelihood of the event under current climate conditions and the second calculates the likelihood under pre–industrial conditions.

One method of conducting this experiment is to use a global climate model (GCM). A GCM is a computer program which models the Earth system by representing the globe on a discrete grid. The model solves equations of motion and radiative transfer explicitly on the same scale as the grid. However, there are also a number of sub-grid scale physical processes, primarily thermodynamic, which are represented by approximations to the physical processes. A GCM runs in simulated time to produce climate variable output such as temperature, precipitation, windspeed and pressure.

In conducting the two PEA experiments, one has climate forcings which replicate the observations, complete with anthropogenic greenhouse gas (GHG) emissions. The other has

climate forcings that are estimated to be as if anthropogenic GHG emissions did not occur. The experiments are conducted very many times for each scenario, and the occurrence of the extreme event is detected in each experiment. An increase in risk of the extreme event occurring due to climate change can then be calculated by taking the ratios of the occurrence of the event in each experiment.

One method of PEA uses GCMs to make statements as to whether extreme weather events experienced in the real world can be linked to the large scale warming caused by anthropogenic climate change. However, until now, there has been no formal validation of the models used for event attribution to determine whether they can adequately represent the distribution of weather and the occurrence of extreme events that are present in the observed record. This thesis addresses this gap by comprehensively testing the suitability of a GCM for making attribution statements.

1.1 Attribution of extreme weather events

Over the past twenty years there has been widespread consensus in the scientific community that the release into the atmosphere of greenhouse gases (GHG) (including CO_2, methane and N_2O) by human activity has caused the global temperature of the Earth to rise. The IPCC Fourth Assessment Report (AR4) (Solomon et al., 2007) estimates this rise to be 0.8°C over the 20$^{\text{th}}$ Century and that it is likely to rise by another 1.1 to 6.4°C over the 21$^{\text{st}}$ Century, dependent on an interacting series of uncertainties. These include the projected emissions of GHG, the emissions of aerosol particulates, the response of the climate system to the elevated levels of GHG in the atmosphere and the heat uptake of the oceans (Solomon et al., 2007). There is now acceptance of anthropogenic global warming (AGW) by the general public, especially in Western Europe, although it has been slow.

While the rise in temperature is not particularly alarming in itself, the change in weather patterns caused by this temperature rise are of the most importance to a range of people and institutions, from the farmer in a developing nation determining when to plant his crops to government and non-government organisations tasked with planning the adaptation of

infrastructure to the changing climate. The change in weather patterns will manifest themselves in changes to the distribution of climate variables, such as temperature, precipitation, snow fall, wind speed and mean sea level pressure. For example, the observed rise in temperature in the IPCC AR4 is expressed as a deviation from the long term mean over a certain period. However, over that period the temperature also has an associated distribution and, although the rise in mean temperature is well known, the change in distribution of temperatures around the mean is less well known. Within the tails of these distributions lie the extreme events: high or low temperatures, high wind speed, cyclones (low–pressure systems) and high levels of precipitation. It is also likely that the change in these distributions will be region dependent and, for example, one region may experience more precipitation whereas another may become more arid (Field et al., 2012). These regional variations may not be easy to predict based upon energy balance arguments, but instead dependent on the changes to the atmospheric flow caused by the warming, and so this motivates the use of GCMs.

Due to the regional effects of the increased warming of the Earth, the elevated levels of GHG may cause both a positive and negative change to regional extreme weather events. There are three possible consequences of changing the distribution of climate variables. Firstly, events that occur with a given frequency in the observational record become greater (lesser) in their magnitude. For example, rainstorms that occur at a 1 in 10 year interval may become more (less) intense in their precipitation rate. Secondly, events of a given size become more frequent. In terms of rainstorms, this could mean that a precipitation rate of 50 mm/day which occurs at a 1 in 10 year interval may now occur at a 1 in 5 year interval (positive change) or 1 in 20 year interval (negative change), doubling or halving the frequency of these rainstorm events occurring. Thirdly, extreme events could become more persistent in their duration. For example, low rainfall could persist for longer, causing a drought or high precipitation could persist for longer, potentially causing flooding.

Attributing the change in the distribution of climate variables to the increase in greenhouse gases in the atmosphere, and its resultant warming, poses a challenge to the scientific community. The possibility of attributing a shift in weather patterns and distributions of climate variables to the change in greenhouse gas concentrations could inform several areas such as

adaptation (in future planning of infrastructure), litigation (in developing nations seeking compensation from industrialised nations) and the public interest (in a general understanding of climate and the effect we are having upon it) (Stott et al., 2013). The attribution of extreme weather events could inform the UNFCCC loss and damage agenda, compensating developing nations for the changes in extreme weather caused by GHG emissions over many years by developed nations (Parker et al., 2015). Work has already commenced examining how major carbon polluters, such as industrial emitters and producers of carbon, may be responsible for the increased risk of occurrence of heat waves in the California Central Valley (Mera et al., 2015).

This presents a key new challenge. If the cost of climate change is to be quantified, as will have to occur if the UNFCCC loss and damage agenda is to go forward, or if, in the USA, class action suits are brought against the major carbon polluters, then an understanding of how climate change is affecting extreme weather is needed, as extreme events are costly. Hurricane Katrina, which devastated New Orleans in 2005, is estimated to have cost $108 billion USD (Blake and Gibney, 2011). Lawsuits or loss and damage compensation would lead to event attribution science being thoroughly scrutinised and so the foundations of the science, including the assessment of the models it is built upon, must be robust.

1.2 Definition of an extreme weather event

An extreme weather event is often defined in an anthropocentric way, in that an event is deemed to be extreme if it is unusual or infrequent, in the context of human experience, and causes significant damage to human infrastructure. Stott et al. (2016) define an extreme event as: "discrete episodes of extreme weather or unusual climate conditions, often associated with deleterious impacts on society or natural systems, defined using some metric to characterize either the meteorological characteristics of the event or the consequent impacts". Often after such an event occurs, such as flooding in the UK (Schaller et al., 2016; van Oldenborgh et al., 2016), there is a large media interest in whether the event was "caused" by climate change. This definition indicates that there are two components to an extreme event, the

meteorological hazard and the loss to life and property that results from the hazard. This thesis is concerned with the first component, the meteorological hazard.

The meteorological hazards, including high (or low) precipitation, high windspeeds, high (or low) temperatures and low MSLP causing storm surges, can occur on a number of temporal scales. These range from very short, high intensity events at the scale of hours, such as convective precipitation events, through days to weeks scale events, such as windstorms and heatwaves, to seasonal long events, such as droughts (Stott et al., 2016). As these events are unusual and occur infrequently, statistically they occur within the tails of the distributions of the variables on the timescale on which they occur. So, a drought will be in the low tail of the distribution of seasonal accumulated precipitation, whereas a windstorm will be in the high tail of the distribution of maximum windspeed.

1.3 Methods of event attribution

As extreme event attribution studies are conducted after the event has occurred, the event is usually defined as the event that occurred in the observations, either as a threshold or as a percentile of the climatological values. Extreme weather events have always occurred and there are many examples prior to the large–scale measurement of the atmosphere which is performed today. Therefore, it is impossible to say whether an extreme weather event was caused by GHG emissions, as it may have occurred anyway, even without any anthropogenic emissions. Instead, event attribution studies attempt to attribute the increased risk of the event occurring due to elevated GHG levels.

Attribution studies present their results in a number of different ways, including probability / likelihood histograms, return time plots and shifts in probability density functions and cumulative distribution functions. One calculation that is performed in Pall et al. (2011) is the fraction of attributable risk (FAR) of an extreme event occurring due to anthropogenic climate change. The FAR is defined by Allen (2003) as:

$$FAR = 1 - \frac{P_0}{P_1}$$

where P_0 is the probability of an extreme event occurring in the non–anthropogenic scenario and P_1 is the probability of the event occurring in the observed climate scenario. This allows the increase in risk of an event occurring due to anthropogenic climate change to be calculated. The valid range of the FAR is between 1 and $-\infty$. The value of 1 occurs when $P_1 \gg P_0$ and indicates that elevated GHG concentrations are entirely responsible for the risk of the event occurring, whereas a value of 0 occurs when $P_0 = P_1$ and implies that elevated GHG concentrations have not changed the risk of the event occurring. Values of the FAR less than 0 occur when $P_0 > P_1$ and indicate that elevated GHG concentrations have decreased the change of the extreme event occurring. Finally, a value of $-\infty$ is more difficult to interpret but it shows that increasing GHG concentrations has not increased the risk of the event occurring.

Stott et al. (2016) identifies five methods of attributing the increase in risk of extreme weather events, which are summarised below:

1.3.1 Methods using coupled ocean–atmosphere models

Coupled ocean–atmosphere global climate models (AOGCM) are used to make attribution statements by using AOGCM runs that have been submitted to the CMIP5 archive (Taylor et al., 2012), for the natural historic forcings (without anthropogenic climate change) and the historical and near–future forcings (observed and RCP4.5 forcings). Models for each scenario are pooled into multimodel ensembles and the distribution of the climate variable of interest is constructed for both scenarios. From these distributions, the change in risk of an event occurring at a threshold of the distribution can be calculated and thus, the change in risk can attributed to the elevated GHG in the historical and near–future scenario.

This method has been used to attribute heatwaves in Australia (Perkins et al., 2014) with a robust result and heavy precipitation also in Australia (King et al., 2013) with a less robust result (Stott et al., 2016). The main advantage of this method is that the model data is already available and so an attribution statement can be made soon after the event has occurred. The disadvantage is that, by using a multi–model ensemble, it is not known how

much of the uncertainty in the attribution statement is due to the variance in the structure of the models.

1.3.2 Methods using large ensembles of atmosphere only models

One way of removing the uncertainty due to structural variance of the CMIP5 models is to use a single atmosphere only global climate model (AGCM), run many times in a large ensemble, perturbing the initial conditions so as to generate divergent weather statistics in each ensemble member. Studies that have used such a setup include Pall et al. (2011), Otto et al. (2012) and Schaller et al. (2014). As with the AOGCM, two forcing scenarios are modelled: one with the observed, anthropogenic levels of GHG, and one with the anthropogenic component removed.

AGCMs require forcing with sea surface temperatures (SST) and so these must be supplied for each scenario. For the anthropogenic scenario, these are taken from an observed dataset, such as HadISST (Rayner et al., 2003). For the non-anthropogenic scenario, the warming pattern due to elevated GHG must be removed. This is estimated by subtracting the SSTs in the observed multi–model ensemble of CMIP5 models from the natural historic SSTs to create SST anomalies which are then subtracted from the observed SSTs to form the non–anthropogenic scenario. Either a multi–model ensemble mean of the SST anomalies can be used, or a range of the SST anomalies can be used to create an ensemble of non–anthropogenic SSTs. The ratios of the number of ensemble members in the anthropogenic ensemble where the event theshold is passed, to the number in the non-anthropogenic ensemble can be computed and used to calculate the FAR or the risk ratio.

This approach has a number of advantages. AGCMs are less prone to biases than AOGCMS (Scaife et al., 2010; Massey et al., 2015) and are less computationally expensive. This allows many ensemble members to be run, enabling high sampling of the tails of the distributions of the climate variables. The volunteer distributed computing method of computing AGCMS can produce many thousands of model runs (Massey et al., 2015; Rowlands et al., 2012), producing an improved signal–to–noise ratio when compared to smaller ensembles. The

main disadvantage of this method is that atmosphere–ocean coupling is not represented in the model, which may lead to a worse representation of extreme events which are particularly influenced by this (Stott et al., 2016).

1.3.3 Analogue based methods

Analogue methods use observations or reanalysis data to find examples, at any time in the historical record, which match the circulation pattern at the time at which an extreme weather event occurs. For example, Cattiaux and Yiou (2012) attributes the European Heatwave of 2011 to anthropogenic GHG emissions by partitioning each day within a 31 day window of the event day within each year in the 20CR reanalysis (Compo et al., 2011) into one of four different circulation types, which are based on MSLP patterns over a Europe–Atlantic region. The 10 circulation patterns that are most similar to the event from the years 1900 to 1960 are found and the temperatures for these days are averaged. Similarly the 10 most similar patterns for 1970 to 2011 are also found and the temperatures are averaged. Then, using the spatial data in the field, the risk of the event increasing due to climate change is the ratio of the probability of the event occurring in the 1970 to 2011 analogues to the probability of it occurring in the 1900 to 1960 analogues.

This method has the advantage that it does not require GCM runs to calculate the attribution, instead using a reanalysis product that is already available. However, it does not allow the explicit exploration of alternative climate scenarios, such as "the world that might have been" if GHG had not been emitted. Rather, it can only attribute to a change in climate, however that change might have occurred. It is also assumed that the time periods are long enough for the internal variability of the climate system to not have an effect.

1.3.4 Empirical and observation based methods

Empirical approaches rely on the notion that an increase in the risk of extreme temperatures occurring is related to the increase in the mean temperature, and that an increase in the risk of extreme (high) precipitation occurring is related to an increase in the mean temperature

allowing more moisture to be held in the air (Stott et al., 2016). One empirical approach (van Oldenborgh et al., 2012) involves fitting a generalised extreme value distribution (GEVD) to observed data and then modifying the location parameter of the GEVD by an amount proportional to the change in global mean temperature from a non-anthropogenic scenario to the current day, GHG emissions influenced climate. From the GEVD, the return periods for a threshold being exceeded for both scenarios can be calculated and the relative increase in risk, or the FAR can be computed.

This method has the advantage that it is computationally simple and relies on little data, allowing an attribution statement to be made very close to the event occurring, or even before it has occurred by using seasonal forecasts. The main disadvantage is that it gives no insight into the atmospheric state which caused the extreme event, and there is the possibility of false positives (or negatives) due to spurious correlations. As with the analogue based method, the empirical method can only attribute any change in risk of an event occurring to the change in climate, not to GHG gases, as the modelling approaches can.

1.3.5 Broad scale methods

Broad scale methods use an adaptation of the methods used in global detection and attribution studies to attribute broad metrics of global change in temperature and precipitation to elevated levels of GHG. These methods avoid the selection bias of the extreme event attribution studies, where the event to study is chosen due to its prominence to populations or media coverage. For example, Zwiers et al. (2011) use the 1961 to 2000 daily minimum (TMin) and daily maximum temperature (TMax) observations to fit GEVDs to the TMin and TMax data, at every grid point in the observations. The shape and location parameters of the GEVD are modified by parameters derived from GCM runs of all forcing and anthropogenic only forcings. They find that globally, the risk of very cold nights has decreased whereas the risk of very hot days has increased.

Such an approach avoids the selection bias inherent in extreme attribution studies and can be performed at any time, as all of the data is already available. They retain a link between

detection attribution methods and extreme event attribution (Stott et al., 2016), and allow future projections by using future RCP projections from CMIP5. However, it is difficult to conduct such a study at regional scales, and the method is not as effective for variables other than temperature (Stott et al., 2016).

1.4 Attribution of extreme events over Europe

The Bulletin of the American Meteorological Society (BAMS) has, since 2012, published a yearly special issue, titled "Explaining the Extreme Events of *yyyy* from a Climate Perspective", where *yyyy* is the previous year. These special issues invite submissions from research groups involved with extreme event attribution to select an extreme event from the previous year and conduct an attribution study into whether the risk of the event occurring has increased (or decreased) due to GHG emissions. The papers may use any attribution methodology they choose (or more than one), although all papers published so far have used one of the five methodologies above. The geographical spread of the attribution studies is wide, with papers on extreme events in North and South America, Europe, Africa, the Middle East, South East Asia, China, Japan, Korea and East Asia and Australia and New Zealand having been published to date, with timescales of extreme events ranging from daily to seasonal events.

The focus of this thesis is validating a model that has been used extensively to conduct attribution studies over Europe. Table 1.4.1 collects and summarises the attribution studies from the four special editions of BAMS, adding four additional papers not in BAMS, but are widely cited as examples of PEA. Attribution studies have been conducted for a number of regions within Europe, and globally, for four climate variables (precipitation, temperature, windspeed and MSLP) over a number of spatial and temporal scales. Here a positive conclusion means that anthropogenic climate change increased the risk of the event occurring, and a negative result means that there is no evidence in the study that the risk of the event occurring has increased due to climate change. All five of the methods outlined in section 1.3 are used, with the AGCM method being used in 13 out of the 23 studies.

Author	**Year**	**Time**	**Event**	**Variable**	**Method**	**Outcome**
Cattiaux and Yiou (2012)	2011	JJA	EU high temp	$T_{mn}(S)$	Analogues	positive
Massey et al. (2012)	2011	ND	UK low / high temp	$T_{mn}(M)$	AGCM	positive
Christidis and Stott (2012)	2011	DJ	UK low temp	$T_{mn}(M)$	AOGCM	negative
de Vries et al. (2013)	2012	DJF	EU low temp	$T_{min}(D)$	Empirical	negative
Dong et al. (2013)	2012	JJA	N EU high precip	$P_{acc}(S)$	AGCM	positive
Tett et al. (2013)	2012	JJA	NW EU high precip	$P_{acc}(S)$	AGCM	negative
Sparrow et al. (2013)	2012	JJA	UK high precip	$P_{acc}(S)$	AGCM	negative
Yiou and Cattiaux (2013)	2012	JJA	EU high precip	$P_{acc}(S)$	Analogues	negative
Trigo et al. (2013)	2012	JJA	Iberian peninsula drought	$P_{acc}(S)$	Empirical / AGCM	positive
Dong et al. (2014)	2013	JJA	EU high temp, low precip	$T_{mn}(S)$, $P_{acc}(S)$	AGCM	positive
Yiou and Cattiaux (2014)	2013	DJF	South EU high precip	$P_{acc}(S)$	Analogues	negative
Schaller et al. (2014)	2013	MJ	Danube and Elba high precip.	$P_{acc}(D)$	AGCM	negative
Añel et al. (2014)	2013	JFM AMJ	High snowfall Pyrenees	$P_{acc}(S)$, $S_{acc}(S)$	AGCM	negative
von Storch et al. (2014)	2013	DJF	EU mid–latitude storm	$MSLP$	Empirical	negative
Christidis et al. (2014)	2013	MAM	UK low temp	$T_{mn}(S)$	AGCM	positive
Christidis and Stott (2015)	2014	DJF	UK high rainfall	$P_{acc}(S)$	AGCM	positive
Feser et al. (2015)	2014	O	EU hurricane transition	$U_{max}(D)$	AOGCM	negative

Author	**Year**	**Time**	**Event**	**Variable**	**Method**	**Outcome**
Vautard et al. (2015)	2014	SON	Cevennes High precip	$P_{acc}(S)$	Empirical	positive
Kam et al. (2015)	2014	Year	Global high temp	$T_{mn}(Y)$	Broad scale	positive
Stott et al. (2004)	2003	JJA	EU high temp	$T_{mn}(S)$	Empirical	positive
Pall et al. (2011)	2000	SON	UK high precip, floods	$P_{acc}(S)$	AGCM	positive
Otto et al. (2012)	2010	JJA	Russian high temp	$T_{mn}(M)$	AGCM	positive
Schaller et al. (2016)	2013	DJF	UK high precip, floods	$P_{acc}(M)$	AGCM	positive

Table 1.4.1: Summary of papers conducting event attribution studies. The time of event is given in first letter of the month abbreviations. Variables are: P_{acc} = accumulated precipitation, T_{mn} = mean temperature, U_{max} = maximum windspeed, T_{min} = minimum temperature, $MSLP$=mean sea level pressure. Letters in brackets denote timescale of variable: (D)=daily, (M)=monthly, (S)=seasonally, (Y)=yearly. Titles in bold do not appear in the BAMS special issues, but in other peer reviewed journals.

1.5 The mid–latitude atmosphere, extra–tropical cyclones and storm tracks

Mid–latitude cyclones are one of the most important meteorological events for Europe, as they bring warm, moist air from the Atlantic Ocean to mainland Europe. These cyclones bring with them heavy precipitation and, potentially, strong winds (Leckebusch et al., 2006), causing extreme weather events and the consequent impacts. As such, cyclones have been studied for over 100 years, for example Loomis (1885), and can cause the type of extreme event (heavy precipitation, high windspeeds) that are of interest to extreme event attribution.

This section gives a brief overview of mid–latitude cyclones; why they occur, their structure and lifecycle. The material follows chapter 6 in Holton (2004) and chapter 8 in Martin (2006).

1.5.1 The structure of the atmosphere

The Earth receives differential amounts of energy (in the form of shortwave radiation) at the equator and poles, due to the lower incidence angle at the pole. This forms a meridional temperature gradient, with the equator being warmer than the poles. However, although the shortwave energy received is strongly coupled to latitude, the outgoing longwave radiation emitted to space is only weakly coupled. There is, therefore, a mechanism of transferring energy from the equator to the poles (Holton, 2004). Hadley, in the 18th century, postulated this mechanism as a large convection cell, where air rises at the equator, moves poleward, sinks again and moves equatorward. The Hadley cell only exists in the tropics, with two other circulation cells (the Ferrel cell and Polar cell) occurring outside of the tropics. The Polar cell acts much like the Hadley cell, with warm air rising and moving polewards and the cooling air sinking and moving equatorward. However, the Ferrel cell is not driven by convection but is driven by the motion of the Polar and Hadley cells creating eddies on either side of the Ferrel cell and causing vertical motion in the opposite direction to the Polar and Hadley cells.

As the air moves towards higher or lower latitudes it is also deflected by the Coriolis force. In the Hadley and Polar cell, the sinking air returning to the equator is deflected to the left, causing the easterly winds known as the trade winds. In the Ferrel cell, the rising air, which is moving polewards is deflected to the right, causing westerly winds. It is the deflection of the North to South moving air within the Ferrel cell which causes the westerly dominant flow which is responsible for much of the mid–latitude weather systems in the northern hemisphere. However, although the flow is prevailingly westerly, this is only strictly true at upper levels where there is no disruption due to the terrain. At the surface, the westerly flow is weaker than the easterly flow in the Polar and Hadley cell and can be disrupted by local weather phenomena. Figure 1.5.1 shows the climatological zonal mean wind at 200 hPa for the northern hemisphere winter and summer. The opposing directions of the mean wind at different latitudes can be seen in the mid–latitudes, close to the poles and tropics.

1.5.2 Storm tracks

The equator to pole temperature gradient is not universal across the northern hemisphere and varies in relation to the longitude due to the distribution of the oceans and land mass, which have different heating rates. This leads to areas with higher mean zonal wind speeds (at 200hPa) than other areas. Additionally, the equator to pole temperature gradient is greater in winter (DJF) than it is in summer (JJA), in the northern hemisphere, whereas in the southern hemisphere the difference between winter and summer pole and equator temperatures is less (Holton, 2004). In the northern hemisphere, the thermal wind equation provides a good approximation to the mean zonal wind, when the mean seasonal temperature is known, and so the seasonal cycle in zonal wind speed is proportional to the meridional temperature gradient (Holton, 2004).

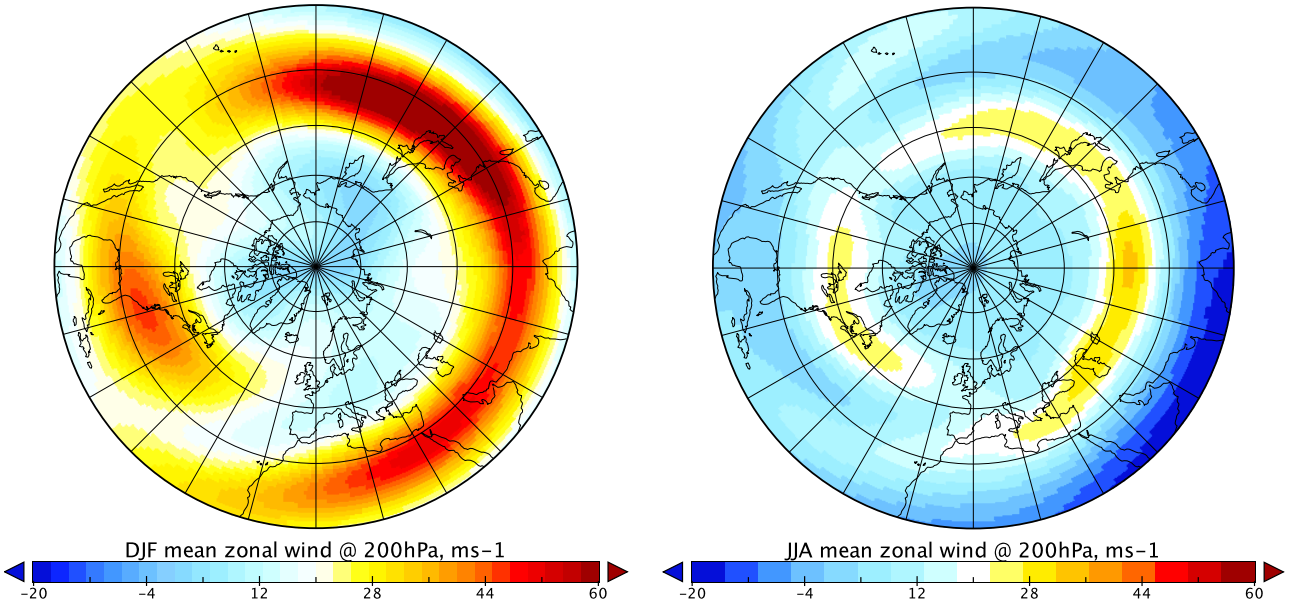


Figure 1.5.1: Climatological seasonal mean zonal wind at 200hPa, for the Northern hemisphere winter (DJF) and summer (JJA), for 1959 to 1999. Data from ERA–40.

Both of these effects can be seen in Figure 1.5.1. Firstly, areas of high zonal wind can be seen in the winter (DJF) near 30°N in the North Atlantic, east of the North American continent and in the East Pacific, east of the Asian continent. There is an additional area of high zonal wind north of the Arabian peninsula. These areas of high zonal wind speed are known as

jets and are also spatially correlated with the storm tracks. The North Atlantic storm track is particularly important as it "steers" low pressure systems from the north–west Atlantic to Europe. The seasonal variation in the strength of these storm tracks can clearly be seen in the contrast between the zonal mean wind speed in DJF and in the summer (JJA). Even in the areas of the storm tracks, the JJA wind speed is less than half of that in DJF. This leads to a seasonal variation in the number of storms forming, and the number which reach Europe before dissipating. It should be noted that orography also has a part to play in the formation of these jets.

1.5.3 The life–cycle of the mid–latitude cyclone

A number of models have been proposed to describe the formation, evolution and dissipation of mid–latitude cyclones. The earliest of these was the Norwegian Cyclone Model (NCM) developed by the University of Bergen in the 1910s to 1920s (Martin, 2006). The major contribution of this work was to identify various stages in the life–cycle of the mid–latitude cyclone, whilst describing the structure of the cyclone at each stage.

- In the NCM, the *initial condition* is described as a knife–like boundary separating cold air to the south and warm air to the north. This is akin to a polar front, separating cold polar air from warm tropical air.

- In the *perturbation stage*, a small disturbance is introduced into the front which deforms it, moves cold air equatorward and warm air poleward. A low–pressure anomaly forms at the centre of the perturbation with a warm front forming to the west of the anomaly and a cold front forming to the east of the anomaly. The positioning of the fronts, where the cold front trails behind the warm front, is due to the Coriolis effect. In the Northern Hemisphere, the rotation is anti-clockwise, drawing cold air down from the poleward direction and warm air from the equatorward direction, ensuring that the cold front is always behind the warm front.

- In the *intensification stage* the low–pressure centre starts to deepen and the cold front rotates anti–clockwise around the anomaly. This causes cold air to move underneath

the warm front, causing uplift of warm air and clouds to form

- In the *mature stage* the low–pressure is at its most intense (deepest) and the strongest winds, due to the large pressure gradient. In this stage the polar front has become so deformed that a cold front extends equatorward and a warm front extends eastward from the anomaly. A region of homogenous temperature between the two fronts exists as the warm zone. Cold air continues to be fed underneath the warm air, leading to more cloud formation.

- By the time the *occluded stage* occurs, the cold front has "caught up" with the majority of the warm–front, and will eventually overtake it. This causes the pressure anomaly to be removed from the warm zone and an occluded front to form connecting the anomaly to the warm zone. This occluded front can be of two types: a cold occlusion, where the cold front ascends the warm front, or a warm occlusion where the cold front undercuts the warm front.

- In the *post–mature* stage the cold front continues to encroach on the warm front and eventually overtakes it, cutting off the supply of warm moist air from the equator side of the front. This eventually causes the low–pressure anomaly to dissipate.

Although the NCM represented a break–through in understanding the evolution of mid–latitude cyclones, it has a number of weaknesses. Firstly, the method by which the perturbation occurs is not discussed. Secondly, the vertical structure of the cyclone is not described, nor the process by which the vertical structure supports the evolution of the cyclone. Finally, the fronts themselves are not knife–like discontinuities of temperature but zones where temperature, density and pressure are continuous. The following section will consider the first of these points more closely.

1.5.4 Baroclinic instability

It is now known that the dominant mechanism by which the perturbations are introduced into the front is baroclinic instability (Holton, 2004), at least in the mid–latitudes. Baroclinic

instability was first discovered, independently, by Jule Charney (in 1947) and Eric Eady (in 1949). In the mid–latitudes, the atmosphere exhibits a strong zonal flow, as seen in Figure 1.5.1, which increases with height, i.e. the winds are stronger aloft than they are closer to the surface. For an atmosphere in hydrostatic and geostrophic balance, the zonal thermal wind equation is given by (Andrews, 2000):

$$f\frac{\partial u}{\partial z} \approx -\frac{g}{T}\frac{\partial T}{\partial y} \tag{1.5.1}$$

where f is the Coriolis parameter, u is the zonal wind, g the gravitational constant, and T the temperature.

From this equation and using an f–plane approximation ($f = f_0 = 2\Omega \sin \phi_0$) and the Boussinesq approximation (the atmosphere is an incompresible fluid, except where density variations are coupled with gravity (Andrews, 2000), for example, where density decreases with height), the thermal windshear equation can be derived (Andrews, 2000):

$$f_0\frac{du}{dz} = \frac{g}{\rho_0}\frac{\partial \rho}{\partial y} \tag{1.5.2}$$

where ρ is the density of the atmosphere and ρ_0 is the reference density. This shows that a zonal wind, which increases with height, is associated with density surfaces which slope both upwards and towards the poles. Without the rotation of the Earth, given by the Coriolis parameter, this slope would not occur, and the density gradients would flatten until density is constant on pressure surfaces. The equator to pole temperature gradient on a rotating Earth has potential energy associated with it, which can be released by baroclinically growing disturbances.

Figure 1.5.2 shows 3 possible situations associated with the exchange of parcels of air from one location to another. The exchange of the denser parcel A with parcel B results in parcel A being surrounded by less dense air and parcel B surrounded by more dense air. Therefore, due to the stratification, these two parcels will be restored to their original positions. Exchanging the less dense parcel C with the more dense parcel D results in the opposite scenario. Parcel C is buoyant (surrounded by less dense air) and will continue to rise whereas D is still

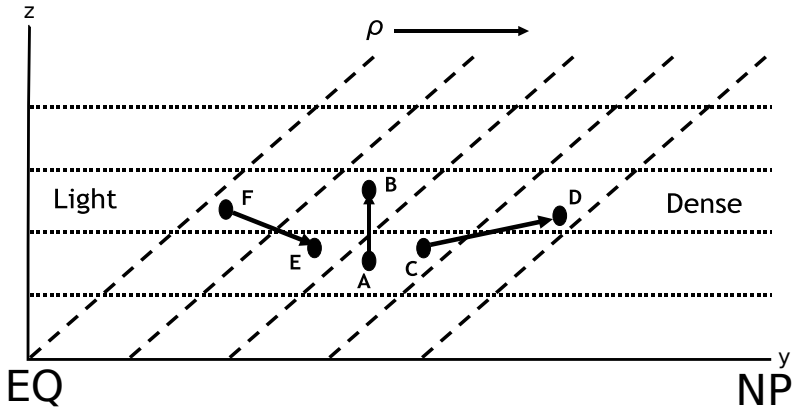


Figure 1.5.2: Schematic demonstrating potential outcomes due to baroclinic instability. Dashed lines show density surfaces, increasing from the equator (EQ) to the North Pole (NP). Dotted lines show height.

more dense than the surrounding air and will continue to sink. The centre of gravity of the system has been lowered, as the lighter parcel has moved higher and the heavier parcel lower and there has been a release of potential energy as kinetic energy. Exchanging the heavier parcel E with the lighter parcel F requires an input of energy and so will not happen without an outside driver. This is because, although each parcel is buoyant with relation to the surrounding atmosphere, the heavier parcel has moved higher and the lighter parcel has moved lower.

The release of potential energy can be evaluated for parcel C and D, assuming each of them occupies a fixed volume V. Then parcel C has mass $m_C = \rho_C V$ and parcel D has mass $m_D = \rho_D V$. If parcel C has height z_C and parcel D has height z_D, then the initial (EP_0) and final (EP_1) potential energy is:

$$EP_0 = g(m_C z_C + m_D z_D) \qquad\qquad EP_1 = g(m_C z_D + m_D z_C)$$

From this, $\Delta EP = EP_1 - EP_0$, gives:

$$\Delta EP = -g(m_D - m_C)(z_D - z_C) = -gV(\rho_D - \rho_C)(z_D - z_C)$$

In this example $\rho_D > \rho_C$ and $z_D > z_C$ so ΔEP is negative and there is a release of potential

energy as kinetic energy, leading to unstable motions (Andrews, 2000). It is this instability that can cause the formation, and then propagate the growth, of mid–latitude cyclones.

1.6 Identification and tracking of mid–latitude storms

The automated tracking of extra tropical cyclones, by identifying and tracking the evolution of storm–indicative features in gridded meteorological data is an active and challenging research topic, with over 15 research groups at institutions around the world having developed and applied their own algorithms (Neu et al., 2013). Although these algorithms have different methods for identifying and tracking the storm–indicative features, they all split the problem into a number of distinct phases. These phases are followed in order and are:

1. The gridded data may be transformed in some way, either via filtering, smoothing or transformation to a different grid or coordinate system. Some algorithms omit this phase and use the untransformed input data.

2. Storm–indicative features in the gridded meteorological data are identified at numerous timesteps by employing an automatic detection algorithm on the regridded data.

3. The identified storm–indicative features are then tracked across the timesteps, via an algorithm which associates features in one timestep with features in subsequent timesteps so as to track the feature as it moves across a domain. This step is known as the correspondence problem (Post et al., 2003).

4. Tracks may be filtered via an inclusion criteria, such as minimum track length, minimum track depth, etc.

5. The tracks are then analysed for spatial position, track length, intensity, depth of depression, etc.

The following sections will discuss these phases of storm–identification and tracking, in the context of the 15 algorithms of Neu et al. (2013) and one other algorithm (Hodges, 1999) which is not part of the inter–comparison study.

1.6.1 Transforming data prior to identification

A common problem facing all tracking algorithms stems from the representation of climate data on a regular longitude–latitude grid. This causes grid boxes to become smaller toward the poles, in effect increasing their resolution and decreasing the spatial scale at which features occur within the data. This results in more feature points per km^2 being identified at high latitudes than at low latitudes. In addition, the data exhibits a singularity at the poles, with the first and last row of data having the same value, which makes searching for localised features difficult.

The 16 tracking methods studied overcome these problems in a number of ways. Four of the methods (Sinclair (1994), Murray and Simmonds (1991), Pinto et al. (2005), Hodges (1999)) transform the data to a polar stereographic projection prior to feature identification. This has the advantage of largely removing the latitudinal bias in the number of features identified, although Sinclair (1997) note that the scale at which features are identified at 60°N is 1.24 times greater than at 30°N. However, the stereographic projection has a disadvantage that features can only be identified in one hemisphere at a time and so, for a global analysis of storm–tracks, the tracking method has to be applied separately to each hemisphere. Other methods transform the data to a rotated coordinates grid (Hewson and Titley (2010), Wang et al. (2006), Rudeva and Gulev (2007)), which is essentially a polar stereographic grid with the pole translated to the mid–latitudes (Jones et al., 2004). This has the advantage of removing the latitudinal bias at the expense of reducing the domain which tracking can be performed over. All projections will introduce some distortion and it is desirable to choose a projection where the distortion is acceptable. For storm–feature tracking the projection should result in grid boxes that are approximately equal in area, which ensures that the features detected are of the same scale throughout the domain.

After transforming the data, filtering of the data is applied in some methods (Pinto et al. (2005), Hewson and Titley (2010), Wang et al. (2006), Murray and Simmonds (1991), Rudeva and Gulev (2007), Sinclair (1994)). This is necessary after remapping the data to another projection as the remapping procedure may introduce some sharp discontinuities which are

not present in the original data. These discontinuities will be identified as features in the next phase of the algorithm, and so have to be smoothed. Additionally, fields which have a high spatial variance, such as vorticity, need to be smoothed so that only features at the scale of interest (i.e. synoptic scale weather features) are identified and not very small scale features (Anderson et al., 2003).

1.6.2 Identification of storm–like features

The methods vary as to which climate variable field they use to identify storm like features. Ten use mean sea level pressure (MSLP), two use geopotential height at 1000hPa (Z1000), one uses geopotential height at 850hPa (Z850) and two use vorticity at 850hPa (VORT850) although each method may use the same field in a different calculation than another method. Hoskins and Hodges (2002) show that a number of climate variables can be used (MSLP, geopotential height, temperature, potential temperature, meridional wind, relative vorticity and vertical wind) to perform identification and tracking, and that there are, qualitatively, similarities in the track positions, but quantitatively differences in the track metrics, such as the lifetime, intensity and propagation speed. The main influence on these track metrics is the choice of climate variable, as that choice will influence the number of cyclones identified and the density of the cyclones. This has a follow–on effect in the tracking algorithm which, in turn affects the resulting track statistics. For example, using MSLP will produce a smaller number of identified features than VORT850, as MSLP features are larger in scale than VORT850 features (Hoskins and Hodges, 2002). It is difficult to interpret the differences in intensity between different climate variables, as the intensity of one variable will have different units and, therefore a different interpretation than another variable. For example, the intensity of MSLP may be expressed as the anomaly from the background field (in hPa), whereas for vorticity it could be the absolute vorticity value (in units of $10^{-5}\mathrm{s}^{-1}$). A normalised anomaly from the background field or climatological mean could be used as a universal intensity metric but would require further investigation to determine its suitability.

Pinto et al. (2005) identifies storm–like features by searching the MSLP field to find a local minimum in the vicinity of a maximum of the Laplacian of pressure (∇_p^2), which is closely

related to the vorticity. A search radius of 1200km is used to find the MSLP minima. This allows a high vorticity to be associated with an actual low pressure system. Benestad and Chen (2006) transform the MSLP data to Fourier space and then search for zero crossings of the first and second differentials in the series, where the first derivative is equal to zero and the second derivative is greater than zero. Hewson and Titley (2010) uses a complicated hybrid method using MSLP, vorticity and temperature to find three feature types: barotropic lows, frontal waves and diminutive waves. Trigo (2006) searches for minima in Z1000 over 3×3 grid points and maximum gradient of Z1000 over 1000×1000km^2.

Sinclair (1994) searches for local minima in the geostrophic relative vorticity field computed from the 1000hPa geopotential. Rudeva and Gulev (2007) find the local MSLP minimum close to the largest local relative vorticity. Hanley and Caballero (2012) simply look for a local minimum within the 8 surrounding grid boxes. Kew et al. (2010) searches for a geo–potential contour which encloses a target area. Wang et al. (2006) find the local minimum MSLP with a MSLP gradient $\geq$ 0.1hPa, and locate the cyclone centre at the maximum value of ∇_p^2. Murray and Simmonds (1991) also find the local minimum MSLP close to the largest local ∇_p^2. Wernli and Schwierz (2006) searches for close MSLP contours that are 0.5hPa higher than the minimum MSLP value.

Hodges (1994) uses a technique developed by Salari and Sethi (1990), to track objects within CCTV images, applied to climate model output data. Thresholding is used to separate the field data into object and background points. Connected component labelling (CCL) and region growing is then used to connect together points that occur within the same object and to grow the objects to include points that belong to the object but that occur outside the region of interest. Once the field data has been segmented in this manner, feature points are detected within the objects by searching for local extrema within the field data and then using the outcome of the CCL to identify feature points occurring within the same object. The position of the local extremum feature is given to be the centroid of the points within the object. The feature point detection in Hodges (1994) is limited to a simple cylindrical projection of the data, in effect a rectangular grid uniform in latitudinal and longitudinal spacing. Hodges (1995) expands the method to work in spherical coordinates on the unit

sphere.

Many of the methods used to identify features in field data documented above, with the exception of Hodges (1994), make no attempt to identify systems of cyclonic or anti-cyclonic activity as objects, relying instead on identifying feature points and equating them with centres of systems. This could lead to two or more feature points being identified per system when, in reality, they belong to the same object. However, even these methods are being used to successfully identify features in climate model output, reanalysis and observational data.

1.6.3 Tracking of storm–like features

Section 1.6.2 results in a number of storm–like features being identified in each timestep of the input data. To produce storm tracks, these features need to be associated with each other across the timesteps. This is known as the correspondence problem (Post et al., 2003) and is solved by the methods in Neu et al. (2013) and Hodges (1999) in various ways.

The **nearest neighbour (NN)** method is the simplest method used by the 15 methods. This method begins (at $t = 0$) by initialising a set of tracks with the feature points in $t = 0$. At $t = 1$, points within a specified search radius are assigned to a track if they are the closest point to the track. If no points are within the search radius of the last point in a track, the track ends. If a point cannot be assigned to a track, as there is no track within the search radius of the point, a new track is created with the point as the first track point.

Some methods (Trigo (2006)) modify this method by having a different search radius westward and eastward, so as to reflect the eastward motion of cyclonic systems in the mid–latitudes. Other methods modify the NN search to locally minimise a cost function (Hewson and Titley, 2010). The main advantage of the NN method is that it is computationally cheap and easy to implement. The main disadvantage is that it does not take into account the storm's previous motion or (in its unaltered state) any of the storm's properties. This can lead to, especially with fast moving systems, a false association between at point at time $t = n$ and $t = n + 1$, occurring when a new point is introduced into the domain at $t = n + 1$

closer to the point at $t = n$ than the system's true point. This can produce very unsmooth tracks, where the system doubles back on itself.

Within the IMILAST inter–comparison, 8 of the 15 methods use some form of NN matching:

- Benestad and Chen (2006) restrict the search distance to 3° of latitude.
- Hewson and Titley (2010) use NN as one of three components in their tracking method.
- Trigo (2006) restrict the search 300km westward, 720km eastward and 540km north/southward for a 6h timestep.
- Wang et al. (2006) restricts the eastward propagation between two 6h timesteps to be 800km and the north/south/westward propagation to 600km.
- Rudeva and Gulev (2007) introduce a three–pass check of forward and backward speed.
- Blender et al. (1997) have a maximum search radius of 1000km per 6h timestep.
- Lionello et al. (2002) restrict cyclones in the next timestep to a box defined by 0.3R in the westward direction, 0.8R in the eastward and 1.4R in the meridional, where R is the average distance of the depression from its centre.
- Bardin and Polonsky (2005) restrict the distance between two 6h timesteps to 800km.

The **predict and match (PAM)** method uses the previous motion of the track to predict the next point in the track at $t = n + 1$ and adds the point, at $t = n + 1$, which is nearest to the predicted point to the track. This method is used by 6 of the 16 identification and tracking methods. The PAM method also starts by initialising tracks with the feature points at $t = 0$. For $t = 1$, there is no previous motion of the track to make a prediction, and so the NN must be assigned to a track. From $t = 2$ and onwards, the two previous points in the track are used to make a projection. For timestep $t = n$, a point is projected from a track using points at $t = n - 2$ and $t = n - 1$, using the positional information from the point. The nearest point at $t = n$ to the projected point is added to the track. Most of the

methods using this technique only project the location information from the point. Within the IMILAST inter–comparison, 6 of the 15 methods use PAM:

- Sinclair (1994), which is an application of Murray and Simmonds (1991), project the pressure and vorticity information and use a weighted sum and cost function minimisation to determine which point to assign to the track.
- Pinto et al. (2005) modify the PAM method by calculating a prediction velocity, which is a weighted sum of the average velocity from previous displacements and a geostrophic steering velocity. The steering velocity is calculated at the surface and from an averaged pressure gradient centred on the cyclone over a radius of 4° in latitude. The resulting velocity is multiplied by a constant so as to account for the increasing wind speed with height.
- Murray and Simmonds (1991) restrict the search radius to 10° of latitude.
- Hanley and Caballero (2012) use a linear projection of the track points at the previous two timesteps to predict the next point. Any point within a certain radius of the projected point is considered a continuation of the track.
- Kew et al. (2010) advects points in an identified structure to the next timestep and their convex hull is found. If an advected hull overlaps more than 10% of an identified structure, then it is considered a continuation of the track.
- Wernli and Schwierz (2006) use previous displacement as first guess for next displacement, with a constant multiplication of 0.75, to account for cyclone slowdown.

The main advantage of the PAM method, over the NN method is that, as the previous motion of the track is used, the doubling back of tracks is not possible, except for possibly very slow moving or static systems. In these cases it is actually desirable for the slow moving system to be able to move against its previous motion as, due to the feature identification process, the feature point may move around on the underlying grid. The main disadvantage is that, at the beginning of the track, for the first two timesteps, it has the same problem as the NN

search, as the NN is used to assign points to tracks when no previous track information is available.

Hodges (1994, 1995, 1999) (HOD) uses the minimisation of a cost function so as to maximise the smoothness of the tracks. Tracks are initialised by using a NN search to build an initial set of tracks. Any tracks which do not last for the full time period of the data are extended with phantom feature points. The cost function is based on the local smoothness of a track, which is a measure of the change in speed and change in direction of a feature point, and tracks subject to an upper bound on the distance moved and the local smoothness. The cost function is minimised, subject to these contraints by a Greedy exchange algorithm that has been adapted from Salari and Sethi (1990). This algorithm iterates forwards and backwards in time, exchanging points in two tracks so as to give the greatest increase in track smoothness.

The main advantage of the HOD method is that a truly optimal, in terms of smoothness, set of tracks is obtained. It is not subject to the double back errors of the NN method, nor is it subject to potential errors at the beginning of tracks like the PAM method. The only disadvantage is that it is more computationally expensive than the other methods.

1.6.4 Track post–processing

After the tracking phase of the algorithm has been completed, there exists a number of tracks, spanning the time frame for which the original data exists. These tracks may be filtered on a number of criteria, such as tracks which do not last for more than 24 hours, do not deviate more than 1000km from their starting point, track over terrain greater than 1500m, do not have a low enough central pressure, or do not have a high enough pressure gradient (Neu et al., 2013).

1.6.5 Track analysis

After filtering, the tracks are analysed, typically for their spatial distribution and density, track length, intensity (MSLP or vorticity value), propagation speed and lifetime (persis-

tence) (Neu et al., 2013). These analyses can be used to construct climatologies of storm tracks or to compare model output to observations or reanalysis data.

In chapter 3, the spatial distribution of the occurrence of cyclones between 1989 and 2009 in the ERA–Interim reanalysis dataset (Dee et al., 2011) is calculated using the same method as in Neu et al. (2013). This calculates the feature density on the original $1.5° \times 1.5°$ grid from ERA–Interim in the following way:

- For each grid box the number of times a feature occurs within the box is counted.
- These counts are then divided by the surface area of the grid box and multiplied by 1000km^2.
- These counts are then further divided by the number of timesteps over the total input data period (1989 to 2009, 6 hourly in the ERA–Interim case here) and multiplied by 100.
- The result is smoothed in the zonal direction by a $4°$ wide smoother with descreasing weights.

This process produces the cyclone centre density (also referred to as the feature density) as the percentage of cyclone occurrence per timestep and per unit area of 1000 km^2 (Neu et al., 2013). This can be interpreted as an expected frequency, i.e.: if the cyclone density is 20% then, over 200 timesteps it is expected to find 10 cyclones in an area of 1000 km^2. A value greater than 100% indicates that, on average, there is more than one cyclone per timestep. Therefore, dividing this number by 100 gives the expected frequency per timestep in a 1000km^2 area.

Although simple to calculate, and used in this thesis so as to facilitate comparisons with the results in Neu et al. (2013), calculating the spatial density of features in this way can lead to a noisy and biased result. The noise is a by–product of counting features. This enables slow moving systems to be counted more than once in a grid box and fast moving systems to be counted in grid boxes that are not adjacent, leaving gaps in the result. To

overcome this problem, the track density can be computed instead. This counts the number of times individual tracks intersect a grid box. Using this method, slow moving systems are only counted once per grid box and, for fast moving systems, the grid boxes between the subsequent feature points are also counted.

However, computing track density does not remove the biases, which Taylor (1986) shows are due to the size and shape of the grid box, and the direction across which the cyclone travels. Taylor (1986) postulates that using circular grid boxes can overcome this bias, a method used by Murray and Simmonds (1991) and Sinclair (1994). Hodges (1996) overcomes this bias by using spherical nonparametric kernel estimators to calculate track statistics directly on the sphere. This allows the construction of probability distributions that are continuous across the sphere, so removing the biases inherent in methods that are essentially calculating a two–dimensional histogram, with discretised and unequal bins, on the sphere. Although out of scope of the thesis, an investigation into how sensitive the track statistics are to the method of computing them would be interesting.

1.6.6 Application of tracking algorithms to climate data

The tracking algorithms detailed above have been used in numerous studies. Some of these are outlined below.

Hoskins and Hodges (2002) use the tracking method of Hodges (1995) and the Eulerian method of Blackmon et al. (1977) to explore the properties of the Northern Hemisphere winter storm tracks represented in the ERA-15 (Gibson et al., 1997) reanalysis dataset and also in the NCEP–NCAR (Kalnay et al., 1996) reanalysis. Several different fields are used, including mean-sea-level pressure (MSLP), geopotential, meriodional wind, relative vorticity, temperature, potential temperature, vertical velocity and potential vorticity. Both the Eulerian and objective tracking methods are applied to the fields for the winter season and averages are taken over a 22 year period (1979-2000) to produce climatological storm track maps for both methods over all of the fields. From these maps a schematic of principal storm tracks for the upper and lower troposphere is generated.

Hoskins and Hodges show, for the mid-latitude storm tracks in the ERA–40 reanalysis (Uppala et al., 2005), that there is good correspondence between the variance statistics obtained by applying the bandpass filter to 850hPa geopotential height data and the density of features identified and tracked by applying the objective feature analysis routines to the same data. A similar study is undertaken in Hoskins and Hodges (2005) with climatological maps of the storm tracks in the Southern Hemisphere produced for all seasons.

Bengtsson et al. (2005) apply the same methods of Hoskins and Hodges (2002) to the output from a coupled ocean and atmosphere climate model to investigate the impact to storm tracks of climate change. The model used is ECHAM5, which is run as a control with pre-industrial quantities of greenhouse gases (GHG), then with observed quantities of GHG from 1860 to 2000. Finally, the model is run to 2100 using the IPCC SRES scenario A1B, which prescribes GHG to increase steadily to 2050 and then decline. The storm tracks in the output from the climate model between 2071-2100 and the storm tracks between 1960-1990 are compared to determine the effect of increasing GHG on the storm tracks. Bengtsson et al. found that, in general, that there is a shift towards fewer, more intense storms in the 21st Century from greater numbers of less intense storms in the 20th Century, although the effect varies depending on the region.

Sinclair (1994) uses their method mentioned above on European Centre For Medium-Range Weather Forecasts (ECMWF) analysis data, from 1980 to 1986, to derive a climatology of cyclonic vorticity for the Southern Hemisphere. Sinclair (1997) uses an improved method, applied to spatially filtered data from 15 years of ECMWF analysis, along with an estimation of cyclone strength to obtain a cyclone climatology for the Northern and Southern Hemispheres from 1980 to 1994. The spatial filtering ensures that each feature identified has the same spatial scale. From this analysis a database of position, pressure, vorticity and circulation of cyclones is built. Sinclair and Watterson (1999) apply the same methods to track centres of cyclonic vorticity in climate models, for models with CO_2 for an industrialised world and also for models with twice this CO_2. A decrease in cyclone activity and a reduction in the number and strength of intense storms is found in the double CO_2 models.

Blender et al. (1997) search for local minima in the 1000hPa geopotential height field of ECMWF weather forecasts over the North Atlantic from East America to Western Europe for winters from 1990 to 1994. Cluster analysis is performed on the climatological average of the storm tracks to derive cyclone track characterisations and a time series of circulation statistics is produced from these characterisations.

Hodges et al. (2003) use the methods of Hodges (1995) to compare synoptic scale weather features in four reanalysis datasets. Good correspondence between the four datasets are found for track statistics in the Northern Hemisphere with the ECMWF reanalysis producing the strongest storm tracks but weaker systems around orographic features. In the Southern Hemisphere, the agreement in position and intensity of the storm tracks is less consistent across the 4 datasets. A comparison of 4 more recent reanalysis products (ERA–Interim, NASA MERRA, JRA–25 and NCEP CFSR)(Hodges et al., 2011) shows a good correspondence in the Northern Hemisphere, whereas in the Southern Hemisphere 3 of the 4 datasets have good agreement, with only the older, lower resolution reanalysis (JRA–25) differing significantly.

Greeves et al. (2007) use the method of Hodges (1995) to assess the representation of winter storm tracks in the Northern Hemisphere in climate models. Storm like features in a variety of climate models from the Hadley Centre at the UK Met. Office are identified, tracked and compared to other measures of storminess, including the Eady growth rate, the band-pass filtering method of Blackmon (1976), an analysis of the diabatic heating and analysis of the components of the energy cycle. The storm tracks in the HadGAM1 model (the state of the model) are compared against ECMWF ERA-40 reanalysis data and the HadAM3 model (the previous generation model). Good correspondence is found between the representation of the storm tracks in HadGAM1 and ERA-40 with an improvement over the previous generation model, HadAM3.

1.6.7 Inter–comparison of tracking methods

Validating a storm tracking algorithm is problematic as there is no universally agreed set of historical storm tracks with which to compare the output (Neu et al., 2013). IMILAST ("Intercomparison of Mid-Latitude Storm Diagnostics") is a tracking–algorithm inter–comparison project which compares the 15 tracking algorithms detailed in this chapter by applying each method to the ERA–Interim reanalysis (Dee et al., 2011) for the period 1989–2009. Even though the 15 tracking algorithms have widely varying methods, they found that there was high consistency across the methods for deep cyclones, which holds for cyclone frequency and lifecycle. However, for propagation speed, there is less agreement across the methods, indicating that some algorithms may not recognise rapid movement, if the search distance between timesteps is set too low. An opposite problem, when the search distance is set too high, is the problem of track aliasing, where shorter, individual tracks are combined into longer tracks, often with a large distance between the start and end points of individual tracks. Input data with a poor time resolution (less than 4 timesteps per day or more than 6 hours per timestep) can also suffer from track aliasing. This is explored more fully in chapter 3, section 3.3..

The spatial density of cyclone centres matches well across the 15 methods, with the exception of the Mediterranean. Differences in the actual total number of cyclones is large, but the normalised distributions are similar across the methods. This large discrepancy is mostly due to the climate variable field and method used to identify the cyclone, rather than the tracking method used. Methods that use a local minima search in either the MSLP or geopotential height field identify between 57 and 95 cyclones in the Northern Hemisphere (NH) winter (DJF), with a mean of 73.5. Method M03 (Benestad and Chen, 2006) identifies only 57 cyclones due to restricting the number of cyclones per timestep in the NH. Methods using the vorticity field or the Laplacian of pressure (which is related to vorticity) identify between 111 to 214 cyclones in NH DJF, with a mean of 147.3. Method M18 (Sinclair, 1994) has the largest number (214) of cyclones, whereas the next largest (M09 (Wang et al., 2006)) has 168, marking M18 as an outlier.

Blender and Schubert (2000) investigate how different spatial and temporal resolution data may influence the results from a feature tracking algorithm. Data from the ECHAM4 climate model is degraded both in spatial and temporal terms, into several different temporal / spatial resolution datasets. For example, the original data is T106 and 2 hourly resolution, and this is degraded to T84,T63,T41 and T21 (the lower the number the lower the resolution). Each of these is then degraded to 4,6,8,10,12, etc. hourly data up to a minimum resolution of 24 hours. The same storm tracking algorithm is then applied to each of these sets of degraded data, and the tracks produced are compared with the original dataset of T106, 2h being taken as the reference track set. They show that temporal resolution is very important, with the more degraded data missing tracks and having a higher proportion of incorrect tracks. By 24h, only 45% of tracks were correct and 26% of tracks were missing altogether. The spatial resolution is also very important, with the lowest resolution (T21, 2h) having 95% missing tracks. By degrading both the time and the spatial resolution to T21, 24h, a balance between correctly identified and missing tracks is achieved, of 53% and 56%.

Raible et al. (2008) compare cyclones detected in the NCEP–NCAR (Kalnay et al., 1996) to those detected in the ERA–40 (Uppala et al., 2005) and find that ERA–40 contains the highest cyclone centre density. Secondly, they compare three tracking methods: Wernli and Schwierz (2006), Blender et al. (1997) and Murray and Simmonds (1991). The tracking schemes show a good correspondence between the different methods, but there are discrepancies in the trend of total cyclone numbers. There are also regional differences in the density of cyclone centres, which can be explained by the parameter settings for the algorithms.

1.7 Definition of sub regions of Europe

The Inter–governmental Panel on Climate Change (IPCC) special report on extreme events (SREX, Field et al. (2012)) defines three regions of Europe over which the projected changes in extreme weather events are analysed. These regions are Northern Europe (NEU), Central Europe (CEU) and Mediterranean (MED) and are defined along geographical and climate type boundaries. Figure 1.7.1 presents the boundaries of the three regions of Europe.

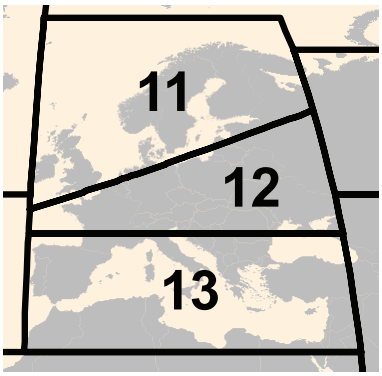


Figure 1.7.1: Regions of Europe used in the analysis of the large ensemble of climate models in Chapters 5,6 and 7. Adapted from Figure 3–1 in Field et al. (2012). Region 11 – Northern Europe (SREX NEU), 12 – Central Europe (SREX CEU), 13 – Mediterranean (SREX MED).

Chapters 5, 6 and 7 use these regions to analyse the representation of European weather in the HadAM3P GCM. Another two regions are also defined, one much larger than the SREX regions, covering the entirety Europe. This has the boundaries of all of the SREX regions combined. The second region is much smaller, and covers just the UK.

Region	Coordinates (Longitude [◦], Latitude [◦]) of region corners			
ALL EU	(10W, 30N)	(10W, 75N)	(40E, 75N)	(40E 30N)
NEU	(10W, 48N)	(10W, 75N)	(40E, 75N)	(40E, 61N)
CEU	(10W, 45N)	(10W, 48N)	(40E, 61N)	(40E, 45N)
MED	(10W, 30N)	(10W, 45N)	(40E, 45N)	(40E, 30N)
UK	(11.25W, 50N)	(11.25W, 58.75N)	(1.875E, 58.75N)	(1.875E, 51.25N)

Table 1.7.1: Regions of Europe used in the analysis of the large ensemble of climate models in Chapters 5,6 and 7. Adapted from Table 3.A-1 in Field et al. (2012), with regions for all Europe and the United Kingdom added. The regions are defined as a four sided polygon, taking the points in the table clockwise.

Using these regions enables the model to be evaluated on several spatial scales and over the areas which extreme weather event attribution studies have been conducted. For example, Massey et al. (2012) and Christidis and Stott (2012) are conducted for the UK region, Cattiaux and Yiou (2012) and de Vries et al. (2013) analyse events over the ALL EU domain, Tett et al. (2013) and Dong et al. (2013) are concerned with the NEU region, Schaller et al. (2014) with the CEU domain and Añel et al. (2014) with the MED domain.

1.8 Aims and scope of this thesis

As identified in section 1.3.2 one method of conducting an attribution study is to use a large ensemble of AGCMs with two climate scenarios, one representing observations and one representing the counter–factual scenario of a world without elevated GHG emissions. Until now, no formal evaluation of a model used in this type of attribution has been performed. This thesis evaluates running an AGCM via a very large ensemble, and determines whether the model can reliably simulate climate conditions at the daily and seasonal timescales so as to gauge whether the model is suitable for use in extreme weather event attribution studies. The model to be validated is the HadAM3P GCM from the UK. Met Office, and is already being used in extreme event attribution studies, including a number of studies in section 1.4 (Otto et al., 2012; Schaller et al., 2016; Massey, 2012; Trigo et al., 2013; Sparrow et al., 2013; Añel et al., 2014). To perform the validation, the climate variables from the model are compared to climate variables from observed datasets and reanalysis products.

As discussed in sections 1.2 and 1.4 extreme weather events occur on a number of temporal scales, from the hourly to the yearly, with the attribution studies mentioned above having scales ranging from daily to seasonal. The weather events occur also on a number of spatial scales, from single points for high temporal events, through country to continent wide for longer timescale events. This motivates using climate variables at different spatial and temporal scales to validate the model. Finally, high precipitation and windspeed events in Europe are intrinsically linked to the occurrence and path of cyclones. Therefore, to fully validate the model, the location, duration and intensity of cyclones is compared to those in reanalysis data.

The thesis will validate the model in the following ways:

- The seasonal means and percentiles of four climate variables: temperature, MSLP, precipitation and upper–level wind are compared to the seasonal means and percentiles of observations and ERA–40 reanalysis. This determines the model's suitability to

studying extreme events that occur on a seasonal timescale.

- The distributions of daily minimum temperature, maximum temperature, maximum windspeed at 10m, mean MSLP and accumulated precipitation are compared to the distributions of the same variables in ERA–Interim reanalysis. This will determine the model's ability to model extreme events that occur on the daily timescale.

- A storm–like feature identification and tracking algorithm is developed that is particularly suited to applying to large ensembles of climate data, as it is computationally efficient.

- The feature identification algorithm is applied to the output of the large ensemble of a climate model and ERA–Interim. Cyclone centre density (Neu et al., 2013) and track metrics are compared between the model and ERA–Interim to determine how well the model represents mid–latitude storms. This will determine the model's suitability for use in attribution studies into windstorms and high precipitation events, as well as validating the general circulation of the model.

- The distributions of precipitation, MSLP and windspeed at 10m within 1000km of a cyclone centre are compared to the same quantities in ERA–Interim. This determines the model's ability to represent the very extreme weather events, which occur at the very edge of the tails of the distributions.

Chapter 2 describes a novel feature detection algorithm that is used to detect the mid–latitude cyclones. This algorithm relies on regridding the climate model data to an equal area triangular mesh and then identifies cyclones by examining the surrounding triangles. These cyclone centres are then grown into features by a region growing algorithm. An optimisation algorithm that employs a cost function is used to form tracks of the cyclones. Using these algorithms it is possible to study the storm tracks in both the very large ensemble and the reanalysis data.

Chapter 3 applies the novel feature identification and tracking algorithms to ERA–Interim 6 hourly data and compares the result with a number of other tracking algorithms which

have partaken in a tracking algorithm inter–comparison project. The feature identification and tracking method is, therefore, validated against other methods.

Chapter 4 details the construction of a very large ensemble of a global climate model, computed between the period 1961 to 2010. The formulation of the model used in this ensemble is described and the climate scenario under which this ensemble is performed is discussed, along with the derivation of the boundary forcing conditions. The initial condition of the model is documented, as is the perturbation of the initial condition in order to fully exploit the internal variability of the model. Finally, the large ensemble is achieved using volunteer distributed computing and this is also described.

Chapter 5 presents the first of the model validation chapters, testing the seasonal biases of the model over the entire Earth. It considers some of the differences between observational and reanalysis datasets and how this can affect a validation exercise. The maximum bias is calculated for the ERA–Interim reanalysis for temperature and precipitation, and this maximum bias is compared to the bias between the very large ensemble computed in chapter 4 and observations for the seasonal means over the whole globe, for temperature and precipitation. A comparison is also made between the model and ERA–40 for seasonal MSLP and upper–level wind.

Chapter 6 presents the validation of climate variables over the five regions of Europe described in section 1.7 on daily timescales. The distribution of five climate variables, maximum and minimum temperature, accumulated precipitation, MSLP and windspeed at 10m are compared to the distribution of the corresponding variables in ERA–Interim reanalysis.

Chapter 7 presents the results of applying the feature tracking algorithms of chapter 2 to the very large ensemble and comparing the storm tracks, intensity of storms and their frequency to the results of applying the algorithms to ERA–Interim. The distributions of minimum MSLP, maximum precipitation and maximum windspeed at 10m within 1000m of the cyclone centres are also compared between the large ensemble and ERA–Interim.

Chapter 8 presents the conclusions of the thesis and discusses how the various methods

presented for validating very large ensembles of a climate model could be used in an attribution of extreme events study. Potential future work describing how the storm tracking algorithms could be used in an attribution study is also discussed.

2 Feature tracking on a hierarchical triangular mesh

As detailed in 1.6.1, tracking storm–indicative features in gridded climate data is prone to a number of challenges, including a latitudinal bias in the number of storms identified, gridded climate data having a singularity at the poles, and requiring that small scale perturbations can be isolated from the background flow. This chapter shows that these problems can be overcome by regridding the regular longitude–latitude data to a hierarchical triangular mesh obtained by repeatedly sub–dividing an icosahedron contained within the unit sphere. Constructing the icosahedron so that a centroid of a triangle corresponds to the locations of the poles ensures that there are no singularities. Additionally, the mesh is equalised via a vertex repulsion algorithm that ensures each triangle has the same area. Climate data can then be regridded to this hierarchical triangular mesh, storm–indicative features can be identified in the regridded data and these features can then be tracked across successive timesteps. This chapter extends and improves Massey (2012) by reimplementing the feature detection and tracking phases of the algorithm.

In chapter 3, the regridding, feature identification and tracking software is applied to reanalysis data, and compared to other tracking algorithms, so as to validate the methods used. Furthermore, the sensitivity of the algorithm to the temporal resolution of the data is also explored.

2.1 Previous work

The (approximately) spherical shape of the Earth lends itself to a gridded representation of a triangular mesh that has been sub–divided from a Platonic solid. Representing climate variable data on such a mesh has a number of advantages. Firstly, the triangles are approx-

imately equal area, and can be equalised further. Secondly, the data can be represented hierarchically at different resolutions. Thirdly, navigation through the hierarchy to a particular location on the mesh is both simple and fast to compute. Lee and Samet (2000) show how the Earth can be mapped to the top–level triangles of an icosahedron and how these 20 triangles can be sub–divided into 4 triangles, which are then subsequently sub–divided. A linear quad tree (a tree without any pointers) is used to store the triangle definitions, and any data within the triangles. Very fast algorithms for locating triangles via a labelling scheme and finding the triangle's neighbours is also presented.

Hodges (2008) uses a spherical quad tree (SQT) to sub–divide an icosahedron of spherical triangles bound by the unit sphere into smaller triangles, so as to estimate confidence intervals for spherical data derived from feature tracking. A labelling scheme is also presented along with a method of traversing the tree to find neighbouring spherical triangles.

The methods presented in this thesis draw on the work of Teanby (2006), Lee and Samet (2000) and Hodges (2008). The sub–division of the icosahedron and data–structures used to store the triangular mesh follow Hodges (2008), although applied to planar triangles rather than spherical triangles. This requires an additional step of normalising the sub–divided vertices to the unit sphere everytime a triangle is split. The storage of data on the mesh uses the hierarchical triangular array (section 2.2.7) which is a form of linear quad tree with tesseral addressing, as used by Lee and Samet (2000).

2.2 Generating a hierarchical equal area triangular mesh

In this thesis, the mesh generation consists of three steps. Firstly, a mesh of triangles is constructed in the shape of an icosahedron contained within the unit sphere. Secondly, the mesh is sub-divided by recursively partitioning each triangle. Thirdly, the mesh is optimised to ensure each triangle has an equal area.

2.2.1 Initialisation of the triangular mesh

The initial phase of creating the mesh follows Teanby (2006) but with improvements to the positioning and data storage of the mesh. The twelve vertices of an icosahedron are created in three–dimensional Cartesian coordinates and are rotated around the x-axis by $\pi/8$ to ensure that no vertex coincides with a pole. The value of $\pi/8$ and a rotation around the x-axis are chosen so that each pole lies on the centroid of a triangle in the mesh, which ensures that a feature detection algorithm searching for local minima can detect them at the pole. Without rotation the poles would coincide with the vertex of a triangle, leading to undefined behaviour for locating minima at the poles. Rotation around the x-axis in a three–dimensional Cartesian Coordinate system is given by (Foley et al., 1990):

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \tag{2.2.1}$$

where θ is the angle of rotation around the x axis.

Multiplying the matrix out and substituting $R_y = \cos\theta$ and $R_z = \sin\theta$ give the equations in the third column of Table 2.2.1 The mesh is stored in two data structures, a point cloud and an array of quad trees. These data structures allow for easy storage and access to the vertices, addition of new vertices, subdivision of the triangles and the labelling of the triangles.

The point cloud

The point cloud used to store the vertices of the mesh is a simple random access array of objects. Each member of the array contains the *x, y, z* of the vertex and a force vector, which is used in the optimisation of the triangle areas. When a new vertex is added to the point cloud, it is compared against the vertices currently in the cloud and only added if it does not already occur. This ensures that, when the triangles are split, no duplicate vertices are created.

Vertex number	*xyz coordinate*	*xyz coordinate after rotation*
1	$(0, \phi, \theta)$	$(0, \phi R_y - \theta R_z, \phi R_z + \theta R_y)$
2	$(0, -\phi, \theta)$	$(0, \theta R_z - \phi R_y, \phi R_z + \theta R_y)$
3	$(0, \phi, -\theta)$	$(0, \phi R_y - \theta R_z, -\phi R_z - \theta R_y)$
4	$(0, -\phi, -\theta)$	$(0, -\phi R_y + \theta R_z, -\phi R_z - \theta R_y)$
5	$(\theta, 0, \phi)$	$(\theta, -\phi R_z, \phi R_y)$
6	$(-\theta, 0, \phi)$	$(-\theta, -\phi R_z, \phi R_y)$
7	$(\theta, 0, -\phi)$	$(\theta, \phi R_z, -\phi R_y)$
8	$(-\theta, 0, -\phi)$	$(-\theta, \phi R_z, -\phi R_y)$
9	$(\phi, \theta, 0)$	$(\phi, \theta R_y, \theta R_z)$
10	$(-\phi, \theta, 0)$	$(-\phi, \theta R_y, \theta R_z)$
11	$(\phi, -\theta, 0)$	$(\phi, -\theta R_y, -\theta R_z)$
12	$(-\phi, -\theta, 0)$	$(-\phi, -\theta R_y, -\theta R_z)$

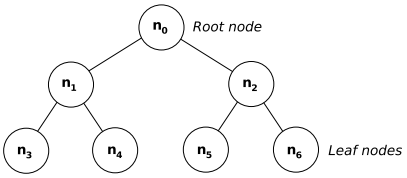
Table 2.2.1: Initial coordinates of the 12 vertices forming an icosahedron, where $\phi = 2\cos(\pi/5)/\tau$, $\theta = 1/\tau$, $\tau = \sqrt{1 + 4cos^2(\pi/5)}$, $R_y = \cos(\pi/8)$ and $R_z = \sin(\pi/8)$

Quad trees

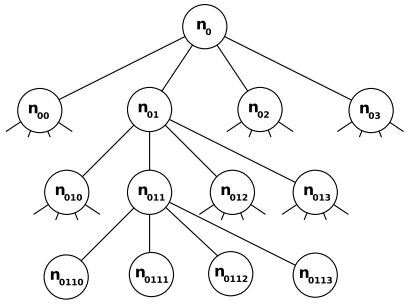
A tree is a data structure consisting of nodes and edges (Sedgewick, 1992), the edges representing connections between the nodes. There can be an arbitrary number of connections between the nodes and these connections are expressed as familial relationships. The nodes n_1 and n_2 which are connected to node n_0 are the children of n_0, node n_0 is the parent node of n_1 and n_2 and nodes n_1 and n_2 are siblings of each other. The root node is the node at the top of the tree and has no parent whereas leaf nodes have no children and are at the bottom of the tree. There can be an infinite number of leaf nodes but only one root node. The level of a node in the tree is analogous to the generations in a family, for example, a node at level 2 will be 2 generations away from the root node, that is, its parent's parent is the root node. Figure 2.2.1a shows a simple balanced tree.

Moving from one node in the tree to another is traversing the tree and visiting every node in the tree, in order from the root node, is walking the tree. Tree walking is a recursive procedure which simply involves starting at the root node and visiting each sub tree, and each sub tree of that tree, in turn until either a leaf node is reached or some search criteria is reached.

A quad tree (Finkel and Bentley, 1974) is a tree in which each node has exactly four children. A balanced quad tree occurs when all the nodes at a certain level have the same number of children. In a quad tree, this will be either 4 or 0. Figure 2.2.1b shows a balanced quad tree with 3 levels. In the hierarchical triangular mesh, 20 quad trees are created, one for each



(a) A simple balanced binary tree



(b) A quad tree with 3 levels. Not all nodes are shown.

Figure 2.2.1: Binary (a) and quad tree (b)

face of the icosahedron and these 20 quad trees are stored in a random access array. Each node in each quad tree contains a reference to 3 vertices in the point cloud. Thus, each node in the quad tree represents a triangle and the root node of each quad tree in the 20 member quad tree array represents one of the triangular faces of the original icosahedron. This is similar to the spherical quad tree (SQT) method of Hodges (2008) but applied to planar triangles that are normalised to the unit sphere.

2.2.2 Subdividing the mesh

Now that the initial vertices and faces of the icosahedron have been created and stored in suitable data structures, they can be subdivided to generate a mesh of higher resolution. The level of the mesh is defined as the number of times each triangle in the mesh has been divided from the original face in the icosahedron. The initial icosahedron has a level of 0 and the level in the mesh is analogous to the level in the quad tree. Subdivision is a recursive procedure and will be performed on each triangle until the desired mesh level is reached.

For a triangle consisting of vertices A, B and C, subdivision proceeds as follows. Firstly, the mid-points of the vectors $\boldsymbol{AB}, \boldsymbol{BC}$ and $\boldsymbol{AC}$ are determined. The mid-points are then

projected to the unit sphere by normalising the mid-point vectors and added to the point cloud. As a feature of the implementation of the point cloud, no duplicate vertices will be created. Finally, four triangles are created from the original 3 vertices and the 3 mid point vertices and the four triangles are added as children of the node containing the current triangle in the quad tree.

Figure 2.2.2 shows the subdivision of the icosahedron and the interaction between the mesh, point cloud and quad tree. Due to the data structures and method used to construct the mesh, triangulation, such as the Delaunay Triangulation provided by Renka (1997), is not needed.

The data structures contain a fully realised representation of the triangular mesh that is guaranteed to have no holes, as it has been subdivided from an original tessellated shape. The mesh is created just once for each level and stored in a mesh definition file. This allows the regridding algorithm discussed later to just read in the mesh definition file and save on computation time.

2.2.3 Labelling the mesh

Labelling the triangles and their corresponding nodes in the quad tree allows for an easy way to uniquely reference each triangle. The labelling scheme for the mesh follows Kunszt et al. (2001). Each of the 20 triangles in the original icosahedron are labelled alphabetically from A to T. The triangles that result from a subdivision have a 1, 2, 3, or 4 appended to the label of the triangle from which they have been divided. The numbers relate purely to the arbitrary order in which the triangles are created. In terms of the quad tree, the root nodes have the labels A to T and the child nodes have a 1, 2, 3 or 4 appended to the label of their parent node. This continues down the tree until the leaf nodes are reached. The mesh labels are also stored in the mesh definition file.

Figure 2.2.3 shows the order of labelling and the relationship between the quad tree and mesh.

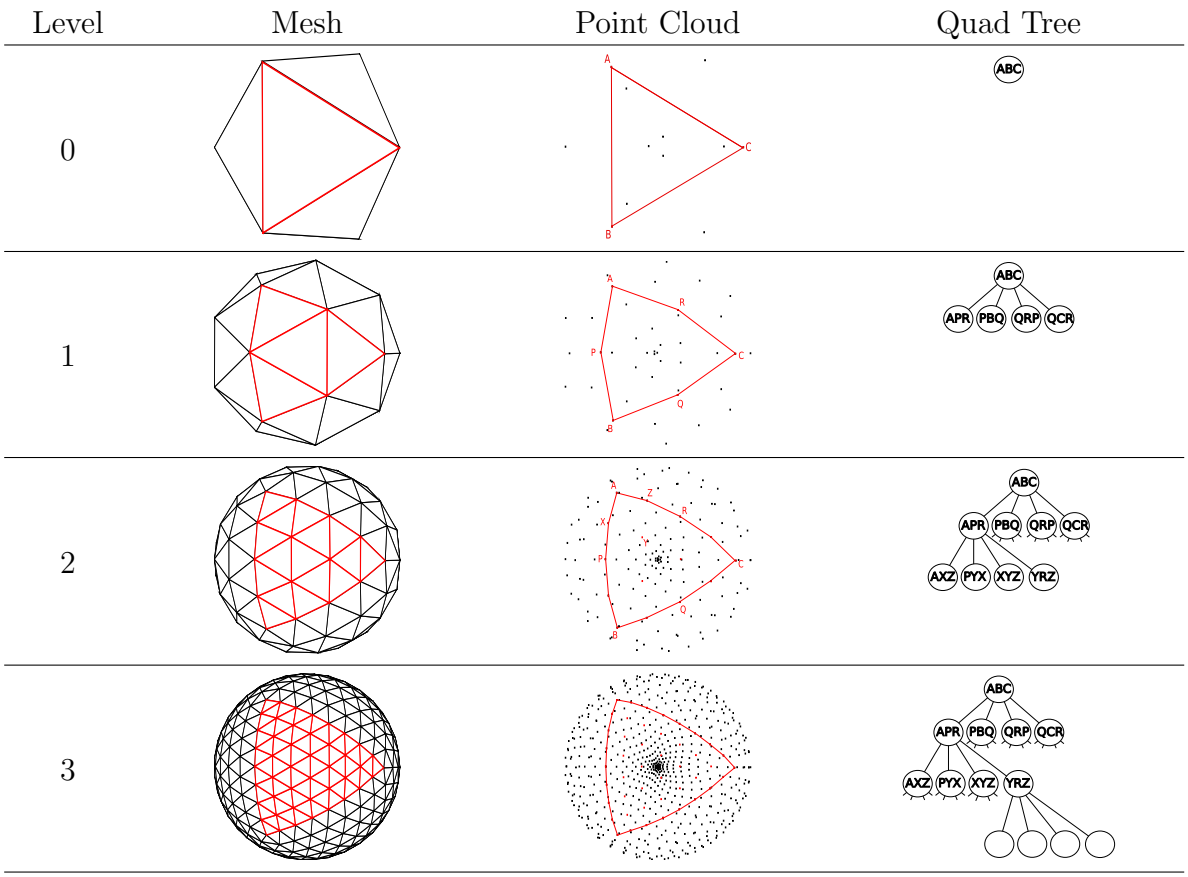


Figure 2.2.2: Recursive sub division of triangles to form the mesh. The mesh column shows the three dimensional shape created at this level of the grid. The point cloud column shows the associated point cloud with this shape and the quad-tree column shows the tree containing the mesh definition, where each node contains three points from the point cloud.

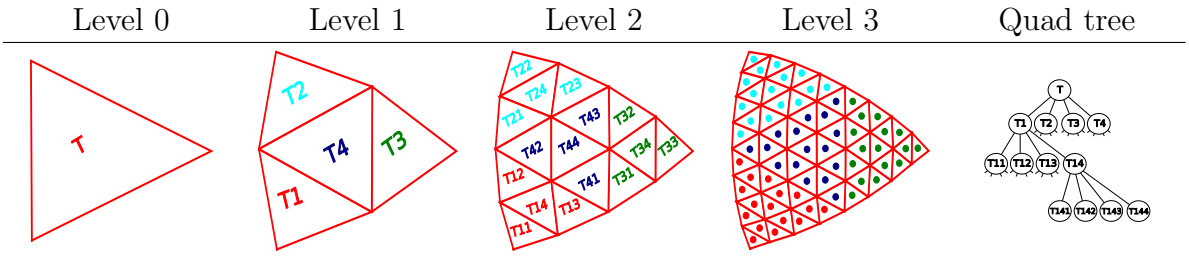
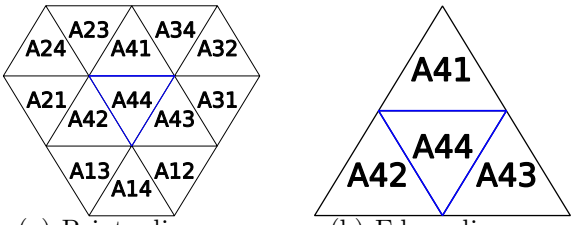


Figure 2.2.3: Labelling of the sub divided mesh and the corresponding quad tree. Colours from level 1 onwards indicate the triangle in level 1 that subsequent triangles are divided from.

2.2.4 Calculating the adjacency relationships of the triangles

To enable objective feature detection or polygonal binning of data regridded onto the mesh, it is useful to calculate which triangles are adjacent. There are two obvious types of adjacency,

one defined by triangles sharing points and one defined by triangles sharing edges. These are referred to as point adjacency and edge adjacency respectively. Figure 2.2.4 shows the two different adjacency types and the triangles in the mesh that correspond to the adjacency. Adjacencies are calculated for every level of the mesh by traversing the tree and



(a) Point adjacency (b) Edge adjacency

Figure 2.2.4: Adjacency types in the hierarchical triangular mesh. *a* shows the point adjacency, where each triangle shares at least one point with another triangle. *b* shows the edge adjacency where each triangle shares one edge with another triangle.

examining the relationships between the vertices in the point cloud that are referenced by the triangle definitions held in the quad tree. An associative container, such as a hash table (Sedgewick, 1992) can greatly simplify finding the adjacencies. An associative container is a data structure composed of pairs of data, a key and a value. The key or value themselves can be a further data structure. A hashing function works on the key to determine the location in the container that the data should be stored in (Sedgewick, 1992).

To determine the point adjacencies in the mesh, an associative container is created with vertices as the keys and a list of labels as the values. The tree is traversed to the level that adjacencies are currently being calculated for, with every node in the tree to that level being visited. Every node in the tree contains three references to vertices in the point cloud and a label. The indexing function is applied to the coordinates of each vertex and the label is added to the list of labels at the location in the data structure that the indexing function points to.

Due to triangles sharing vertices in the point cloud, this algorithm builds lists of triangles (via the labelling scheme) that contain a common point. For point adjacency this means that each triangle in the list is adjacent to the other triangles in the list. The point adjacency

map is then created by refactoring each adjacency list in the associative container into a number of one to many mappings of labels indicating which triangles are adjacent to each other by sharing a common point.

The edge adjacency map is built is a similar manner, but the indexing function acts upon edges. These edges are the pair of vertices in the point cloud that represent the edge of a triangle. To enable the indexing function to work, the equivalence of edges has to be defined. Edges are equal when each edge contains the same vertices and the direction of the edge is unimportant. The point and edge adjacencies for all levels in the mesh are generated after the mesh has been labelled and is stored in the mesh definition file.

2.2.5 Optimisation of the mesh to equal area triangles

The method discussed so far has largely followed the algorithm of Teanby (2006) with the refinement of using quad tree and point cloud data structures, the labelling scheme of Kunszt et al. (2001) and pre-calculating point and edge adjacencies.

However, due to the triangle subdivision process only creating new vertices on great circle diameters, the triangles in the mesh are not equal area. Figure 2.2.5 shows that, for a level 3 mesh, the area of the triangles can vary by almost 30%. This section describes a method of equalising the areas of the triangles in the mesh. In three dimensional Cartesian coordinates, where the vertices of a triangle are defined by the 3 vectors $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$, the area of the triangle is defined to be:

$$A = \frac{1}{2}|\boldsymbol{AB} \times \boldsymbol{BC}| \tag{2.2.2}$$

If the vectors $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ connect vertices in the point cloud and ABC is a triangle defined in the quad tree, then Equation 2.2.2 suggests that, if the distances between the vertices in the point cloud and their nearest neighbours are equalised, then the triangles in the mesh will have equal area. Note that Equation 2.2.2 is the area of a planar triangle, rather than a spherical triangle. This is a consequence of applying the area equalisation algorithm which will cause triangle vertices to move away from great circle diameters and so the triangles will cease to be spherical triangles.

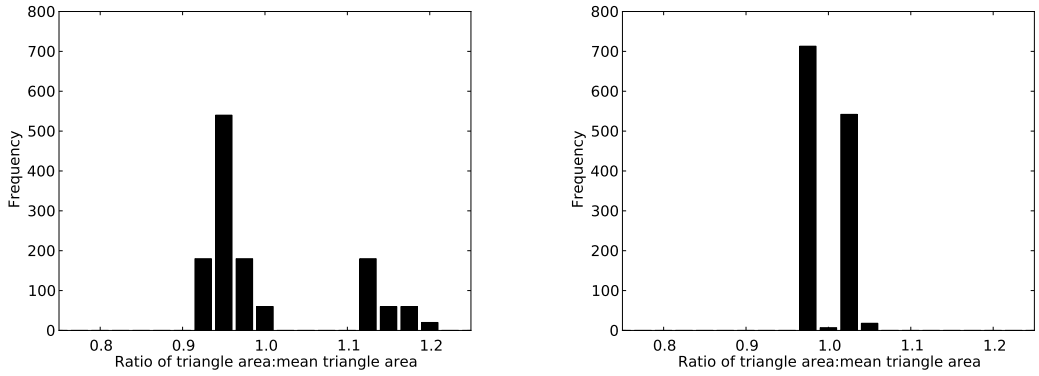


Figure 2.2.5: (a) Frequency of the ratios of triangle area to mean triangle area in a level 3 mesh after subdividing from a level 0 mesh, and (b) the frequency of the ratios after applying 2000 iterations of the equalisation algorithm.

This equalisation is performed by an algorithm that iteratively applies a force vector to each vertex in the point cloud. This vector is a force similar to Coloumb's law of electrostatic repulsion in that it repels the vertices away from each other with a force that is inversely proportional to the distance between them. The force for a single vertex is defined to be:

$$\boldsymbol{F_i} = nG \sum_{j=0}^{j=n} \frac{\boldsymbol{V_i} - \boldsymbol{V_j}}{|\boldsymbol{V_i} - \boldsymbol{V_j}|^2} \ , \qquad i \neq j \tag{2.2.3}$$

where $\boldsymbol{F_i}$ is the force vector, n is the number of vertices in the point cloud, G is a constant of repulsion and $\boldsymbol{V_i}$ and $\boldsymbol{V_j}$ are vertices in the point cloud. Rewriting $\boldsymbol{V_i} - \boldsymbol{V_j}$ as a unit vector $\widehat{\boldsymbol{V}}_{\boldsymbol{ij}}$, then $\widehat{\boldsymbol{V}}_{\boldsymbol{ij}}|\boldsymbol{V_i} - \boldsymbol{V_j}| = \boldsymbol{V_i} - \boldsymbol{V_j}$, and equation 2.2.3 becomes:

$$\boldsymbol{F_i} = nG \sum_{j=0}^{j=n} \frac{\widehat{\boldsymbol{V}}_{\boldsymbol{ij}}}{|\boldsymbol{V_i} - \boldsymbol{V_j}|} \ , \qquad i \neq j$$

There is only a linear dependence on the distance between the two vectors. This has the effect, when compared to the inverse squared relationship of Coloumb's law, of reducing the force vector per iteration.

Adding the force vectors calculated in Equation 2.2.3 to all the vertices in the point cloud, i.e. for $i = (0 \ldots n)$, will move the vertices away from each other, the maximum distance

moved is constrained by the constant G. Performing this iteratively will eventually lead to the vertices settling into a steady state, where the force vectors from the other vertices in the point cloud cancel out. Ideally, in this state the areas of the triangles will be the same as the mean area of all the triangles and the standard deviation of the triangle areas will be 0.

Figure 2.2.6 shows the convergence of the mesh to a steady state over 2000 iterative applications of the equalisation algorithm. As can be seen the standard deviation does not quite reach 0. This is thought to be due to the imprecision of representing real numbers on a computer as running the algorithm significantly more than 2000 times does not result in a smaller standard deviation as can be seen in Table 2.2.2. The implementation of the algorithms in this thesis use the 32 bit IEEE representation of floating point numbers.

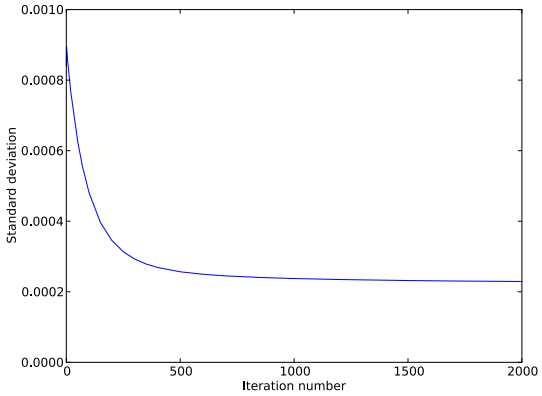


Figure 2.2.6: Standard deviation from the mean area of the triangles over 2000 iterations of applying the area equalisation algorithm to a level 3 mesh.

Iteration number	0	5	10	500	1000	5000	10000
Normalised standard deviation	1.0	1.02655	0.98339	0.30545	0.28271	0.27000	0.27048

Table 2.2.2: Table of normalised standard deviations from the mean area of the triangles in a level 3 mesh at intervals of applying the equalisation algorithm over 10000 iterations.

Although the algorithm does not cause the triangles to have truly equal area, Figure 2.2.5b

shows that the difference in triangle area is significantly smaller after 2000 iterations, when compared to Figure 2.2.5a. In fact, the areas are within ±5% of the mean area and the majority of the triangle areas are within ±2.5% of the mean area.

2.2.6 Locating points on the mesh

To regrid data from a grid defined by latitude and longitude onto the hierarchical triangular mesh two steps are taken. Firstly, the location of the data points are transformed from latitude / longitude coordinates to three dimensional Cartesian coordinates using the standard transformation for spherical to Cartesian coordinates (Kreyszig, 1993), with an adjustment to starting the coordinate system at $\theta = 90°$:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = r \begin{pmatrix} \sin(90-\theta)\cos\lambda \\ \sin(90-\theta)\sin\lambda \\ \cos(90-\theta) \end{pmatrix} \tag{2.2.4}$$

where θ is latitude in degrees, λ is longitude in degrees and r is the radius of the sphere to transform the coordinates to. As the initialisation and subdivision of the icosahedron is defined to always be contained within the unit sphere, $r = 1$.

Secondly, to determine which triangle a datum point is in, a traversal of the quad tree along with a point inclusion test is used. To determine if a point, $\boldsymbol{p}$ is within a triangle, ABC, the following test is used: (Kunszt et al., 2001):

$$\begin{aligned} &(\boldsymbol{A} \times \boldsymbol{B}) \cdot \boldsymbol{p} > 0 \quad \textbf{and} \\ &(\boldsymbol{B} \times \boldsymbol{C}) \cdot \boldsymbol{p} > 0 \quad \textbf{and} \\ &(\boldsymbol{C} \times \boldsymbol{A}) \cdot \boldsymbol{p} > 0 \end{aligned} \tag{2.2.5}$$

To find which triangle the datum point is in, the point inclusion test is firstly applied to each of the triangles represented by the root nodes of the 20 quad trees, i.e. the 20 faces of the initial icosahedron. This determines which quad tree the triangle containing the datum point will be in. The point inclusion test is then applied to the children of the root node to

determine which branch of the tree to follow. This is repeated until a leaf node is reached. The triangle represented by this leaf node contains the datum point and the label of the triangle can be read out of the node data structure.

2.2.7 Storing data on the mesh

Storing data that is to be regridded on the mesh could be performed by adding a datum field to each node, with the algorithm in section 2.2.6 determining which node the datum should be stored in, given the point at which the datum is located. However, to store more than one field (and more than one timestep) of data would require the duplication of the tree, leading to unnecessary memory usage. A more efficient way to store the data is to use a non-clashing hash table. This is similar to the hash table discussed in section 2.2.4 but guarantees that the location for each key is unique, i.e. that there is a one to one mapping between key and location, whereas a clashing hash table has a many to one mapping.

To store data on the hierarchical triangular mesh, a one dimensional array is used, referred to as the hierarchical triangular mesh array (HTMA). The HTMA is analogous to the linear quad tree used by Lee and Samet (2000). This array stores the data, its the location in the hash table, whereas the key is given by the label of the node in the quad tree which references a triangle. The number of elements necessary in this array to store data for every triangle in the mesh at all levels is given by the following formula:

$$N_e = \sum_{i=0}^{l} F4^i \tag{2.2.6}$$

where N_e is the number of elements in the HTMA, F is the number of faces in the initial shape (20 for the icosahedron) and l is the number of levels in the mesh.

The hashing function (Sedgewick, 1992), which determines the mapping from key to location in the HTMA, uses the label of a triangle, as described in section 2.2.3, as keys. Treating the

label L as an array of characters, the hashing function can be summarised by the following:

$$I = \Big(\sum_{i=0}^{l} 4^i\Big)\big(\Lambda(L[0]) - \Lambda(\text{“A”})\big) + \sum_{i=1}^{l}\Big\{1 + \Big(\sum_{j=0}^{i} 4^j\Big)\big(\Lambda(L[i]) - \Lambda(\text{“1”})\big)\Big\} \tag{2.2.7}$$

where I is the index into the HTMA, l is the number of levels in the mesh, the operator $[\ \ ]$ retrieves a single character from the label string and the function Λ converts a single character to a number. The HTMA is organised as a series of “blocks” of contiguous memory, one block containing the data for one of the 20 original triangles in the icosahedron and its children. The first term in the sum of equation 2.2.7 locates the beginning of the block by determining the index of the array the original icosahedron, identified by the letter part of the triangle label. Here, “A” is the first label and will have index 0, all other labels will be offset from “A” by a multiple of $\sum_{i=0}^{l}(4^i)$. The second part of the sum identifies the index of the child (or grand–child, etc.) of the original triangle. Again, the HTMA is organised as blocks of contiguous memory containing the triangle and all its ancestors.

Figure 2.2.7 shows how the data is arranged in a hierarchical triangular mesh array for a level 3 mesh. The hashing function allows for data to be stored on the mesh when the label of the triangle that the data point is located in is known. Determining the label for a data point is performed by applying the algorithm in section 2.2.6. Note that it is possible to store data for all levels in the mesh.

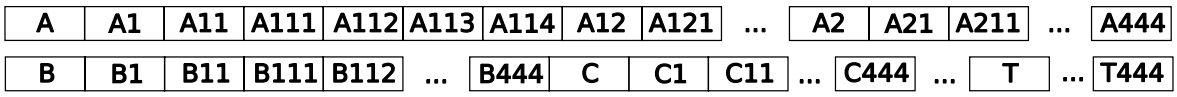


Figure 2.2.7: Order of data storage in the hierarchical triangular mesh array for the nodes in a level 3 mesh that represent the labelled triangles. Not all nodes are shown.

To enable the HTMA to store a timeseries of data on the hierarchical mesh, a two–dimensional array is used, with the number of first elements being equal to the number of timesteps and the number of second elements being the HTMA as above. In effect, there is one HTMA per timestep.

2.2.8 Regridding regularly gridded data onto the mesh

Sections 2.2.6 and 2.2.7 present a method for determining where a data point lies on the mesh and a data structure for storing data on the mesh. These methods can be applied to regrid data output from a climate model that has a grid defined in latitude and longitude spacing.

For a grid defined by $S = (\phi_S, \lambda_S)$, $E = (\phi_E, \lambda_E)$ and $\Delta = (\delta_\phi, \delta_\lambda)$ where S indicates the start of the grid, E is the end of the grid, Δ is the spacing of the grid, ϕ is the latitude and λ indicates the longitude of the grid, the data on that grid can be defined as:

$$\begin{aligned} D &= \{d_{\phi,\lambda}\} \\ \phi &\in \left\{\phi_S + n\delta_\phi, n \in \mathbb{I} : 0 \leq n \leq \frac{\phi_E - \phi_S}{\delta_\phi}\right\} \\ \lambda &\in \left\{\lambda_S + m\delta_\lambda, m \in \mathbb{I} : 0 \leq m \leq \frac{\lambda_E - \lambda_S}{\delta_\lambda}\right\} \end{aligned} \tag{2.2.8}$$

Regridding the data to the triangular mesh uses an area weighted averaging method:

- Create a hierarchical triangular mesh array (HTMA), as defined in section 2.2.7, where each element of the array contains two items, the value and the weight.
- Loop over every ϕ and λ in D and determine which triangle on the mesh the point lies in by using the method in section 2.2.6.
- For each of the data points:
  - Find the element in the HTMA which corresponds to the label of the triangle by applying the hashing function of section 2.2.7.
  - Calculate the weight (W) that the data point should contribute to the value for the triangle. The weight is the ratio of the grid box area in the original grid to the triangle area. Add the weight to the weight item of the element in the HTMA.
  - Add the data point value $d_{\phi,\lambda}$ multiplied by the weight W to the value item of the element in the HTMA.

  - The triangle found is at a leaf node but the data can be propagated up the tree by finding the triangle's parent node in the quad tree, applying the hashing function to its label, calculating the weight and adding the weight and value to the element in the HTMA found by the hashing function. This can be continued until the root node of the tree is reached, thus regridding the data at a number of different resolutions.

- Now that each data point in the original grid has been assigned to a triangle, via the HTMA, loop over every element in the HTMA and divide the sum of the values by the sum of the weights.

The weight W, in the algorithm above, is calculated as the ratio of the surface area of the box where the data point occurs in the original grid to the surface area of the triangle in the mesh. The surface area of a grid box at latitude ϕ the grid defined by Equation 2.2.8 is:

$$\begin{aligned} A_d &= \delta_\lambda \left\{ \sin\left(\phi - \frac{\delta_\phi}{2}\right) - \sin\left(\phi + \frac{\delta_\phi}{2}\right) \right\} \\ &= 2\delta_\lambda \cos(\phi) \sin\left(\frac{\delta_\phi}{2}\right) \end{aligned}$$

For small values of δ_ϕ, $\sin(\delta_\phi) \approx \delta_\phi$, and so the above equation can be rewritten as:

$$A_d = 2\delta_\lambda \delta_\phi \cos(\phi) \tag{2.2.9}$$

From Equation 2.2.2 the area of a planar triangle is:

$$A_T = \frac{1}{2}|\boldsymbol{AB} \times \boldsymbol{BC}|$$

where $\boldsymbol{AB}$ and $\boldsymbol{BC}$ are vectors representing two sides of the triangle. Then, the weight W can be defined to be:

$$W = \frac{A_d}{A_T} \tag{2.2.10}$$

Although the regridding method above will regrid the data onto a triangular mesh, errors in the regridding can occur. These fall into two groups. Firstly, a triangle that is not assigned

a value due to it falling between the spacing in the original grid. That is, there are triangles in the mesh that no data points fall into, when transformed using Equation 2.2.4.

Secondly, a data point may not be assigned to a triangle at a leaf node, due to the condition in Equation 2.2.5 not being met, even though the condition is met for a leaf node's parent triangle. During the grid creation process, the triangles are created as spherical triangles with vertices at the 3 points where 3 great circle diameters intersect. During the equalisation phase, the points are free to move away from these intersections, and so the vertices can now lie on up to 6 great circle diameters. This causes problems for Equation 2.2.5 as it only holds for spherical triangles (Kunszt et al., 2001), that is, triangles defined as having points at the intersection of 3 great circle diameters.

The first error can cause holes in the regridded data to appear due to triangles in the mesh not having data assigned to them. This error is overcome by using the edge adjacency list derived in section 2.2.4 to get the values of the surrounding triangles and assign the mean of these values to the triangle with missing data.

The second error does not have as large an effect on the regridded data as the first error. However, it does lead to the regridded data not being as representative of the original data as it would be if all data points were included in the regridding. Also, the more a mesh is optimised towards a mesh with truly equal area triangles, the more frequently this error occurs.

To ensure that all data points are used in the regridding algorithm, an additional, slower method is used to determine which triangle in the mesh the point is closest to. Firstly, Equation 2.2.5 is used to find which of the 20 root node triangles the point lies in. The root node triangles are all equal area and spherical triangles by the nature of their creation and so Equation 2.2.5 is still usable. Then the tree is traversed to all of the leaf nodes, i.e. the triangles at the highest level of the mesh, and the distance between the triangles and the point is calculated. The point is assigned to the triangle which has the shortest distance to the point, that is, the triangle that is closest to the point. The distance between a triangle

and point is given by:

$$d = |\boldsymbol{P} - \boldsymbol{C_T}|$$

where $\boldsymbol{P}$ is the point and $\boldsymbol{C_T}$ is the centroid of a triangle. Using the algorithm outlined above, data output from a climate model, or from a different source such as satellite retrieval data, can be regridded to the hierarchical triangular mesh.

2.3 Objective feature detection on the mesh

Detecting the storm–indicative features in the regridded data deviates considerably from Massey (2012), where a simple threshold test and local extrema search were used to determine whether a triangle contained a feature point. This method was found to be problematic when applied to reanalysis data and climate–model output as it did not correctly identify winter storms in the Mediterranean. It is thought that was due to using a simple threshold on the mean sea–level data (MSLP), below which value a cyclone (low pressure system) would be detected. The method proved sensitive to the threshold value and, a threshold value which successfully allowed the detection of MSLP minima in Northern Europe would be too low for detecting minima in Southern Europe. In the revised method presented here, the background flow of the MSLP field is removed and minima are detected as negative anomalies to this field.

Anderson et al. (2003) show that storm–tracking methods are sensitive to the type of background field removed prior to feature identification and that using the hemispherical mean pressure as a background field to subtract before the search for local MSLP minima leads to a lower density of features. This can be improved by using a time–averaged mean (climatology) at the expense of weakening the low MSLP features, due to the features themselves contributing to the climatological mean. A further improvement can be made by removing the large scale flow from the field before identifying the local features, as the density of features remains similar to that if the time–averaged mean is removed, but the intensities of the identified low MSLP features are not reduced. Anderson et al. (2003) uses a spectral spatial filter to remove the large scale flow. However, the hierarchical nature of the triangular mesh

L	0	1	2	3	4	5	6	7	8	9	10
D	4650	2213	991	480	238	118	59	29	14	7	3
A	19431357	5447253	1339579	326373	80204	19859	4939	1232	308	77	19

Table 2.3.1: Distance between the centroid of adjacent triangles and surface area (A) of triangles at different levels of the triangular grid. Distance is given in km, to the nearest 1 km. Surface area is given in km^2, to the nearest 1 km^2.

allows data at a lower resolution of the mesh to be used as the background field, for example data at level 3 can be removed to identify features at level 6 in the mesh. Table 2.3.1 shows the surface area and spatial scale of triangles at various levels in the mesh.

Using a background field, rather than partitioning the data using a threshold value, allows the perturbations to the mean flow to be identified as features. This means that the spatial structure of the field, such as the Aleutian low and Atlantic low in MSLP, or the equator to pole gradient in temperature, is taken into account before any feature identification is performed. The high spatial resolution of regional data can be problematic when locating local extrema as multiple features can be identified that actually belong to the same feature. For example, multiple low MSLP features that belong to the same extra–tropical cyclone. Other tracking methods (Murray and Simmonds, 1991; Sinclair, 1994; Hodges, 1994; Benestad and Chen, 2006) apply a filter to effectively reduce the spatial resolution before the feature search. However, this has the effect that the footprint of the feature cannot be recovered. The revised feature identification method presented here uses three levels of the hierarchical mesh to identify features. Firstly, the background field, occurring at a low level of the mesh (*the background field level*) is removed . Secondly, features are identified at a spatial resolution lower than the resolution of the original data. This is done by identifying features at a higher level than the background field, but a lower level than the maximum level in the mesh. This is *the identification level*. Finally, these features are refined to a higher spatial resolution, that is, the highest level of the mesh (*the refinement level*). As in Massey (2012) the method identifies feature objects, rather than a single feature point. A feature object is a collection of triangles on the mesh that are deemed to belong to the same low MSLP system and this definition is also used in the tracking routine.

2.3.1 Background field removal

The feature detection process commences with the removal of the background field. This can be thought of as the large scale flow, and can be approximated from numerous sources, including time averaged fields, such as the climatological monthly mean for a timeseries of observations (Anderson et al., 2003), or the ensemble mean of a large ensemble climate modelling experiment. In this revised algorithm, the background field is taken to be the data at a lower mesh level in the regridded data (*the background field level*). Due to the way the triangular mesh is structured, and how the regridding algorithm works, this data is a low resolution version of the data at the identification level. Using a regridded MSLP field, perturbations to this background flow can be detected by subtracting the data at the *background field level* from the data at the *identification level*. Negative values in the resulting data can then be identified as cyclones.

Simply subtracting the value in a triangle n–levels lower than the identification level, where n = *identification level - background field level*, and the lower triangle is found by taking the parent of the *identification level* triangle recursively n times, produces sharp edges in the resulting data. These sharp edges occur at the edge of the triangles in the *background level* and lead to the erroneous identification of cyclones along the edges of those triangles. In order to avoid this edge–effect, the data at the *background level* must be smoothed to the *identification level*.

Smoothing commences by creating a new HTMA (Hierarchical Triangular Mesh Array – section 2.2.7), with the same size as the regridded data, and copying the regridded data at the *background field level* into the HTMA. Then, for each triangle in this level, the four child triangles are considered. Three of these triangles will occur with points at the corners of the parent triangle (triangles a to c), whereas one triangle will occur with only the mid-points of the parent triangle (triangle d). For triangles a to c, the triangles in the *background level* which share a point with these triangles in the *background level + 1* level are found. Triangle a is then assigned the mean of the values of the triangles which share a point with it, and the same is done for triangles b and c. Triangle d is assigned the value of the parent triangle.

This procedure is repeated for subsequent levels, all the way up to the *identification level*, and is performed for each timestep in the regridded data. Figure 2.3.1 shows the process of this smoothing for one level.

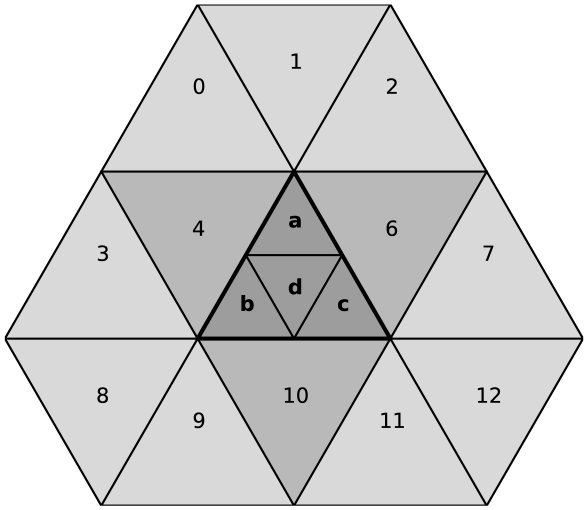


Figure 2.3.1: Smoothing procedure applied to triangles (a, b, c, d) at level L in the hierarchical triangular mesh. Triangles 0 to 12 occur at level L-1 in the mesh. Triangle a is assigned the mean of triangles 0,1,2,4,5,6, triangle b is assigned the mean of triangles 3,4,5,8,9,10 and triangle c the mean of triangles 5,6,7,10,11,12. Triangle d is assigned the value of triangle 5.

At the end of this procedure, a new HTMA contains the smoothed data at the levels between the *identification level* and the *refinement level*, for each timestep in the regridded data. Another HTMA is then created containing the result of subtracting the smoothed data from the regridded data, for all levels and timesteps. This HTMS is then contoured – i.e. the data is rounded to the nearest H hPa. The feature point detection algorithm then works upon this HTMS which contains the anomalies from the background field.

2.3.2 Feature object detection

Once the background field has been removed, features can be detected as anomalies from the mean flow, by using one of the adjacency lists to consider the triangles surrounding each triangle at a particular level in a mesh. The user selects whether they wish to consider point adjacencies or edge adjacencies. Data regridded to the hierarchical triangular mesh is

available at a number of spatial resolutions at the different levels in the grid and this enables the user to select the spatial scale that they wish to initially identify features at, by selecting the level of the mesh at which to start the process. At this level, features are identified by comparing the value for this triangle in the HTMA with the values of the adjacent triangles. If the triangle value is less than the surrounding triangle values and the value is also less than a user defined maximum difference from the mean flow, then the triangle is considered as a feature. To create the feature object, the labels of the descendants of the initial triangle ($\mathbf{T_i}$) at the *refinement level* in the mesh are added to a list and the location of the object is set to be the centroid of $\mathbf{T_i}$.

Not all of the triangles at the *refinement level* that are added to the list of labels should belong to the object, as they will have values that are smaller than the largest anomaly from the mean flow. To account for this, the largest anomaly in the object $\left[\underset{anom}{\mathbf{max}}\right]$ is found by examining the value (in the HTMA) of each triangle in the list of triangle labels in the object. Triangle labels are then removed from the list of labels if the value of the triangle is not equal to $\left[\underset{anom}{\mathbf{max}}\right]$.

At this stage, the feature object identified will not extend beyond the bounds of the triangles at the *identification level*, even though the feature itself may extend into triangles that are bound by other triangles at the *refinement level*. To counteract this, an object growing routine is employed which begins by constructing the *concentric shell* of the object. The concentric shell is a list containing the labels of all the triangles adjacent to the object and is constructed from the edge adjacency list of the individual triangles in the object. For each triangle in the object ($\mathbf{T_o}$) if a triangle label in the adjacency list of $\mathbf{T_o}$ is not in either the list of labels in the object or in the list of labels of the current concentric shell, then it is added to the concentric shell.

Using the concentric shell, the feature object is grown by examining the values of the triangles referenced by the labels in the concentric shell. If a triangle value is within one contour level of $\left[\underset{anom}{\mathbf{max}}\right]$ then it is added to the feature object by appending its label to the list of triangle labels within the object. An added restriction here is that objects are not allowed to grow

to be more than 1000km in radius. This has the consequence that non–closed systems will have an artificial closure 1000km away from the triangle with $\left[\mathbf{max}_{anom}\right]$.

This approach to feature detection and object growing is analogous to the connected component labelling (CCL) of Samet (1990) scheme employed by Hodges (1994). Hodges (1994) uses a quad tree to decompose gridpoint data into a hierarchy of "blocks" of gridpoints at different effective resolutions. This quad tree is then used to label features according to how connected they are and gridpoints outside the initial search space which are also connected to the feature are added through region growing. The algorithm presented is this thesis is largely similar, but uses a hierarchical triangular representation rather than the "blocks" of Hodges (1994).

After the object growing routine has completed, the identified objects are merged. Merging is performed by comparing the lists of triangle labels in each object. Overlapping objects are defined as satisfying at least one of the following conditions:

- A label in object $\mathbf{O_1}$ is also contained in object $\mathbf{O_2}$
- A label in the concentric shell of $\mathbf{O_1}$ is contained in object $\mathbf{O_2}$
- A label in object $\mathbf{O_1}$ is contained in the concentric shell of $\mathbf{O_2}$
- A label in the concentric shell of $\mathbf{O_1}$ is contained in the concentric shell of $\mathbf{O_2}$

Objects are merged by appending the triangle labels of $\mathbf{O_2}$ to those of $\mathbf{O_1}$ and deleting the $\mathbf{O_2}$ object.

2.3.3 Deriving the feature point from the object

The feature tracking algorithm in section 2.4 requires a single point, as well as the object definition, to track the evolution of the feature objects across multiple timesteps. To derive

this point a weighted average of the locations of the triangles in the object is used:

$$\begin{aligned} \overline{\boldsymbol{P}} &= \frac{1}{\sum_{i=0}^{n} w_i} \sum_{i=0}^{n} w_i \boldsymbol{P_i} \\ w_i &= 1 - \frac{V_i}{\left[\mathbf{max}_{\boldsymbol{anom}}\right]} \end{aligned}$$

where $\overline{\boldsymbol{P}}$ is the central feature point of the object, $\boldsymbol{P_i}$ is the location of the centroid of the ith triangle in the object, V_i is the value of the ith triangle in the object, n is the number of triangles in the object and $\left[\mathbf{max}_{\boldsymbol{anom}}\right]$ is the maximum anomaly in the object. The point $\overline{\boldsymbol{P}}$ is then converted to latitude and longitude coordinates using the standard transformation. The object also records the intensity of the MSLP low, which is set to be $\left[\mathbf{min}_{\boldsymbol{MSLP}}\right]$, the minimum MSLP value in the object, and the delta of the MSLP low, which is $\left[\mathbf{min}_{\boldsymbol{MSLP}}\right] - \left[\mathbf{max}_{\boldsymbol{MSLP}}\right]$. Figure 2.3.2 provides a graphical summary of the feature detection routine.

2.4 Feature tracking

After the feature identification routine of section 2.3 has been applied to the regridded data, a list of feature objects exist for each timestep in the regridded data, each object containing a central feature point in latitude–longitude coordinates. It is the job of the feature tracking algorithm to associate these objects across the timesteps so as to track the progress of the features as they form, move and dissipate within the region. This is performed by creating an initial set of tracks, each containing one of the feature objects occurring in the first timestep. For subsequent timesteps, feature objects are assigned to the existing tracks based on inclusion criteria. If a feature object cannot be assigned to an existing track, then a new track is created. The revised feature tracking method in this thesis deviates from that in Massey (2012), which required the assigning of weights to four terms in a cost function. In practice, deriving values for these weights proved difficult, with different values required for different regions. Also, the algorithm was prone to erroneous assignment of objects to tracks, as the inclusion criteria was based on a single value for the cost. For example, an object may have been assigned to a track due to its proximity to the previous object in the track, in

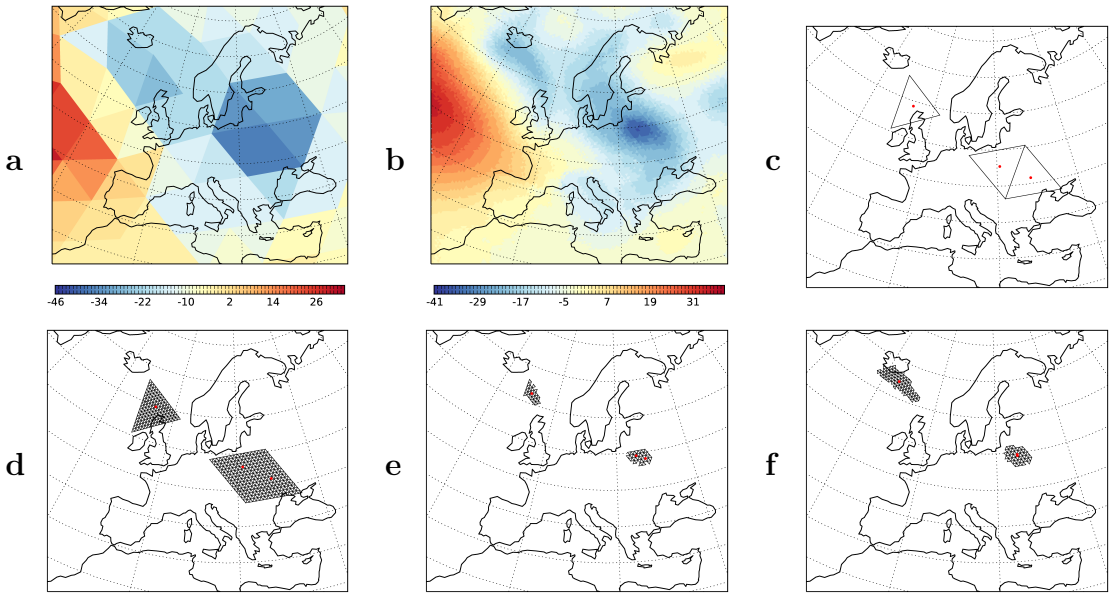


Figure 2.3.2: Illustrative depiction of feature identification routine. **(a)** MSLP anomaly from the background field at a single timestep at mesh level 3. **(b)** MSLP anomaly from the background field at mesh level 7. **(c)** MSLP minima detected at mesh level 3. **(d)** Feature object composed of triangles at level 7 descended from those triangles in **c** identified as minima at level 3. **(e)** Refinement of the triangles in the feature object in **d** by discarding triangles whose values are not within one contour level of the maximum anomaly in the object. **(f)** Feature objects after the growing of the objects to encompass all triangles whose values are within one contour level of the maximum anomaly. The objects are also merged.

spite of the new object being in the opposite direction to the tracks previous movement. To counteract this, the new method uses a series of hierarchical rules to assign object to tracks.

2.4.1 Geostrophic wind as a steering vector

Using a steering vector to predict the motion of a feature based on physical characteristics of the field can reduce the number of erroneous point to track assignments in a tracking algorithm (Murray and Simmonds, 1991). Here the geostrophic wind, derived from the

geopotential height in spherical coordinates, is used. It is defined by:

$$u_g = \frac{g}{-fR}\frac{\partial z}{\partial \phi} \qquad (2.4.1)$$
$$v_g = \frac{g}{fR\cos(\phi)}\frac{\partial z}{\partial \lambda}$$
$$f = 2\Omega\sin(\phi)$$

where u_g and v_g are the geostrophic wind components, g is the acceleration due to gravity, f is the Coriolis force at a latitude, R is the radius of the Earth, z is the geopotential height, ϕ is the latitude and λ is the longitude. To approximate the geostrophic wind from the geopotential height field of regional climate data, a spline is fit through the latitude–longitude point and the partial derivatives are estimated from the derivatives of the spline.

The geopotential height at 700hPa is used to calculate the geostrophic wind throughout this thesis and is used instead of the model output of the U and V winds for two reasons. Firstly, the geostrophic wind represents a smoothed version of the U and V winds, with fewer local disturbances caused by topography. Therefore the geostrophic wind represents the large–scale flow of the atmosphere. Secondly, geostrophic wind can be calculated from a single climate variable, whereas the U and V wind components are two variables. This reduces the bandwidth and storage burden when carrying out climate model runs via volunteer distributed computing, as is done in chapter 4 in this thesis. The full winds are used, rather than a fraction of the wind, as the tracking rule associated with the geostrophic wind (the steering rule, see section 2.4.2) uses only the bearing between the geostrophic wind and the current point, rather than a projection of distance, to determine whether a feature point is a good candidate to add to a track.

2.4.2 Tracking rules

The rules to determine whether a feature object should be added to an existing track are based on five criteria:

1. The distance from the end of a track to the candidate object.

2. The amount the last object in the track overlaps the candidate object.

3. The angle between the vector formed from the last object in the track and the candidate object and the vector formed from the last object in the track and the point predicted from the steering vector.

4. The difference in intensity between the last object in the track and the feature object.

5. The change in direction of the track if the candidate point is added.

Distance rule

The user can specify a maximum distance between the last object in a track ($\mathbf{O^{t-1}}$) and the candidate object ($\mathbf{C^t}$), which should be based upon the number of hours per timestep in the source data (h) and how far the object is expected to move within that time period. This term is named the search radius. Extra–tropical cyclones in the Northern Hemisphere have a mean propagation speed of up to $30ms^{-1}$ (Hoskins and Hodges, 2002), and so it is suggested to set this search radius to be upwards of $125h$km.

An addition to the revised method is that an adaptive constraint (Hodges, 1999) on this search radius is calculated from the geostrophic wind steering vector. If the magnitude of this vector multiplied by h is greater than the user input search radius, then the search radius is set to be this value. The distance between the last track object and the candidate track object is defined to be:

$$\Delta = \mathcal{H}(\mathbf{C^t}, \mathbf{O^{t-1}}) \tag{2.4.2}$$

where $\mathcal{H}$ is a function that calculates the haversine distance (Sinnott, 1984) between $\mathbf{O^{t-1}}$ and $\mathbf{C^t}$. This measures the distance between two points (ϕ_1, λ_1) and (ϕ_2, λ_2), both in latitude and longitude coordinates, by travelling over the skin of the sphere, rather than through it,

and is given by:

$$\begin{aligned} \triangle\phi &= \phi_2 - \phi_1 \\ \triangle\lambda &= \lambda_2 - \lambda_1 \\ a &= \sin^2\left(\frac{\triangle\phi}{2}\right) + \cos(\phi_1)\,\cos(\phi_2)\,\sin^2\left(\frac{\triangle\lambda}{2}\right) \\ d &= 2R\arcsin(\sqrt{a}) \end{aligned} \tag{2.4.3}$$

Overlap rule

An advantage of identifying feature objects, rather than single feature points, is that their spatial characteristics can be used during the tracking phase. In the Northern Hemisphere, slow moving cyclones that are prevented from moving eastwards by high–pressure blocking systems may vacillate around a common area (Hodges, 1999). During this vacillation, the central point of the object may change direction sharply, so that the steering rule and curvature rule are not met. However, the triangles in $\mathbf{O^{t-1}}$ may overlap with triangles in $\mathbf{C^t}$. If more than a user specified percentage of an object's triangles overlap with the triangles in $\mathbf{O^{t-1}}$, then the object has passed the rule. In order to assess this rule in terms of minimising a cost function the rule is expressed as:

$$\beta = 100 - \mathcal{O}(\mathbf{C^t}, \mathbf{O^{t-1}}) \tag{2.4.4}$$

where $\mathcal{O}$ is a function that measures the percentage of overlap between the candidate object $\mathbf{C^t}$ and the object $\mathbf{O^{t-1}}$.

The threshold for overlap is selected by the user. Throughout this thesis, a threshold of 10% is used, i.e. objects have to share at least 10% of the same triangles from one timestep to the next to be considered overlapping. This tracking rule is primarily targeted at slow moving systems which may fail the curvature or steering rule, as mentioned above. Small systems with very sharp pressure gradients and fast moving systems are not likely to overlap, but will instead rely on the curvature and steering rules for track assignment.

Steering rule

The steering rule uses the geostrophic wind (Equation 2.4.1) and the central point of $\mathbf{O^{t-1}}$ to predict the position of the next object in the track, in latitude–longitude coordinates:

$$\begin{aligned} P_\lambda^t &= O_\lambda^{t-1} + \frac{360 u_g N_H}{F_\phi} \\ P_\phi^t &= O_\phi^{t-1} + \frac{180 v_g N_H}{F_R} \\ F_\phi &= 2\pi E_R \cos(\phi) \end{aligned} \tag{2.4.5}$$

where $P_{\lambda,\phi}^t$ is the projected latitude–longitude point, $O_{\lambda,\phi}^{t-1}$ is the central point of the last feature object in a track ($\mathbf{O^{t-1}}$), u_g, v_g is the geostrophic wind vector from Equation 2.4.1, F_R is the meridional circumference of the Earth and F_ϕ is the zonal circumference of the Earth at a latitude ϕ, E_R is the radius of the Earth and N_H is the number of seconds per timestep, i.e. for 6 hourly data $N_H = 6 \times 60 \times 60$. These equations use the u_g and v_g component of the geostrophic wind, which are expressed in Cartesian coordinates and have units of ms^{-1}. These wind components are converted into degrees latitude and degrees longitude per timestep and a projection of the expected feature position in the next timestep is made.

Once the point has been projected, the difference in the bearing between the latitude–longitude coordinate of the central point of the candidate object ($C_{\lambda,\phi}^t$) and $O_{\lambda,\phi}^{t-1}$ and the bearing between $P_{\lambda,\phi}^t$ and $O_{\lambda,\phi}^{t-1}$ is measured. The bearing is defined as (Jennings, 1994):

$$\begin{aligned} c &= \sin(\Delta\lambda)\cos(\phi_B) \\ d &= \cos(\phi_A)\sin(\phi_B) - \sin(\phi_A)\cos(\phi_B)\cos(\Delta\lambda) \\ \angle(A_{\lambda,\phi}, B_{\lambda,\phi}) &= \mathrm{atan2(c, d)} \end{aligned} \tag{2.4.6}$$

and so the cost function for the difference in bearings for the steering rule becomes:

$$\Psi = \angle(O_{\lambda,\phi}^{t-1}, C_{\lambda,\phi}^t) - \angle(O_{\lambda,\phi}^{t-1}, P_{\lambda,\phi}^t) \tag{2.4.7}$$

Objects with bearing differences of $-90°$ to $90°$ are considered to have passed the rule. This ensures that features that are moving in approximately the same direction as the geostrophic wind, rather than against the wind, are considered as candidates. Due to the local optimisation procedure, the candidate which has the lowest steering rule cost will be assigned to the track, if no rules higher in the hierarchy are passed.

Intensity rule

The intensity of an object (in this case the value of the MSLP minimum associated with the object) can be a useful metric when tracking, as it acts as a fingerprint of the object. MSLP minima are likely to change between timesteps but, as long as the temporal resolution of the input data is sufficiently high (such as 6 hourly), the change is likely to be small. This metric is especially useful when tracking very low depressions where other, smaller, depressions are introduced into the domain. Using the intensity rule will allow the correct allocation of the very low depression to the track containing the path of the depression, even if allocating the smaller depressions would meet some of the rules. The intensity cost is defined as:

$$\chi = \left|\mathcal{I}(\mathbf{C^t}) - \mathcal{I}(\mathbf{O^{t-1}})\right| \tag{2.4.8}$$

where $\mathcal{I}$ is a function that returns the intensity value of the object, as described in section 2.3.3. Objects with intensity differences less than or equal to 100hPa are considered to have passed the rule.

Curvature rule

The curvature rule aims to maximise smoothness within tracks. Like the steering rule, it uses the definition of the bearing given in Equation 2.4.6. The curvature between a candidate point and the track is then measured as:

$$\mathcal{C} = \angle(O^{t-1}_{\lambda,\phi}, C^{t}_{\lambda,\phi}) - \angle(O^{t-2}_{\lambda,\phi}, O^{t-1}_{\lambda,\phi}) \tag{2.4.9}$$

where $O_{\lambda,\phi}^{t-2}$ is the central point of the penultimate object in the existing track. From this measure of the curvature, and the distance measure of Equation 2.4.2, the curvature cost is defined to be:

$$\Xi = w_1 \left(\frac{\mathcal{C}}{90} \right) + w_2 \left(\frac{\mathcal{D}}{sr} \right) \tag{2.4.10}$$

where $\mathcal{D}$ is the distance cost, w_1 is the weight of the curvature term in the cost function, w_2 is the weight of the distance term in the cost function and sr is the search radius, as defined in section 2.4.2, in metres. The curvature term $\mathcal{C}$ is divided by 90 so as to normalise the cost for this term to the maximum permitted value of the curvature (90°). Similarly, the distance term $\mathcal{D}$ is divided by the search radius to normalise the cost for this term to the maximum permitted distance between two track points. Assigning weights in a cost function could incur the same problems that beset the weight tuning in Massey (2012). However, the weights here are set to $w_1 = 0.5$ and $w_2 = 0.5$, i.e. each term is equally weighted. This ensures that candidate objects that have a small curvature, but are far away from the last track object, are penalised equally, or more, than those candidate points which are closer but have a larger curvature.

To pass the curvature rule, the curvature cost function must evaluate to less than 1.0. This allows curvatures of up to 180°, with zero distance between the track objects, and distances up to 2000km, with zero curvature between the track objects.

The hierarchy of rules

To remove the reliance on a cost function, the optimisation routine that forms the core of the tracking algorithm uses a hierarchy of rules instead. When an object passes one or more of the rules, and is to be considered a candidate point for adding to a track, the rules it passes are considered in the order of the rule hierarchy. All candidate objects have to pass the distance rule requirement (to ensure that the point is within the search radius) and then the hierarchy is:

5 Curvature rule

4 Intensity rule

3 Steering rule

2 Overlap rule

1 Distance rule

Using the hierarchy has the advantage that, during the optimisation process candidate objects can be compared to each other to see which has passed the highest rule. If two objects have passed the same rule, then the costs for that rule for both objects are calculated and the object with the lowest cost can be assigned.

The order of the hierarchy is based upon two criteria. Firstly, how much information is available for each rule to work upon. This places the curvature rule at the top of the hierarchy, as it uses the most information in the positions of the feature objects in the two previous timesteps. The intensity rule is next, as it uses information directly from the data, and not derived from the data. The steering rule and overlap rule use information derived from the data: the geostrophic wind calculated from the geopotential height for the steering rule and the outer contour of the MSLP anomaly for the overlap rule. Finally, the distance rule is the lowest in the hierarchy as it is the only rule which has to be passed and essentially relieves the computational burden by reducing the number of feature objects that have to be searched through. The second criterion is the smoothness of the tracks, which is also explicitly maximised in the global optimisation routine. The main effect of this criterion is to order the steering rule and overlap rule, with the steering rule being higher in the hierarchy due to its potential to maximise track smoothness.

Which rule a feature object has passed is encoded in a bitfield. This is a single integer where the higher the number the higher the rule passed in the hierarchy. Additionally the encoding allows a logical AND operator to be used to determine which rules the feature object has passed. The rules are encoded in the bitfield as:

- Curvature rule (10_{16})
- Steering rule (08_{16})

- Intensity rule (04_{16})
- Overlap rule (02_{16})
- Distance rule (01_{16})

Where N_{16} indicates that the number is in hexadecimal. For example, a feature object that passes both the curvature and intensity rule would have a hexadecimal value of 14_{16} (20 in decimal). Logical AND with either 10_{16} (curvature) or 04_{16} (intensity) would result in a non-zero answer, whereas logical AND with any other bitfield mask would result in zero. Hence, it can be determined exactly which rules have been passed from a single number.

2.4.3 Assigning objects to tracks

The tracking routine commences with constructing a new track for each feature point in the first time step, i.e. $t = 0$. In subsequent timesteps, objects for that timestep are added to existing tracks based on whether they pass the inclusion rules and subject to a local optimisation protocol. If an object cannot be be added to an existing track then a new track is created, which has that object as the first object in the track.

The local optimisation routine adds feature objects in timestep t to tracks whose last timestep occurs in timestep $t-1$. For each timestep, all feature objects in timestep t are added to a queue of objects. In turn, these objects are popped from the front of the queue and assessed against the tracks that end in timestep $t-1$ ($\mathbb{T}_{0\ldots n}^{a\leq t-1}$), using the rules in section 2.4.2. If an object passes the distance rule and at least one other rule then it is added to track $\mathbb{T}_{i}^{a\leq t-1}$ as the candidate object. It is possible that an object might pass the rules for multiple tracks, in which case it is added to the track for which it minimises the value that passes the highest rule in the hierarchy. For example, if an object passes the curvature rule for more than one track then it is added to the track for which it has the lowest curvature value. To avoid the ordering of objects influencing the tracking outcome (Hodges, 1999) and to produce a locally optimum track, feature objects that are still in the queue are allowed to replace the current candidate object if they are deemed to be a more locally optimal solution. This is done by following the hierarchy of rules for the new object ($\mathbf{N^t}$) and the current candidate

object ($\mathbf{C^t}$).

If, after assessing $\mathbf{N^t}$ against $\mathbf{O^{t-1}}$ (the last point in a track), $\mathbf{N^t}$ passes a rule higher up in the hierarchy than the rule which $\mathbf{C^t}$ passes, then $\mathbf{N^t}$ becomes the candidate object and $\mathbf{C^t}$ is returned to the queue of feature points. If the two objects have both passed the same rule, then the costs for the highest rule are calculated for each object and the object with the highest or lowest value is assigned as the candidate object. For example:

- If both objects have passed the curvature rule then the object with the lowest curvature value in Equation 2.4.9 is assigned as the candidate.
- If both objects have passed the intensity rule then the object with the lowest intensity value in Equation 2.4.8 is assigned.

If the candidate object is replaced it is added back to the queue so that it may be assigned to other tracks. During this optimisation procedure, objects may be assigned to a track, replaced, and then assigned to another track as the procedure continues until there is no more assignment of objects. If an object cannot be assigned to a track (as it does not pass any of the rules, or another object passes the rules with a lower cost) then a new track is started in timestep t, with the feature object as the first object in the track.

2.4.4 Track optimisation

Following the procedure in section 2.4.3 produces a set of tracks of feature objects that are locally optimal at each timestep in the track. However, this approach may encounter several problems which prevent a globally optimal solution to the correspondence problem being found. These problems include the false assignment of a point to a track, particular for the first two timesteps in the track, where the lack of previous track information forces a reliance on the distance, intensity and steering rules (Hodges, 1999). A second problem occurs when a feature object obscures another feature object or the feature detection routine fails to correctly identify one feature. This may cause a track to end at timestep $t-1$ and another track to begin at timestep $t+1$, due to the feature identification routine not identifying an

object at timestep t, even though the track trajectory indicates that the two tracks should be just one track.

To alleviate these problems, the track optimisation routine adds phantom feature objects to the beginning and end of a track and then exchanges sections of overlapping tracks so as to reach a globally optimum set of tracks. Hodges (1996) introduces phantom feature points into his tracking algorithm and here phantom feature objects are added at the beginning and end of tracks by projecting the track backwards and forwards to determine the location of the phantom object For a track $\mathbb{T}_a^b$, that begins at timestep a and ends at timestep b, the forward projection is given by:

$$\begin{aligned} T_\lambda^{b+1} &= T_\lambda^b + (T_\lambda^b - T_\lambda^{b-1}) \\ T_\phi^{b+1} &= T_\phi^b + (T_\phi^b - T_\phi^{b-1}) \end{aligned} \tag{2.4.11}$$

and the backward projection is given by:

$$\begin{aligned} T_\lambda^{a-1} &= T_\lambda^a - (T_\lambda^{a+1} - T_\lambda^a) \\ T_\phi^{a-1} &= T_\phi^a - (T_\phi^{a+1} - T_\phi^a) \end{aligned} \tag{2.4.12}$$

where $T_{\lambda,\phi}^t$ is the projected point at timestep t and $T_{\lambda,\phi}^{t-l}$ is the central point of the feature object $\mathbf{T^{t-l}}$ at timestep $t-l$. Phantom feature objects are identified by having a bitfield representation of FF_{16}, (255 in decimal).

The optimisation process has the advantage over the track assignment routine of section 2.4.3 that it operates on a near optimal set of complete tracks. Many of the tracks will contain more than one timestep and, therefore, a measure of track smoothness over a number of timesteps can be used as the cost function in the optimisation process. This cost is the mean curvature cost of all of the feature objects in the track. For a track, $\mathbb{T}_a^b$, that begins at timestep a and ends at timestep b, this mean curvature cost is:

$$\overline{\Xi} = \frac{1}{b-a} \sum_{i=a}^{i \leq b-2} \Xi(T_{\lambda,\phi}^i, T_{\lambda,\phi}^{i+1}, T_{\lambda,\phi}^{t+2}) \tag{2.4.13}$$

where $\Xi(a, b, c)$ is the curvature cost function of three feature objects, as defined in Equation 2.4.10 and $T^t_{\lambda,\phi}$ is the central point of the feature object $\mathbf{T^t}$ at timestep t.

Also required for the track optimisation is the total track length, defined for the track $\mathbb{T}^b_a$ as:

$$\mathcal{L} = \sum_{i=a}^{i \leq b-1} \mathcal{D}(T^i_{\lambda,\phi}, T^{i+1}_{\lambda,\phi}) \tag{2.4.14}$$

where $\mathcal{D}$ is the distance between two feature objects as defined in Equation 2.4.2

After the algorithm in section 2.4.3 has run, a set of locally optimal tracks exist. The track optimisation process loops over each of these tracks and performs an optimisation in the following manner:

1. The current track, $\mathbb{A}^b_a$, begins in timestep a and ends in timestep b. The mean curvature ($= \overline{\Xi}(\mathbb{A}^b_a)$) and length of the track ($= \mathcal{L}(\mathbb{A}^b_a)$) are found using Equations 2.4.13 and 2.4.14.
2. Phantom feature objects are added to the beginning and end of the track using Equations 2.4.11 and 2.4.12. The resulting track is $\mathbb{A}'^{b+1}_{a-1}$.
3. From $\mathbb{A}'^{b+1}_{a-1}$, a set of track–segments is constructed, where every permutation of a and b is used to produce a segment. A segment of track $\mathbb{A}'^{b+1}_{a-1}$ is here denoted $\mathbb{Q}^d_c$, where $c \geq a-1$ and $c \leq d \leq b+1$.
4. A set of tracks which overlap in time with segment $\mathbb{Q}^d_c$ is found. Tracks are defined as overlapping if they share at least one time step. A track $\mathbb{B}^t_s$ overlaps with segment $\mathbb{Q}^d_c$ if either of these three conditions are met: $c \geq s \wedge c \leq t$ or $d \geq s \wedge d \leq t$ or $c \leq s \wedge c \geq t$, where $\wedge$ indicates logical and.
5. Phantom feature objects are also added to $\mathbb{B}^t_s$ to produce $\mathbb{B}'^{t+1}_{s-1}$.
6. Track segments of $\mathbb{B}'^{t+1}_{s-1}$ are constructed where every permutation of t and s are used. A segment of $\mathbb{B}'^{t+1}_{s-1}$ is here denoted $\mathbb{R}^v_u$ where $u \geq s-1$ and $u \leq v \leq t+1$.
7. If track segments $\mathbb{Q}^d_c$ and $\mathbb{R}^v_u$ can be concatenated together (i.e. $v = c-1$ or $d = u-1$) then they are joined to form a new track $\mathbb{C}^v_c$ or $\mathbb{C}^d_u$, depending on the order of the concatenation. $\mathbb{C}^v_c$ will be used to denote this track from now on.

8. The track length ($= \mathcal{L}(\mathbb{C}_c^v)$) and mean curvature ($= \overline{\Xi}(\mathbb{C}_c^v)$) are calculated and recorded.
9. This procedure is repeated for all possible track–segment combinations of track $\mathbb{A}_a^b$ and all possible overlapping tracks $\mathbb{B}_s^t$. The indices of the segment $\mathbb{Q}(d, c)$, the track number of $\mathbb{B}_s^t$ and the indices of the track–segment $\mathbb{R}(v, u)$ are recorded into a table, along with the mean curvature and track–length of the concatenated track $\mathbb{C}$.

After all possible track–segment combinations have been evaluated for track $\mathbb{A}$ and the tracks which overlap it, the combination which maximises the track length and minimises the mean curvature is found. If the length is the same or greater than the original length ($\mathcal{L}(\mathbb{A}_a^b)$) and the mean curvature is the same or less than the original mean curvature ($\overline{\Xi}(\mathbb{A}_a^b)$), then tracks $\mathbb{A}$ and $\mathbb{B}$ are replaced by up to 5 tracks:

1. The concatenation of track–segments $\mathbb{Q}_c^d$ and $\mathbb{R}_u^v$: $\mathbb{C}_c^v$.
2. If $a < c$, the segment of track $\mathbb{A}_a^b$ preceding c: $\mathbb{A}_a^{c-1}$.
3. If $b > d$, the segment of track $\mathbb{A}_a^b$ succeeding b: $\mathbb{A}_{d+1}^b$.
4. If $s < u$, the segment of track $\mathbb{B}_s^t$ preceding u: $\mathbb{B}_s^{u-1}$..
5. If $t > v$, the segment of track $\mathbb{B}_s^t$ succeeding v: $\mathbb{B}_{v+1}^t$.

After the merging process, any phantom feature objects that occur as the first or last feature object in a track are removed from that track. Any phantom feature objects which occur within the track are interpolated to a position half way between the preceding and succeeding feature object. The objects intensity and delta values are set to be the mean of the preceding and succeeding object's intensity and delta values. However, the objects are still flagged as being phantom via their bitfield and by having no triangle labels associated with them.

To obtain a globally optimal set of tracks, this optimisation algorithm is applied to all tracks that are derived from the locally optimal tracking algorithm in section 2.4. Track segments are swapped between overlapping tracks until the total curvature of all the tracks is minimised, without reducing the length of the tracks. More than one iteration of the algorithm may be needed before the optimiser converges to a solution. Note that the optimisation

procedure can swap single points between the start or end of tracks, as the minimum size of a track segment is 1 feature object and that it allows the merging of tracks that begin and end one timestep apart, i.e. at $t = a - 1$ and $t = a + 1$, by projecting a phantom feature object to $t = a$ from either track. The major advantage of the optimisation scheme is that it allows points and segments at the beginning of a track to be exchanged. These points are particularly susceptible to error as there is no information in preceding timesteps to guide the tracking (Hodges, 1999).

2.4.5 Post–processing of tracks

After the completion of the tracking algorithm, the tracks may be filtered on various criteria, including a minimum number of timesteps the track persists for, how far the track moves from genesis to lysis, the length of the track or the minimum value of MSLP reached along the track. However, this filtering is left to the user as there is no post–processing directly provided by the suite of algorithms.

2.5 Summary and conclusions

This chapter presents a novel method of identifying storm–like features and tracking their path through a domain. It uses a hierarchical triangular mesh to regrid climate data to a number of spatial resolutions, each resolution occurring at a different level in the mesh. The triangular mesh undergoes an iterative area–equalisation phase so as to remove any bias due to spatial inconsistencies in identifying features. Using a triangular mesh solves several problems related to feature tracking.

1. Features can be tracked close to and over the poles, and across the Equator. This contrasts with methods that use a cylindrical projection, where the grid boxes at the poles have the same value for each longitude, which prevents identifying and tracking features close to or over the poles. Methods that project the data onto a polar projection suffer from only being able to track features in one hemisphere at once, and

features that cross the Equator will be split into two separate tracks.

2. All features in the data regridded to the triangular mesh have the same spatial scale at the same level in the mesh. This overcomes the problem that regularly spaced grids have where the resolution of the data effectively increases towards the poles, and the scale of each grid box decreases. Additionally, in a polar stereographic grid, the grid spacing decreases toward the equator, being 80% less at 30° than at 60°. In the triangular mesh this does not occur, as the area equalisation phase of the mesh construction ensures that all grid spacing between centroids of the triangles is equal.

3. The hierarchical mesh has data available at a number of resolutions, for each timestep. This allows data at a lower resolution (lower grid level) to be removed from data at a higher resolution before feature identification is performed. This is analogous to removing the background field from the data and doing this dynamically has an advantage over using a static background field, such as removing a constant value, or a climatological mean, as it does not weaken the MSLP low features (Anderson et al., 2003).

Climate data is regridded to the hierarchical mesh and the background field is removed. This is achieved by smoothing a lower resolution version of the data, from a lower level in the mesh, to the higher level at which feature identification is performed and subtracting the smoothed data from the higher resolution data. The resulting data is then contoured to 1hPa levels, which simplifies the search for minima in the data. Feature identification locates low pressure features by examining adjacent triangles at an *identification level* and refining these to a higher level. From these triangles a feature object is derived by expanding the triangles via a concentric shell so as to encapsulate a cyclone (low pressure system) as a feature object. Central points are obtained for these feature objects by a weighted average of the positions of the individual triangles in the object.

In this thesis, feature tracking is a two phase process. Firstly a set of locally optimal tracks is derived by using a hybrid prediction–optimisation scheme which builds tracks in real time, as feature objects become available. To add feature objects to existing tracks a number

of hierarchical rules need to be passed and the tracks are optimised locally by finding the object which passes the highest rule with the lowest cost when added to the track. To fix the occasional wrong assignment of a feature–object at the beginning and end of a track, a global optimisation scheme is employed which exchanges segments between tracks so as to maximise track lengths and minimise track mean curvature across all tracks.

The methods presented in this chapter are computationally efficient and suitable for running on many thousands of ensemble members. The most computationally expensive component is the global optimisation of the tracks. However, the user can select how many iterations of this optimisation they wish to apply to the tracks and so the run–time can be adapted to the computing resources available. In the next chapter the performance of the feature–tracking algorithm will be evaluated by applying it to ERA-Interim data (Dee et al., 2011) and comparing it to 15 other tracking algorithms. The sensitivity of the algorithms to the temporal resolution of the input data is also investigated.

3 Validation and sensitivity of a storm–tracking algorithm

Validating a storm tracking algorithm is problematic as there is no universally agreed set of historical storm tracks with which to compare the output from a storm tracking algorithm with (Neu et al., 2013). Instead, a tracking–algorithm inter–comparison project, IMILAST ("Intercomparison of Mid-Latitude Storm Diagnostics"), compares fifteen different storm tracking methods by applying each method to the ERA–Interim reanalysis (Dee et al., 2011) for the period 1989–2009. The outputs of the 15 methods are then analysed to determine the discrepancies in the spatial distribution of the cyclone centres, in both the Northern Hemisphere and Southern Hemisphere. A set of track metrics (intensity, lifetime and propagation speed) are also calculated from the output of each tracking methods, and the resulting distributions of the metrics are also compared.

In this chapter, the IMILAST inter–comparison protocol is followed, by applying the full process of storm–like feature–identification and tracking detailed in sections 2.2 to 2.4 to the $1.5° \times 1.5°$ resolution ERA-Interim dataset, from 1989 to 2009. The track metrics from IMILAST are also calculated and compared to the 15 methods detailed in Neu et al. (2013). The provides a relative validation of this algorithm when compared to other commonly used methods.

An important and well used tracking algorithm, Hodges (1999), is not part of the IMILAST inter–comparison. However, it forms the feature–identification and tracking component of the XWS extreme windstorm catalogue Roberts et al. (2014), where it is also applied to a higher resolution ($0.77° \times 0.77°$) version of the ERA-Interim data, from 1987 to 2013. To make comparisons to individual storms in the XWS catalogue, the method developed in this thesis is applied to the $0.77° \times 0.77°$ ERA-Interim data, and the individual storms produced

are compared to those in XWS.

To investigate the sensitivity of the method developed in this thesis to the temporal resolution of the input data, the ERA-Interim data (at $1.5^\circ \times 1.5^\circ$ resolution) is resampled to 12 and 24 hour resolution and the tracking method of this thesis is applied to each temporal resolution. Comparisons to the original 6 hourly data are made to determine whether the algorithm can successfully use data of a coarser temporal resolution.

In order to aid discussion and comparison of the various storm tracking algorithms, Neu et al. (2013) assigns numbers to the 15 methods, from M02 to M22 (the numbers are not contiguous). In this chapter, the method developed in chapter 2 will be referred to as MAS, whereas the method of Hodges (1999) will be denoted HOD.

3.1 Application to ERA-Interim data

Neu et al. (2013) analyses the disparities between 15 different storm tracking procedures by applying the algorithms to the ERA–Interim reanalysis data product (Dee et al., 2011). Total cyclone centre densities for all 15 procedures are compared, along with histograms of the track metrics: intensity, lifetime and propagation speed. To determine the relative performance of MAS, it is applied to ERA-Interim data for the Northern and Southern hemisphere for their respective winter months, from 1st January 1989 to 31st August 2009. This follows the inter–comparison protocol of Neu et al. (2013). The mean sea level pressure (MSLP) variable was used to identify features, with the geopotential height at 500hPa (Z500) used to determine the steering vector.

The full process of applying MAS to the ERA-Interim reanalysis data is:

1. The ERA-Interim data has a spatial resolution of $1.5^\circ \times 1.5^\circ$ and a temporal resolution of 6 hours.
2. Generate a mesh, using the routine in section 2.2 and the global MSLP data from ERA-Interim as the input data, which subdivides to a level 5 mesh.

3. The global MSLP for December, January and February (DJF) 1989 to 2009 is regridded to the mesh using the procedure in section 2.2.8. The global MSLP for June, July and August (JJA) for the same period if also regridded to the mesh.
4. Low pressure features in the regridded MSLP are identified using the procedure in section 2.3. Level 3 is used as the *background field level*, level 4 as the *identification level* and level 5 as the *refinement level*. The anomalies from the background field are contoured to the nearest 1hPa, and a minimum anomaly of -2hPa is required for each low pressure feature.
5. The low pressure features are then assembled into tracks using the track identification and optimisation routines in section 2.4. A 1000km search radius is used, with a minimum overlap of 10%.
6. As in Neu et al. (2013), tracks are filtered so only those that last 24 hours or more (5 timesteps and above) are included in the analysis.

3.1.1 Northern Hemisphere winter

After applying MAS to the DJF ERA-Interim data, 43,310 tracks are identified of which 14,351 had a persistence of 24 hours or more. This is directly comparable to the number of tracks produced by the 15 algorithms compared in Table 2 of Neu et al. (2013), where the number of tracks identified in DJF ranges between 5,700 and 20,500, with a mean of 12,393 and standard deviation of 4,504.

Figure 3.1.1 shows the density of cyclone centres per timestep and 1000km^2 for the ERA-Interim reanalysis data after applying MAS. This Figure applies the same method of calculating the cyclone density as Figure 1 in Neu et al. (2013), and so is directly comparable with the results presented there. The calculation is the percentage of timesteps at which a cyclone occurs within a grid box and per unit area of 1000km^2.

The MAS tracking procedure performs as well as the 15 methods in Neu et al. (2013), qualitatively agreeing with the identification of the areas of high storm activity east of Greenland, around the Northern edge of the Scandinavian peninsula, the two centres in the

North Pacific and on the Eastern side of the Rocky mountains. MAS also identifies a large frequency of occurrence of storms over the Mediterranean, which Neu et al. (2013) note has a large discrepancy between methods. The MAS algorithm compares most favourably to those methods which also use MSLP or geopotential height minima for the identification of cyclone centres. These include methods M03 (Benestad and Chen, 2006), M08 (Trigo, 2006), M13 (Hanley and Caballero, 2012) and M14 (Kew et al., 2010). The earlier method of Massey (2012) failed to identify any features over the Mediterranean, a shortcoming of the algorithm which is related to using an absolute threshold, rather than an anomaly to the background field, as in the improved method of MAS. There is no attempt in Neu et al. (2013) to standardise the depression depth or intensity that a cyclone is required to have to pass the threshold of each individual algorithm and the MAS method is somewhat inclusive, requiring only a depression of 2hPa away from the large scale flow to register as a storm like feature.

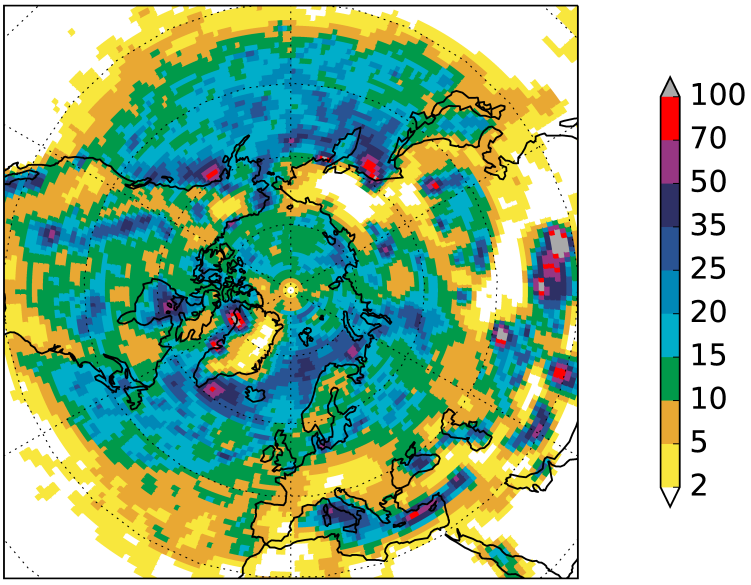


Figure 3.1.1: Total cyclone centre density (percentage of cyclone occurrence per 1000km^2 per 6 hour timestep) for the Northern Hemisphere winter (DJF) in the ERA-Interim reanalysis data for 1989 to 2009. The MAS feature–identification and tracking algorithms are applied to the ERA-Interim data at 1.5° x 1.5° resolution and cyclones lasting less than 24 hours are filtered out.

Figure 3.1.2 shows the relative frequencies of three storm track metrics. This Figure is directly comparable to Figure 3 (a,c,e) in Neu et al. (2013), who produce the same rela-

tive frequency plots for the distribution of the 15 storm tracking methods, with the mode, standard deviation and range over the 15 methods being shown. Again, MAS compares favourably with the 15 methods studied in Neu et al. (2013). For the intensity, the fraction of unit for each range bin is within the standard deviation of the 15 other tracking methods. Very deep depressions, with a central pressure < 960 hPa, occur with a frequency of 3.88%, which is within the 2% to 8% range of occurrence in the 15 methods in Neu et al. (2013). For lifetime, the same is true, even though the standard deviations in Neu et al. (2013) have a much narrower range. Neu et al. (2013) note that there is a much wider spread in the 15 methods for short–lived cyclones, and this is apparent in MAS, which is towards the higher percentiles of the distribution of the occurrence frequency of cyclones which last a single day. Again, for propagation speed MAS compares well with the 15 other methods, except that there is a higher frequency of occurrence for tracks with speeds above 50 km/hr. Neu et al. (2013) note that there are a range of behaviours for very high system velocities, above 80 kmh^{-1}.

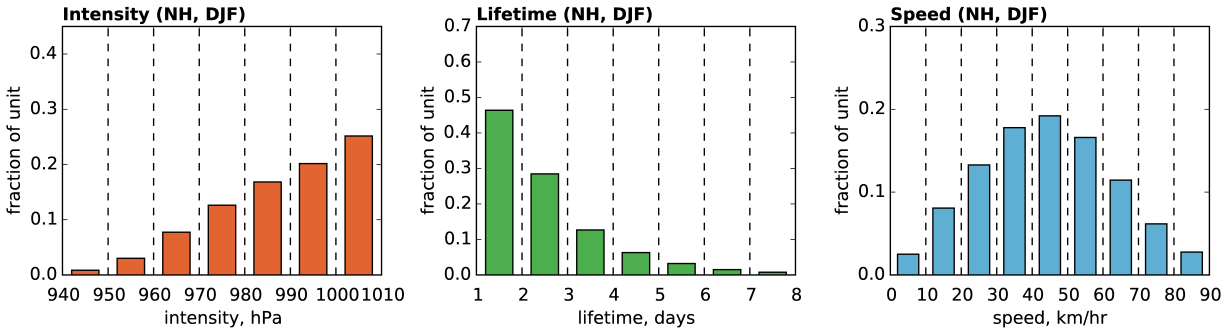


Figure 3.1.2: Normalised distribution of three track metrics for the cyclones identified by applying the feature detection and tracking algorithm to the ERA-Interim 1989 to 2009 Northern Hemisphere winters (DJF). The metrics are cyclone intensity (the lowest depression encountered along the cyclone track), cyclone lifetime (the number of days the cyclone persists for) and speed (the mean propagation speed of the cyclone centre along the track).

3.1.2 Southern Hemisphere winter

The tracking procedure outlined in section 3.1 is repeated by applying MAS to the June, July and August (JJA) ERA-Interim data for the Southern Hemisphere winter. There are 96,884

tracks in the Southern Hemisphere of which 12,707 last 24 hours or more. This compares with the tracks generated by the 15 methods in Neu et al. (2013) which have between 3800 and 24,400 tracks, with a mean of 11,040 tracks and a standard deviation of 5335 tracks (Table 3 of Neu et al. (2013)) and so is within the range of tracks in the IMILAST comparison. This indicates that the mean number of tracks, and the variance between the 15 methods is similar in the Southern Hemisphere winter (JJA) as it is in the Northern Hemisphere winter (DJF).

Figure 3.1.3 shows the density, per timestep and 1000km^2, of the occurence of cyclone centres in the tracks produced from applying MAS to the ERA-Interim JJA data for 1989 to 2009, in the Southern Hemisphere only. Comparing Figure 3.1.3 to Figure 2 in Neu et al. (2013) shows a reasonably good correspondance in the spatial distribution of the cyclone centres between the 15 tracking methods and MAS. All 15 methods show an Equator to Pole gradient of cyclone centres, with more centres close to the Antarctic coast and fewer towards the Equator. This is also present in MAS in Figure 3.1.3, showing the method can capture the large scale climatology of the mid–latitude cyclones in the Southern Hemisphere. The method shows four areas of high cyclone activity: the Andes, which is also shown in most of the 15 methods; on the Ross Ice Shelf, which is also seen in 5 methods (M02, M10, M12, M14 and M21); on the Amery Ice Shelf, seen in just one other method (M02) and on the Pacific coast of Antarctica, seen in 4 methods (M02, M10, M12, M18). Additionally there is a higher level of activity on the Atlantic coast side of Antarctica, in the Weddel Sea, also seen in methods M02, M09, M12, M13, M16 and M21. In general, apart from the Equator to Pole gradient, there is much less correlation between methods in the spatial distribution of cyclone centres in JJA in the Southern Hemisphere than there is for DJF in the Northern Hemisphere. This makes validating an algorithm for the Southern Hemisphere more difficult than for the Northern Hemisphere. However, the method presented here does not look significantly different to any of the other 15 methods presented in Neu et al. (2013) and so must be considered as good a method as any of those 15.

As with the Northern Hemisphere, the tracks in the Southern Hemisphere are analysed and three track metrics are calculated. These are the distribution of the relative frequencies of

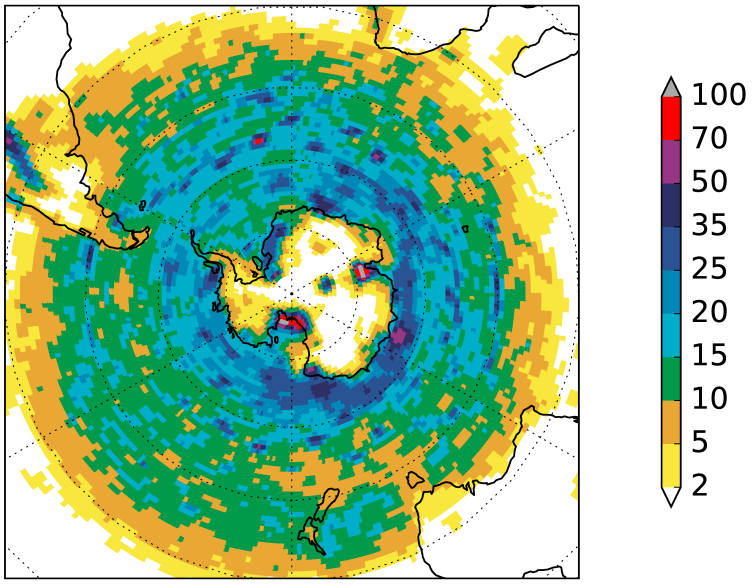


Figure 3.1.3: Total cyclone centre density (percentage of cyclone occurrence per 1000km^2 per 6 hour timestep) for the Southern Hemisphere winter (JJA) in the ERA-Interim reanalysis data for 1989 to 2009. The MAS feature–identification and tracking algorithms are applied to the ERA-Interim data at 1.5° x 1.5° resolution and cyclones lasting less than 24 hours are filtered out.

the intensity (the lowest MSLP value of the cyclone centres along the track), the lifetime of the track, in days, and the mean propagation speed of the cyclone centres along the track. Figure 3.1.4 shows the track metrics for the Southern Hemisphere winter (JJA) and is directly comparable to Figure 3 (b,d,f) in Neu et al. (2013). Again, as with the Northern Hemisphere, the intensity distribution closely matches that of the 15 methods in Neu et al. (2013), with the fraction of unit for each range bin being within the standard deviation of the 15 other tracking methods. Very low depressions, defined as being < 950 in the Southern Hemispheres, occur with a frequency of 8.6%, which compares favourably with the 4% to 12% range in the 15 methods. The distribution of track lifetimes shows some discrepancy, in that the percentage of tracks lasting between 1 and 2 days is outside the standard deviation from the mean of the 15 methods. However, this lifetime shows the widest range of all of the lifetime bins (between ~ 0.15 and ~ 0.65 for the 15 methods). The relative frequency of tracks lasting 1 to 2 days in the MAS method is within this wide range, indicating that, although it is outside the standard deviation of the 15 methods, it is within the range exhibited by those methods, which indicates that MAS is still a comparable algorithm. There are similar disparities in

the other relative frequencies of lifetimes, except for tracks which last for 2 to 3 days. The distribution of the relative frequencies of propagation speeds in MAS also does not match the distribution of the propagation speed of the 15 methods. Here the distribution is wider, with more fast moving cyclone systems. Again, for each of the bins in MAS, the relative frequency is within the range of the relative frequencies in the 15 methods, but outside the standard deviation from the mean.

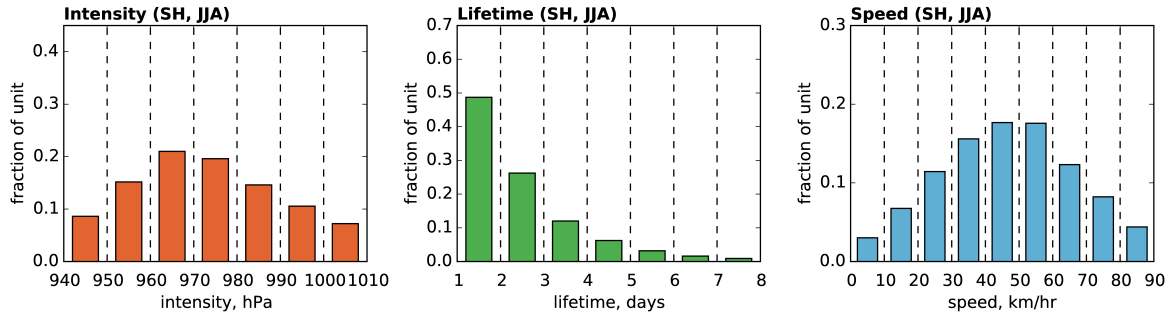


Figure 3.1.4: Normalised distribution of three track metrics for the cyclones identified by applying the feature detection and tracking algorithm to the ERA-Interim 1989 to 2009 Southern Hemisphere winters (JJA). The metrics are as in Figure 3.1.2

3.1.3 Summary of validation of algorithm on ERA-Interim data

The algorithm of MAS performs well in identifying the position of cyclone centres (low pressure systems) in the ERA-Interim (Dee et al., 2011) data from 1989 to 2009, with the spatial distribution closely matching that identified in the 15 storm tracking methods presented in Neu et al. (2013), for both the Northern Hemisphere winter (DJF) and Southern Hemisphere winter (JJA). The MAS algorithm also produces distributions of track metrics that compare favourably with those in the 15 methods for the Northern Hemisphere winter, with each metric being within the standard deviation from the mean of the 15 methods. However, the distributions of track metrics in the Southern Hemisphere do not match the standard deviation from the mean of the metrics in the 15 methods, with the exception of cyclone intensity.

The two metrics which have discrepancies, lifetime and propagation speed, are closely influ-

enced by the tracking algorithm in section 2.4, whereas the cyclone intensity metric is only influenced by the feature–identification algorithm of section 2.3. The tracking algorithm requires some free parameters to be selected by the user, most noticeably the search radius to find track points to add to existing tracks in subsequent timesteps. Tuning these parameters, and using the distribution of track metrics for the Southern Hemisphere in Neu et al. (2013) as the distribution to fit against, may improve the algorithm's representation of these track metrics.

However, the rest of this thesis concentrates on validating a model over the Northern Hemisphere and Europe, in particular. The algorithm has been shown to work well over this domain and so no further investigation of tuning the algorithm for the Southern Hemisphere will be undertaken here.

3.2 Application to individual storm test cases

While section 3.1 shows that the MAS feature–identification and tracking algorithm, when applied to the ERA-Interim data, produces climatologies of storm tracks that are comparable to other storm tracking methods applied to the same data, it is instructive to study individual test cases of storms to determine the accuracy of the tracks produced by the tracking algorithm are.

Applying the feature–identification and tracking procedure to relatively coarse ERA-Interim data, as in section 3.1 allows for a comparison to the other tracking methods in Neu et al. (2013). However, the algorithms have been designed to work on data with a range of resolutions. To demonstrate this, in this section the MAS tracking procedure is applied to a higher resolution version of the ERA–Interim reanalysis (Dee et al., 2011) and compared to tracks that are derived from the same reanalysis data but using a different tracking algorithm, that of Hodges (1999).

The XWS Extreme Windstorm Catalogue (Roberts et al., 2014) applies the storm feature identification and tracking method of Hodges (1995, 1999) to the 0.77° resolution version of

the ERA–Interim reanalysis data. Storm tracks for the period 1981 to 2013 are identified and a catalogue of the storms responsible for the highest insurance losses over the 33 year period is then created, with information for the insured loss, maximum windspeed, lowest depression, maximum vorticity and a storm severity index. In XWS, the storm tracks identified in the ERA-Interim data are then overlaid onto output from a regional climate model (RCM): the Met Office Unified Model at 0.22° (24km) resolution, with a domain over Europe and the North Atlantic, forced at the boundaries by the ERA–Interim reanalysis dataset. This allows the catalogue to provide more fine resolution windspeed, MSLP and storm severity indices data so that the impacts of the storms on infrastructure and insurance losses can be studied. XWS contains 20 named storms, which are considered to be the most destructive (in terms of insured losses) over the 33 year period.

Following a similar design, this section applies the MAS feature–identification and tracking algorithm from chapter 2 to the 0.77° resolution version of the ERA–Interim reanalysis. The triangular mesh is generated from the ERA-Interim grid, with a maximum mesh level of 6. This mesh has four times as many triangles as the mesh generated for section 3.1, which was based on the 1.5° resolution ERA-Interim reanlysis. When identifying features, the *background level* is 3, the *identification level* is 4 and the *refinement level* is 6. The *background level* and *identification level* are the same as those chosen in section 3.1 when identifying features in the lower resolution data. This is to maintain consistency in the parameters chosen in this section and section 3.1. The same tracking parameters as in section 3.1 are used during the tracking phase and, once the storm–tracking procedure has been applied to the ERA-Interim data, the tracks produced are compared to those in the XWS catalogue.

During the lifetime of any individual XWS storm, many storm tracks can exist in the MAS track–set concurrently. To determine which of these tracks most closely matches the XWS storm, a track matching procedure is performed between the MSLP track of interest in XWS and the tracks in the MAS track–set which occur concurrently with it. This matching procedure calculates the sum of the distances between points in the two tracks at each timestep which the tracks share. Any timesteps which occur in one track but not in the

other are penalised in the sum with a distance of 1000km. For two tracks, $\mathbb{S}^{a\leq t\leq b}$ and $\mathbb{T}^{c\leq t\leq d}$, where track $\mathbb{S}$ occurs between timesteps a and b and track $\mathbb{T}$ occurs between timesteps c and d the distance measure can be summarised by:

$$\begin{aligned}\Lambda &= \sum_{i=\min(a,c)}^{i\leq\max(b,d)} \Theta(S^i_{\phi,\lambda}, T^i_{\phi,\lambda}) \\ \Theta &= \begin{cases} \Delta(S^i_{\phi,\lambda}, T^i_{\phi,\lambda}) & \text{if}(i \geq a \wedge i \geq c \wedge i \leq b \wedge i \leq d) \\ 1000 & \text{otherwise} \end{cases}\end{aligned}$$

where Δ is the haversine distance, as given by Equation 2.4.3, $S^i_{\phi,\lambda}$ is a point in track $\mathbb{S}$ at timestep i, $T^i_{\phi,\lambda}$ is a point in track $\mathbb{T}$ at the same timestep and $\wedge$ indicates logical and. After the track distance measure has been applied to all concurrent tracks, the track in the ERA-Interim track–set with the smallest distance sum is then selected as most closely matching the XWS track.

XWS (Roberts et al., 2014) provides a table of storms over Europe, ranked by their insured loss. Using this loss as a proxy for the most powerful storms, the six most destructive storms in the 0.77° ERA-Interim output are shown in Figure 3.2.1. These storms are Daria 25/01/1990, Lothar 26/12/1999, Kyrill 18/01/2007, Great Storm of 87 16/10/1987, Vivian 26/02/1990 and Klaus 24/01/2009

Figure 3.2.1 shows, for each of the six most destructive storms, the paths from the XWS track–set for tracks identified from both the MSLP field and the vorticity field. For each storm, the path of the mostly closely matching MAS track is also shown. This track has been identified as that which most closely matches the path of the MSLP track for the storm in the XWS track–set using the track matching procedure above. Below each track is the evolution of the MSLP depth along the track of the storm for the MSLP track in the XWS track–set and the corresponding track in the MAS track–set.

As seen in Figure 3.2.1 the MAS tracking algorithm is largely able to successfully identify and track the low pressure features belonging to the six most destructive storms and that their tracks closely resemble, in terms of position, lifetime and MSLP intensity, the tracks

in the XWS catalogue. The most noticeable discrepancies are:

1. Daria (25/01/1990, **a**) begins 24 hours before its equivalent in XWS and deviates from the XWS MSLP path towards the end of the track. However, it should also be noted that the XWS vorticity path also deviates from the MSLP path towards the end of the track. The MSLP intensities are very similar, apart from around the 48 to 60 hour timesteps.

2. Lothar (26/12/1999, **b**) begins 18 hours before the XWS track. It also deviates from the path of both the MSLP and vorticity path around the 36 to 54 hour timesteps. There are discrepancies in the MSLP intensities for these time steps as well.

3. Kyrill (18/01/2007, **c**) begins 24 hours earlier and ends 12 hours sooner. However, the position is very consistent with the XWS MSLP track and the MSLP intensities are also very close to those in the XWS track.

4. Great Storm 87 (16/10/1987, **d**) does not match the XWS track for the first 96 hours of its lifetime, where the track is position in the Atlantic west of France and Spain. However, after the algorithm identifies the storm over the British Isles, the path and MSLP intensity are very consistent with the paths in the XWS tracks, as are the MSLP values.

5. Vivian (26/02/1990, **e**) also begins later than the XWS equivalent track, in this case nine hours later than the XWS track. The track in MAS also ends sooner than the XWS track, with a positional difference towards the end of the track as it moves over Finland. The position is consistent over the British Isles, North Sea and Norway, with MSLP values close to those in the XWS track.

6. Klaus (24/01/2009, **f**) matches the path and MSLP values of the track in the XWS dataset very closely, although it is identified 18 hours earlier than in XWS.

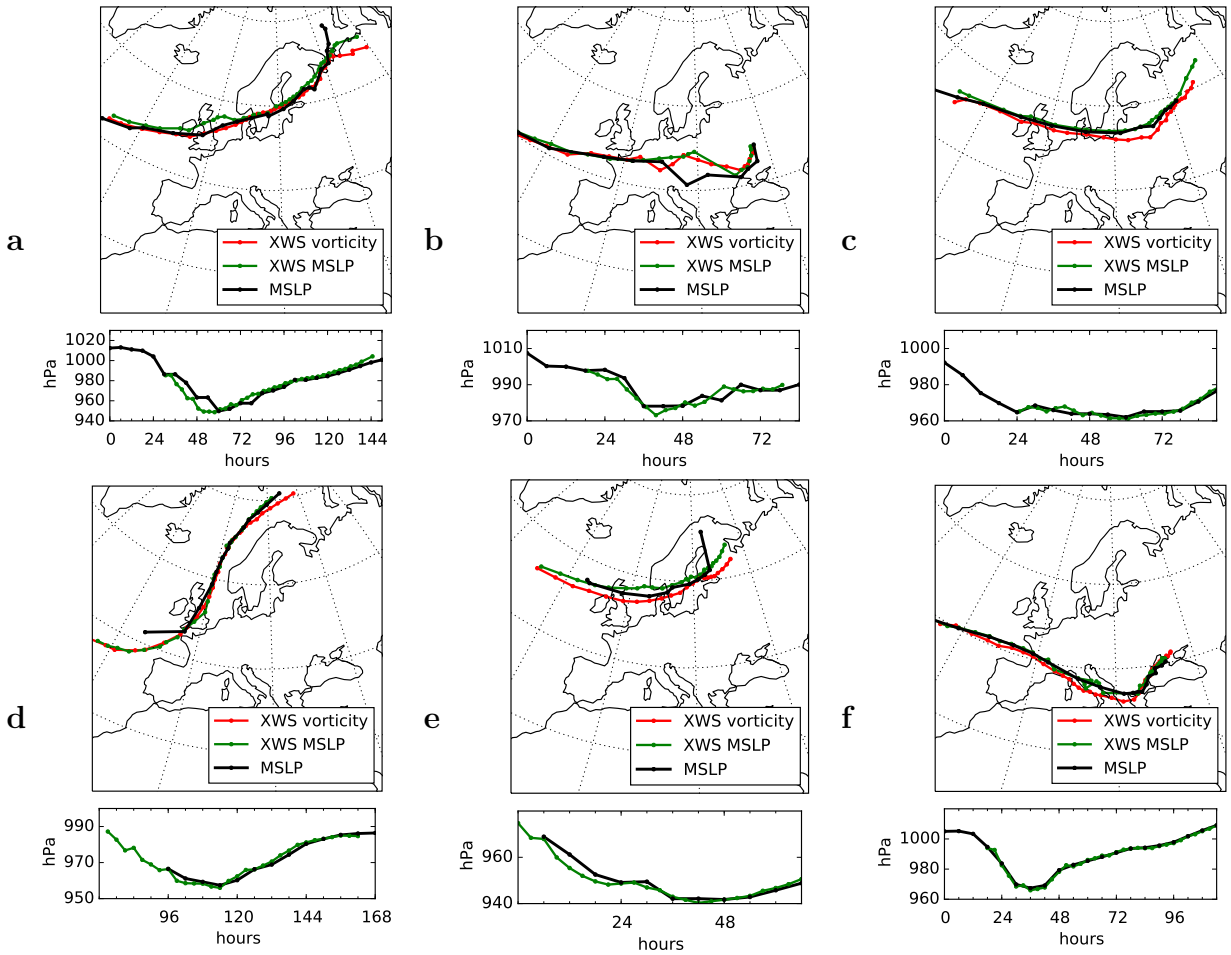


Figure 3.2.1: The six most destructive storms (in terms of insured loss) in the XWS catalogue and the corresponding storms identified in the ERA-Interim data by the storm MAS identification and tracking procedure. MAS MSLP tracks are shown in black XWS tracks for vorticity and mean sea–level pressure are shown in red and green respectively. The storms are: **(a)** Daria 25/01/1990, **(b)** Lothar 26/12/1999, **(c)** Kyrill 18/01/2007, **(d)** Great Storm of 87 16/10/1987, **(e)** Vivian 26/02/1990 and **(f)** Klaus 24/01/2009

To provide a more quantitative assessment of the difference between the tracks in the XWS catalogue and those produced by applying MAS to the 0.77° ERA-Interim reanalysis, the track matching procedure is applied to every track in the XWS catalogue. Track metrics such as the persistence, track length and minimum sea–level pressure along the track are derived. Additionally the distance measure used by Blender and Schubert (2000) (Equation 6) is also calculated, taking the XWS track as one track and the track produced by MAS as the other. This distance measure is a spatiotemporal measure and does not require tracks

to even share timesteps, although, for this analysis, the track matching procedure ensures that they do. The distance measure of Blender and Schubert (2000) is given by:

$$D_{12}^2 = \frac{1}{AA'}\left[\sigma_{12}^2 - \frac{1}{2}(\sigma_{11}^2 + \sigma_{22}^2)\right] \tag{3.2.1}$$

where σ_{12}^2 measures the variance spanned by two tracks:

$$\sigma_{12}^2 = \frac{1}{AA'}\int_0^A da \int_0^{A'} da' \times \{\alpha([x(a) - x'(a')]^2 + [y(a) - y'(a')]^2) + \beta[t(a) - t'(a')]^2\}$$

where track A lasts for A timesteps and has a position in space and time of $[x(a), y(a), t(a)]$. Therefore, this variance metric is measuring the difference between the position and timesteps of a track. The constants α and β are set so that each term has a physically equivalent weight. Here, typical values for the distance between subsequent points in a track and the timescale are used, and so $\alpha = 1$ and $\beta = 10$ where 10ms^{-1} is the synoptic velocity scale, i.e. the typical propagation speed of a mid–latitude cyclone. In Equation 3.2.1 the addition of the σ_{11}^2 and σ_{22}^2 terms is so that when comparing a track to itself, the sum is 0.

These track metrics are shown in Table 3.2.1, where the distance measure (D_{12}) is multiplied by 100 for ease of formatting and reading. This Table shows quantitatively that there is a good agreement between the tracks in the XWS catalogue and those produced by applying the storm–tracking procedure to the ERA-Interim reanalysis data. Only two storms (Xylia and Stephen) have a comparatively high D_{12}. Of these storms, the reasons are:

1. Xylia (28/10/1998). The track in MAS is farther south and continues for longer than the XWS track.Additionally, the MSLP intensity in MAS matches that in XWS well for the first 12 hours of the track, but shows a large difference between hours 12 and 36, with up to 20hPa difference between the MAS track and the XWS track. This suggests that an incorrect feature object has been assigned to the track, or the correct feature object has not been identified.
2. Stephen (26/12/1998). The track in MAS is identified 18 hours before the track in XWS and dissipates more quickly, 42 hours before the track in XWS ends. However,

Storm name	Persistence (hrs)		Minimum SLP (hPa)		Length (kms)		Distance measure
Anatol	96	66	956	963	5982	4228	2
Daria	114	114	948	959	5510	5337	0
Emma	126	132	959	960	5811	6780	8
Erwin Gudrun	78	78	959	961	4171	4584	13
Fanny	57	48	966	976	3747	3275	1
Great87	87	72	955	957	5173	4095	24
Gero	96	114	947	949	5097	4484	15
Herta	96	48	989	993	6502	3035	9
Jeanette	105	138	974	975	5406	7175	12
Klaus	105	114	965	967	6001	6872	17
Kyrill	78	90	961	964	5432	6275	13
Lore	99	42	962	964	6582	2987	4
Lothar	60	60	973	976	5259	4037	25
Martin	135	108	967	976	7303	5367	0
Oratia Tora	99	126	948	955	3275	4250	3
Stephen	87	60	946	953	4147	3933	65
Vivian	75	60	940	940	3955	3471	18
Wiebke	102	96	970	975	5822	5499	6
Xylia	108	60	960	961	2843	4537	45
Yuma	51	48	971	976	3693	3537	7

Table 3.2.1: Comparison of the top insurance loss storms occurring in the XWS dataset with the same storms identified by applying the MAS tracking method to the ERA-Interim reanalysis data at 0.77° resolution. The resultant tracks are matched to those in XWS by a minimum distance matching algorithm. The track distance measure is taken from Blender and Schubert (2000)

during the 18 to 60 hour timesteps, there is good agreement in the position and MSLP values of the MAS track when compared to the XWS track.

3.3 Sensitivity of the tracking algorithm to temporal resolution

Using volunteer distributed computing (VDC) to generate large ensembles of a climate model, as described in chapter 4 and Massey et al. (2015), places a number of burdens upon both the citizen scientists donating their computing time, and the computing infrastructure of the project which collects the results from the volunteers. Using VDC ensures that the computing

time is essentially free, however climate models are able to generate vast outputs of diagnostic variables and, to keep volunteers engaged with the project, the output diagnostics should be carefully chosen so as to not place too high a demand on the volunteer's bandwith. On the project side, requesting too many diagnostics places a strain on computing servers, both in terms of bandwith when receiving results and in storage capacity when storing those results.

The storm tracking algorithms in Neu et al. (2013) and the result from the new MAS method presented in chapter 2 all use 6 hourly instantaneous or 6 hourly mean fields of climate variables to identify and track the features in, with the exception of method M06 in Neu et al. (2013). This section investigates whether a coarser resolution could be used with the new tracking method. If this is possible it will reduce the burden on the volunteers and project servers and possibly allow for a larger ensemble to be computed.

3.3.1 Degrading 6 hourly data to 12 and 24 hours

The results presented in sections 3.1 and 3.2 use 6 hourly instantaneous data from ERA-Interim Dee et al. (2011). That is, the data is the state of the reanalysis at 00:00, 06:00, 12:00 and 18:00 with no averaging across time of the state of the reanalysis between these times. There are two obvious ways to degrade this 6 hourly data into 12 and 24 hourly data:

1. Sample every 2nd timestep (for 12 hourly data) or every 4th timestep (for 24 hourly data). These will be referred to as 12 hourly instantaneous (12HI) and 24 hourly instantaneous (24HI). The original 6 hourly data will be referred to as 6HI.
2. Calculate the mean of 2 timesteps: 00:00 to 06:00 and 12:00 to 18:00 for 12 hourly data or the mean of 4 timesteps for 24 hourly data. These will be referred to as 12 hourly mean (12HM) and 24 hourly mean (24HM).

After degrading the data the feature–identification routine of MAS is applied to each of the four new datasets, using the same parameters as used in section 3.1.1. However, for the tracking of the feature objects, the search radius needs to be adjusted to reflect that the objects will be further apart in subsequent timesteps in the 12 and 24 hour data, when

Percentile	Distance (km)
95^{th}	824
96^{th}	863
97^{th}	905
98^{th}	951
99^{th}	1006

Table 3.3.1: Upper percentiles of the distance between track points in the tracks of MSLP minima in the 6 hourly ERA–Interim data.

compared to the 6 hour data. To determine the value to use, Table 3.3.1 shows the 95^{th} to 99^{th} percentile values of the distribution of the distance between subsequent points in the 6 hour track data. The 99^{th} percentile occurs with a value greater than 1000km as the optimisation procedure and the cost function of Equation 2.4.10 allows distances between points of up to 2000km, if the curvature between the points is sufficiently low. Blender and Schubert (2000) show that midlatitude cyclones exhibit a non–linear growth of the displacement between timesteps and, when searching for cyclones in coarser temporal data, a linear increase in the search radius should not be used. Instead, the search radius should increase at a rate of 0.75. Here the search radius for the 12 and 24 hour data is set to be:

$$S_t = S_{t_0} \left(\frac{dt}{dt_0} \right)^{0.75}$$

where S_{t_0} is the search radius at 6 hourly resolution, $dt_0 = 6$ and $dt = 12$ or $dt = 24$. Using the 98^{th} percentile ($\sim$ 950km) value as S_t, this gives search radii for the 12 hourly data of ($\sim$ 1600km) and for the 24 hourly data of ($\sim$ 2687km). Blender and Schubert (2000) use a nearest–neighbour approach to tracking, defining the search radius as a number of grid boxes. They show how the number of grid boxes must be increased in higher latitudes, as the distance between grid box centres becomes less towards the poles. In the MAS method this is unnecessary as the search is carried out purely in latitude–longitude space, with the search radius controlling the possible number of latitude and longitude degrees between potential track points.

After applying the MAS feature–identification and tracking algorithm to the four sets of degraded data, the spatial distribution of the cyclone centres and the track metrics are

calculated in the same manner as in sections 3.1.1 and 3.1.2. The tracks are again filtered so that all tracks last for at least 24 hours. In the 12 hourly track sets, this requires tracks to last for at least 3 timesteps, whereas for the 24 hourly track sets, the tracks have to last for at least 2 timesteps.

3.3.2 Comparing tracks in different temporal resolution data

Analysis in this section concentrates solely on the Northern Hemisphere. After applying the MAS algorithm to the four sets of temporally degraded ERA-Interim DJF data for 1989 to 2009, the spatial distribution of the cyclone densities for the 12HI, 12HM, 24HI and 24HM data are calculated per timestep and per 1000km^2. The cyclone densities calculated for the 6HI data (as shown in Figure 3.1.1) are then subtracted from these cyclone densities, to obtain four sets of differences in cyclone densities between the 6HI data and the four degraded data sets, i.e. 12HI−6HI, 12HM−6HI, 24HI−6HI and 24HM−6HI.

Figure 3.3.1 shows the differences between the cyclone densities in the 6HI data and the 12HI, 12HM, 24HI and 24HM data, as identified by MAS. Generally, there is little difference between the 6HI cyclone density and the 12HI and 12HM cyclone density. Many locations have a difference of only ±2 to 5 cyclones per 1000 km^2 per timestep. The most noticeable differences are over areas with high topography, such as the Himalayas and the Rocky Mountains. There are also some differences in the 12 hour datasets in locations that show high activity of cyclones in Figure 3.1.1, such as the northwest tip of Greenland and the northern Sea of Japan.

Applying the MAS algorithm to the 12HI data results in an over–identification of cyclone centres, having a mostly positive anomaly, whereas apply it to the 12HM data results in under–identification of the cyclone centres, having a mostly negative anomaly. This is shown in Table 3.3.2 where the 12HI data has 13.36% of grid boxes positive and 10.44% negative. In contrast, the 12HM data is 10.64% positive and 12.57% negative. This is likely due to the temporal averaging of the field smoothing out short–lived features in the field, shortening tracks by one or two timesteps and, therefore, being filtered out by the minimum persistence

(of 24 hours) filter. Both the 12HI and 12HM cyclone densities have a similar RMSE score when compared to the 6HI cyclone densities.

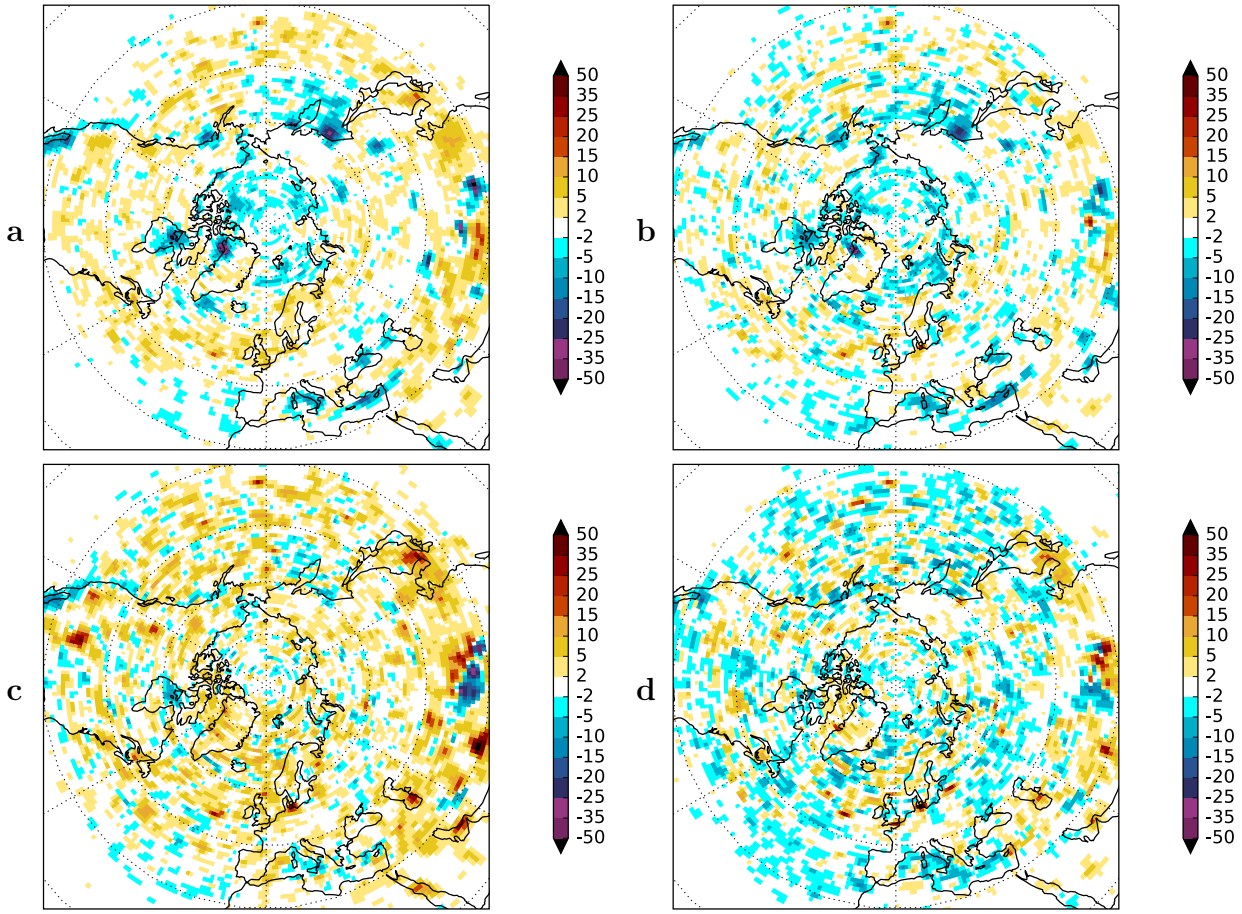


Figure 3.3.1: Difference in spatial distribution of the total cyclone centre density in the NH winter, for cyclones lasting 24 hours or more. The differences are between the 6 hour instantaneous data and (**a**) 12 hour instantaneous data, (**b**) 12 hour mean data, (**c**) 24 hour instantaneous data and (**d**) 24 hour mean data.

Applying the MAS algorithm to the 24HI and 24HM data, results in more locations with discrepancies between the 24 hour cyclone density and the 6HI data. Applying the MAS algorithm to the 24HI data is more likely to lead to an over–identification of the cyclone centres (21.39% positive, vs 6.91% negative, Table 3.3.2, whereas application to the 24HM data leads to a more even distribution of the errors between over– and under–identification (12.33% vs 13.81%). Again, this is likely due to the temporal smoothing of the field. Both the 24HI and 24HM show more areas where there is large over–prediction of cyclone centres,

being an over–prediction of 25% or more, than the 12HI and 12HM data. This is confirmed by the Root Mean Square Error (RMSE) values for the 12 and 24 hour temporal resolutions. Table 3.3.2 shows that the 24HI has a higher RMSE score than the 24HM and both RMSE scores for the 24 hour data are greater than they are for the 12 hour data. However, the overall picture is that the total cyclone density is not that different in the degraded data to that calculated from the original 6 hour data.

Difference	RMSE	% negative	% positive	% zero
12HI - 6HI	2.01	10.44	13.36	76.20
12HM - 6HI	1.88	12.57	10.64	76.79
24HI - 6HI	2.74	6.91	21.39	71.70
24HM - 6HI	2.20	13.81	12.33	73.85

Table 3.3.2: Errors in the track cyclone density for the degraded 12 and 24 hourly data, when compared to the 6 hourly data. Root mean square error (RMSE) between the degraded and 6HI data is shown, as well as the percentage of grid boxes in the differences that are positive, negative or zero.

This similarity between the feature objects identified in the degraded data and in the 6 hourly data is confirmed by Figure 3.3.2. This shows the distribution of lowest depressions, i.e. the lowest value of MSLP, encountered along a track for the original 6 hour data and the degraded 12 and 24 hour data. For each intensity bin, the percentage of tracks with that minimum MSLP is approximately the same whether the tracks are derived from the original 6 hour data, or the 12 or 24 hour data or the instantaneous or mean data.

The MAS algorithm shows the ability to work on temporally degraded data and still identify cyclones in approximately the same locations with approximately the same frequency. Additionally, the cyclones also have the same distribution of intensity (minimum MSLP value) between the 6 hourly data and the temporally degraded data. However, this part of the algorithm is static, identifying features in single timesteps of data. Therefore, as long as the temporal degradation does not reduce the number of time samples within the data too much the algorithm should successfully identify features.

The major challenge for a tracking algorithm working on degraded data comes with assigning these features to tracks (the correspondence problem, see section 2.4). As the data becomes

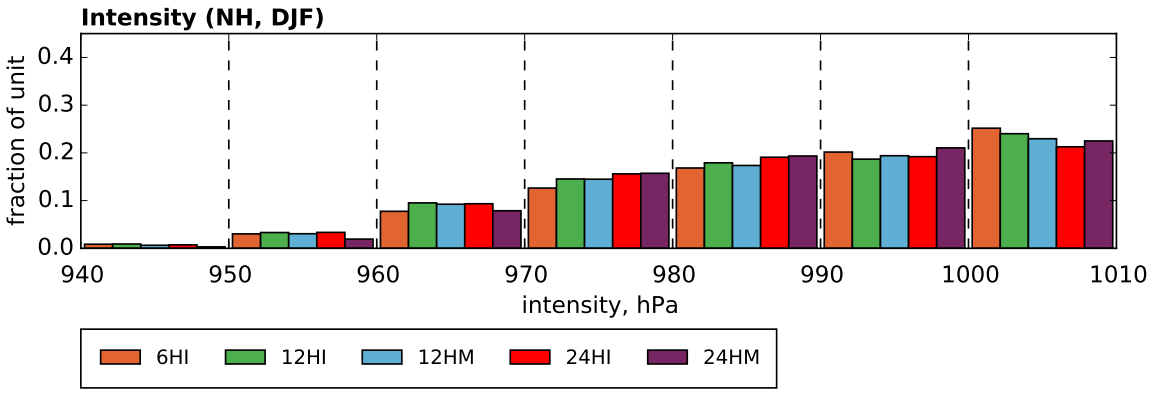


Figure 3.3.2: Normalised distribution of the intensity metric, the lowest depression encountered along a track, for the 6 hourly data (6HI), the 12 hour instantaneous data (12HI), the 12 hour mean data (12HM), the 24 hour instantaneous data (24HI) and the 24 hour mean data (24HM).

more temporally degraded, the feature points belonging to a track become more widely spaced out. Increasing the search radius to compensate for this allows more false assignments to be made, as there is the possibility of many more feature points being present in the enlarged search radius, which pass the series of hierarchical rules set out in section 2.4.2.

Figure 3.3.3 shows the effect this enlarged search radius can have on the distributions of the propagation speed, track length, lifetime and point displacement speed. The propagation speed measures the mean speed of the cyclone centre as it passes along the track from genesis to lysis. Figure 3.3.3 **a** shows that, by temporally degrading the data and increasing the search radius to enable systems of equivalent speeds to be tracked, the distribution of propagation speed is altered, resulting in more cyclone tracks with high propagation speeds at coarser temporal resolution. Applying the MAS algorithm to the 12 hour data and the 24 hour data results in a similar alteration of the distribution, with more systems with propagation speeds greater than 60kmh^{-1} and fewer systems with propagation speeds below 40kmh^{-1}, when compared to MAS applied to the 6 hour data.

This change in distribution is also reflected in the track lengths, as seen in Figure 3.3.3 **b**. Applying MAS to the 12 hour degraded data results in distributions with fewer short tracks than in the 6 hour data, and more long tracks. Both the 12HM and 12HI consistently over–

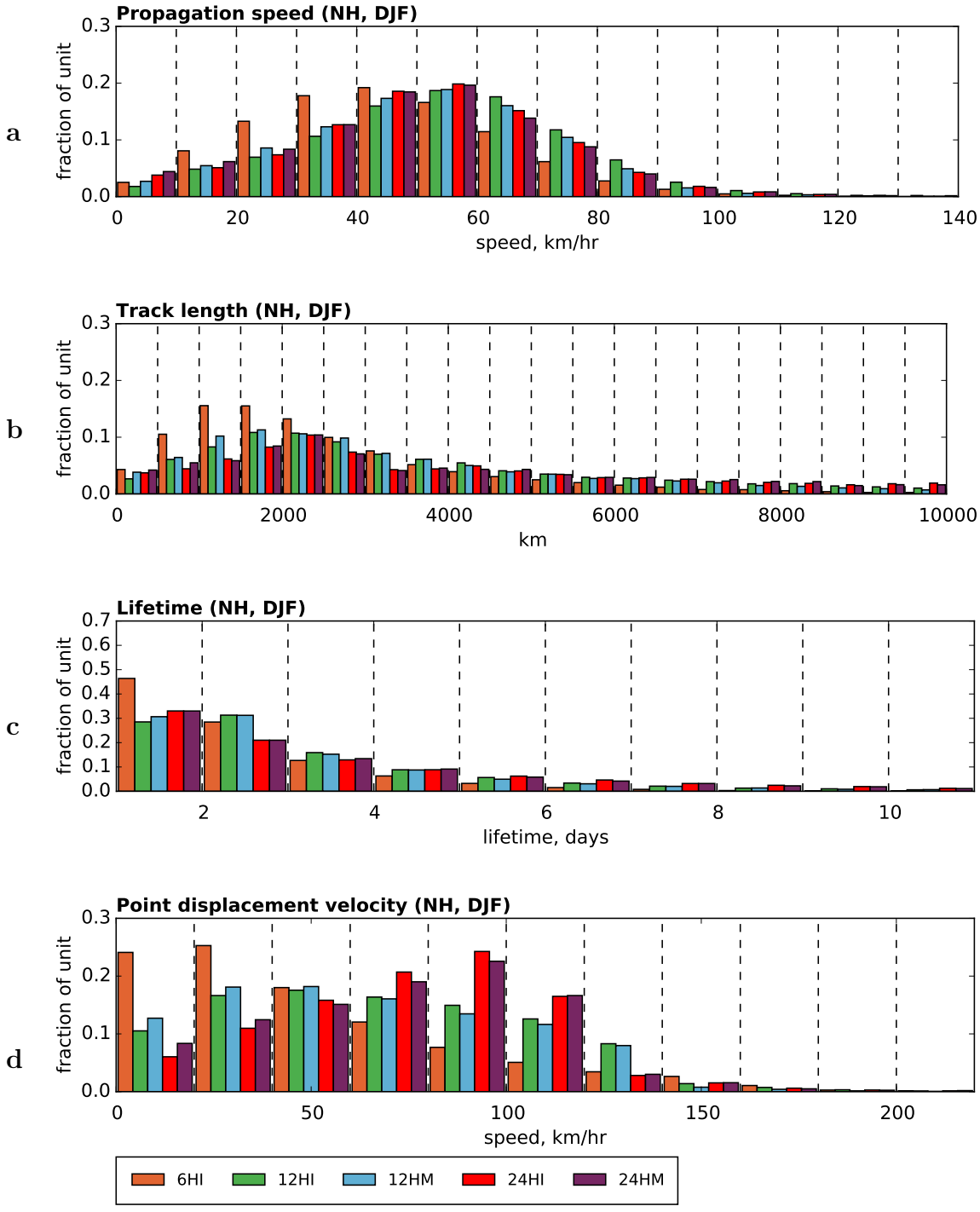


Figure 3.3.3: Normalised distribution of four track metrics for the cyclones identified by applying the feature detection and tracking algorithm to the ERA–Interim 6 hourly data for 1989 to 2009 Northern Hemisphere winters (DJF) and for the temporally degraded versions of the data. The metrics are mean propagation speed of the cyclone centre, the track length, the number of days the cyclone lives for and the point displacement speed (the speed at which a cyclone moves within one timestep).

estimate the number of tracks with a length ¿ 4000km and under–estimate the number of tracks with lengths between 0 and 2000km. Track lengths between 2000 and 4000km have more agreement with the 6 hour data. MAS applied to the 24 hour data produces track length distributions that are similar to those for the 12 hour data, especially for tracks with lengths $>$ 2000km. However, there are even fewer tracks with lengths less than 2000km in the 24 hour data than there are in the 12 hour data and many fewer than in the 6 hour data.

A large search radius encourages longer tracks by assigning feature objects to tracks where the feature object has no association to the track beyond the interpretation of the tracking algorithm. i.e the feature object is another cyclone system that has formed independently of the cyclone in the track, and is actually far away from the track, but the tracking algorithm has assigned it to the track as it is within the search radius and has passed the hierarchical rules.

Applying the MAS method to the degraded data results in a mismatch with the results from applying the method to the 6 hour data. The more degraded the data the greater the mismatch, with the 24 hour data having very different distributions of lifetime, track length, propagation speed and point displacement velocity, when compared to the 6 hourly data. This is consistent with Blender and Schubert (2000), who examined the differences between applying their tracking method to a sequence of temporally degraded data. They found some agreement between tracks produced from 6 hourly and 12 hourly data, but agreement in only 45% of tracks produced from 24 hourly data. Of the 15 methods studied in Neu et al. (2013), only one method (M06, Hewson and Titley (2010)) uses 12 hourly data, although they also use a "half–time" step, where each candidate feature is projected back to a half–time position using a steering wind. All of the other methods use 6 hourly data and so the main conclusion of this chapter is that, for the MAS method to perform well 6 hourly data must be used. Here, performing well means matching the spatial distribution of the 15 methods in Neu et al. (2013) (as seen in Figure 3.1.1) and matching the distributions of track metrics (as seen in Figure 3.1.2).

Figure 3.3.3 **d** shows the reason behind the differences in the tracks identified by the MAS

algorithm in the 6 hour data and those identified in the degraded data. The increase in the search radius, to compensate for the temporal degrading of the data, produces more fast moving systems and fewer slow moving systems. This is especially apparent in the 24HI and 24HM data where the mode is 120 to 140 kmh^{-1}. This is a very fast moving system, and only occurs in about 2.5% of tracks in the 6HI data. As discussed above, this is due to a too large search radius permitting feature objects to be added to tracks where they are not the same cyclone system, they are two separate cyclone systems with similar physical features, such as MSLP depth.

3.4 Summary and conclusions

This chapter investigates how well the feature–identification and tracking algorithm described in chapter 2 performs in relation to other feature tracking algorithms. The new method is applied to mean sea–level pressure (MSLP) data from the ERA–Interim reanalysis dataset Dee et al. (2011) for the Northern Hemisphere winter (DJF) and the Southern Hemisphere winter (JJA). The resulting cyclone centre density, cyclone intensity, lifetime and persistence are compared to the same metrics from applying 15 other tracking algorithms (Neu et al., 2013) to the same reanalysis. The new method is found to perform well, being consistent with the 15 other methods. Taking the output of the 15 methods as an ensemble, the output from the new method is statistically indistinguishable from the other ensemble members.

The new method is also applied to a higher resolution version of the ERA–Interim data from 1979 to 2009, over Europe and the North Atlantic only, and compared to the XWS windstorm catalogue (Roberts et al., 2014). The storm tracks in XWS are derived in a similar manner: the algorithm of Hodges (1999) is applied to the higher resolution ERA–Interim data. Comparing the tracks produced by the new method shows that there is a good match, both spatially and in terms of the MSLP depth along the track, for the six most destructive storms in the catalogue. Using the distance measure of Blender and Schubert (2000), it is shown that the method is able to successfully identify all but 2 of the 20 named tracks in the XWS catalogue.

Finally, the sensitivity of the tracking algorithm to the temporal resolution of the input data is investigated by degrading the ERA–Interim reanalysis to 12 hourly and 24 hourly data. The algorithm performs reasonably well on the 12 hourly data, but not as well as it does on 6 hourly data. However, on 24 hourly data it performs very poorly, allowing the false association of feature objects to tracks and permitting very fast moving tracks. The purpose of this investigation is to determine whether a smaller volume of data can be returned by the climate model running under volunteer distributed computing, so as to relieve the burden of the data transfer and storage. However, due to the reduction of the performance of the algorithm on degraded data, 6 hourly data will be returned from the distributed climate model. To reduce the burden of transfer and storage, the 6 hourly data will be restricted over a domain covering Europe and the North Atlantic, rather than the entire globe.

3.5 Future work

This chapter has shown that the proposed feature tracking algorithm works as well as the 15 methods detailed in Neu et al. (2013) and that, due to its computational efficiency, the algorithm is suited to deploying on the large ensemble of the HadAM3P climate model. However, three further investigations and improvements could be made as future work.

The first of these is to investigate the sensitivity of the tracking algorithm to the number of rules used in the hierarchy detailed in section 2.4.2, and the individual contribution to the overall tracking result of each rule. This investigation could be carried out by performing the same tracking analysis a number of times but omitting combinations of rules for each iteration. Calculating the same track statistics as in sections 3.1.1 and 3.1.2 could then show how each rule contributes to the overall tracking result.

Secondly, section 3.3 investigates the tracking algorithms' sensitivity to the temporal resolution of the input data and shows that, as the timestep increases, the false assignment of points to tracks also increases, leading to very long tracks with very fast propagation speeds. The section investigates the difference between 6 hourly, 12h and 24h data. However, Blender and Schubert (2000) shows that there is some false assignment of points when

moving from a 2h timestep to 4h and 6h timestep, in that there is a probability of false assignment of 0.06 and 0.15 respectively. Therefore it is reasonable to assume that there is still some false assignment of points even in the tracks produced from the 6h data in this thesis. To investigate this, the forecast component of ERA–Interim could be used as well as the analysis component to produce tracks, as the forecast component is available at 3h timesteps (Dee et al., 2011). Taking the tracks produced from this 3 hourly dataset as the reference track set, rather than the 6h tracks used in this chapter, would then allow the number of false assignments to be determined in the 6h track set.

Thirdly, the algorithms are designed to be computationally efficient, as they are used across a very large ensemble of climate model output. However, the efficiency could be further improved by removing the reliance on trigonometric functions in the tracking rules documented in section 2.4.2. Currently, the feature identification algorithm in 2.3 outputs the feature point locations in latitude–longitude coordinates, even though the identification is performing on the triangular mesh in Cartesian coordinates. The features are identified in Cartesian coordinates which are then transformed to latitude–longitude coordinates at the end of the algorithm and saved into the output file. Rather than doing this conversion, saving the feature point locations in Cartesian coordinates, essentially as points on the unit sphere, would allow more computationally efficient methods for determining distance and direction to be used, rather than the trigonometric functions currently used.

For example, Equation 2.4.3 (the haversine distance) could be replaced by the geodesic norm (Hodges, 1995):

$$d = \arccos\left(\boldsymbol{P_1} \cdot \boldsymbol{P_2}\right)$$

where $\boldsymbol{P_1}$ and $\boldsymbol{P_2}$ are the feature points and $\cdot$ denotes the scalar product of two vectors. To determine the angle between two points on the unit sphere, the tangent vector method of Hodges (1995) can be used to replace Equation 2.4.6 (the bearing angle) in Equations 2.4.7 and 2.4.9. Also, the projection to latitude–longitude coordinates of the geostrophic wind, performed by Equation 2.4.5 would also be unnecessary.

4 Design of a climate model validation experiment using a very large ensemble

Probabilistic event attribution studies that use GCMs and employ two climate scenarios, such as those by Otto et al. (2012), Pall et al. (2011) and Massey (2012), and as discussed in Section 1.3.2, rely upon large ensembles of a single climate model, with perturbations to the initial conditions and, sometimes, the boundary forcings, to explicitly simulate the statistics of extreme events. Extreme events, by their definition, occur in the tails of the distributions of climate variables and may only occur once in every 100 (or more) years. Using large ensembles allows a single year to be simulated over and over again and, by employing a 1000 member ensemble, a 100 year event should occur 10 times (on average). Therefore, using large ensembles, confidence can be gained in the statistics of these variables, particularly for the rare events occurring in the tails of the distribution.

However, if these large ensembles of a single climate model are to be compared to observed quantities of climate variables, their performance in capturing the distribution of weather events over a historical period should be evaluated. To do this, an ensemble of a comparable size to that used by the probabilistic event attribution study should be computed for every year in the historical period. This chapter describes the construction and execution of a very large ensemble of a single climate model covering a fifty–two year period from 1959 to 2010. By comparing the output from this ensemble with observed quantities, an evaluation of the suitability of using the model in probabilistic event attribution studies can be made.

4.1 The HadAM3P climate model

The model used in the validation exercise is the HadAM3P atmosphere only general circulation model (AGCM) from the Hadley Centre at the UK Met. Office. HadAM3P is based upon the atmospheric component of HadCM3 which is the Hadley Centre's coupled atmosphere-ocean general circulation model (AOGCM). The atmospheric component of this model is discussed in Pope et al. (2000) and hereafter referred to as HadAM3. The study by Pall et al. (2006) and Pall et al. (2011) uses a higher resolution version of HadAM3 than used in this study. However, Pope and Stratton (2002) showed that there is no benefit in increasing the horizontal resolution, if the number of vertical levels remains constant. Also, the model used in this study has some noticeable differences and improvements to the physical parameterisations when compared to both the standard resolution HadAM3 and the higher resolution version used by Pall et al. (2006).

An atmosphere only model has a number of advantages over a coupled ocean–atmosphere climate model, when run as part of a distributed volunteer computing project. Firstly, as the sea–surface temperatures (SSTs) are prescribed, there is no need to spin–up the ocean component of the model to reach a stable state. In other climate modelling volunteer computing projects, spinning up the ocean has represented a large proportion of the overall computing amount (Frame et al., 2009). Secondly, by prescribing the SSTs, the large biases in the atmospheric circulation in climate models that can be linked to the biases in the ocean are removed (Scaife et al., 2010). Finally, and related to the first point, it requires less computation time to run an integration of an atmosphere only climate model for a model year, than it does to run a coupled ocean–atmosphere climate model. This allows more ensemble members to be run, so that the tails of the distributions of the climate variables can be sampled more thoroughly. It also allows a higher resolution atmospheric component of the climate model to be run, which can provide an improvement on the representation of the atmospheric circulation (Pope and Stratton, 2002).

HadAM3P is a grid point model that divides the globe into a number of latitude and longitude points, which are spaced evenly in latitude and evenly in longitude. The height of the

atmosphere is represented by a number of vertical levels. The initial state of the model is supplied by an initial condition file and the model is then iterated forward on a timestep by timestep basis. The model solves equations of motion (for a quasi-hydrostatic atmosphere) and radiative transfer explicitly on the same scale as the grid. However, there are also a number of sub-grid scale physical processes, primarily thermodynamic, which are represented by physical parameterisations. These are approximations to the actual physical process and are usually represented by a "bulk formula" that empirically represent the physical behaviour of the variable being parameterised. It is the changes in these parameterisations that form the majority of the differences between the standard HadAM3 and HadAM3P.

One major change from the atmospheric component of HadCM3 is an increase in the horizontal resolution from N48 ($2.5 \times 3.75°$) to N96 ($1.25 \times 1.875°$). Pope and Stratton (2002) and Stratton (2004) show that, by increasing the horizontal resolution, the simulation of the North Atlantic storm track improves in terms of the position and intensity of storms. Improvements are also seen in the precipitation, which moves from a convective-precipitation dominated scheme in the N48 resolution to having more large scale precipitation in the mid-latitudes at the higher resolution. The vertical resolution of 19 levels remains unchanged from HadCM3 as Pope et al. (2001) found that increasing the vertical resolution to 30 levels improves the representation of moisture in the upper troposphere and stratosphere at the expense of a degradation in the tropical convective rainfall. As HadAM3P is to be used to study mainly surface and near surface climate variables, the upper troposphere and the stratosphere are of less interest and so the vertical resolution is not increased. This also has the effect of not further increasing the computational burden. HadAM3P has a timestep of 15 minutes, which is half that of HadAM3 at N48, as a shorter timestep is required when the horizontal resolution is increased in order to maintain numerical stability (Pope et al., 2000).

The method of solution to the equations of motion and radiative transfer remain the same between HadAM3 to HadAM3P, but there are a number of improvements to the sub-grid scale physical parameterisations and the modelling of atmospheric chemistry.

Firstly, the calculation of the cloud cover is improved, from a scheme that assumes clouds fill the entire volume of a grid box at a given model level, to a scheme that calculates the cloud fraction at three separate sub-levels within the grid box. The cloud fraction amount at the sub-levels are obtained by vertically interpolating a saturation variable between the levels of the model and the overall cloud fraction amount is taken to be the maximum cloud fraction of each sub-level within the grid box. This results in larger values of cloud fraction, with a corresponding reduction in the optical depth of the cloud, which more closely matches observations than the values given by HadAM3 (Massey et al., 2015).

Secondly, the calculation for the relative humidity threshold at which clouds form is improved. In HadAM3, the threshold is calculated by a quadratic spline interpolation that relies on a fixed value (Rh_{crit}) which varies between model levels but does not vary with time, geography or pressure. This is improved by calculating Rh_{crit} from the q_c prognostic variable, which is the difference between the total water in the grid box and the saturation vapour pressure. HadAM3P calculates Rh_{crit} for each grid box from the standard deviation of q_c in the 3×3 surrounding grid boxes, multiplied by a function of pressure (Massey et al., 2015).

Thirdly, the radiative effects of convection benefit from an improved calculation. In HadAM3, there is no accounting for the presence of anvil clouds as the cloud fraction is assumed constant between the base and top of the cloud, which leads to errors in the estimation of the optical thickness of the cloud. HadAM3P partitions the cloud into two types, deep clouds and non-deep clouds. Deep clouds are defined as those whose base extends to the boundary layer and their tops extend to above the freezing layer. For these types of cloud, the scheme increases the cloud fraction linearly with height from the freezing layer to the top of the cloud. This represents the anvil. To represent the convective tower, the cloud fraction is decreased to a constant value below the freezing layer. For the non-deep clouds the scheme remains as in HadAM3 (Massey et al., 2015). In addition, an improvement is made to the calculation of the cloud water amount by excluding convective precipitation from the water path in the radiation scheme and by scaling the concentration of cloud water to account for it being a small fraction of the cloud from the convective updraught (Massey et al., 2015).

Fourthly, while HadAM3 includes a sulphur cycle scheme which models the interactions of both natural and anthropogenic sulphate aerosol emissions, the version of HadAM3 used by Pall et al. (2006) and Pope and Stratton (2002) do not. HadAM3P incorporates the sulphur cycle from HadAM3 and makes several improvements to it. The model has five extra prognostic variables to model the distribution of sulphate aerosol, which are the mass mixing ratio (MMR) of sulphur dioxide (SO_2), the MMR of dimethyl sulphide (DMS), sulphate dissolved in clouds and two other sulphate modes. These modes are assumed to have log-normal distributions and represent the Aitken mode, with a mean particle radius of 24nm and the accumulation mode with a mean particle radius of 95nm (Massey et al., 2015). Transport of these five variables via horizontal and vertical advection, convection and turbulent mixing is simulated. The variables needed by the chemical reaction equations which handle the oxidisation of DMS to SO_2 and SO_2 to sulphate are contained within three dimensional fields of the hydroxyl radical (OH), hydrogen peroxide (H_2O_2) and the peroxide radical (HO_2). These fields are obtained by off-line runs from the STOCHEM model described in Collins et al. (1997) and are supplied as a year-long cycle of monthly values.

The standard inputs to the HadAM3P sulphur cycle scheme are the surface anthropogenic emissions of SO_2, elevated emissions (e.g. from a chimney stack) of SO_2 which are released at a higher model layer than the surface, natural emissions into the stratosphere of SO_2 from regularly erupting, but small scale, sources and natural emissions of DMS, primarily from phytoplankton in the ocean. An extra modification to HadAM3P for this project is the addition of large scale volcanic eruptions emitting large quantities of SO_2, such as the Mount Pinatubo eruption in 1991. These large natural emissions are modelled as a change in optical depth in the stratosphere for four latitude bands, with a value prescribed for each band for each model year. As these eruptions are modelled as stratospheric aerosol loading, they do not feature in the interactive sulphate chemistry scheme in the model, but just as values that are injected into the stratosphere and decay as prescribed in the input file. Input files are required for the four emission types above, as well as the 3D fields for the oxidisation variables and the stratospheric concentration of volcanic sulphates.

The sulphur cycle in the model affects the radiation budget by simulating the direct effect of

aerosols, which reduces the radiation reaching the lower atmosphere and surface by absorption and scattering consistent with Mie theory, and by simulating the first indirect effect. This reduces incoming radiation by increasing cloud albedo, causing more of the incoming short wave radiation to be reflected back to space.

The first indirect effect, also called the Twomey effect after Twomey (1974), is caused by sulphate aerosols creating larger numbers of cloud condensation nuclei (CCN), and causing an increase in cloud albedo. This has a negative effect on the amount of radiation reaching the surface and so has a net cooling effect on the atmosphere. The second indirect effect, which is a reduction in the amount of precipitation formed within a cloud, due to the sulphate aerosols (Houghton et al., 2001) is not modelled in HadAM3P.

Finally, small changes to the land surface scheme are implemented. In HadAM3, heat is transported from the land surface to the underlying soil using a conduction term. In HadAM3P this is changed to a radiative term, which leads to an improved diurnal cycle and a more realistic distribution of minimum temperatures associated with winter soil freezing. In HadAM3P, the upward surface long wave radiation flux is updated every 15 minutes (every model timestep) as opposed to every 3 hours (every 12 model timesteps) in HadAM3. This has the effect of removing hot spots in very arid regions of the model, where the temperatures are allowed to increase unrealistically due to a lack of cooling via evaporation (Massey et al., 2015).

4.2 Forcings at the boundary layer

As HadAM3P is an atmosphere only model, it requires forcing conditions to be applied at the boundary between the ocean and the atmosphere. These forcings are the sea-surface temperatures (SST), which are the mean temperatures of the sea at the surface of the grid boxes, and sea ice fractions (SIF), which are the proportions of the grid boxes covered by ice.

4.2.1 Sea–surface temperatures

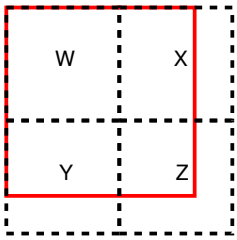
The sea-surface temperatures (SST) used in the model over the 52 year validation period are derived from HadISST v1.1. HadISST is a gridded dataset from the UK Met. Office Hadley Centre containing global coverage of monthly mean SSTs on a $1^\circ \times 1^\circ$ grid, from 1870 to the present day. The HadISST dataset is updated monthly, and lags 2 months behind the current day. The input data and methods for constructing the global coverage are described in close detail in Rayner et al. (2003) but the main points are outlined here.

From 1870 to 1981, only in-situ measurements of SST are used. These are taken from individual ships' observations and do not have global coverage at any point in the time period. These data are bias corrected, passed through a quality control check and then reconstructed onto the $1^\circ \times 1^\circ$ grid. From 1982 onwards, the in-situ measurements are combined with satellite data from the advanced very high resolution radiometer (AVHRR). These data are also bias corrected and are then combined with the in-situ data for the corresponding period, to construct a continuous dataset from 1870 to the present day.

Although HadISST is designed to drive atmospheric models (Rayner et al., 2003), and has been used in the ERA-40 analysis (Uppala et al., 2005) for the 1958 to 1981 period, some extra work is required to enable HadISST SSTs to be used in the HadAM3P climate model. These include temporal interpolation, spatial regridding and dealing with discrepancies in the land sea mask between HadISST and HadAM3P.

Spatial regridding

HadISST is defined on a $1^\circ \times 1^\circ$ grid, whereas HadAM3P employs a $1.875^\circ \times 1.25^\circ$ grid, in the longitude and latitude, and requires its input forcing field data to be at the same resolution. In order to preserve the large scale averages from the original data, an area-weighted averaging method is used, as shown in Figure 4.2.1. For each box in the HadAM3P grid, the proportion of each of the grid boxes in the original grid which intersect the HadAM3P grid box is calculated, and this proportion forms the weight in the averaging function.



A_H = Area of HadAM3P grid box
A_W = Overlapping area of W grid box from HadISST grid
V_W = Value of data in W grid box
W_W = Contributing weight of W grid box = $\frac{A_W}{A_H}$
$V_H = W_W \times V_W + W_X \times V_X + W_Y \times V_Y + W_Z \times V_Z$

Figure 4.2.1: Contributing weights for area-average regridded from HadISST data (black grid) to HadAM3P grid (red grid), where V_H is the value of the regridded data in the HadAM3P grid

Temporal interpolation

The SST fields in HadISST are given as monthly mean values. Woollings et al. (2010) show that the North Atlantic storm track in high resolution atmosphere only climate models is sensitive to the temporal resolution of the forcing SSTs, and that increasing the temporal resolution, from monthly mean values of SST to weekly mean values, leads to an improvement in storm track density, more closely matching ERA-40. This is thought to be due to an increase in horizontal SST gradients leading to greater baroclinicity and also to an increase in the variance of the air-sea contrast. As the SST values in HadISST are monthly, the actual weekly values of the SSTs cannot be recovered or reconstructed. However, Sheng and Zwiers (1998) present an interpolation method that recovers some of the variability in a time series from periodic means, when interpolating from the means to points more frequent than the period.

Given a time series of monthly mean SSTs at a single grid box, an AGCM will typically interpolate the monthly values to daily values using a simple linear interpolation scheme. Such a scheme has two disadvantages: the mean of the interpolated daily values will not equal the original monthly mean and the interpolated values are smoothed, losing inter-month variability (Sheng and Zwiers, 1998). The method of Sheng and Zwiers (1998) overcomes these problems by adjusting the values of the monthly means, then linearly interpolating between the adjusted values. The adjusted values are derived by constraining the interpolated values over a month so that the mean of interpolated values for that month is equal to the original

monthly mean. Given the original temperatures as a vector, $\vec{T_o}$, the adjusted temperatures $\vec{T_t}$ are found, subject to the constraint that:

$$\int_{j-1}^{j} T(t)dt = \bar{T}_j \quad \text{for} \quad j = 0 \ldots N \tag{4.2.1}$$

where $\bar{T}_j$ is the original monthly mean for month j in the timeseries and $T(t)$ is a function that interpolates data for a time t, where $0 \leq t \leq N$ and N is the number of months in the original timeseries.

Denoting the vector of adjusted values as:

$$\vec{T_t} = (\tilde{T}_1, \tilde{T}_2, \ldots, \tilde{T}_N)^T$$

and the vector of original observed SST values as:

$$\vec{T_o} = (\bar{T}_1, \bar{T}_2, \ldots, \bar{T}_N)^T$$

These two equations can be related using the equation:

$$A\vec{T_t} = \vec{T_o} \tag{4.2.2}$$

To find A, the constraint of Equation 4.2.1 along with a simple linear interpolation function can be used. This function is typically used to interpolate monthly mean data to sub-monthly values and is given by Sheng and Zwiers (1998) as:

$$T(t) = \begin{cases} \tilde{T}_1 & t \in \left[0, \frac{1}{2}\right] \\ \tilde{T}_j + (\tilde{T}_{j+1} - \tilde{T}_j)(t - j + \frac{1}{2}) & t \in \left[j - \frac{1}{2}, j + \frac{1}{2}\right] \\ \tilde{T}_N & t \in \left[N - \frac{1}{2}, N\right] \end{cases}$$

Integrating this equation, as in the constraint in Equation 4.2.1, gives:

$$\int_{j-1}^{j} T(t) = \begin{cases} \frac{7}{8}\tilde{T}_1 + \frac{1}{8}\tilde{T}_2 & j = 1 \\ \frac{1}{8}\tilde{T}_{j-1} + \frac{3}{4}\tilde{T}_j + \frac{1}{8}\tilde{T}_{j+1} & 2 \leq j \geq N-1 \\ \frac{1}{8}\tilde{T}_{N-1} + \frac{7}{8}\tilde{T}_N & j = N \end{cases}$$

which, in turn, gives A in Equation 4.2.2 as an $N \times N$ matrix:

$$A = \begin{bmatrix} \frac{7}{8} & \frac{1}{8} & & & & & & \\ \frac{1}{8} & \frac{3}{4} & \frac{1}{8} & & & & & \\ & \frac{1}{8} & \frac{3}{4} & \frac{1}{8} & & & & \\ & & \cdot & \cdot & \cdot & & & \\ & & & \cdot & \cdot & \cdot & & \\ & & & & \cdot & \cdot & \cdot & \\ & & & & & \frac{1}{8} & \frac{3}{4} & \frac{1}{8} \\ & & & & & & \frac{1}{8} & \frac{7}{8} \end{bmatrix}$$

As A is a tridiagonal matrix, Equation 4.2.2 can be solved to find $\vec{T_t}$ by backwards substitution and Gaussian elimination (Burden and Faires, 2001). This gives the adjusted values for each monthly mean, which can then be linearly interpolated between.

To use the interpolation scheme on SSTs, a modification has to be made. Certain grid boxes in the SST and SIF data will fluctuate between having zero sea ice cover and some sea ice cover over the period of a year. When a grid box has sea ice cover of over 40% the HadAM3P model will set the SST for that grid box to -1.8°C (the average freezing point of sea water) and the HadISST data set reflects this, also setting grid boxes with sea ice cover to -1.8°C. If the above scheme is applied to data where the original means have a constant value of -1.8°C for a number of months, then the first adjusted mean value could be below -1.8°C. This can be pragmatically corrected by resetting all values below -1.8°C to -1.8°C. To further prevent this having an effect on the interpolated data, open water periods are identified in the data for each grid box. These are time periods where there is no

sea ice cover and the value of N in the above scheme is set to the length of the open water period, with the calculation of adjusted values just being applied to the open water period. Although this prevents the algorithm preserving the mean over the entire timeseries, it does ensure that the mean is preserved in each open water period, and that there are no values below $-1.8°$C.

The interpolation scheme of Sheng and Zwiers (1998) is applied to regridded SSTs from HadISST, on a grid box by grid box basis, to generate SSTs at 5 daily intervals. This is analogous to the weekly SSTs that led to an improvement in storm track density in Woollings et al. (2010). Figure 4.2.2 shows an application of the Sheng and Zwiers interpolation scheme and a linear interpolation scheme to a single grid box of data from HadISST over a five year period, for two different locations. Figure 4.2.2(a) shows a location that has open water all year round. The error plot shows that the Sheng and Zwiers method has a lower mean error more consistently than the linear method. Figure 4.2.2(b) shows the results from a location that has periodic ice cover. Again, the Sheng and Zwiers method performs better overall, albeit with noticeable errors around the start and end of an open water period.

To investigate how the constraint of Equation 4.2.1 improves the preservation of the monthly mean, the monthly mean of the interpolated values are computed for both the Sheng and Zwiers interpolation method and the simple linear interpolation. As Equation 4.2.1 involves an integral, the means are calculated from data interpolated from HadISST with an increasing number of interpolation points. Figure 4.2.3 shows the sum of the error between the monthly means of the interpolated values and the actual HadISST monthly means for the same data period in 4.2.2(b) (the location with intermittent ice cover) for an increasing number of interpolation points. As expected, increasing the number of interpolation points decreases the error in the monthly mean of the interpolated values for the Sheng and Zwiers method but not for the simple interpolation scheme. Choosing the number of interpolated values to use in the distributed computing experiments is a balance between the size of the SST and SIF files used, which must be downloaded by the volunteers, and the error in the interpolation method. To place less burden on the volunteers, a small download size is required whereas, to improve the interpolation, a high number of interpolation points is

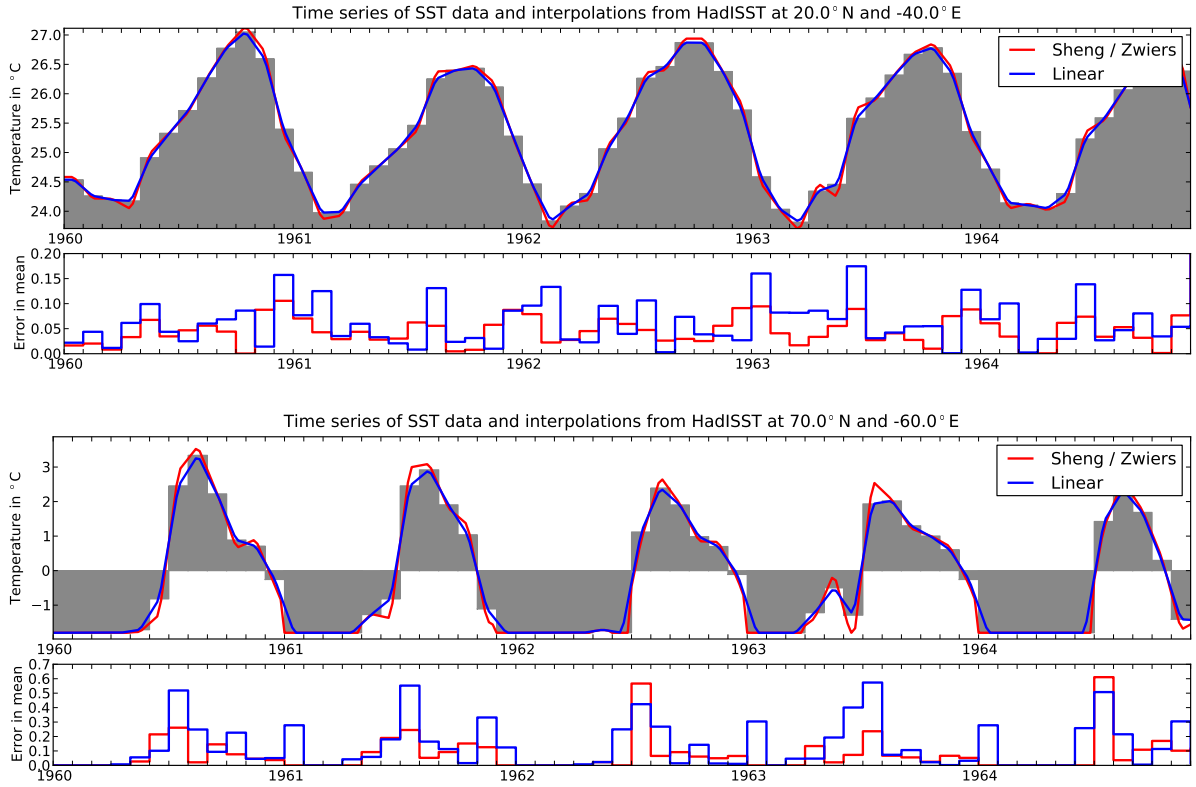


Figure 4.2.2: Timeseries of 5 years of monthly mean SST (grey) and the interpolation to 5 day intervals by the Sheng and Zwiers method (red) and a simple linear interpolation (blue) for *(a)* a location that has open water all year round and *(b)* a location that has periodic ice cover. The grey histogram shows the original monthly mean temperatures at the location. The square root of the mean squared error is shown below each location.

needed. In this thesis, 6 interpolation points are used. This reduces the error while also keeping the download size of the SSTs and SIFs to a manageable number.

Land sea mask considerations

The HadISST data set and the HadAM3P model are both defined on a global grid and rely on a land sea mask (LSM) to indicate which part of the grid represents ocean and which part represents land. The LSM is a simple two dimensional array whose size matches the grid and contains 0s to mark sea and 1s to mark land. The difference in resolution between the HadISST and HadAM3Ps grids can lead to a problem where latitude / longitude coordinates marked as land in the HadISST grid are marked as ocean in HadAM3P, even after the

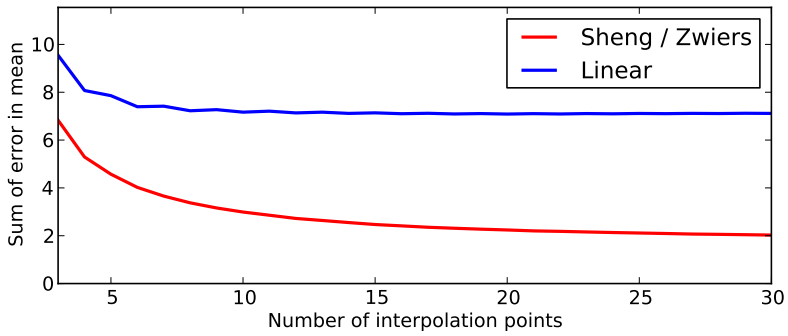


Figure 4.2.3: Error in the monthly mean of the interpolated values against the actual mean from HadISST for increasing numbers of interpolation points.

regridding technique of Figure 4.2.1 has been applied. This can lead to missing data in the ocean grid boxes in the HadAM3P grid as no values have been assigned to them from the HadISST grid.

A pragmatic solution to this problem is used so that where missing data occurs in the HadAM3P grid, the average of the 8 surrounding grid boxes is assigned to the grid box with missing data, as long as the surrounding grid boxes do not contain missing data. If one of the surrounding grid boxes is missing data, then the 7 surrounding, non-missing data grid boxes are used and so on. This technique generally works well as the missing data is mostly along coastlines or lake shores so at least one of the surrounding grid boxes is defined as ocean and not land and will have a value to use. However, HadAM3P does contain ocean grid boxes that have 8 surrounding land grid boxes and, hence, for those ocean grid boxes the above approach will not work. For these special cases, a scan is made along the latitude line that the grid box belongs to, in both the East and West direction, until a grid box is found in each direction that contains a non-missing value. The mean of these two values is used in the ocean grid box with missing data. One caveat with this method is that it will assign values to lakes which are derived from sea-surface temperatures and does not consider topography or lake depth when assigning these values. Therefore, the values at these grid boxes are likely to be incorrect. However, this affects very few grid boxes over the entire dataset and so is considered an acceptable compromise in light of the lack of the data which would be required to compute these values correctly.

4.2.2 Sea ice fraction

HadAM3P also requires sea ice fractions (SIF) to be prescribed at every grid point over the time period of the experiment. The SIF is the proportion of the grid box that is covered by ice and HadISST v1.1 is used again to supply this information. As with the SSTs, in HadISST the SIF is defined as monthly means on a $1^\circ \times 1^\circ$ global grid and so some extra work is again required to temporally interpolate the SIF and also to regrid from the HadISST grid to the HadAM3P grid.

In HadISST several different sources of sea ice concentrations are used together to produce global coverage sea ice fraction maps. These include digitised hand drawn sea ice charts, such as the Walsh Northern Hemisphere Sea Ice Concentration Charts, which record month end sea ice concentrations in the Northern Hemisphere from 1901 to 1995. Other digitised sea ice charts from the National Ice Centre are used, along with Great Lakes ice charts and Antarctic ice climatologies (Rayner et al., 2003). From 1978 onwards, passive microwave measurements from satellites are also used. These datasets are combined, along with the SST datasets used in section 4.2.1 to produce global coverage of monthly mean SIF (Rayner et al., 2003).

To use the HadISST SIF dataset in the HadAM3P model, it is first regridded to the HadAM3P grid using the area-averaging method detailed in Figure 4.2.1 and is then temporally interpolated using the Sheng and Zwiers method outlined in section 4.2.1. A final consistency check is applied to both the SST and SIF regridded and interpolated datasets. This checks that ice is present in grid boxes with a temperature of -1.8°C and, if it is not, the SIF for that grid box is set to the average of the SIF in the surrounding 8 grid boxes. Also, a check is made to ensure that grid boxes with a SIF above a certain threshold (set to 0.9 in this experiment, to match the value used in the model) have a temperature of -1.8°C.

4.2.3 Sea–surface temperatures and sea ice fraction forcings summary

The use of the HadISST dataset ensures that the SSTs and SIF for the period the experiment covers is as close to observations as possible. Using the Sheng and Zwiers algorithm for temporal interpolation adds some variability back into the dataset while retaining the monthly means of the SSTs and SIF. The SST and SIF datasets produced are used in the very large ensemble experiment to force the climate model at the boundary between ocean and atmosphere.

4.3 Atmospheric composition

To model the period from 1959 to 2010, HadAM3P also requires the specification of the atmospheric composition. This includes greenhouse gases, sulphate concentrations, ozone and other trace gases. This section details the data sources for these gases, and how their interactions are modelled in HadAM3P.

4.3.1 Greenhouse gas concentrations

The absorption and emission of electromagnetic radiation by the atmosphere and surface are represented in HadAM3P by the radiation scheme, which models the solar spectrum, the absorption by radiatively active gases (including anthropogenically emitted greenhouse gases and water vapour), the terrestrial radiation cascade (as a series of discrete layers), scattering due to atmospheric gases and aerosols and the albedo effect due to clouds and ice. To carry out this modelling, a number of inputs are needed, such as the solar spectrum and the abundance of the radiatively active atmospheric gases. The following sections detail the specification of these gases in the ensemble validation experiment.

In order to model future climate change, the Intergovernmental Panel on Climate Change (IPCC) make available projections of greenhouse gas concentrations, subject to various sce-

narios. The scenarios detailed in the Special Report on Emission Scenarios (SRES, Nakicenovic and Swart (2000)) are used in the IPCC Third Assessment Report (Houghton et al., 2001) and the Fourth Assessment Report (Solomon et al., 2007). The scenarios are based on modelling human activity in terms of economic development, energy use, land use and population. This thesis uses the A1B scenario for greenhouse gases on the recent past. This is necessary as, although observations of CO_2 concentrations are available for the period from 1959 to 2000, they are not for other greenhouse gases for more recent years, or for CO_2 from 2000 to 2010, and so the scenario is used to complete the dataset. The emissions scenarios are continuous, so that when the period where observations are available stops (the year 2000 for CO_2), the modelled scenario continues the emissions smoothly. Greenhouse gas emissions increase at a similar rate as 1986 to 1990 and so A1B can be thought of as a "business as usual" scenario, which is suitable for near-term projections.

Carbon dioxide

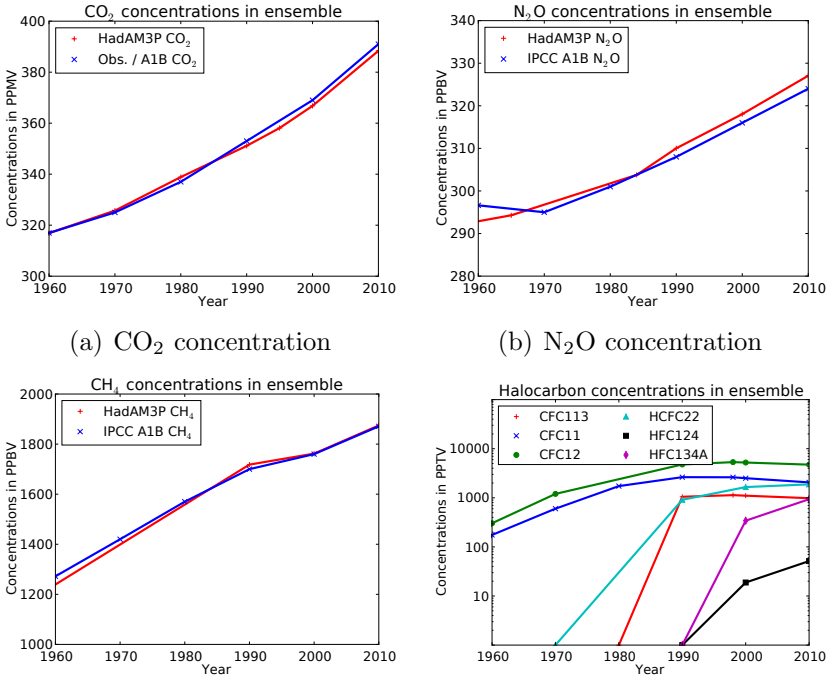
Carbon dioxide (CO_2) is the most abundant greenhouse gas (GHG) but also the least powerful in terms of the absorption of longwave radiation per unit mass. However, CO_2 is still the greatest contributor of the GHGs to anthropogenic global warming due to its abundance (Solomon et al., 2007). Measurements of the amount of CO_2 in the atmosphere have been taken at the Mauna Loa observatory since 1956 and show that concentrations of CO_2 have increased by about 17% over a 35 year period, from 325 PPMV in 1970 to 379 PPMV in 2005 (Solomon et al., 2007).

HadAM3P requires the concentration of CO_2 in the atmosphere to be specified as a mass mixing ratio (MMR), which is simply the ratio of the mass of CO_2 in the atmosphere to the mass of the other gases in the atmosphere. The IPCC assessment reports (such as Houghton et al. (2001), Solomon et al. (2007)) define the atmospheric concentration of gases in parts per million by volume (PPMV) or parts per billion by volume (PPBV). Therefore, to use IPCC figures for CO_2 concentrations in HadAM3P, they must be converted from PPMV to MMR, using the formula:

$$MMR = \frac{PPMV}{10^6} \times \frac{M_{CO_2}}{M}$$

where M is the molar mass of dry air, equal to 28.97 (Andrews, 2000), and M_{CO_2} is the molar mass of CO_2, equal to 44.01 (Houghton et al., 2001).

The concentration of CO_2 used in the large ensemble experiment has been supplied by the UK Met Office Hadley Centre (David Hassell, *pers. comms*) but is consistent with both observations and the IPCC SRES scenario A1B. Figure 4.3.1a shows the concentrations, in PPMV, of CO_2 for both the values used in the ensemble and the values of observations from 1960 to 2000 (from the Mauna Loa observatory: Tans and Keeling (2011)) and the IPCC SRES A1B scenario thereafter (Nakicenovic and Swart, 2000).



(a) CO_2 concentration

(b) N_2O concentration

(c) CH_4 concentration

(d) Halocarbon concentrations

Figure 4.3.1: Trace gas concentrations in the ensemble modelling experiment. *(a)* CO_2 concentrations in the ensemble and compared to IPCC SRES A1B. *(b)* N_2O concentrations and compared to A1B. *(c)* CH_4 concentrations and compared to A1B. *(d)* concentrations of the six halocarbons represented in the ensemble.

Nitrous oxide

Although Nitrous oxide (N_2O) is much less prevalent in the atmosphere than CO_2, with concentrations measured in parts per billion by volume (PPBV) rather than per million (PPMV), its contribution per unit mass to the radiative forcing of the atmosphere is greater than CO_2 (Solomon et al., 2007). Therefore, N_2O is a very important GHG.

N_2O concentrations for the validation experiment are again obtained from the UK Met Office Hadley Centre. Observed values for N_2O are not as readily available as for CO_2, and so values from the SRES A1B scenario are used for later dates in the modelled period. Figure 4.3.1b shows the Met. Office supplied values plotted against the values from the IPCC SRES A1B scenario from 1970 to 2010 and a single observed value for 1960 from the IPCC First Assessment Report Table 2.5 (IPCC, 1990). Again, the abundances of N_2O used in the validation experiment are largely consistent with values from the observations and SRES A1B.

Methane

Methane (CH_4) is the second most important GHG (Solomon et al., 2007), due to its relatively large abundance ($\sim$10 times that of N_2O but still $\sim$100 times less than CO_2). Figure 4.3.1c shows the concentrations of CH_4 used in the ensemble validation experiment, plotted against the values from IPCC SRES A1B for 1970 to 2010 and the single value from IPCC (1990) Table 2.5 for 1960. Once again, the values of CH_4 were obtained from the UK Met. Office Hadley Centre and show good agreement with the IPCC SRES A1B values, and single observation.

Halocarbons

The halocarbons are a group of very powerful GHGs, including varieties of chlorofluorocarbons (CFCs), hydrofluorocarbons (HFCs) and hydrochlorofluorocarbons (HCFCs). They are also responsible for the depletion of the ozone layer and their emission is controlled under the Montreal Protocol (Solomon et al., 2007). HadAM3P requires the specification of 6 halocar-

bons: CFC113, CFC11, CFC12, HCFC22, HFC124 and HFC134A. Although halocarbons are very powerful absorbers and emitters of long wave radiation (Solomon et al., 2007), their abundance in the atmosphere is measured in parts per trillion (PPTV), as opposed to PPMV in the case of CO_2. Figure 4.3.1d shows the concentrations of the halocarbons used in the ensemble validation experiment. These concentrations were, once again, obtained from the UK Met Office Hadley Centre. They compare favourably with Figure 2.6 from the IPCC 4th Assessment Report (Solomon et al., 2007).

4.3.2 Ozone

Ozone (O_3) is an important gas in the stratosphere, as it produces localised heating via absorption of both infra-red and ultra-violet radiation (Andrews, 2000). O_3 has seen large changes in the period modelled, due to the emission of chlorofluorocarbons (CFCs) which deplete O_3 via a breakdown of the CFC into molecular chlorine (Cl_2), which is in turn photo-dissociated into atomic chlorine (Cl), which then destroys the O_3 in a fast catalytic reaction (Andrews, 2000). This has created a "hole" in the ozone layer appearing above Antarctica in the spring. The decrease in O_3 around this hole may have climatic effects, including the cooling of the stratosphere and troposphere (Andrews, 2000) and so, a realistic O_3 profile should be used in any modelling of extreme weather events (Pall, 2006).

HadAM3P requires O_3 to be supplied as zonal mean values for each level in the model. The O_3 profile was obtained from the UK Met Office Hadley Centre and its time mean profile for two periods is shown in Figure 4.3.2. The ozone hole can be seen in the 1990-2000 period in the O_3 concentration at the levels from 109hPa to 13hPa at latitudes south of 55°S.

4.3.3 Natural and anthropogenic sulphate concentrations

As noted in section 4.1, the HadAM3P climate model features an interactive sulphate aerosol scheme. Interactive, in this sense, indicates that the atmospheric chemistry associated with sulphate gases is modelled in a time varying manner. Emissions from four different sources of sulphate gases are specified as monthly mean values for each month in the period modelled.

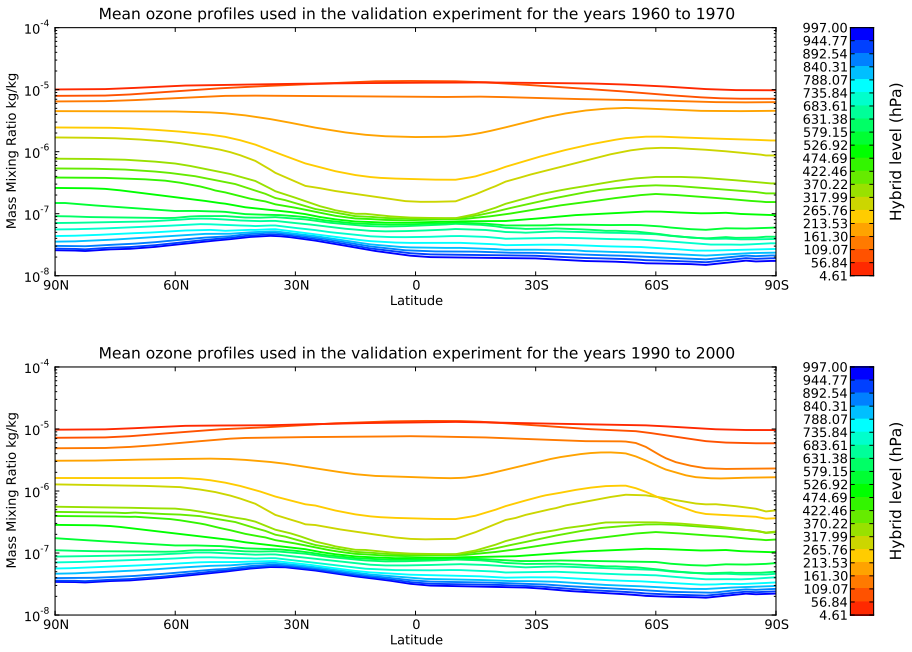


Figure 4.3.2: Ozone profiles used in the ensemble validation experiment. Decadal means for the periods *(a)* 1960-1970 and *(b)* 1990-2000 are plotted for each latitudinal strip and each vertical level in the model.

These sources are sulphur dioxide (SO_2) from anthropogenic emissions, SO_2 from small scale natural sources, SO_2 from large scale volcanic emissions and dimethyl sulphide (DMS) from natural sources. The effects of emissions from biomass burning are not represented in HadAM3P, and so no emissions for biomass burning are required.

Anthropogenic sulphate aerosols are associated with the burning of coal and natural gas which emit SO_2 during its combustion. SO_2 oxidises to sulphate aerosol in the atmosphere, which has a cooling effect by increasing the albedo of clouds and, therefore, increasing the reflection of incoming solar radiation back to space (Solomon et al. (2007). In HadAM3P, to model this effect of sulphate aerosol on the atmosphere, the monthly mean emission rate of SO_2 (in $kgm^{-1}s^{-1}$) is required for each point on the model grid for each month in the period to be modelled. HadAM3P requires these emissions at two levels: the near surface (1.5m) and high level at 930 hPa, which is the level of tall chimneys. These emissions have been supplied by the UK Met Office Hadley Centre and follow observations and the A1B scenario.

The burning of fossil fuels is not the only source of sulphate aerosols in the atmosphere. When volcanoes erupt, they emit large quantities of SO_2, along with CO_2 and water vapour (H_2O). The version of HadAM3P used in the ensemble validation project includes a modification that allows for the effect of the emission of these gases on the atmosphere to be modelled. The modification requires that the optical depth at 0.55 microns in the stratosphere be specified for four equal area latitude bands, each band being contained within the latitude bands 90°S to 30°S, 30°S to 0°, 0° to 30°N and 30°N to 90°N. The optical depth is the inverse of the transmittance, that is, it measures how much solar radiation is prevented from passing along a path, in this case, from the top of the atmosphere to the surface. The optical depths were taken from the Sato et al. dataset (Sato et al. (2011) for the data, Sato et al. (1993) for the methodology) and are, when averaged to the four latitude bands, shown in Figure 4.3.3.

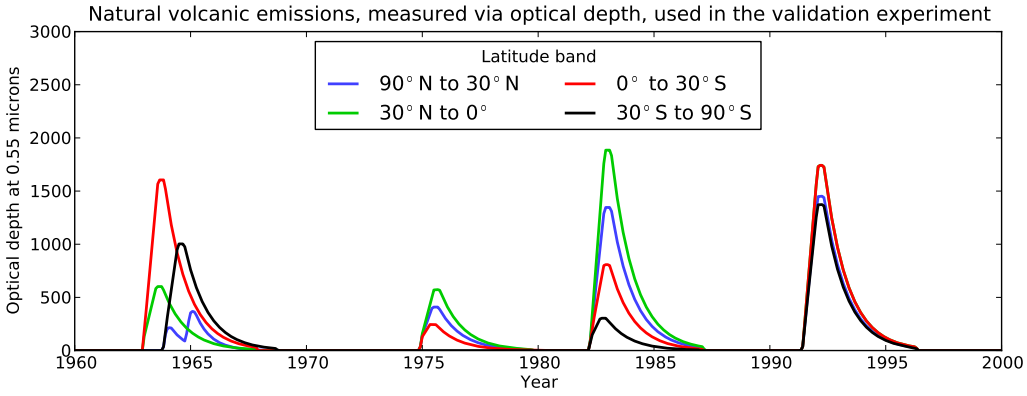


Figure 4.3.3: Optical depth in the stratosphere at 0.55 microns due to natural volcanic emissions, as used in the ensemble validation experiment.

The large spikes in the optical depth in Figure 4.3.3 show the major volcanic activity over the period. These correspond to eruptions of Agung (1963), Fuego (1975), El Chichon (1982) and Pinatubo (1991).

Dimethyl sulphide ($(CH_3)_2S$ - DMS) is a naturally occurring gas that is emitted by phytoplankton in the oceans (Charlson et al., 1987). HadAM3P requires DMS concentrations to be supplied at a single surface level for every point on the model grid, at monthly means. DMS concentrations were supplied by the UK Met Office Hadley Centre.

4.3.4 Orography

The HadAM3P model also requires the specification of the orography. This is described in five fields within the model, the height within a grid box, the standard deviation of the height, and three gradient components in the xx, xy, and yy directions. These fields are based on an improved dataset (GLOBE, Webster et al. (2003)) which is a digitally generated elevation map at 30″ resolution, which equates to about 1km at the equator. The GLOBE dataset resolution is reduced to 10′ (20km) by calculating the arithmetic mean of the 30″ data for each 10′ grid box, and also calculating the standard deviation of the data for each grid box. The resulting data is smoothed, so as to avoid instabilities in the model, manifest as the model "blowing–up" with unphysical quantities cascading through the model (Webster et al., 2003). The 10′ GLOBE orography is then interpolated to the HadAM3P model's N96 grid, and the five fields are stored in the initial conditions file.

4.4 Solar forcing

As mentioned in section 4.1, the HadAM3P climate model used in this study is modified to allow anomalies to the solar constant to be specified. The solar constant is the flux of short wave radiation emitted by the Sun that is incident upon a plane perpendicular to the direction of the emittance and present just outside the Earth's atmosphere. In effect, this is the energy available to the atmosphere, oceans and land surface of the Earth and in the HadAM3 climate model this is assigned a single value of 1359.86 Wm^{-2}. However, this value does vary from year to year, for example with the 11-year solar cycle, and so, to represent atmospheric conditions as accurately as possible, it should also vary within the model.

The modification allows changes to the solar constant to be specified as anomalies to the above value on a yearly basis. These anomalies are based on the difference between the mean value and absolute value of the reconstructed Total Solar Irradiance (TSI) Krivova et al. (2007) dataset for the years 1960 to 2000 and are shown in Figure 4.4.1. Figure 4.4.1 clearly shows the 11 year cycle of the solar output. The maximum amplitude of the anomaly

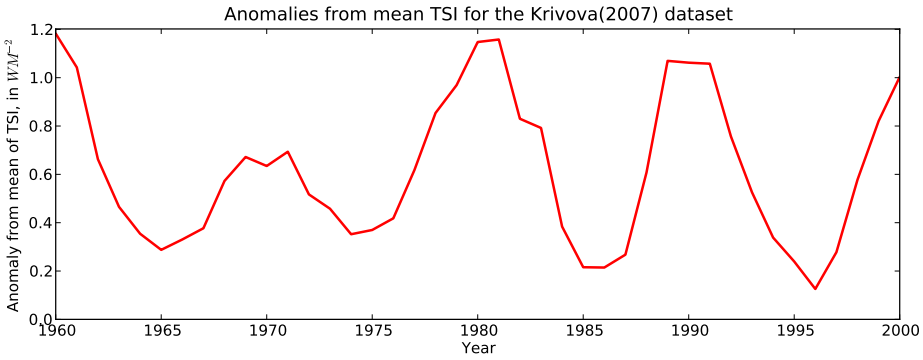


Figure 4.4.1: Anomalies from the mean Total Solar Irradiance for the Krivova et al. (2007) dataset from 1960 to 2000 as used in the validation experiment.

is just 1.2 WM^{-2}, a very small percentage of the total solar constant. Also, HadAM3P does not have an interactive ozone scheme, utilising prescribed ozone values only, so there is no modelled influence of ozone on stratospheric temperatures. Therefore, it seems unlikely that this small variation in the solar constant will have much influence on the Earth's climate, but the anomalies are included for completeness.

4.5 Initial conditions for the model

HadAM3P requires a file of initial conditions (IC) from which to start the model, that is, a file of the state of the atmosphere at timestep zero. This file contains 66 prognostic variables covering the entire global domain, some of which are multi-level fields. Christensen et al. (2005) found that there is a tradeoff between the length of time a climate model runs in a volunteer distributed computing (VDC) project, and the number of successfully completed model runs. The rate of failure of model runs is called the attrition rate and failure may be caused by poorly configured hardware or software or the volunteer growing tired of a long–integration time model and cancelling the model run.

In order to minimise the attrition rate, it was found that the optimum time taken to complete a model run is under a week (Christensen et al., 2005). Therefore, continuous 50 year model integrations cannot be performed and, instead, time–slice experiments are. These consist

of two–year model runs which start from a year in 1959–2010. The model runs overlap, e.g. 1959–1961 is computed, as is 1960–1962. This allows the first year of integration to be discarded as this may be affected by the "spin-up" of long–memory processes which, for an atmosphere only model, is primarily the representation of soil moisture (Pope et al., 2000).

These model runs are computed in parallel, so as to fully exploit the large parallel nature of VDC. An alternative method would be to compute two year's worth of integration and then restart another model run from the final conditions of the initial model run. However, this does not fully exploit the number of processors available under VDC, which can be in the order of 10,000, and would take longer to compute the ensemble than computing all years in parallel. Additionally, the logistics of restarting the model from 1000s of different ICs would be logistically difficult.

In order to produce a large ensemble this IC file must be supplied to each member of the ensemble and, as the ensemble is to cover a 50 year period, an initial condition file is generated for each year in the 50 year period. Also, to produce an ensemble with a wide range of potential weather events and to fully sample the internal variability of the model, a number of IC perturbations are generated.

Figure 4.5.1 depicts the process of generating the IC files, generating the IC perturbations and computing the runs via VDC, as detailed in sections 4.5.1 to 4.6.

4.5.1 Generation of the initial conditions

To generate the ICs, five overlapping decadal runs of the HadAM3P model were run "in-house" on desktop PCs running the Linux operating system. These runs all start from the same IC, which is the state of the modelled atmosphere in December 1958 from a long run of a coupled atmosphere-ocean climate model (HadCM3) provided by the UK Met. Office.

The five decadal runs cover the periods 1958-1970, 1969-1980, 1979-1990, 1989-2000 and 2000–2010 and the forcings and boundary conditions supplied to the model are the same as outlined in sections 4.2 and 4.3. The runs overlap as it can take a full year for the

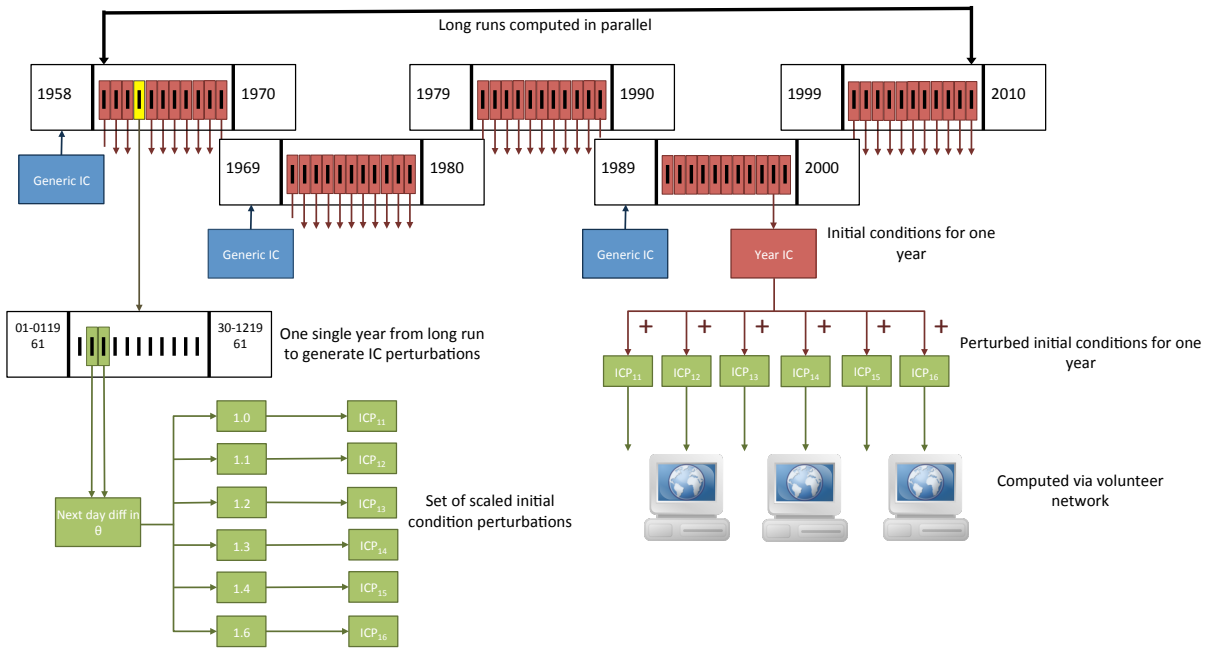


Figure 4.5.1: Schematic of generating initial conditions from 5 long runs of the HadAM3P climate model, generating initial condition perturbations and distributing the runs via volunteer distributed computing. See sections 4.5.1 to 4.6 for full details.

soil variables of the model to "spin-up", that is, become consistent with the forcings and boundary conditions (Massey et al., 2015). By overlapping, the first year of the run can be discarded, with the last year of the previous decade's run replacing it. The state of the atmosphere in the model is output every December in these runs, and these files form the ICs for the ensemble. The ICs are obtained in this manner solely to speed up the generation of the ICs.

4.5.2 Initial condition perturbation

In probabilistic event attribution, the major focus is on the tails of the distribution of the climate, as that is where the extreme weather events occur (Pall, 2006). Integrating a number of instances of a single deterministic climate model forward from a single initial condition (IC) would produce the same temporal and spatial output of climate variables. To fully sample the potential distribution of climate variables, in order to accurately represent the tails of the distribution, the full sampling of the internal variability of the model is required.

To do this, a perturbation is applied to the initial conditions for each ensemble member which produces a spread in the output variables of the climate model. However, the means of these variables are not affected, as the means are solely dependent on the forcings, which vary between years but remain the same for each model which is modelling a particular year.

When computing the long in-house decadal runs that produce the ICs, numerous climate variables are also output, including daily mean potential temperature. The initial condition perturbations (ICP) are produced by taking the next day differences in potential temperature over a year chosen arbitrarily. Potential temperature is simply the temperature a parcel of air will have if it is instantly transported from its level in the atmosphere to a reference level, usually sea-level (Andrews, 2000). A model year is 360 days long and so this produces 348 three dimensional fields that cover the entire globe and all 19 model levels. The differences between the last day of a month and the first day of the next month are not used, hence 348. A scaling function is applied to these next day differences for all levels above a certain level in the model's atmosphere. This is to ensure that there is no perturbation at the top of the atmosphere and so the top of atmosphere flux is not influenced too greatly. The scaling function has the form:

$$S = 1 - \frac{\sin(Z_c - Z_0)}{N_z - Z_0}$$

where Z_c is the level to apply the scaling at, Z_0 is the level above which no perturbation occurs and N_z is the number of levels in the model's atmosphere. The maximum amplitude of the perturbation is also limited to 5K to minimise the risk that a large perturbation in potential temperature could lead to an instability in the model, causing it to crash or produce a spurious result.

Finally, to create a much larger set of initial condition perturbations, global scaling factors of 1.1, 1.2, 1.3, 1.4 and 1.6 are applied to the potential temperature perturbations to generate a set of 1740 initial condition perturbations. The HadAM3P model is modified to accept one file of perturbations before the integration begins. The values of potential temperature perturbations in the perturbation file are added to the values of potential temperature in the IC file, and the integration begins.

Perturbing a single grid box of potential temperature in a coupled atmosphere–ocean climate model is a technique used by both the QUMP project (Murphy et al., 2004) and the climateprediction.net project (Stainforth et al., 2005; Rowlands et al., 2012). Here this is expanded from a single grid box to a perturbation of the entire atmosphere. This allows for a quicker divergence of the individual ensemble members, which is required as the model is being integrated for two years, rather than 150 years as in Stainforth et al. (2005). The scaling of the ICP allows ensemble members to diverge at different rates. The ICPs are based on next day differences of a model run and so they have statistics which are consistent with the daily variability of the potential temperature in the model. Using these ICPs is akin to starting a model run using an IC from a different day of the model. However, adding the ICP to the IC reduces the number of files required, as it is the sum of the number of IC over the 50 years and the number of ICP, rather than the product.

Figure 4.5.2 shows the effect this ICP has on the daily mean near surface temperature at a single grid box over the first thirty days of the integration. The model simulations are reasonably consistent for about 7 days but then quickly diverge and by the 14th day the spread of temperatures is approximately 15K. Although 15K might seem a large range, the model is starting from a state that has not been initialised from observations, and there has been a, large for some ensemble members, perturbation upon this state. The range of 15K is consistent with the variability of the temperature in the grid box itself, as some ensemble members show such a range over six days. By amplifying the initial condition uncertainty to such a large degree, and allowing very quick divergence of the climate variables, it is likely, given 1000s of ensemble members that the full potential distribution of climate variables will be extensively sampled.

4.6 Computation via Distributed Computing

Sections 4.2 to 4.5 detailed the construction of forcings, boundary conditions and initial conditions used by a climate model to generate a very large ensemble. This ensemble spans 40 years and has 1740 initial condition perturbations. To compute such an ensemble on a

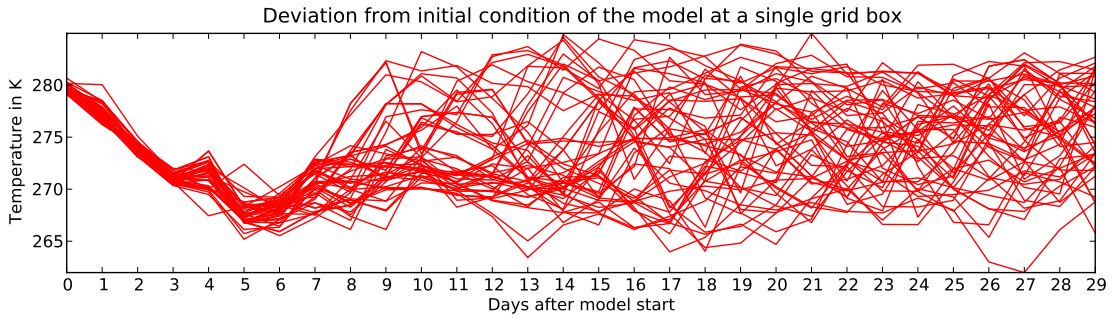


Figure 4.5.2: Deviation of the daily mean near surface temperature from the initial state of the model for 100 model integrations at a single grid box in the model.

single desktop or even a small cluster would take a very long time. The four parallel long runs used to generate the yearly initial conditions in section 4.5 took approximately 6 days to complete each model year, running on a reasonably up to date single processor desktop machine. In order to complete a 40 year, 1740 initial condition perturbation ensemble would take around 1143 years. Therefore, it is obvious that some other method of computing the ensemble is needed.

Volunteer distributed computing (VDC) offers a solution to calculating such a large ensemble. VDC enables members of the public to download an *application* to run on their home PC, the *client* . This application could perform any task, but the majority of VDC projects are based around a scientific application, such as searching for an indication of extraterrestrial life via examining extraterrestrial radio signals in SETI@home (Anderson et al., 2002), determining the shape of proteins from DNA in Rosetta@home (Das et al., 2007) and exploring the effect of varying the parameters in the sub-grid-scale parameterisations of a climate model in climate*prediction*.net (Frame et al., 2009). The originators of these applications are usually scientists who are conducting an experiment that requires large amounts of computing power, which is termed the *project*. These scientists also define the *workunits* that describe the work that should be done by the application, for example, a snippet of radio signal or a string of DNA or the forcings, boundary conditions and initial conditions of a climate model. Each of these workunits can contain unique work or contain duplicates of other workunits in order to

reach a quorum of results that agree with each other. The application is assigned a workunit, which is then computed on a volunteer's client and the results are returned to the project. Volunteers may complete many results over the time they dedicate to the project.

Berkeley Open Infrastructure for Network Computing (BOINC; Anderson (2004), Anderson et al. (2006)) offers an open source infrastructure that enables scientists and engineers to create VDC projects with less effort than writing their own custom VDC software. BOINC is based around a client-server architecture, with members of the public donating computing resources on their home computers (the client) and the scientists maintaining the infrastructure (the server) needed to marshal work, make files available for download by the client and accept uploaded results from the client. BOINC provides libraries to enable scientists to adapt their applications to run on a client computer via a VDC network and also allows the output from the scientific application to be visualised as the calculation progresses. BOINC also includes a suite a programs that are loaded onto the project's servers which handle the downloading of the application, the assignment of workunits to the client, the downloading of the workunit and the uploads of results returned by the client. Participants register via a website (also provided by BOINC) and are rewarded for their computation time with credits, which can be used to compete as to which client has achieved the most computations. A further feature of the website is a user-forum, where users can ask both the project scientists and other users questions about the project, be the questions technical (e.g. problems running the application) or scientific (e.g. why does the application exhibit certain behaviour given the inputs).

Using the BOINC software and libraries, climate*prediction*.net (CPDN; Frame et al. (2009)) has computed over 123 million model years of mixed-layer slab ocean-atmosphere, coupled ocean-atmosphere and atmosphere only climate models. The CPDN project maintains a number of servers to handle the downloading and uploading of data by clients and also to make available the results from the various modelling experiments. This project uses the CPDN infrastructure to compute the experiment laid out in sections 4.2 to 4.5.

The HadAM3P model was originally designed to be run on a multi-processor computer,

where the long computation times can be reduced by decomposing the domain spatially and computing different areas of the model on individual processors. Such an approach requires fast data connections between the individual processors, something that BOINC cannot offer as it relies on client computer's network connections for communications. These connections are far too slow for communication on a timestep by timestep basis, especially considering the amount of data which has to be passed between the boundaries of the sub-domains. Instead, the domain is decomposed temporarily and a *time-slice* experiment is created.

In the time-slice experiment, the period to be covered (1959 to 2010) is split into individual years. The initial conditions generated by the long run in section 4.5 are output yearly and so workunits are generated for every year in the period. These workunits are supplied with the forcing and boundary conditions for a two year run, along with an initial condition perturbation. Therefore, the workunits overlap by a year and an ensemble covering the entire period is returned, even if the first year of each individual ensemble member has to be discarded due to "spin-up" effects (see section 4.5 for more details of spin-up). Although, each member of the ensemble for a given year will start with the same initial conditions, due to the perturbation they will diverge from one another within a few days (Pall (2006), section 4.5.2). These workunits are then sent to client computers via BOINC, computed and the results returned.

4.7 Summary

This chapter has discussed the challenges faced in computing a very large ensemble of a single climate model and has presented a method of computing this ensemble via volunteer distributed computing (VDC). This ensemble has been completed with over half a million ensemble members being returned via VDC. The following chapters will present the results from analysing this data to determine the suitability of a large ensemble approach to probabilistic event attribution.

5 Validation of a very large ensemble of a climate model on seasonal timescales

Chapter 4 described the design of a very large ensemble climate modelling experiment, using observations of greenhouse and other atmospheric gases, along with observed sea-surface temperatures (SST) and sea-ice fractions (SIF) to model the period 1959 to 2010. Using volunteer distributed computing, approximately 2500 ensemble members per year have been generated. As discussed in section 1.3.2, such large ensembles allow the occurrence of extreme events in the upper percentiles of the distribution of modelled climate variables to be assessed with a high level of confidence. To be able to make attribution statements requires confidence that the model adequately represents the extreme weather events that occur in the historical observed and reanalysis datasets, and also requires an ensemble size similar to that used in attribution studies.

This chapter presents methods commonly used to verify climate models, numerical weather prediction systems and other forecast models and applies them to the very large ensemble. These methods rely on comparing the forecast quantities, made before the observation is available, with the observed or reanalysis quantities, after the observation is available. In this study, a period of history has been modelled and so a direct comparison can be made with the observed and reanalysis quantities. However, as the first section shows, care must be taken when choosing the observed or reanalysis dataset.

5.1 Choosing a dataset to verify against

Verifying a global climate model (GCM) against observations first requires an observational dataset to be chosen. Although this sounds straightforward, several issues arise.

For each climate variable that is to be verified, several datasets exist, each one originating from a different research group who may use different analysis methods and input data to derive their observed dataset. For example, air temperature at the near surface is recorded in the HadCRUT3 dataset (Brohan et al., 2006) from the UK Met. Office Hadley Centre, the GISTEMP dataset (Hansen et al., 2010) from the NASA Goddard Institute of Space Sciences and in the Global Historical Climatology Network-Monthly (GHCN-M) dataset (Lawrimore et al., 2011) from the National Climatic Data Center at NOAA. Each of these datasets have their own errors, including station errors, sampling errors and bias errors (Brohan et al., 2006). In each of these datasets the networks of observing stations vary, both in their overall number and in their location, in their persistence through the time period recorded and in their quality. All three datasets have their own methods of dealing with poor quality station recordings, outliers and missing data in the record. The GISTEMP dataset also uses satellite measurements of nighttime light radiancy to adjust for the urban heat island effect, which the other two datasets do not.

Although the three datasets largely agree with each other, and all show the significant trend in warming from 1980 onwards, the differences in their input data and methodologies lead to some discrepancies. GISTEMP and GHCN-M both have 2005 as the warmest temperature on record, whereas HadCRUT3 has 1998 as the warmest. Hansen et al. (2010) postulates that this is due to the different interpolation schemes and ocean datasets used. There are also some differences in the spatial distribution of the anomalies from the mean-temperature, as seen in Figure 12 in Hansen et al. (2010), where the positive anomaly over the west coast of North America is greater in GISTEMP than in HadCRUT3.

A further problem is that the GCM to be validated outputs its climate variables on a global grid at a particular resolution, with no missing data at the surface. As the observed datasets

are based upon recordings from weather and climate observing stations, they have data missing at certain geographical locations. This is most noticeable in the Antarctic. Also, HadCRUT3 has coverage over the ocean from the HadSST2 (Rayner et al., 2006) ship and buoy collected dataset, GISTEMP has ocean coverage from the HadISST1 (Rayner et al., 2003) satellite, ship and buoy dataset, whereas GHCN-M does not have coverage over the ocean. In order to assess the GCM over the full domain of its output, some form of spatial interpolation is needed. CRUTS 3 (Mitchell and Jones, 2005) is a dataset that is largely based upon HadCRUT3 for temperature, and employs the interpolation method of New et al. (1999), which uses a 3 dimensional thin-plate spline to interpolate over missing values in the data in latitude, longitude and elevation to produce a globally gridded dataset at 0.5° resolution, over land only. However, as the true elevation of the missing data grid point is not used, the elevation cannot be corrected for. New et al. (1999) themselves note that the method tends to underestimate the lapse rate in mountainous regions with sparse data coverage, which will lead to a positive temperature bias in those areas.

The datasets discussed above observe just one variable - temperature. Other datasets, such as the Global Precipitation Climatology Project (GPCP) measure other climate variables, but there are few datasets that provide both temperature and precipitation. One that does is the CRUTS 3 dataset (Mitchell and Jones, 2005). There is a physical link between temperature and precipitation on the regional scale (Solomon et al., 2007), and the Clausius-Clapeyron (C-C) equation shows the theoretical link between temperature and saturation water vapour pressure (Andrews, 2000), which can, under the correct atmospheric conditions, provide an estimate of the upper limit for precipitation. In practice, Allen and Ingram (2002) show that, under a warming climate, the global increase in precipitation amount does not follow C-C, but increases linearly at approximately 3.4% per Kelvin. By using two different datasets, from different research groups, likely using different analysis methods, there is no guarantee that the physical link between climate variables will be maintained. The CRUTS 3 dataset does not attempt to guarantee these links, instead treating each climate variable as though it is an independent dataset, albeit using the same analysis techniques for the interpolation (New et al., 1999).

A final major shortfall of these datasets is that they only provide data at a monthly temporal resolution. Extreme weather events can occur on a sub weekly basis, especially for precipitation driven events like the UK floods of 2000 (Pall et al., 2011), or on a multi-weekly basis, such as the Russian heatwave of 2010 (Otto et al., 2012). In order to make attribution statements on sub-monthly extreme weather events the model needs to be validated to ensure it represents sub-monthly events, and so a corresponding sub-monthly observed dataset, provided on a global grid, is required to do this.

From 1979, measurements of atmospheric temperature have been taken by a number of satellites using microwave sounding units (MSU) (Spencer et al., 1990) or radiometers (such as the Advanced Very High Resolution Radiometer - AVHRR) (Kerr et al., 1992), following a polar orbit. The data produced by these instruments has good spatial and temporal coverage, with a horizontal resolution of 2.5° or greater and a temporal resolution of 6 hours (Spencer et al., 1990). However, these instruments do not directly measure temperature, but instead measure radiances from microwave bands, which must be converted to temperature using mathematical methods (Andrews, 2000). This enables different research groups to each produce a different analysis of the data, using distinct mathematical methods, and so produce different temperature series from the same data. Also, the time series obtained from this data is not fully homogeneous, as it is a product of several similar, but different, instruments (Christy et al., 2007).

In summary, in order to successfully validate the GCM for sub-monthly extreme weather events a dataset is required to validate against that has daily coverage and is globally gridded with no missing data. Observed datasets are unable to provide this. However, a class of datasets that matches these criteria are reanalysis products. Reanalysis datasets are related to operational Numerical Weather Prediction (NWP) systems, and use the assimilation phase of operational NWPs to produce a historical dataset of a number of climate variables, including temperature and precipitation. The crucial difference with NWP systems is that the assimilation process is kept constant across the historical period, whereas an operational NWP system will have its assimilation process improved due to technological advances (Uppala et al., 2005).

The ECMWF ERA–40 reanalysis product combines, at 6 hourly intervals, the observations from a number of sources, with some background state of the atmosphere. This background state is obtained from running a short-term forecast, with a T–159 resolution climate model (Uppala et al., 2005). Observations are obtained from radiosondes, surface based weather observing stations, satellite measurements, in-flight measurements from aircraft, ocean buoys and ship measurements (Figure 1, Uppala et al. (2005)). This observed data is combined using a 3D-var analysis and, every 6 hours, further combined with the background state from the forecast model to produce 6 hourly outputs of various climate variables, including temperature, mean sea-level pressure, precipitation and winds, at the surface and at 60 levels in the atmosphere (Uppala et al., 2005). The spatial resolution of these variables is T–159 in spectral space which is approximately 1.125° in latitude–longitude space. The use of the forecast model, with the same topography for each climate variable, ensures that the physical relationships between the variables are preserved to be the same as in the analysis. Combined with the high spatial and temporal resolution, this means that reanalysis products are ideal datasets to validate a model against.

At the time this work commenced, ERA–40 was the most up-to-date and complete ECMWF reanalysis product, spanning the years 1957 to 2002, and so the experimental setup described in chapter 4 was initially produced to model the same time period as ERA–40. More recently an improved ECMWF reanalysis product has become available: ERA–Interim, which uses a more recent version of the ECMWF analysis system, an improved vertical and horizontal resolution, a more frequent assimilation timestep, improved model physics and new and improved observation datasets (Dee et al., 2011). However, ERA–Interim is available only from 1979 to the present day. In this chapter, which determines the performance of the model on seasonal timescales, ERA–40 is used as the reanalysis dataset, as its relatively long timespan allows more samples of each season than ERA–Interim, with its shorter timespan. In chapters 6 and 7 ERA–Interim is used, as these chapters examine daily data, of which there are plenty of samples in ERA–Interim. Using ERA–Interim also allows the same time period over which the tracking methodology was validated in chapter 3 to be used.

5.1.1 Near surface air temperature

Validating a GCM against datasets based on either observations or reanalysis data is problematic, in that it is not known which of the datasets most closely resembles reality. This section highlights the differences between an observed dataset and two reanalysis datasets, and presents a method which can be used to verify the near surface temperature of a GCM, taking into account the shortcomings of the validation datasets.

Figure 5.1.1 illustrates some of the differences between an observed dataset and two different reanalysis products for the near surface air temperature. The observed dataset is CRUTS 3.10 (Mitchell and Jones, 2005), which is an optimally interpolated version of HadCRUT, giving total grid point coverage over land only. The reanalysis datasets are the ECMWF ERA–40 reanalysis product (Uppala et al., 2005), which gives total grid point coverage globally and the NCEP/NCAR reanalysis product (Kalnay et al., 1996) which also gives global coverage. The bias between each dataset for the December, January, February (DJF) seasonal average is shown, for both the actual value and the percentage difference between the datasets. The bias is defined as:

$$b = \frac{1}{n}\sum_{t=1}^{n}(m_t - o_t) \tag{5.1.1}$$

where m_t is the modelled climate variable value at some time t and o_t is the observation at the same time. In the case of the first plot, m_t is the seasonal average for DJF from the ERA–40 dataset, o_t the DJF average from the CRUTS dataset and t is every DJF season from 1961 to 2000. The biases reflect the mean difference between the two datasets over the time period. For biases expressed as percentages of the observed value, the bias is defined as:

$$b' = \frac{1}{n}\sum_{t=1}^{n}\frac{(m_t - o_t)}{o_t} \times 100 \tag{5.1.2}$$

As can be seen, the reanalysis datasets largely agree with the observations, except in specific regions. ERA–40 is colder over Greenland than CRUTS, colder than CRUTS over Northern

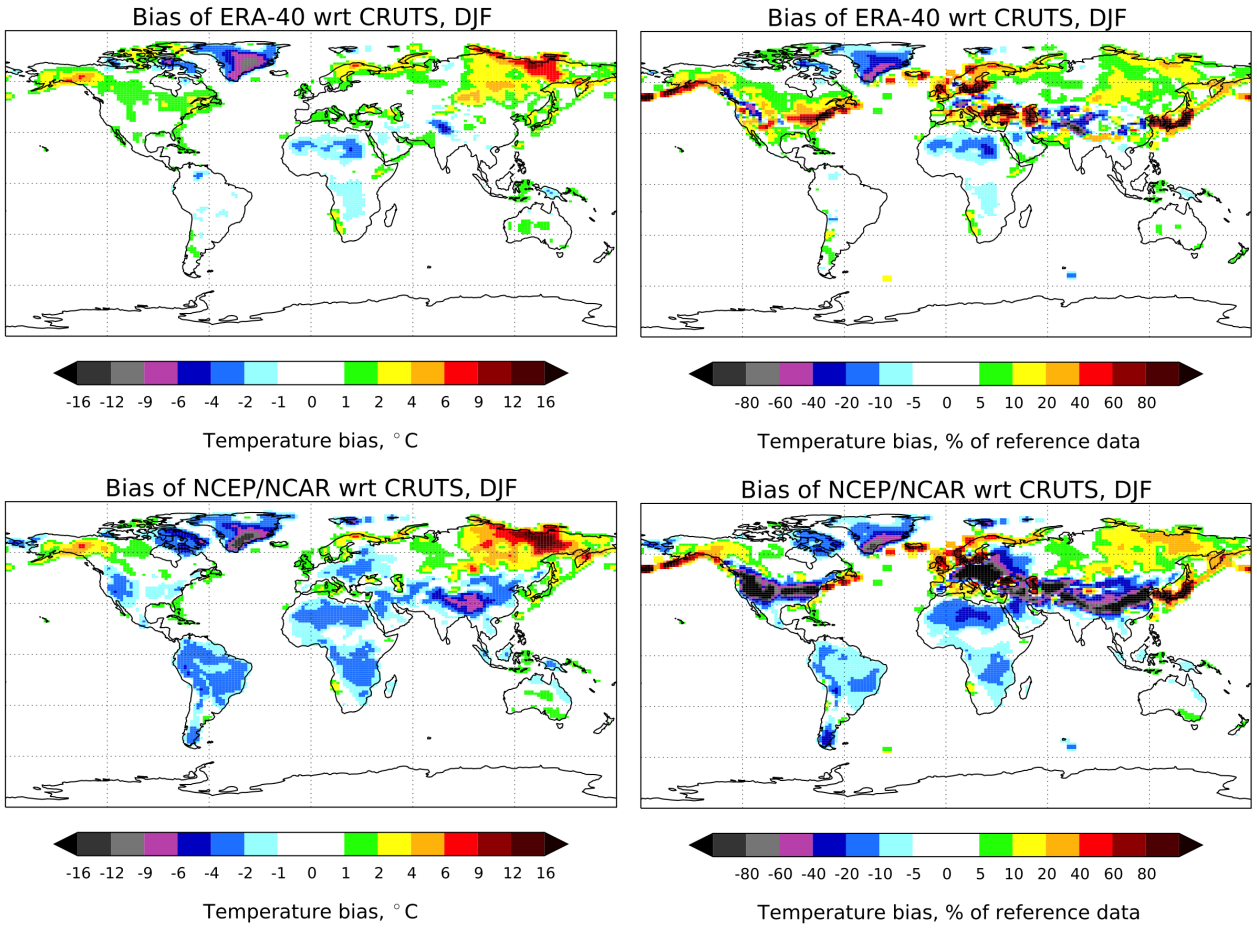


Figure 5.1.1: The biases, with respect to each other, of one observed and two reanalysis datasets, for the DJF seasonal mean temperature at 2m for the period Dec. 1960 to Nov. 2000. Left to right, they are: bias of ERA–40 with respect to CRUTS, actual values; bias of ERA–40 with respect to CRUTS, percentages; bias of NCEP/NCAR with respect to CRUTS, actual values; and the bias of NCEP/NCAR with respect to CRUTS, percentages.

Africa and warmer over Alaska. Both ERA–40 and NCEP/NCAR are warmer than CRUTS over Northern Siberia, but have a similar bias to each other in that region. NCEP/NCAR has a cold bias over the Himalayas when compared to CRUTS, with additional cold biases in Central Africa, Brazil and Argentina. It has a smaller cold bias over Greenland than ERA–40, which is mostly confined to the Southern and Eastern coastline.

Examining the biases as percentages with respect to the observed dataset shows biases in similar locations, with the bias in Northern Siberia being relatively lower in the percentage

bias than in the actual value bias. However, there is an increase in the temperature bias, when expressed as a percentage, in latitudes between $\sim 45°N$ and $\sim 60°$N, with the NCEP/NCAR reanalysis product showing a greater bias than the ERA–40 data.

The higher bias in NCEP/NCAR than in ERA–40 is confirmed by Table 5.1.1. This Table shows, for the entire globe and for the sub–regions detailed in section 1.7 the root mean square error (RMSE) of the bias. This is defined, for every grid box location in a region as:

$$RMSE = \sqrt{\frac{1}{TYX}\sum_{t=1}^{T}\sum_{y=1}^{Y}\sum_{x=1}^{X}\left(\cos(\theta_y)\left(m_{(t,y,x)} - o_{(t,y,x)}\right)\right)^2} \tag{5.1.3}$$

where T is the number of seasons in the datasets, X is the width of the region, in number of grid boxes and Y is the height of the region. $m_{t,y,x}$ is the value of a grid box in the model data at time t and grid box location y, x, θ_y is the latitude of the grid box and $o_{t,y,x}$ is the value of a grid box in the observations.

Table 5.1.1 shows that, when calculating the RMSE across the entire globe, ERA–40 has a lower bias for each individual season (DJF, MAM, JJA and SON) and for all seasons (ALL), with much lower errors when considering the bias in terms of percentage of the CRUTS observation (25.31% in NCEP/NCAR vs 14.74% in ERA–40 for DJF, 29.83% vs 18.41% for SON, etc.). ERA–40 also has a lower RMSE than NCEP/NCAR in the European region (all EU), across each individual season and for all seasons. However, the differences are not quite as pronounced as for calculating the RMSE across the whole globe. This is due to the areas where NCEP/NCAR has a higher bias than ERA–40 (Northern Siberia, the Himalayas, North America in particular) being outside of the European domain. This is also true for the Central European (SREX CEU) region, where NCEP/NCAR has a much larger bias in DJF temperatures (10.83% vs 4.50%) than ERA–40 but has similar RMSE for the other individual seasons. For the Mediterranean (SREX MED), Northern European (SREX NEU) and UK regions (SREX UK), the RMSE for NCEP/NCAR and ERA–40 are very similar across each individual season and for all seasons. These regions are much closer to the sea and so are closely coupled to the sea–surface temperatures. NCEP/NCAR and ERA–40 both use sea–surface temperature observations from the UK Met Office up to the year 2001

(Kalnay et al. (1996), Uppala et al. (2005)) and so it is expected that regions close to the sea will have very similar temperatures, as the lower boundary is being forced by the same dataset. Over a more continental region (such as SREX CEU), the model component of the reanalysis is able to run more freely and so larger discrepancies in the surface temperatures are produced.

	global		all EU		SREX CEU		SREX MED		SREX NEU		UK	
DJF actual	0.98	1.43	0.20	0.28	0.05	0.16	0.12	0.16	0.15	0.16	0.04	0.05
percentage	14.74	25.31	7.89	13.02	4.50	10.83	4.70	5.26	4.47	4.94	1.50	1.91
MAM actual	0.68	1.02	0.11	0.16	0.04	0.08	0.08	0.09	0.07	0.10	0.02	0.01
percentage	18.65	29.28	5.39	8.20	1.29	1.67	0.63	0.68	5.20	8.00	0.23	0.18
JJA actual	0.63	0.92	0.12	0.14	0.05	0.06	0.09	0.11	0.04	0.07	0.01	0.01
percentage	10.78	15.41	0.67	0.76	0.37	0.31	0.39	0.43	0.40	0.55	0.10	0.06
SON actual	0.65	1.12	0.12	0.21	0.05	0.12	0.08	0.13	0.07	0.10	0.03	0.03
percentage	18.41	29.83	5.71	7.05	0.92	2.58	0.55	0.72	5.61	6.52	0.33	0.31
ALL actual	0.75	1.14	0.14	0.20	0.05	0.11	0.09	0.13	0.09	0.11	0.03	0.03
percentage	15.97	25.62	5.58	8.47	2.39	5.63	2.39	2.69	4.43	5.73	0.78	0.97

Table 5.1.1: Biases of the near–surface temperature in two reanalysis products with respect to the CRUTS observed temperature dataset. Root mean square error (RMSE) values are shown for ERA–40 (in the grey column) and NCEP/NCAR (in the white column) for the entire globe and five regions discussed in section 1.7. RMSE scores are shown for actual temperature values, in °C, and for scores expressed as percentages of the observed dataset.

It is encouraging that the three datasets largely agree with each other, especially in areas that are of interest in probabilistic event attribution studies, such as Europe, Africa, the USA, the Indian subcontinent and South America. This result is in agreement with Simmons et al. (2004), which examines trends and low frequency variability in reanalysis products and CRUTS. The areas with large discrepancies, such as Antarctica, Northern Siberia, the Arctic and Greenland are likely to have such large inter-dataset biases due to the sparsity of measuring stations, as seen in Figure 1 in Brohan et al. (2006), which shows there are very few land station observations in North Siberia, none in the Arctic, a few surrounding the Himalayas (but not at high altitude), two in Greenland and a handful on the coast of Antarctica. This means that, in the observed dataset CRUTS 3, much of the coverage of these locations is a product of interpolation. The differences between the two reanalysis products, ERA–40 and NCEP/NCAR, could be due to different data-assimilation methods

being used, different sets of input data for the assimilation, different bias corrections being applied to the common input data prior to the assimilation, or differences in the model formulation used for the forecast.

5.1.2 Precipitation at the surface

Determining the differences between observed and reanalysis datasets for the seasonal accumulated precipitation at the surface follows the same method as that for temperature, except the observed dataset used is the monthly Global Precipitation Climatology Project at 2.5° (GPCP, Adler et al. (2003)). As GPCP only covers the years 1979 to the present, the biases are calculated for 1979 to 2000.

Figure 5.1.2 shows the bias in the seasonal accumulated precipitation for the December, January, February (DJF) season for two reanalysis products, when compared to an observed precipitation dataset. The two reanalysis products are ERA–40 and NCEP/NCAR and the observed dataset is GPCP. The biases are calculated using Equation 5.1.1 for the actual values (in mm) and Equation 5.1.2 for the biases expressed as percentages of the observed values. Precipitation in ERA–40 is available only as a forecast variable and Uppala et al. (2005) notes that the precipitation amount is improved by using the forecast valid at 24 hours after the assimilation time. These are the data that are used for ERA–40 precipitation values throughout this thesis. For NCEP/NCAR, the precipitation is derived completely from the 6 hour forecast of the model (Kalnay et al., 1996). For the reanalysis datasets the 6 hourly accumulated precipitation is used and summed to daily, monthly and seasonal values, depending on the analysis to be carried out.

For the DJF season, there is mostly good agreement between the two reanalysis products and the observed dataset, with most areas having a bias of less than 100mm per season. The notable exception is over the oceans in the tropics, which has a large bias for the ERA–40 reanalysis, of up to 350mm per season and a smaller bias in NCEP/NCAR of up to 200mm per season. Uppala et al. (2005) notes that ERA–40 has a bias in rainfall over the tropical oceans. This is due, partly, to an attempt to correct too-dry regions that results

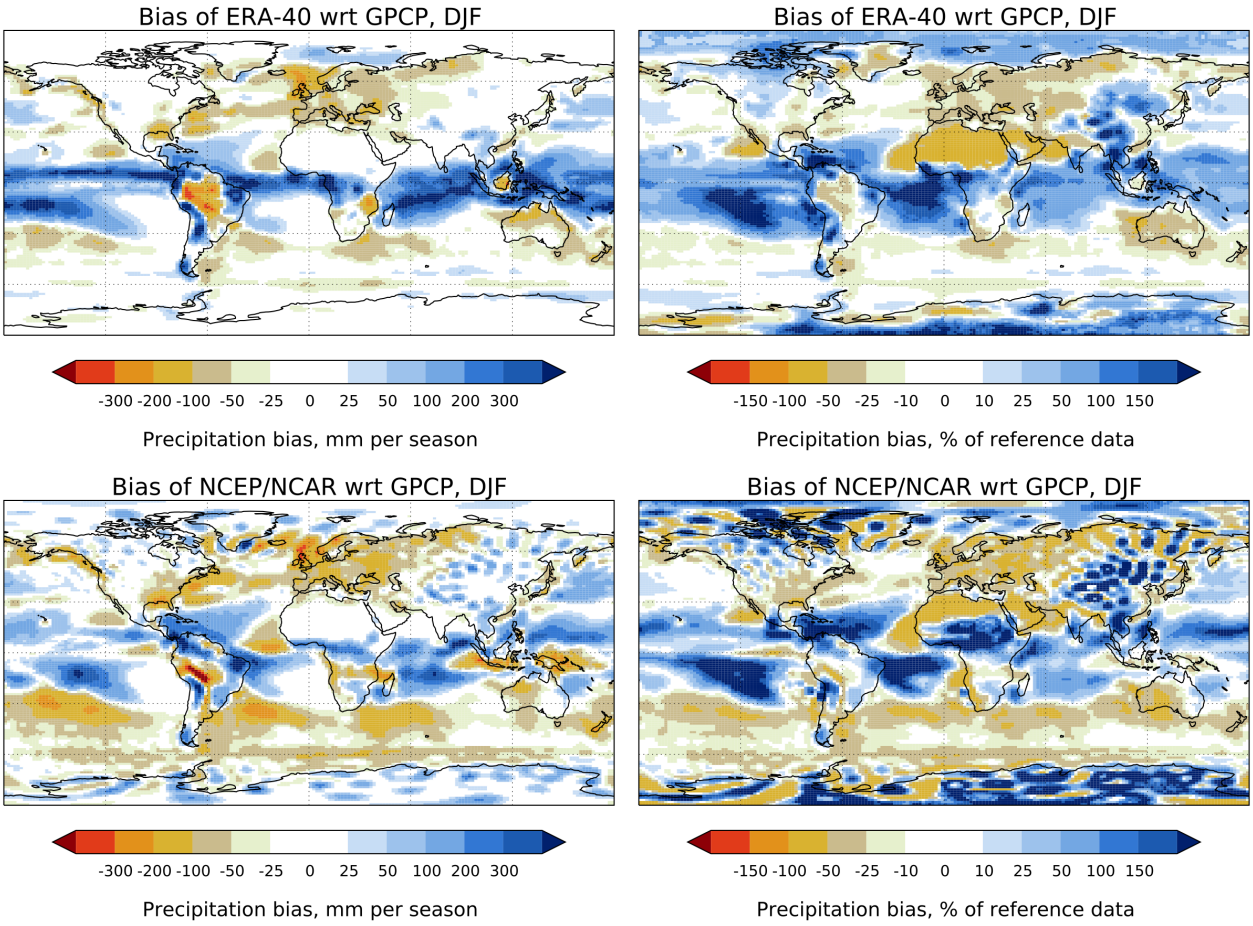


Figure 5.1.2: The biases, with respect to each other, of one observed and two reanalysis datasets, for the DJF seasonal accumulated precipitation for the period Dec. 1979 to Nov. 2000. Left to right, top to bottom, they are: bias of ERA–40 with respect to GPCP, actual values; bias of ERA–40 with respect to GPCP, percentages; bias of NCEP/NCAR with respect to GPCP, actual values; and the bias of NCEP/NCAR with respect to GPCP, percentages.

in neighbouring regions, which are already close to saturation, being artificially moistened further. There is also a feedback effect, with the excess latent heating in the artificially moist regions altering the tropical circulation leading to more dry regions being further moistened by the assimilation method and the saturated neighbour regions being further moistened as before. A secondary contribution to the excess precipitation is from the effects of volcanic aerosols on satellite data after the Mount Pinatubo eruption in 1991. The biases in the infrared radiances resulting from these aerosols, are uncorrected and lead to humidity

changes in the tropical troposphere (Uppala et al., 2005).

Examining the biases in terms of the percentage of the observed value reveals that much of North Africa is drier in both of the reanalysis products and that much of Europe also has a dry bias. A notable bias outside of the tropics is evident over the North Atlantic and UK in the DJF season, with a bias of up to 300mm per season over Scotland. This is equivalent to approximately 50% of the observed rainfall.. This bias is present in both of the reanalysis products, ERA–40 and NCEP/NCAR being too dry in this area, with NCEP/NCAR having a greater bias than ERA–40.

In order to quantify which reanalysis dataset has the lowest bias across all seasons, and across the five regions detailed in section 1.7, the RMSE values are calculated using Equation 5.1.3 and shown in Table 5.1.2. RMSEs are calculated both for the actual seasonal accumulated precipitation and for the precipitation expressed as a percentage of the observed dataset (GPCP). The Table shows that, over the entire globe and for every season, ERA–40 has a higher overall bias for the actual accumulated precipitation than NCEP/NCAR, but when this bias is expressed as a percentage of the observation, ERA–40 has lower overall RMSEs than NCEP/NCAR. For the five regions, ERA–40 has a lower bias than NCEP/NCAR for each region and each season, except for SON in the Mediterranean (SREX MED). That ERA–40 has a higher bias over the globe and a lower bias over the five regions is due to the large biases in the tropics in ERA–40 not being included in any of the five regions. Over the UK, the RMSE values are largely similar between ERA–40 and NCEP/NCAR with a maximum discrepancy of around 1.5mm per season, between the two reanalysis products.

5.1.3 Conclusions and validation method for global data

This section has concentrated on highlighting the differences between two reanalysis products and one observation dataset for temperature and precipitation. Each of these could be used to validate a model against, with the outcome that the validity of the model could be dependent on which of the datasets are chosen. sections 5.2.1 and 5.2.2 validate the output from the large ensemble of the climate model generated by the experiment described in chapter 4, but

	global		all EU		SREX CEU		SREX MED		SREX NEU		UK	
DJF actual	160.67	127.38	16.41	20.81	6.96	9.93	8.96	9.60	11.86	15.56	6.06	7.47
percentage	97.39	127.65	6.84	8.32	2.83	4.16	5.11	5.63	3.56	4.50	1.26	1.67
MAM actual	162.66	126.57	9.95	12.38	4.73	5.46	6.09	6.77	6.30	8.81	3.19	4.05
percentage	85.22	109.48	6.62	7.95	2.37	2.99	4.91	5.98	3.76	4.30	1.14	1.53
JJA actual	153.63	144.79	10.61	16.56	5.52	10.81	4.19	5.99	8.04	11.02	3.41	4.52
percentage	91.20	128.47	10.98	14.50	2.48	4.69	10.08	12.80	3.57	4.96	1.30	1.78
SON actual	158.80	130.10	14.85	18.05	6.29	7.93	8.90	8.27	10.09	13.94	5.08	6.33
percentage	87.65	110.47	6.69	9.05	2.62	3.27	5.48	7.49	2.80	3.90	1.21	1.54
ALL actual	158.98	132.42	13.25	17.22	5.93	8.78	7.32	7.78	9.31	12.61	4.59	5.76
percentage	90.48	119.36	8.00	10.31	2.58	3.84	6.74	8.47	3.44	4.43	1.23	1.63

Table 5.1.2: Biases of the seasonal accumulated precipitation in two reanalysis products with respect to the GPCP observed dataset. Root mean square error (RMSE) values are shown for ERA–40 (in the grey column) and NCEP/NCAR (in the white column) for the entire globe and five regions discussed in section 1.7. RMSEs are shown for actual accumulated precipitation amounts (in mm), and for errors expressed as percentages of the observed dataset.

rather than using just a single dataset to validate against, two datasets per climate variable are used, one observed dataset and one reanalysis product. ERA–40 is shown, in sections 5.1.1 and 5.1.2 to have a lower bias, in relation to CRUTS and GPCP, and so ERA–40 is used as the reanalysis dataset. The method proceeds as follows:

- Firstly, the bias of the ensemble mean from the model, with respect to the observed dataset (CRUTS for temperature and GPCP for precipitation), is calculated for each seasonal mean in the period 1961-2000, for the near-surface temperature and seasonal accumulated precipitation, for each gridpoint.

- Then, for each season, and again for each gridpoint, the absolute value of the bias is compared to the absolute maximum bias of the climate variable in the ERA–40 reanalysis. Here the maximum bias is the maximum value (per gridpoint) of the bias of individual seasons from 1961 to 2000 (for temperature) and 1979 to 2000 (for precipitation).

- If the bias between the ensemble mean and the observation is greater than the maximum bias between ERA–40 and the observation then this is considered a significant bias and is marked on the map–plots accordingly.

The definition of "significant bias" here is used as a non–parametric test of whether the model can be distinguished from the reanalysis data, in that the model's bias is within the range of biases between the observed and reanalysis datasets. A more traditional parametric significance test is difficult to implement due to having essentially only two samples for the reanalysis datasets, being the difference between the ERA–40 reanalysis and the observations and the difference between the NCEP/NCAR reanalysis and the observations. With so few samples it is impossible to know the underlying distribution of the biases, and so a traditional significance test (such as the two way Kolmogorov Smirnoff test used in Chapter 6) cannot be used. The method presented here can be used to determine whether the model is as good as the reanalysis datasets in representing observed seasonal means of climate variables without knowing the underlying distributions of the climate variables.

5.2 Validating seasonal means of climate variables

Section 5.1 presents three near-surface temperature datasets, one of which is an observed dataset and two of which are reanalysis datasets. It also presents three precipitation at the surface datasets, again, one of which is observed and two which are reanalysis. The differences between these datasets, are shown and a method of comparing the model output from the very large ensemble of climate models described in chapter 4 is presented. The following sections show the results of applying this method to the very large ensemble.

Sections 5.2.1 to 5.2.4 use the ensemble mean of the climate variables from the very large ensemble. Due to the large amount of computing power available using the distributed volunteer computing infrastructure, as discussed in section 4.6, approximately 2500 integrations of the model were calculated per year, for each year between 1961 and 2000. The ensemble means analysed in the following sections are calculated from the mean of the ensemble members at every grid point in the model and every month in the period 1961 to 2000. This produces, for every location in the grid, a timeseries of the mean state of the model, around which each individual ensemble member represents a part of the variability of the model's climate.

5.2.1 Validating seasonal mean temperature

Figure 5.2.1 shows the bias between the seasonal mean of the ensemble average and the seasonal mean of CRUTS, for the two opposite seasons of DJF and JJA, calculated for each year between 1961 and 2000. The values of the bias (in °C) are shown and superimposed on each of these biases is how the value of each individual bias compares to the bias between ERA–40 and CRUTS, as shown in Figure 5.1.1 for DJF and summarised in Table 5.1.1 for all seasons. Stippling indicates that the bias between the ensemble mean and CRUTS is greater than the bias between ERA–40 and CRUTS. Without the stippling the bias is within the range of the biases shown in ERA–40 and the ensemble can be considered to be indistinguishable from ERA–40. The presence of a circle, therefore, indicates a significant bias as, for that location, the ensemble is distinguishable from ERA-40. ERA–40 on its own is used, as opposed to the mean of the two reanalysis products, so as to have a dataset that is internally consistent across variables, such as temperature and precipitation, and to maintain the relationships between those variables.

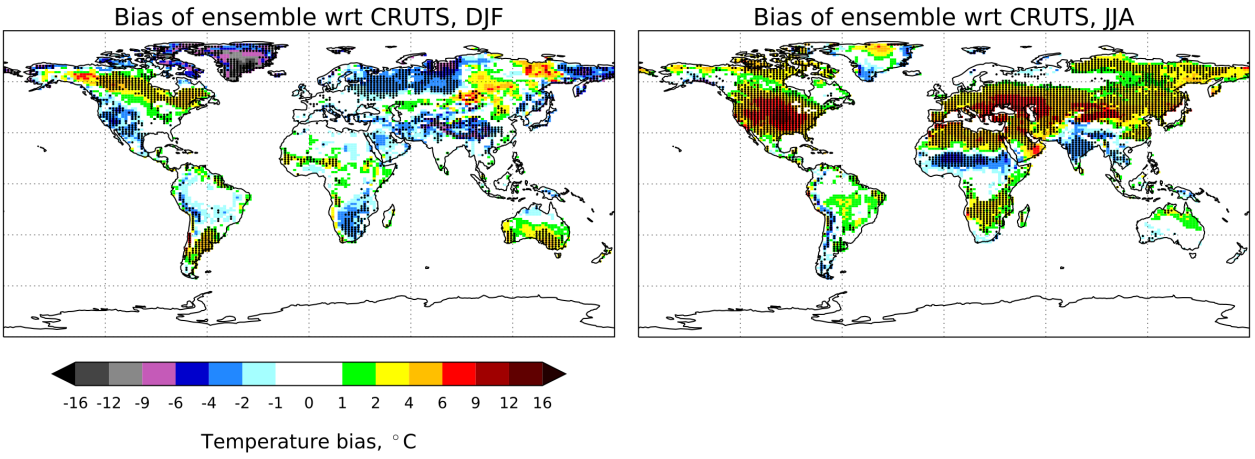


Figure 5.2.1: The biases of the seasonal mean temperatures at 2m in the HadAM3P ensemble mean, with respect to CRUTS, for the DJF (left panel) and JJA (right panel) seasons from Dec. 1961 to Nov. 2000. The stippling indicates whether the value of the bias, for that grid box, is greater than the maximum bias between ERA–40 and CRUTS. Grid boxes with stippling are considered to be significantly biased, as the bias is greater than the maximum bias in ERA–40.

As can be seen, when examining the biases, there are many areas that produce a bias of between -1° and 1°C where there is no significant bias. In the DJF season this is most noticeable over Western Europe, areas of North and South America and North Africa. The model shows large cold biases over Greenland (up to 16°C) Northern Russia (4 to 9°C) and the Himalayas (up to 9°C). There is a slight warm bias over Canada (up to 6°C) but a cold bias on the West Coast of America (4°C). The biases in these regions are all considered to be significant as they are greater than the bias between ERA–40 and CRUTS.

The large biases seen over European Russia, Central Asia and China in the JJA season are consistent with Otto et al. (2012), who found a bias of 3°C, in the same model, in the 2010 summer temperatures over Moscow, Russia. There is also a large bias over North America in JJA. Although the model appears to perform quite poorly, there is one area which has little bias and that is over the UK and Western Europe. This is encouraging for studies that have so far concentrated on attribution of weather events over the UK, such as Massey (2012).

	global	all EU	SREX CEU	SREX MED	SREX NEU	UK
DJF	1.62	0.41	0.14	0.28	0.27	0.07
MAM	1.27	0.25	0.06	0.21	0.11	0.02
JJA	**1.86**	**0.68**	0.46	0.49	0.09	0.03
SON	1.17	0.23	0.09	0.17	0.14	0.04

Table 5.2.1: Biases of the seasonal mean temperature in the large ensemble of HadAM3P with respect to the CRUTS observed dataset. Root mean square error (RMSE) values are shown for the actual bias (in °C) for the entire globe and five regions discussed in section 1.7. Values highlighted in bold have a higher RMSE than the RMS of the maximum bias in ERA–40 with respect to CRUTS for the same region and season.

Table 5.2.1 calculates the root mean square error (RMSE) (Equation 5.1.3) for all seasons over the five regions detailed in section 1.7. Entries in the table highlighted in bold have a higher RMSE than that of the maximum bias in ERA–40 with respect to CRUTS. Values in bold are considered to be significantly biased as they are distinguishable from the values obtained by comparing ERA–40 and CRUTS.

Considering the value of the RMSEs (in °C) the only regions with significant biases are the entire globe (global) and the entirety of Europe (all EU) but only in the JJA season.

This corresponds to the large warm biases seen in JJA in North America and Europe in Figure 5.2.1. These results also agree with the validation of the model in Massey et al. (2015).

That there are only two regions within Europe which show significant biases indicates that the model is well formulated for probabilistic event attribution studies over Europe, such as those detailed in section 1.4. The two regions which show significant bias may be able to be used in an attribution study if bias correction is performed. This is developed further in section 6.2.1 where the distribution of daily values of temperature in the model is compared to the distribution of daily values of temperature in ERA–Interim (Dee et al., 2011) and a simple bias adjustment is made by correcting the mean of the model. Also, in section 5.3.1 the extremes of seasonal mean temperatures are compared to the extremes in ERA–40 by considering the difference between an extreme percentile (95th, 99th) and the 50th percentile, in effect bias correcting the model by comparing anomalies from the median of the model and ERA–40.

5.2.2 Validating seasonal mean precipitation

Probabilistic event attribution studies which attempt to attribute the change in frequency of occurrence of a high precipitation event to an increase in greenhouse gas concentrations in the atmosphere concern the changes in return time of extreme precipitation events. These studies, detailed in section 1.4 (such as Pall et al. (2011), Schaller et al. (2016), Yiou and Cattiaux (2013)) first define a threshold of precipitation which, when exceeded, an extreme event is deemed to have occurred. This threshold is defined in terms of the event they are attempting to attribute (99th percentile of precipitation in Schaller et al. (2016), 41mm of runoff in Pall et al. (2011)) and in terms of the observed record. For example, an event may be considered extreme if it occurs at or above the 90th percentile of the observed record which, if averaged over a certain time–period is the climatology of the observations. To gain confidence that the model used in these probabilistic event attribution studies can adequately represent the climatology of real–world precipitation, this section compares the ensemble mean of precipitation from the very large ensemble to observed and reanalysis data

products.

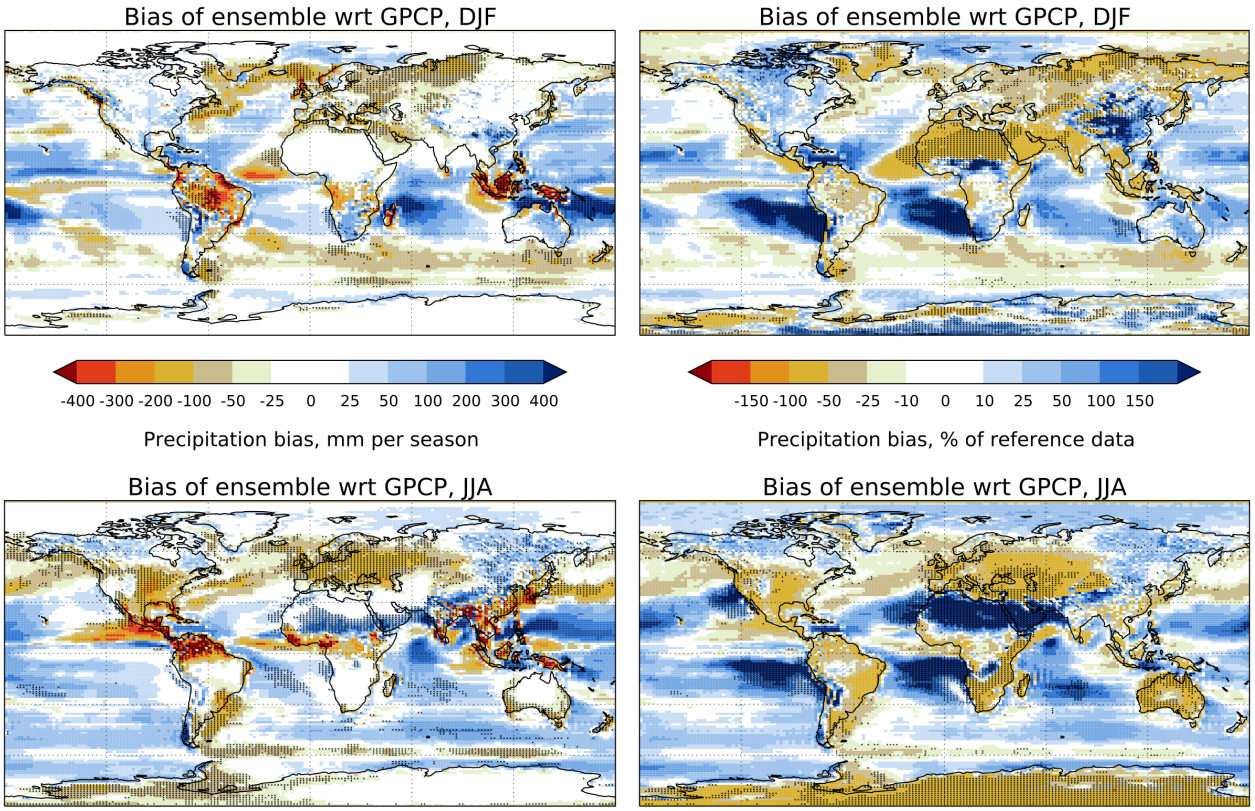


Figure 5.2.2: The biases of the seasonal accumulated precipitation in the HadAM3P ensemble mean, with respect to GPCP, for the DJF and JJA seasons from Dec. 1979 to Nov. 2000. The actual value (in mm/season) is shown in the left panel and the bias expressed as a percentage of the observation is shown in the right panel. The stippling indicates whether the value of the bias, for that grid box, is greater than the maximum bias between ERA–40 and GPCP. Grid boxes with stippling are considered to be significantly biased.

Figure 5.2.2 shows the bias of the ensemble mean of the seasonal accumulated precipitation from the very large ensemble for the years 1961 to 2000 for the DJF (December, January, February) and JJA (June, July, August) season. The bias is calculated with respect to the Global Precipitation Climatology Project (GPCP). The circles indicate whether the absolute value of the bias is greater than the maximum bias of ERA–40 with respect to GPCP over the years 1979 to 2000, as calculated in section 5.1.2. The actual value of the bias in (mm/season) is shown in the left panel, whereas the bias expressed as a percentage of the observation from

GPCP is shown in the right panel.

The most apparent difference between Figure 5.2.2 and Figure 5.2.1 is that, compared to the temperature, there is less bias in the precipitation, relative to the maximum bias of ERA–40, especially in the Northern Hemisphere. In the DJF season, when examining the actual value of the bias, there are significant biases over parts of the UK and over central Europe and Russia. When considering the bias as a percentage of the observed data (GPCP) these significant biases largely disappear over these three regions. These regions have relatively high precipitation in DJF and, although the bias in the actual value may be quite high, the contribution of this bias to the overall total precipitation is not significant, as shown when considering the bias as a percentage of GPCP. The dry biases in the actual values over Brazil, Indonesia, Malaysia and Papua New Guinea, which are all regions with high precipitation, are also not significant when considered as percentages of the observed precipitation. The opposite is true for North Africa and the Sahara Desert, an area of which receives very little precipitation. In DJF, the bias in the actual values of seasonal accumulated precipitation shows very little bias ($<$ 25mm per season) over the Sahara. However, when the bias is expressed in terms of the percentage of the observations the Sahara region is shown to be too dry by up to 100% of the GPCP value. This is a significant bias as it is greater than the bias of ERA–40 with respect to GPCP. This shows that, for regions with very low precipitation, the overall contribution of even a small bias is very important.

In JJA there are dry biases in the large ensemble which coincide, spatially, with the warm biases shown in Figure 5.2.1. These are over North America, which has a dry bias of up to 300 mm/season, corresponding to a warm bias of up to 12°C, and over Southern and Central Europe, where a dry bias of up to 200 mm/season corresponds to the warm bias of up to 6°C. Additionally, the cold bias of up to 6°C over the Sahel corresponds to a wet bias of up to 300 mm/season. Expressing these biases as percentages of the observed values reduces the number of locations with significant dry biases in North America and Western Europe. However, the dry bias over Central Europe remains, with the driest biases corresponding, spatially, to the hottest biases in Figure 5.2.1. The wet bias over the Sahara and Sahel becomes even more pronounced when expressed as a percentage of the observation, receiving

over 200% of the observed rainfall in the large ensemble.

To quantify the biases in the large ensemble, and to investigate how the biases may affect a probabilistic event attribution study over regions of Europe, the root mean square error (RMSE) with respect to GPCP is calculated, using Equation 5.1.3, for each of the five regions discussed in section 1.7 and for each season. These RMSEs are shown in Table 5.2.2. Any RMSEs which are greater than the RMS of the maximum bias for the corresponding season in ERA–40 and GPCP are highlighted in bold, and are considered to be significantly biased, as the large ensemble is distinguishable from ERA–40.

	global		all EU		SREX CEU		SREX MED		SREX NEU		UK	
DJF	125.45	(81.68)	17.76	(7.93)	6.27	(2.64)	10.89	(6.67)	12.55	(3.39)	6.05	(1.21)
MAM	116.54	(89.35)	11.04	(6.89)	4.55	(2.20)	7.56	(5.66)	6.63	(3.26)	3.72	(1.26)
JJA	137.77	(107.20)	17.40	(24.56)	**12.92**	**(5.72)**	7.15	(23.58)	9.19	(3.85)	5.22	(1.98)
SON	113.30	(110.55)	17.17	(7.58)	8.25	(3.73)	8.95	(5.70)	12.10	(3.34)	6.91	(1.60)

Table 5.2.2: Biases of the seasonal accumulated precipitation in the large ensemble of HadAM3P with respect to the GPCP observed dataset. Root mean square error (RMSE) values are shown for the actual bias (in mm / season, in the grey column) and expressed as a percentage of the observation (in the white column) for the entire globe and five regions discussed in section 1.7. Values highlighted in bold have a higher RMSE than the RMS of the maximum bias in ERA–40 with respect to GPCP for the same region and season

As can be seen in the Table, the only region and season to have a significant RMSE in precipitation is JJA (June, July, August) in central Europe (SREX CEU) which, as can be seen in Figure 5.2.2 is a dry bias. This is the region with the highest RMSE for temperature, and the dry bias is likely linked to this high temperature bias as enhanced evaporation caused by the high temperature will remove soil moisture from the region and reduce the available precipitation. Other regions and seasons with significant biases in temperature (global, JJA; all EU, JJA; SREX MED, JJA) do not show a corresponding significant bias. However, the biases for the SREX MED region (7.15 mm / season, 23.58%) and all EU region (17.40 mm / season, 24.56%) are quite high, which indicates that the maximum bias in ERA–40 is even higher.

The overall picture of the representation of the precipitation in the model on seasonal timescales is that, for European regions, the model produces a realistic representation for

most regions of Europe and for most seasons. The noticeable exception is the JJA season, which has significant biases in central Europe and high (but not significant) biases in the Mediterranean and over European domain if taken as a whole.

5.2.3 Validating seasonal mean MSLP

Localised low values of mean sea–level pressure (MSLP) which track from the warm, moist atmosphere of the Western Atlantic to Western Europe are a major cause of extreme weather events in Europe, bringing with them high precipitation, high windspeeds and coastal surges (Christidis et al., 2014). Probabilistic event attribution studies using large ensembles of a single climate model have been conducted which attempt to attribute the extreme weather brought by these systems to an increase in greenhouse gas (GHG) concentrations caused by anthropogenic emissions. These include attribution studies of high precipitation over the UK (Pall et al. (2011), Schaller et al. (2016)) and heavy precipitation of the Elbe and Danube basins (Schaller et al., 2014).

There is a negative pressure gradient from Equator to Pole and these cyclones (low pressure systems) can be considered anomalies with regard to this background pressure gradient. Due to the differential heating rates between the land and the ocean and the effect of the Coriolis force, the gradient is not zonally uniform, but differs with longitudinal position. Furthermore, the position of the three circulation cells and the Westerly Polar Jet positioned at the Northern side of the Ferrel Cell can affect the pressure gradient. Therefore, in order to conduct an attribution study into weather events associated with low pressure systems, the model used must be able to represent MSLP in a realistic manner.

This section compares seasonal means of the MSLP from the large ensemble of HadAM3P to the seasonal mean MSLP from ERA–40. Biases are calculated using Equations 5.1.1 for the value of the MSLP bias (in hPa) As the bias is, in effect, calculating the difference in seasonal means between the large ensemble and ERA–40, the model's ability to represent the observed background pressure field is being investigated, as this background pressure is synonymous with the climatological mean of the MSLP field. Cyclones are anomalies to this background

field, and their associated localised pressure gradients are the cause of high winds, whereas the transport of moisture from warm to cold regions causes the associated high precipitation. The representation of these anomalies in the HadAM3P model is investigated in chapter 7, where the feature tracking algorithm developed in chapter 2 is applied to the large ensemble of HadAM3P model runs.

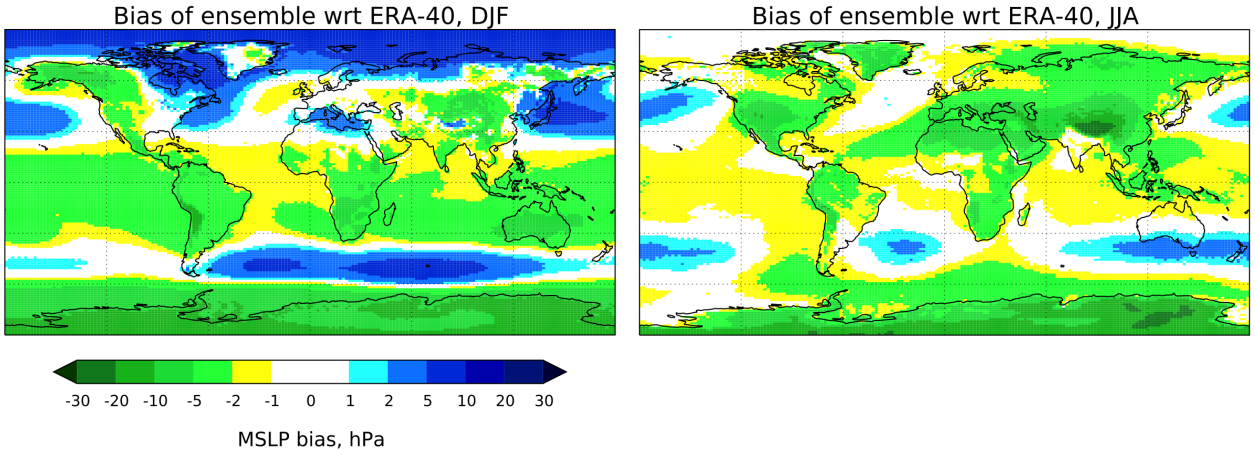


Figure 5.2.3: Biases of the seasonal mean MSLP in the HadAM3P ensemble, with respect to ERA–40, for the DJF (left panel) and JJA (right panel) seasons from Dec. 1961 to Nov. 2000.

Figure 5.2.3 shows the bias between the seasonal mean MSLP in the large ensemble of HadAM3P and the seasonal mean MSLP in ERA–40, for the DJF (December, January, February) season and the JJA (June, July, August) season between Dec. 1961 and Nov. 2000. In DJF, there is a discrepancy of up to 10hPa over high Northern latitudes between the ensemble and ERA–40. The resulting shallower Equator to Pole pressure gradient in the large ensemble occurs off the Eastern coast of the USA, which is an area of cyclogenesis (Martin, 2006). The reduction in the gradient could result in a reduction in the occurrence of baroclinic instability and the formation of fewer extra–tropical cyclones (and their associated low pressure centres). In addition, there is a positive bias over the Mediterranean sea in DJF, which is an area and season of high storm activity. This positive bias could also lead to a reduction in the number of cyclones. This will be investigated further in chapter 7.

By contrast, there is a negative bias in MSLP over the South Pole and Antarctica in both

the DJF and JJA season. This could lead to an increase in the the occurrence of low pressure systems in the Southern Hemisphere. However, this thesis concerns itself with the validation of a model over Europe, and so no further investigation into the Southern Hemisphere will be made.

To quantify the biases in the large ensemble the root mean square error (RMSE) of MSLP with respect to ERA–40 is calculated, using Equation 5.1.3, for each of the five regions discussed in section 1.7 and for each season. These RMSEs are shown in Table 5.2.3.

	global	all EU	SREX CEU	SREX MED	SREX NEU	UK
DJF actual	6.22	0.34	0.06	0.23	0.25	0.05
MAM actual	6.01	0.24	0.04	0.15	0.19	0.01
JJA actual	3.40	0.82	0.36	0.72	0.15	0.04
SON actual	2.94	0.10	0.23	0.31	0.06	0.02

Table 5.2.3: Biases of the seasonal mean MSLP in the large ensemble of HadAM3P with respect to the ERA–40 reanalysis dataset. Root mean square error (RMSE) values are shown for the bias in hPa for the entire globe and five regions discussed in section 1.7.

As can be seen in the table, the five regions of Europe do not show any large biases in MSLP, despite there being a $\sim$ 5hPa bias globally. The largest bias is attained when calculating the RMSE over the entire region of Europe in JJA. This coincides, spatially, with the large warm bias in this region and season seen in section 5.2.1. For the Northern hemisphere winter (DJF) the largest biases occur over the Mediterranean (SREX MED) and Northern Europe (SREX NEU) regions. For the SREX MED region this bias is manifest as a positive bias over the Mediterranean Sea, whereas in the SREX NEU region it is the reduced Equator to Pole gradient causing the bias. Although the other regions show smaller biases, these are only biases in the mean flow and are not necessarily indicative of a good representation of low–pressure systems in the large ensemble of the HadAM3P model over these regions. Low–pressure systems are often formed outside of the region and advected into it. This is the motivation behind using the storm identification and tracking methodology developed in chapter 2.

Overall, there is some bias in MSLP in the large ensemble of HadAM3P. Of major note are

the shallower Equator to Northern Pole gradient in DJF, which is the season with the highest incidence of storms, and the positive bias of MSLP in the Mediterranean in DJF, also an area of storminess in winter. How these MSLP biases affect the frequency and intensity of low–pressure systems over Europe will be investigated further in chapter 7.

5.2.4 Validating seasonal mean upper–level wind

As discussed in sections 5.2.3, the Polar jet stream plays a large role in European weather, as the position of it can steer low pressure systems as they track west to east over the Atlantic, transporting warm and moist air from the East coast of America and Central America to Europe. The position and strength of the jet stream varies with the Equator to Pole temperature gradient of upper level air, weakening if the gradient is shallow and strengthening if the gradient is deeper. The position of the jet stream is affected by the temperature gradient over land varying compared to the gradient over the oceans, producing a wave like meandering across the Earth.

A strong and persistent jet stream over the UK was the main driver of the floods in the UK in winter (DJF) 2013/2014 (Huntingford et al., 2014; Schaller et al., 2016). The jet stream steered a rapid succession of deep cyclones across the Atlantic and over the UK for the three months of winter, bringing associated high precipitation leading to widespread flooding in Southern UK (Huntingford et al., 2014). Schaller et al. (2016) conducted a probabilistic event attribution study of this event and found the likelihood of the event has increased under anthropogenic greenhouse gas forcing, defining the event as when the magnitude of rain in the model in the DJF season exceeds the amount seen in the observations in DJF 2013/2014.

In order to be confident in making attribution statements about extreme weather events caused by a strong and persistent jet stream, the model used for probabilistic event attribution studies should be able to represent the jet stream in a realistic manner. This section examines how well the HadAM3P model represents the Polar jet stream by comparing the Westerly wind at 200hPa (U–wind @ 200hPa) in the model with the same quantity

in ERA–40. Unfortunately, there is no global dataset with upper level wind data, and so the reanalysis is used. Biases of the U–wind @ 200hPa, in terms of the actual value (in ms^{-1}) and in terms of the percentage of the ERA–40 value, are calculated using Equations 5.1.1 and 5.1.2, and the RMSE of the U–wind @ 250hPa is also calculated for the five regions detailed in section 1.7.

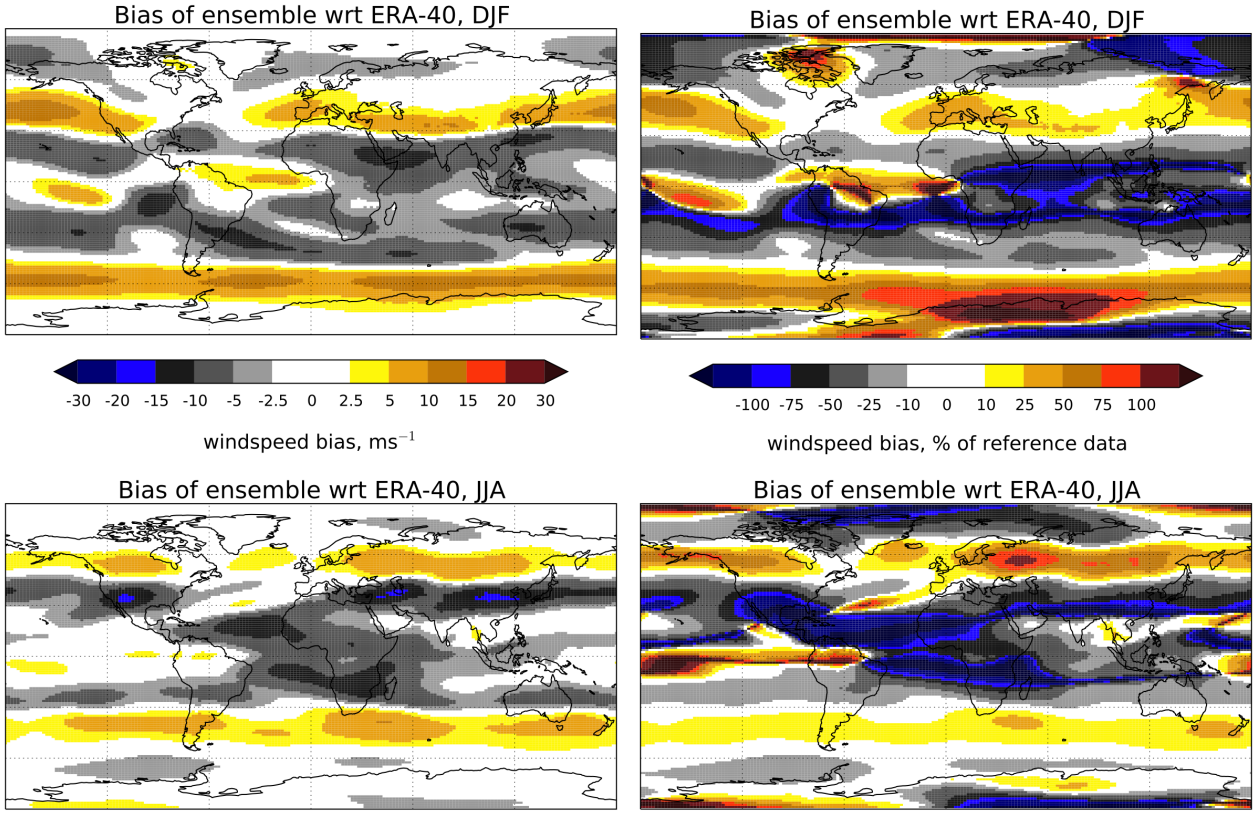


Figure 5.2.4: Biases of the seasonal mean Westerly wind @ 200hPa in the HadAM3P ensemble, with respect to ERA–40, for the DJF and JJA seasons from Dec. 1961 to Nov. 2000. The actual value (in ms^{-1}) is shown in the left panel and the bias expressed as a percentage of the observation is shown in the right panel.

Figure 5.2.4 shows the bias, with respect to ERA–40, of the U–wind @ 200hPa, i.e. the bias of the upper level wind in the large ensemble of HadAM3P. The actual values of the bias (in ms^{-1}) are plotted, as are the values of the bias expressed in terms of the percentage of the ERA–40 values. As can be seen in the Figure, in the DJF season the large ensemble of HadAM3P has a relatively low bias in the Northern Hemisphere, with a bias of between ± 2.5

and ±10 ms^{-1} and large areas of bias below ±2.5. The strongest bias over Europe is in the Mediterranean and over the Iberiam peninsula, where positive anomalies of up to 10 ms^{-1} are seen. When expressed as percentages of the U–wind values in ERA–40, this equates to up to 50%. Expressing the bias in terms of percentages reveals large biases to the west of Greenland (positive over 100%) and over the Arctic north of the West Pacific (negative up to 150%). However, the European region and the North Atlantic, the area of interest where the Polar Jet is located (between 30°N and 60°N), has smaller biases in DJF.

The bias in the upper level Westerly wind is quantified by using Equation 5.1.3 to calculate the root mean square error (RMSE), with respect to ERA–40, for each season, over the whole globe and the five regions detailed in section 1.7. These RMSEs are shown in Table 5.2.4. As can be seen in the Table, the highest RMSEs occur for the entire globe in each season

	global		all EU		SREX CEU		SREX MED		SREX NEU		UK	
DJF actual	5.70	(38.65)	0.69	(4.31)	0.33	(2.47)	0.54	(2.89)	0.30	(2.02)	0.11	(0.67)
MAM actual	4.42	(29.17)	0.66	(4.25)	0.24	(2.24)	0.59	(3.31)	0.16	(1.44)	0.08	(0.68)
JJA actual	5.13	(41.42)	1.03	(7.26)	0.38	(4.15)	0.90	(4.51)	0.32	(3.89)	0.08	(0.60)
SON actual	4.98	(34.02)	0.29	(1.97)	0.12	(0.87)	0.26	(1.70)	0.07	(0.46)	0.01	(0.09)

Table 5.2.4: Biases of the seasonal mean Westerly wind @ 200hPa in the large ensemble of HadAM3P with respect to the ERA–40 reanalysis dataset. Root mean square error (RMSE) values are shown for the actual bias (in ms^{-1}, in the grey column) and expressed as a percentage of the observation (in the white column) for the entire globe and five regions discussed in section 1.7.

with errors of between 4.42 and 5.70 ms^{-1} (29.17 to 41.42%). Over the European region these errors decrease and further decrease for each sub–region of Europe. The UK region, which also has the smallest domain of all the regions, has the smallest error across all of the seasons. Across the regions, the JJA season has the largest error of all the seasons. How these biases affect the frequency of occurrence and intensity of cyclones over Europe will be investigated further in chapter 7.

5.2.5 Summary and conclusions of validating seasonal means.

This s has compared four climate variables from the large ensemble of HadAM3P to both observations and reanalysis data. Seasonal mean air temperature has significant biases, with

respect to CRUTS and ERA–40, for the JJA season over the entire globe and also over three of the five European regions. However, over the UK and Northern Europe in JJA no significant biases are seen, and no significant biases are seen in the other regions for all seasons. Comparing precipitation in the large ensemble to GPCP and ERA–40 reveals a significant bias in the Central European region in JJA. This is also the region with the highest bias in JJA temperature and there is likely to be a causal relationship between these, with a reduction in soil moisture affecting the available precipitation.

Two atmospheric circulation related variables are compared to ERA–40. These are the mean sea level pressure (MSLP), where positive biases in the Equator to Pole pressure gradient and over the Mediterranean Sea are found in DJF and negative biases are found over Central Europe and the Mediterranean in JJA. The other atmospheric circulation variable is the Westerly wind at 200hPa, which can be used as a measure of jet stream strength. Here, very low biases are found for the Polar Jet region.

The HadAM3P model, run as a very large ensemble with initial condition perturbations, appears to realistically model the mean state of the climate system over Europe for all seasons except for JJA. All other seasons are indistinguishable from ERA–40 for temperature and precipitation as the RMSE with respect to their observed datasets is less than the RMS of the maximum bias of ERA–40 for those variables. However, the mean state does not reveal much about how the model represents the tails of the distributions of climate variables, which is where extreme events occur and what probabilistic event attribution is most concerned with. The next section (5.3) will assess HadAM3P in terms of how well it models the deviation from the mean state of the climate on seasonal timescales. Chapter 6 examines how well the model represents the distribution of climate variables on a daily timescale, including the tails of the distribution.

5.3 Validating extremes of seasonal means

Section 5.2 compares the bias, with respect to ERA–40, in the seasonal mean of the large ensemble of the HadAM3P climate model for four climate variables. However, probabilistic

event attribution studies are concerned more with the tails of the distributions of climate variables than they are with the mean value. This section compares the high and low percentiles of seasonal mean values of climate variables with those in observed and reanalysis datasets to determine whether the model can represent the distribution of climate variables and, in particular, the extreme values.

Sections 5.2.1 to 5.2.4 show that the HadAM3P model has a number of biases, when compared to observed and reanalysis data products. Most notable of the these is the warm bias over Europe in the June, July, August (JJA) season. However, although it is worthwhile to ensure the model realistically represents the observed quantities of climate variables, probabilistic event attribution studies are concerned with the change in the percentiles of these variables between two ensembles of the climate model, one run with boundary forcings matching those from the world with anthropogenic forcing and one run with boundary forcings with the anthropogenic component removed. Therefore, the ability of the model to represent extreme values away from the mean state is important. Here, the inter–percentile range is used. For a percentile p, the inter–percentile range is defined as:

$$I_p = V_p - V_{50} \tag{5.3.1}$$

where I_p is the inter–percentile range, V_p is the value at the pth percentile and V_{50} is the value of the 50th percentile. The inter–percentile range is the difference between the value at that percentile and the median. This effectively removes the bias from the model, as all values are now expressed as anomalies from the median state. So, for the HadAM3P model, which is biased warm in the JJA season, if the 90th percentile is biased as warm as the mean state then the inter–percentile range will closely match the inter–percentile range in the CRUTS observed dataset (Mitchell and Jones, 2005).

It is worth noting that extreme seasons are being evaluated in this section. An extreme season may be one that is anomalously warm (or cold) or wet (or dry) or have a low mean sea level pressure, when the entire season is taken into account. Seasonal means are calculated from 1961 to 2000 (1979 to 2000 for precipitation) and the percentiles are calculated, per grid

box and for each season, from these 40 seasonal means (22 for precipitation). For the large ensemble of HadAM3P, the percentiles are calculated from all ensemble members, i.e. 2500 members per year for each season, so there are effectively 100,000 years of a season from which to calculate the percentiles. For the observations and reanalysis, there is only one member for each year. Chapter 6 evaluates how well the large ensemble of the HadAM3P model represents the distribution, including the extreme percentiles, of daily climate values.

5.3.1 Validating extremes of seasonal mean temperature

Figure 5.3.1 shows the bias in the inter–percentile range of seasonal mean temperature for the 5th (I_5) and 95th (I_{95}) percentiles for the DJF and JJA seasons respectively. For DJF, the 5th percentile is chosen so as to show anomalously cold winters in the Northern Hemisphere. In DJF, there is very little bias except for Greenland, which has a cold bias in the inter–percentile range and the West of North America. In Figure 5.2.1 Greenland has a very cold bias of up to 12°C, whereas North America has a mix of warm and cold biases of up to 6°C.

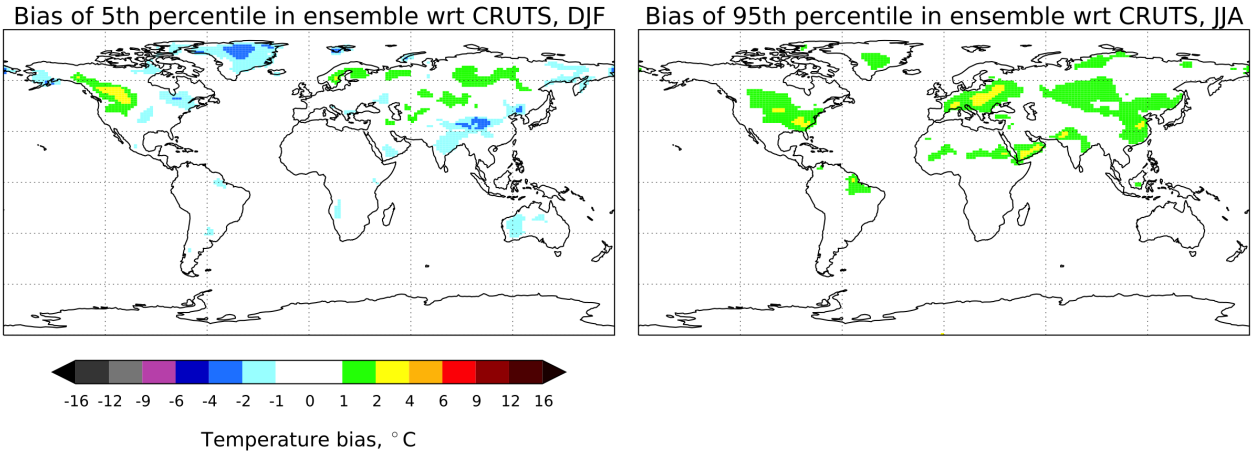


Figure 5.3.1: Biases in the inter–percentile range for the seasonal mean temperatures in the HadAM3P ensemble, with respect to the CRUTS observations. The bias in the inter–percentile range of the 5th percentile is shown for DJF, and the bias in the inter–percentile range of the 95th percentile is shown for JJA.

For JJA, the areas of bias in I_{95} also correspond to the areas where there are large biases in the actual temperature values. These are over North America, which has a warm bias in I_{95} of up to 4°C and Central Europe, which has a warm bias of up to 12°C, as shown in

Figure 5.2.1. The 95th percentile is chosen so as to show anomalously warm summers in the Northern Hemisphere.

	global	all EU	SREX CEU	SREX MED	SREX NEU	UK
DJF 1st	0.71	0.17	0.10	0.10	0.09	0.02
DJF 5th	0.55	0.13	0.07	0.09	0.06	0.01
DJF 95th	0.78	0.31	**0.23**	0.21	0.04	0.03
DJF 99th	0.58	0.14	0.09	0.10	0.05	0.01
MAM 1st	0.47	0.12	0.05	0.07	0.08	0.02
MAM 5th	0.35	0.08	0.04	0.06	0.04	0.01
MAM 95th	0.57	0.23	**0.16**	0.17	0.03	0.02
MAM 99th	0.40	0.09	0.06	0.06	0.03	0.01
JJA 1st	0.44	0.08	0.04	0.05	0.05	0.01
JJA 5th	0.33	0.09	0.05	0.06	0.03	0.00
JJA 95th	0.46	0.18	0.15	0.09	0.05	0.01
JJA 99th	0.39	0.10	0.07	0.05	0.04	0.01
SON 1st	0.68	0.14	0.08	0.08	0.08	0.01
SON 5th	0.51	0.16	**0.10**	0.11	0.05	0.01
SON 95th	0.67	**0.27**	**0.22**	0.13	0.09	0.03
SON 99th	0.55	0.14	0.11	0.08	0.04	0.02

Table 5.3.1: Biases in the inter–percentile range for the seasonal mean temperatures in the HadAM3P ensemble, with respect to the CRUTS observations. Root mean square error (RMSE, °C) values are shown for the entire globe and the five regions discussed in section 1.7 for each season and for two low percentiles and two high percentiles. Values highlighted in bold are greater than the corresponding RMSE value for the region and season in Table 5.2.1.

Table 5.3.1 presents the results of calculating the RMSE, using Equation 5.1.3, of the inter–percentile range of seasonal mean temperature in the large ensemble of HadAM3P with respect to CRUTS, for the five regions described in section 1.7 and for percentiles in both tails of the distribution of temperature. RMSE values that are greater than the RMSE for the season and region in Table 5.2.1 are highlighted in bold. This indicates that the RMSE of the inter-percentile range is greater than the RMSE of the ensemble mean of the actual temperature value, and can be considered to show that the inter–percentile range is significantly biased.

As can be seen in Table 5.3.1, the only sub–region where the RMSE of the inter–percentile range is greater than the RMSE of the ensemble mean is over Central Europe (SREX CEU).

This corresponds to the region with the highest bias in the temperature in Figure 5.2.1 and Table 5.2.1. However, the season which has the largest warm bias (JJA) has a RMSE for the inter–percentile range that is lower than the RMSE for the actual temperature value. The overall reduction in error when comparing RMSEs of the ensemble mean and the inter–percentile range shows that by expressing the temperature values as anomalies from a mean or median state, a probabilistic attribution study can still be relevant even if the mean state is itself severely biased.

5.3.2 Validating extremes of seasonal mean precipitation

Figure 5.3.2 shows the bias in the inter–percentile range of the seasonal accumulated precipitation for the 5th (I_5) and 95th (I_{95}) percentiles for the JJA and DJF seasons respectively. For DJF, the 95th percentile is chosen so as to show anomalously wet winters in the Northern Hemisphere. In DJF, there is very little bias over Western Europe, with the exception of the UK and Ireland, which shows a dry bias of up to 100mm per season and the Iberian Peninsula which shows a similar bias. Mountainous regions, such as the Alps, show a slightly wet bias of up to 50mm. However, the large area of dry bias seen over Northern Europe and Russia in Figure 5.2.2 is not present when considering I_{95}. Areas that have a low bias in Figure 5.2.2 still have a low bias in Figure 5.3.2 with an improvement in the bias over Indonesia, Malaysia and Papua New Guinea. However, a large dry bias is present over the middle of the Pacific Ocean. Although this does not affect any areas which are of interest to probabilistic event attribution studies, the lack of extreme rainfall in this area may cause changes to global circulation patterns, which could in themselves have some influence on extreme weather over regions of interest. However, an investigation into this is outside the scope of this thesis.

For JJA, the 5th percentile is chosen so as to show anomalously dry summers over Europe in the Northern Hemisphere. There are very few areas of bias for I_5, with Central Europe having a wet bias of up to 100mm per season. This is the area which corresponds to the large warm bias of up to 12°C seen in Figure 5.2.1 and a bias of up to 4°C for I_{95} seen in Figure 5.3.2. That an area with a high temperature bias should also be too wet in I_5

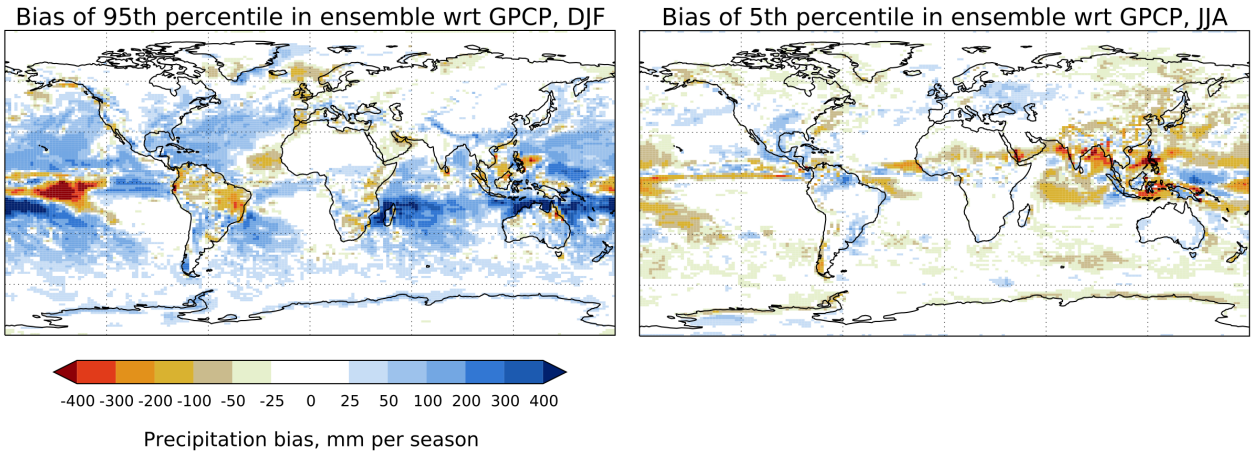


Figure 5.3.2: Biases in the inter–percentile range for the seasonal accumulated precipitation in the HadAM3P ensemble, with respect to the GPCP observations. The bias in the inter–percentile range of the 95th percentile is shown for DJF, and the bias in the inter–percentile range of the 5th percentile is shown for JJA.

for seasonal accumulated precipitation, which is akin to drought conditions, is surprising as the high temperatures should remove soil moisture and reduce the amount of available precipitation. The wet bias in I_5 could be related to the convection scheme used in HadAM3P generating relatively high amounts of precipitation (the precipitation is still low) due to the high temperatures. However, it is beyond the scope of this thesis to investigate this. There is also a wet bias of up to 50mm over the UK and Ireland which, coupled with the dry bias in I_{95} for DJF may indicate that the large ensemble of HadAM3P has not adequately captured, over the UK, the width of the distribution of seasonal accumulated precipitation that is present in the observations.

Table 5.3.2 presents the results of calculating the RMSE, using Equation 5.1.3, of the inter–percentile range of seasonal mean accumulated precipitaion in the large ensemble of HadAM3P with respect to GPCP, for the five regions described in section 1.7 and for percentiles in both tails of the distribution of precipitation. RMSE values that are greater than the RMSE for the season and region in Table 5.1.2 are highlighted in bold. This indicates that the RMSE of the inter-percentile range is greater than the RMSE of the ensemble mean of the actual seasonal accumulated precipitation, and can be considered to show that the

	global	all EU	SREX CEU	SREX MED	SREX NEU	UK
DJF 1st	81.18	7.98	3.03	5.81	4.55	2.14
DJF 5th	72.57	5.68	2.89	3.50	3.42	1.73
DJF 95th	78.39	5.95	4.09	3.12	3.00	1.53
DJF 99th	74.52	7.50	3.14	4.68	4.95	2.33
MAM 1st	66.29	7.05	2.54	5.28	3.91	2.06
MAM 5th	59.13	4.66	2.39	2.85	2.81	1.74
MAM 95th	62.94	4.89	3.33	2.62	2.44	1.45
MAM 99th	59.21	6.21	2.92	4.11	3.63	1.50
JJA 1st	110.23	9.80	3.03	7.14	5.98	3.10
JJA 5th	103.15	5.99	3.17	3.49	3.69	1.15
JJA 95th	99.82	5.86	3.33	3.75	3.02	1.40
JJA 99th	85.80	8.04	3.12	6.20	4.06	1.69
SON 1st	166.59	10.88	4.55	8.17	5.56	2.88
SON 5th	164.92	9.20	5.11	6.09	4.61	1.19
SON 95th	157.79	8.61	5.27	5.39	4.16	1.77
SON 99th	137.65	12.55	4.48	**10.40**	5.40	2.29

Table 5.3.2: Biases in the inter–percentile range for the seasonal accumulated precipitation in the HadAM3P ensemble, with respect to the GPCP observations. Root mean square error (RMSE, mm/season) values are shown for the entire globe and the five regions discussed in section 1.7 for each season and for two low percentiles and two high percentiles. Values highlighted in bold are greater than the corresponding RMSE value for the region and season in Table 5.1.2.

inter–percentile range is significantly biased.

As can be seen in the Table, the only sub–region where the RMSE of the inter–percentile range is greater than the RMSE of the ensemble mean is over the Mediterranean in I_{99} for the September, October, November (SON) season. The large dry bias for Central Europe (SREX CEU) in the JJA season seen in Table 5.1.2 has been removed for all percentiles by considering inter–percentile ranges rather than calculating the bias between actual values in the observations and their corresponding values in the large ensemble of HadAM3P.

5.3.3 Validating extremes of seasonal mean MSLP

Figure 5.3.3 shows the bias in the inter–percentile range of seasonal mean MSLP for the 5th (I_5) and 95th (I_{95}) percentiles for the DJF and JJA seasons respectively. For DJF, the 5th percentile is chosen so as to show winters with anomalously low mean MSLP in the

Northern Hemisphere. This may be caused by a persistent low over an area or by a succession of cyclones arriving in the domain. In I_5 for DJF, the positive bias in the Equator to Pole pressure gradient has been reduced slightly, from a bias of up to 10hPa in Figure 5.2.3 to a bias of up to 5hPa for I_5. There is still a positive bias east of Iceland and north of Norway. A negative bias over Alaska in the DJF seasonal mean MSLP has now become a positive bias in the I_5 MSLP. However, the positive bias of 2 to 5hPa over the Mediterranean in Figure 5.2.3 is not present in I_5.

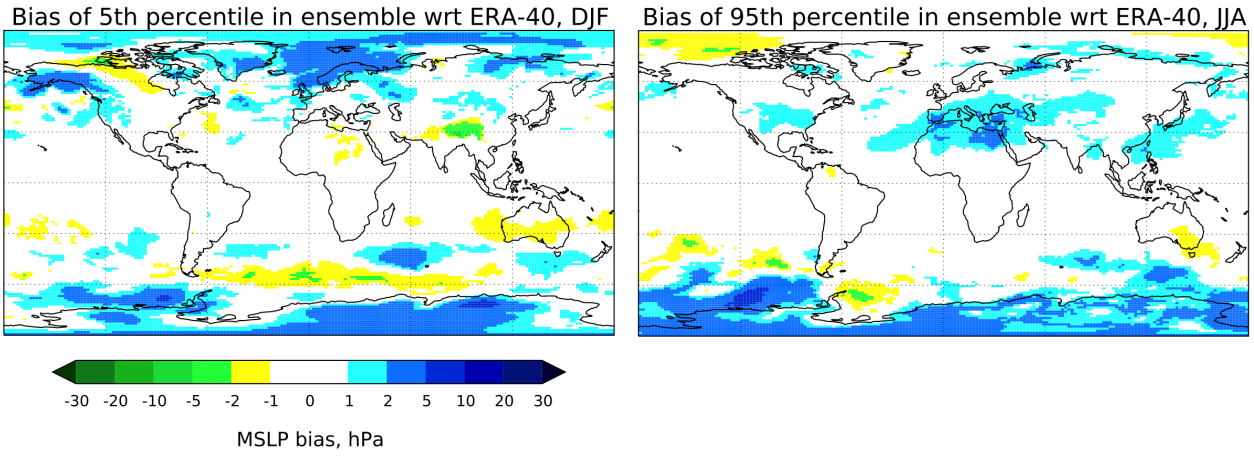


Figure 5.3.3: Biases in the inter–percentile range for the seasonal mean MSLP in the HadAM3P ensemble, with respect to the ERA–40 reanalysis. The bias in the inter–percentile range of the 5th percentile is shown for DJF, and the bias in the inter–percentile range of the 95th percentile is shown for JJA.

In JJA for I_{95}, by considering the inter–percentile range much of the bias, especially the negative bias over land seen in Figure 5.2.3, has been removed. A small negative bias has been introduced north of Alaska and Canada, and a negative bias of up to 5hPa is now present over the Mediterranean and North Africa. However, the overall bias is much improved when compared to the seasonal mean of MSLP.

Table 5.3.3 shows the RMSE, calculated using Equation 5.1.3, of the inter–percentile range of seasonal mean MSLP in the large ensemble of HadAM3P with respect to ERA–40, for the five regions described in section 1.7 and for percentiles in both the high and low tails of the distribution of MSLP. RMSE values that are greater than the RMSE for the season

	global	all EU	SREX CEU	SREX MED	SREX NEU	UK
DJF 1st	1.42	0.29	**0.10**	0.16	0.22	0.04
DJF 5th	1.12	0.25	**0.18**	0.12	0.12	**0.06**
DJF 95th	1.16	0.21	**0.08**	0.15	0.12	0.01
DJF 99th	0.92	0.21	**0.10**	0.12	0.14	**0.06**
MAM 1st	0.96	**0.27**	**0.07**	0.07	**0.25**	0.06
MAM 5th	0.76	0.17	**0.11**	0.07	0.11	0.02
MAM 95th	0.78	0.14	**0.06**	0.11	0.07	0.02
MAM 99th	0.74	0.11	**0.05**	0.06	0.08	0.03
JJA 1st	0.95	0.26	0.13	0.09	**0.20**	0.04
JJA 5th	0.68	0.15	0.09	0.11	0.06	0.02
JJA 95th	0.89	0.25	0.11	0.22	0.05	0.01
JJA 99th	0.65	0.15	0.06	0.11	0.08	0.04
SON 1st	1.10	**0.21**	0.12	0.09	**0.15**	0.04
SON 5th	0.98	**0.22**	0.13	0.12	**0.12**	0.06
SON 95th	1.46	**0.34**	0.15	0.29	**0.10**	0.02
SON 99th	1.01	**0.23**	0.09	0.17	**0.12**	0.03

Table 5.3.3: Biases in the inter–percentile range for the seasonal mean MSLP in the HadAM3P ensemble, with respect to the ERA–40 reanalysis. Root mean square error (RMSE, hPa) values are shown for the entire globe and the five regions discussed in section 1.7 for each season and for two low percentiles and two high percentiles. Values highlighted in bold are greater than the corresponding RMSE value for the region and season in Table 5.2.3.

and region in Table 5.2.3 are highlighted in bold. This indicates that the RMSE of the inter-percentile range is greater than the RMSE of the ensemble mean of the actual seasonal MSLP, and shows that the inter–percentile range is more biased than the seasonal mean MSLP.

Globally, considering the inter–percentile range has a large improvement in the RMSEs, when compared to the seasonal mean MSLP for each season, with mean reductions in the RMSE of 81% for DJF, 86% for MAM, 77% for JJA and 61% for DJF. However, unlike temperature and precipitation, considering the inter–percentile ranges does not reduce the RMSE for nearly all regions / season / percentile combinations, with 20 out of the 96 (21%) combinations having a higher RMSE in the inter–percentile data. This indicates that, for a sizeable number of regions and seasons, correcting the bias by considering the inter–percentile range is not sufficient.

Cyclones (low pressure systems) are responsible for a large proportion of the extreme weather experienced in Europe, such as high winds, high precipitation and storm surges. Persistent high pressure can lead to heat waves in summer (JJA) or cold spells in winter (DJF). Examining the seasonal mean of the MSLP field for a season can give some indication as to what the prevailing conditions for that season were. For example, a lower than climatology MSLP over Europe in DJF may indicate that there many low pressure events occurred in that season. However, to know for certain the frequency and magnitude of those events requires an analysis at a finer time resolution than seasonal. Chapter 6 examines the distribution of climate variables, including MSLP, in the large ensemble of HadAM3P on a daily timescale, whereas chapter 7 uses the feature identification and tracking method of chapter 2 to detect individual low pressure events and compare the frequency, intensity and spatial distribution of these events to those in ERA–40.

The feature identification and tracking method of chapter 2 identifies features as anomalies from the background field, which is a lower resolution version of the MSLP data. Using this method ensures some degree of bias correction, as the large scale bias in the MSLP field will be removed. However, the model still has to represent the depth of the cyclone satisfactorily, with respect to the background field, for the identification method to firstly detect the cyclone and secondly to calculate the depth of the depression and compare to the depths in ERA–40. Chapter 7 carries out this analysis and determines how well the model represents cyclones in light of the biases presented here.

5.4 Variability of seasonal mean temperature

One measure used to test ensemble forecast systems is to compare the ensemble spread (standard deviation) to the root mean square error (RMSE) of the forecast when compared to the observations (Jolliffe and Stephenson (2003), Wilks (2011)). If the reference data (observations or reanalysis) cannot be distinguished from the ensemble data, then the RMSE of the ensemble must equal the ensemble standard deviation (Jolliffe and Stephenson, 2003). In this section, this concept is adapted to the large ensemble of HadAM3P, by comparing

the RMSE of the ensemble to the standard deviation of the ensemble.

Figure 5.4.1 shows the RMSE and standard deviation for the seasonal mean temperature at 2m in the large ensemble of HadAM3P, for both DJF and JJA. As can be seen, in DJF, the areas with the highest RMSE also have a correspondingly high standard deviation, except for over Greenland, Russia and Alaska. In JJA, the relationship between RMSE and standard deviation is much more pronounced with areas with very high RMSE, such as North America and Southern Europe, also having high variability with large standard deviations.

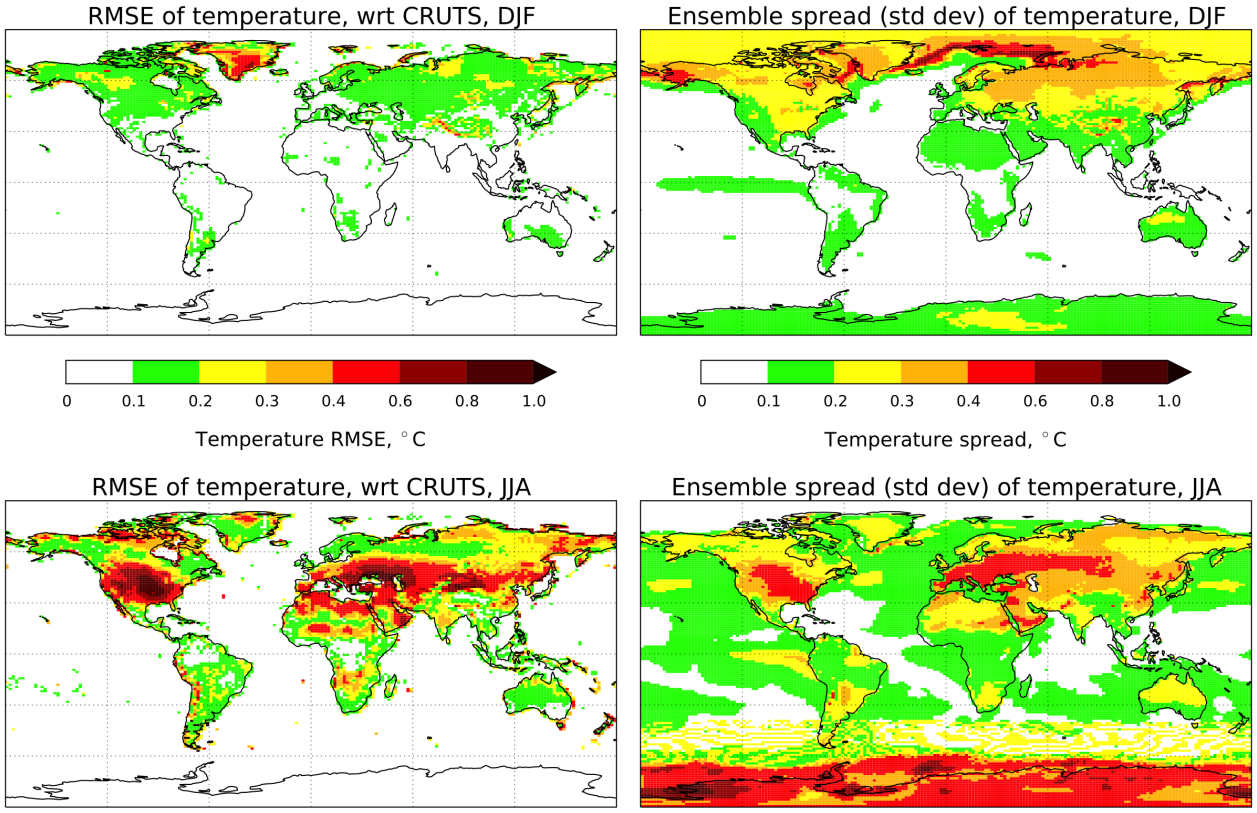


Figure 5.4.1: Root mean square error (RMSE, °) of seasonal mean temperature at 2m in the HadAM3P ensemble, with respect to CRUTS and the standard deviation of the seasonal mean temperature in the ensemble.

Table 5.4.1 quantifies the correlation between the RMSE and standard deviation in the ensemble for 2m temperature by calculating the correlation coefficient from fitting a linear regression model (Wilks, 2011) to the values of RMSE and standard deviation for all grid boxes in a region. The five regions are defined in section 1.7.

	global	all EU	SREX CEU	SREX MED	SREX NEU	UK
DJF	0.793	0.794	0.968	0.779	0.734	0.779
MAM	0.573	0.857	0.988	0.902	0.760	0.881
JJA	0.468	0.900	0.942	0.946	0.838	0.950
SON	0.597	0.839	0.960	0.830	0.775	0.821

Table 5.4.1: Correlation between the RMSE (°)of seasonal mean temperature at 2m in the HadAM3P ensemble, with respect to CRUTS and the standard deviation of the temperature in the ensemble. Correlations are computed for all grid boxes in the five regions described in section 1.7.

As Table 5.4.1 shows, the highest correlation coefficients are for the regions which have the most bias, as shown in Figure 5.2.1 and Table 5.2.1. The highest correlation values are over the Central Europe (SREX CEU) region and the Mediterranean (SREX MED), which correspond to the areas with the highest bias, especially in JJA, where the very high biases around the Mediterranean and Caspian Seas occur. The high correlation between the RMSE and standard deviation shows that, although the model is biased in its year to year modelling of the surface temperature, the error is within the spread of the ensemble. This gives confidence that, in an extreme event attribution study, the model has sufficient variability to capture the potential high temperature events within a season.

5.5 Variability of seasonal mean precipitation

Figure 5.5.1 shows the RMSE, with respect to GPCP, and standard deviation for the seasonal accumulated precipitation in the large ensemble of HadAM3P, for both DJF and JJA. As can be seen, for both DJF and JJA, the areas with the highest RMSE also have a correspondingly high standard deviation. For precipitation, these are largely confined to areas over the tropical oceans. This is a similar error to that displayed by ERA–40 over the tropics (Uppala et al., 2005) and explains why there is little significant bias over the tropical oceans in Figure 5.2.2, as the model actually has a lower bias over the tropical oceans.

Over land, in DJF, the areas with the highest RMSE, such as the western coast of North America, the UK and Norway and the southern coast of Greenland also have the highest standard deviation. Two areas which have high RMSE but relatively low standard deviation

is over Brazil and the South East Asia islands (Indonesia, Borneo, etc.). These are areas with large amounts of rainforest and so the model may not be correctly representing the land–surface processes associated with precipitation over rainforests. However, as this thesis concentrates on extreme weather events over Europe, no further analysis will be made. In JJA, the high RMSE values are restricted to the oceans, with a small error over Europe (RMSE of 0.1) largely being matched by the standard deviation.

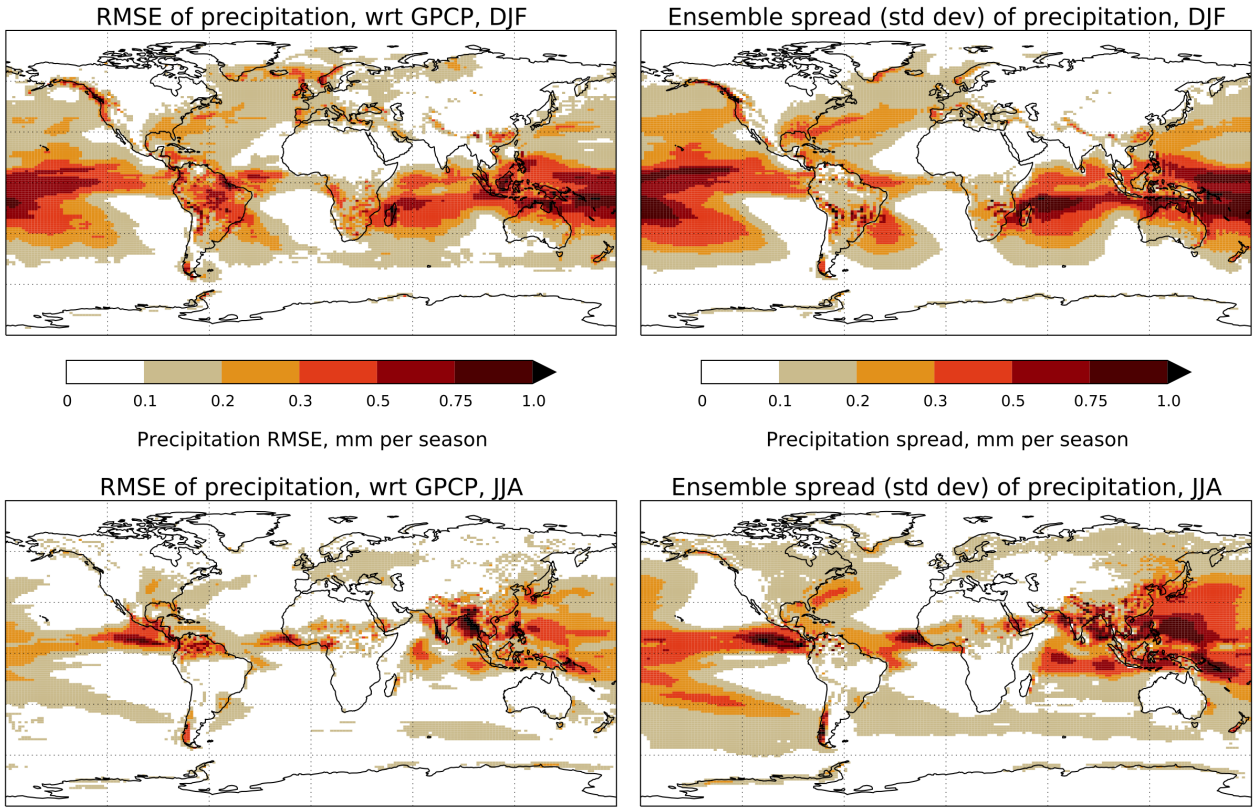


Figure 5.5.1: Root mean square error (RMSE, mm/season) of seasonal accumulated precipitation in the HadAM3P ensemble, with respect to GPCP and the standard deviation of the seasonal accumulated precipitation in the ensemble.

Table 5.5.1 shows the correlation between the RMSE and standard deviation in the ensemble for accumulated precipitation by calculating the correlation coefficient for the values of RMSE and standard deviation for all grid boxes in the five regions, which are defined in section 1.7.

As Table 5.5.1 shows, unlike temperature, the regions with the highest correlation coefficients

	global	all EU	SREX CEU	SREX MED	SREX NEU	UK
DJF	0.886	0.942	0.968	0.877	0.961	0.987
MAM	0.910	0.955	0.976	0.935	0.966	0.966
JJA	0.860	0.944	0.899	0.900	0.971	0.997
SON	0.889	0.969	0.970	0.963	0.980	0.994

Table 5.5.1: Correlation between the RMSE (mm/season) of seasonal accumulated precipitation in the HadAM3P ensemble, with respect to GPCP and the standard deviation (mm/season) of the precipitation in the ensemble. Correlations are computed for all grid boxes in the five regions described in section 1.7.

do not coincide with the regions which have the largest RMSE in Table 5.1.2. The Mediterranean region (SREX MED) has one of the largest errors, across all seasons, in Table 5.1.2, but has the lowest correlation coefficients of the five European regions. This is also true for the Central European region (SREX CEU), which has a high RMSE in Table 5.1.2 but a relatively low correlation in Table 5.5.1. However, for the Northern European region, for DJF and SON, the opposite is true. These seasons have high RMSE, and a high correlation value, which would indicate that the spread of the ensemble closely matches the error.

These results indicate that, for some regions, the error in the precipitation may not be within the ensemble spread and that the model may not be able to capture the full spread of the ensemble, leading to an under–representation of very wet events or very dry events. This is further explored in section 6.2.3 and section 7.5.2.

5.6 Summary and conclusions

This chapter starts by discussing the differences and biases between one observed and two reanalysis datasets for climatological seasonal mean near–surface temperature and accumulated monthly precipitation. Section 5.1.3 presents a method which uses these differences to validate a very large ensemble of the HadAM3P climate model, the details and experimental design of which is presented in chapter 4.

Section 5.2.1 shows that there are large, and significant biases in the seasonal mean temperature at 2m between the large ensemble of HadAM3P and the CRUTS observed dataset,

especially over Europe in JJA. However, section 5.3.1 suggests that, by considering the difference between high or low percentiles (95th or 5th) and the 50th percentile, much of this bias disappears and that a probabilistic attribution study can still be relevant even if the mean state is itself severely biased. This also suggests that correcting the mean bias of the model would remove much of the bias in the entire distribution of seasonal mean temperatures.

Similarly, section 5.2.2 shows that there are biases in the accumulated precipitation for a season, particularly over Central Europe and North Africa in JJA and over North Africa in DJF. Taking the difference between percentiles in the tails of the distribution and the median reduces these biases in all season and region combinations, as shown in section 5.3.2.

Section 5.2.3 and 5.3.3 shows that the bias in the seasonal mean MSLP cannot be removed by considering the differences between percentiles. How this affects the daily MSLP values will be investigated in chapter 6 and how this affects the representation of extra tropical cyclones in the model in chapter 7.

Finally, sections 5.4 and 5.5 investigate how the errors in the large ensemble of HadAM3P relate to the ensemble spread, by comparing the RMSE with the standard deviation for temperature and precipitation. For temperature, the RMSE and standard deviation have a good correlation for all season–region combinations, indicating that the model can capture the variability in temperature that is necessary to capture the high and low temperature events that are of interest in extreme event attribution. However for precipitation, there are some season–region combinations where the RMSE and standard deviation do not correlate closely, indicating that the model may not be able to capture the full spread of the ensemble, leading to an under-representation of very wet events or very dry events

Overall, this chapter has shown that there are biases in the seasonal means of the temperature, precipitation, MSLP and high level winds and that the bias can be corrected by removing the mean bias of the model. The bias can be corrected either considering all values as anomalies from the seasonal mean, or adding the observed seasonal mean to the anomalies to produce climate variable values, rather than anomalies. Applying this correction may fix the distribution of the output variables from the model, but it cannot fix the structural bias

in the model. As the model is a combination of non–linear processes, fixing the bias for the mean state may not ensure that all potential states of the model are corrected. It is quite surprising that the model performs as well as it does in JJA precipitation, given the high biases in JJA temperature and the links between temperature and precipitation related to evaporation and the Clausius–Clapeyron relationship (Allen and Ingram, 2002). Chapter 6 will explore this further, by applying a bias correction to the daily climate variables.

6 Validation of variables on a daily timescale from a very large ensemble of a climate model

Chapter 5 presents the validation of the very large ensemble of a climate model (the formulation of which is described in chapter 4) on a seasonal timescale. However, it is likely that extreme weather events occur on a much shorter timescale than the seasonal or monthly. For example, the floods in the UK that occurred in Autumn 2000 were the product of 2 weeks of above average rainfall, thanks to an Eastward displacement of the North Atlantic jet stream (Pall et al., 2011). The flooding of Tewkesbury in the UK in 2007 was the result of 2 days of heavy rain, with 132.7mm falling during the 19th and 20th July on already saturated ground which was a result of heavy rainfall in June and July (Prior and Beswick, 2008). During the European heatwave of 2003, France recorded 7 days where the temperature exceeded 40°C (Stott et al., 2004). However, the Russian heatwave of 2010 lasted for a month and a half (Otto et al., 2012) showing the importance of testing the validity of the model on different timescales.

This chapter presents the results from using verification techniques to test the validity of the output from the very large ensemble. The analysis concentrates on daily values of the accumulated precipitation, minimum and maximum near-surface temperature, mean sea level pressure (MSLP) and maximum windspeed at 10m climate variables. The large ensemble of HadAM3P computed using the methods in chapter 4 produces a very large volume of data, in the order of 10 terabytes (TB). To reduce the amount of data to be analysed, the output of the large ensemble is first sub–set to the five SREX regions described in section 1.7, and then an area weighted mean is taken of the sub–set of data, which further reduces the data volume. As a result of this processing a time series of daily climate variables is produced for

each region.

An alternative method would be to not take an area weighted average, but to include all the points within the region in the daily time series. However, this would have increased the data size by between 42 (for the UK region) and 972 (for the entire EU region) times. Chapter 7 examines points close to the centres of cyclones by finding the maximum precipitation and windspeed within 1000km radius of the central low pressure. This provides a method of validating the extreme values of precipitation and windspeed on a point by point basis.

This chapter uses ERA–Interim reanalysis data (Dee et al., 2011) from 1989 to 2009 as the verification dataset. ERA–Interim is used in preference to the previously used observed datasets (CRUTS and GPCP) for two reasons. Firstly, ERA–Interim data is available at daily temporal resolution, which enables the performance of the model on daily timescales to be assessed. Secondly, using ERA–Interim ensures a comparison can be made between the storm tracking metrics derived by applying the methods of chapter 2 to ERA–Interim, as is done in chapter 3. These storm tracks and the precipitation, windspeed and pressure associated with the storm tracks are assessed in chapter 7 with respect to the ERA–Interim values. The 1989 to 2009 period is chosen so as to match the period used in the verification of the storm–tracking algorithm in chapter 3 and the analysis of the associated variables in chapter 7. ERA–Interim contains data from 1979 to the present day, so a longer period could be used with the caveat that a comparison to the analyses in chapter 3 and 7 could not be made. ERA–Interim brings the ECMWF reanalysis product up to date, with additional data being added in a quasi–operational way, whereas ERA–40 stops in 2002. Additionally, improvements have been made to the output of the reanalysis system, in terms of climate variables matching the observed datasets (Dee et al., 2011). This is especially true for precipitation, where ERA–Interim now much more closely matches GPCP, when compared to ERA–40, especially for rainfall over the tropical oceans. This improvement is due to the introduction of a completely revised humidity analysis scheme (Dee et al., 2011). The two reanalysis products considered in chapter 5 (ERA–40 and NCEP/NCAR) also contain daily data values. However, ERA–Interim is preferred for the analysis in this section due to the improvements in the humidity analysis, which is likely to have an impact on daily

precipitation values, as well as the time period covered.

6.1 Methods used to validate daily data

This chapter presents the results from applying a single verification technique to the output of the very large ensemble, by comparing the value at the percentiles of the distribution of climate variables produced by the large ensemble of HadAM3P with the values in the observations at corresponding percentiles. To do this quantile–quantile plots (Q–Q plots) are used (Wilks, 2011) as they provide an easy to interpret graphical representation of the non–parametric goodness of fit between two populations. The observed values are one population, whereas the values from the large ensemble of HadAM3P are the other population. Additionally, the root mean square deviation (RMSD) of the values at the percentiles between the two populations is calculated, and a two sample Kolmogorov–Smirnoff (KS) test is used to determine how close the two populations are and whether they are, in fact, statistically indistinguishable from each other.

In order to use these techniques, the two or three spatial dimensions of the model output must be reduced to a single dimension. The variables considered in this section are daily minimum and maximum near-surface temperature, daily accumulated precipitation, maximum windspeed and mean sea level pressure, all of which are two dimensional fields. These are reduced to five sets of one dimensional data by taking the area-weighted average of the variables over a domain that encloses the regions described in section 1.7 . The area-weighted value of the variable over the region is:

$$V = \frac{1}{s} \sum_{i=1,j=1}^{i=m,j=n} w_{i,j} v_{i,j} \ , \qquad s = \sum_{i=1,j=1}^{i=m,j=n} w_{i,j}$$

where m is the number of columns in the region, n is the number of rows, $w_{i,j}$ is the weight at row i and column j, $v_{i,j}$ is the value of the field at row i and column j and s is the sum of the weights. The weights are calculated for each grid box from the latitude position of the

grid box and the longitude and latitude spacing of the grid:

$$w_{i,j} = r^2 \Delta\lambda \cos(\phi_j)$$

where r is the radius of the Earth, $\Delta\lambda$ is the longitudinal spacing, and ϕ_j is the latitude of a grid box. All terms in the equation remain constant, except for ϕ_j, and so, when combined with the averaging function, the weight of a grid box at position i, j in the grid is $w_{i,j} = \cos(\phi_j)$. Both the output from the very large ensemble and the ERA–Interim reanalysis dataset are averaged over the region in the same way. Even though the output from the model and from ERA–Interim are at different resolutions, this allows direct comparison of the variables. The analysis concentrates on using the daily minimum, maximum, mean or accumulation of these area–weighted averages.

6.1.1 Bias correction

Sections 5.2 to 5.3.3 show that there are biases in the climatological seasonal mean temperature, accumulated precipitation, MSLP and mean high level wind in the very large ensemble of HadAM3P. Sections 5.3.1 to 5.3.3 show that, by considering values at low (1st and 5th) and high (95th and 99th) percentiles as anomalies from the 50th percentile, much of the bias in the temperature, precipitation and MSLP can be corrected and that the inter–percentile range of the large ensemble represents the inter–percentile range of the observations.

All models are biased (Box, 1976) but still have utility if their biases can be corrected without affecting the variance of the variable around the mean of the distribution. For example, Figure 5.2.1 shows that the large ensemble of the HadAM3P model has a warm bias in Europe. However, if it can be shown that this bias is constant across the years and forcing inputs for which the model is run and, that removing the bias has no effect on the temperature range, then the model can still be used for extreme weather event attribution studies. Attribution studies of this nature implicitly perform a bias correction by considering the difference between two runs of the model with different forcing scenarios for the return time or fraction of attributable risk (FAR) of an extreme value of a climate variable. Here the

bias in the model is unimportant, so long as the bias is constant for both forcing scenarios.

However to successfully perform statistical testing on the variance of the climate variable, using the two sample KS test, Q–Q plots and RMSD of percentile values, the mean of the distribution of the large ensemble must be equal to the mean of the distribution of the observations. For example, for two distributions that have the same standard deviation, with one distribution having a constant offset (i.e. a constant bias) from the other distribution, a two sample KS test will return a negative result (the distributions are not the same), despite the distributions being identical except for the constant offset.

Throughout this section, a bias correction is used that corrects the mean of the climate variable in the large ensemble of HadAM3P to the mean of the corresponding variable in the observations. This bias correction does not assume any underlying distribution for the climate variables (it is non–parametric) but merely corrects any constant offset between the distributions. A population **B** can be corrected to match the population **A** by applying the following correction:

$$\mathbf{B}' = \mathbf{B} + \mu(\mathbf{A}) - \mu(\mathbf{B}) \tag{6.1.1}$$

where μ the mean of the variable. The effect of this bias correction will be examined for each climate variable, to determine whether a variable in the HadAM3P ensemble has a different distribution to the corresponding variable in ERA–Interim

6.1.2 Non–parametric estimate of confidence intervals

Even though a very large ensemble of a climate model is used to gain a good estimate of the true value of the climate variables, the actual true value can not be known. To gain knowledge of the uncertainty that the estimated value has, a non-parametric estimate of the confidence interval of the value is used. Non–parametric simply refers to the fact that no attempt is made to fit a distribution function to the data, and the confidence interval is defined as the width of the 5th to 95th percentiles.

To estimate these confidence intervals, the individual ensemble members are used to calculate the required quantity and then the distribution of these quantities is used to calculate the interval. For example, to calculate the confidence interval for the 50th percentile of daily minimum temperature, the 50th percentile of minimum temperature is calculated for each ensemble member. From the resulting distribution of values at the 50th percentile of minimum temperature, the 5th and 95th percentiles are calculated and this gives the confidence interval. As 2500 ensemble members per year have been calculated over 21 years, this effectively gives 52,500 years of data from which to calculate the confidence intervals.

6.1.3 Non–parametric goodness of fit test

The two-sided Kolmogorov-Smirnov (K-S) test can be used to test whether two sets of data are samples from the same population and have the same underlying distribution function (Wilks, 2011). Two values are calculated by performing the test, the K-S statistic, which measures the maximum difference between the two empirical cumulative distribution functions (ECDF), and the p-value, which measures the probability that the two datasets have the same distribution. In a K-S test, the null-hypothesis is held that the two samples are drawn from the same population and have the same distribution function. For this to be rejected, with a confidence of 95%, the p-value must be less than 0.05. For values of greater than 0.05, the null-hypothesis cannot be rejected, with a 95% confidence. A K-S statistic close to 0 indicate that the two datasets have the same distribution, as the maximum difference between the two ECDFs is low. The two-sided K-S test is used in this section to determine whether the distribution of climate variables in the large ensemble of HadAM3P is the same as the distribution for the corresponding variable in ERA–Interim and that they are drawn from the same population, i.e. they have the same ECDF.

6.1.4 Model verification plots

Figures 6.2.1 to 6.2.5 show the results of calculating the quantiles for the daily climate variables output from the very large ensemble that have first been area–averaged to the five regions detailed in section 1.7. The quantiles for the corresponding variable in ERA–Interim

are also calculated and the Q–Q plot is created. The analysis is performed for each season and for each region, and a plot per season for a single region (Northern Europe) for each variable is shown.

The Q–Q plot (Jolliffe and Stephenson, 2003), shows a direct and non-parametric comparison between the distribution of the climate variable values in the observed dataset (ERA–Interim) and the corresponding variable values in the very large ensemble, over the 1989 to 2009 period, by plotting the value of the 1st to 99th percentile from the observation (ERA–Interim) against the value for the corresponding percentile from the very large ensemble. The ERA–Interim data is treated as a single ensemble member, with each daily value from 1989 to 2009 being treated as a data point, giving 7560 data points, as the data has been transformed to a 360 day year to match the ensemble's 360 day year. For smoothly varying fields, such as temperature, this transformation is done by fitting a spline to the 365 days of data, and taking 360 values along the full length of the spline. For non-smoothly varying fields, such as precipitation, the 365 data points are subsampled in order to produce 360 unique data points, effectively missing out the 72nd, 145th, 218th, 291st and 364th days. From these 7560 data points, the value at each percentile from 1st to 99th is calculated.

For the very large ensemble, the percentiles of the entire ensemble are first calculated. This is done by taking all the daily values from all of the 2500 ensemble members for all years from 1989 to 2009. This results in 1.89×10^7 points, as each of the 2500 members has 7560 data points, 360 for each year from 1989 to 2009, as in the ERA–Interim dataset. This is called the *pooled ensemble*. These percentiles from the entire ensemble are plotted against the percentiles from the ERA–Interim data as the thick line shown in Figures 6.2.1 to 6.2.5.

The bootstrapping procedure outlined in section 6.1.2 is then used to estimate the confidence intervals of the values of the percentiles. The ensemble is composed of a number (2500) of individual members for each year in the period 1989 to 2009. To calculate the uncertainty in the pooled ensemble, a timeseries of the climate variable over the same period is required. This is necessary as climate drivers and internal variability over the 21 year period will have an effect on the values at the percentiles. For example, calculating the percentile values for

the temperature at the beginning of the period will result in lower values than at the end of the period, due to the warming of the climate over the 21 years. Therefore, to produce a timeseries that is statistically comparable to the pooled ensemble requires a continuous series of ensemble members to be constructed across time.

To generate a timeseries of the climate variable over the 21 year period, the individual ensemble members for a single year are joined with other ensemble members for subsequent years. This is done in the order in which the members appear in the ensemble. For example, the 1st member in 1989 is joined to the 1st member in 1990 and the 1st member in 1991 (and so on). The 2nd member in 1989 is joined to the 2nd member in 1990 and so on. Doing this in the order in which they appear in the ensemble ensures a random joining of the ensemble members and ensures that each ensemble member is used. This procedure constructs 2500 unique ensemble members that span the time period 1989 to 2009, the same period as the pooled ensemble.

For each of these 2500 ensemble members the values at each percentile, from the 1st to the 99th, are calculated and stored. This gives 2500 values for the 1st percentile, 2500 values for the 2nd, etc. For each percentile the confidence interval is then determined by taking the 5th and 95th percentile of the 2500 values, the range between which forms the confidence interval. From these confidence intervals the envelopes seen in Figures 6.2.1 to 6.2.4 are plotted. A perfect fit between the ensemble data and the observed data would result in the thick line, generated from all ensemble members, lying directly on the 1:1 line, drawn in black on the plot.

6.1.5 Root mean square deviation of percentiles

Producing a Q-Q plot for each region, season and variable would require 100 different plots, making comparisons between the regions and seasons difficult. In order to quantify and tabulate the differences in how well the modelled distribution of the climate variable matches the observed distribution, the root mean square difference (RMSD) of the values at the 1st to 99th percentiles is calculated for each region, season and climate variable. In addition, the

RMSD for the low percentiles (1 to 10th) and high percentiles (90th to 99th) are calculated to determine how well the large ensemble of HadAM3P performs in the tails of the distributions of climate variables. The RMSD is defined as:

$$RMSD = \frac{1}{99}\sqrt{(\sum_{p=1}^{p=99}(o_p - e_p)^2)}$$

where o_p is the value at the pth percentile in the observations and e_p is the value at the same percentile in the ensemble. In Tables 6.2.1 to 6.2.5 this RMSD is calculated for the percentile values computed from the pooled ensemble, i.e. the solid line in Figures 6.2.1 to 6.2.5.

6.2 Daily data over five European Regions

6.2.1 Daily maximum temperature

Daily maximum temperature ($\mathsf{T}_{\mathsf{Max}}$) is an important variable for probabilistic event attribution studies that determine the contribution of anthropogenic greenhouse gas (GHG) emissions to temperature related events, such as heatwaves or unseasonably warm weather. Heatwave attribution studies (Otto et al. (2012), Stott et al. (2004), Cattiaux and Yiou (2012)) can use the variable directly in the study, determining whether an increase in the risk of a high value of $\mathsf{T}_{\mathsf{Max}}$ occurring can be attributed to anthropogenic GHG, or not. Alternatively, heatwave indices (Russo et al., 2014; Jacob et al., 2014; Frich et al., 2002) may be used, where $\mathsf{T}_{\mathsf{Max}}$ is used as one of the input variables to the heatwave index. These indices may also use additional variables such as relative humidity to calculate the apparent temperature (Fischer and Schar, 2010) or the maximum temperature may be required to pass a threshold for a number of days before a heatwave is detected.

This section applies the methods of sections 6.1 to 6.1.5 to $\mathsf{T}_{\mathsf{Max}}$ from the large ensemble of HadAM3P, over the five regions described in section 1.7.

Figure 6.2.1 shows the Q–Q plots for $\mathsf{T}_{\mathsf{Max}}$ over the Northern European (SREX NEU) re-

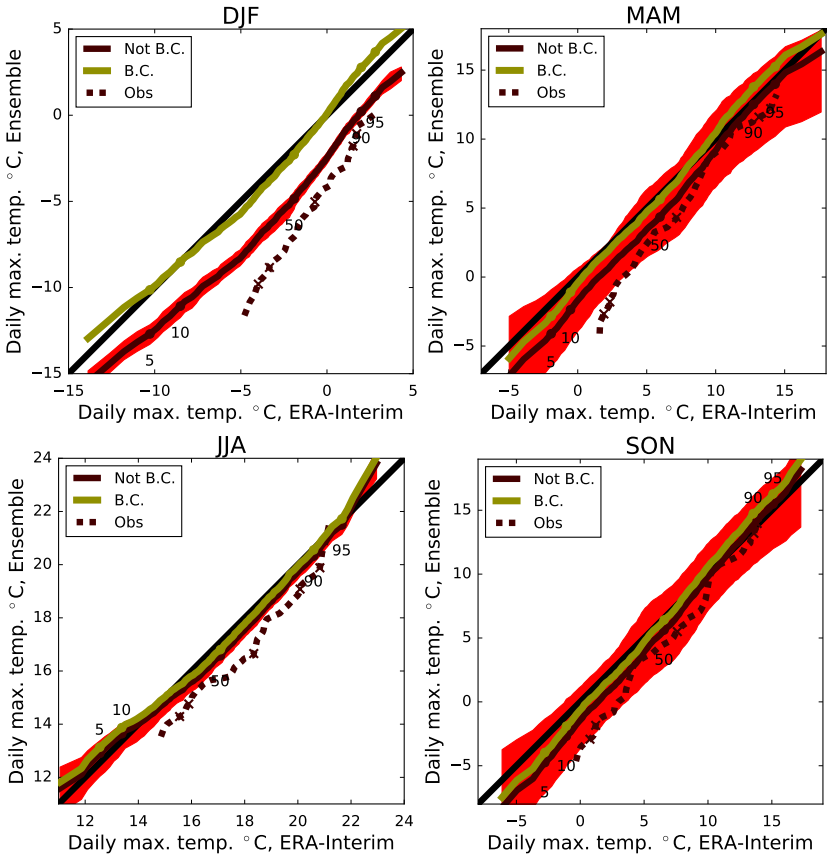


Figure 6.2.1: Quantile–Quantile plots of the daily maximum temperature, over the Northern European (SREX NEU) region, for each season. Quantiles from the large ensemble of HadAM3P for the pooled ensemble are plotted against the quantiles from ERA–Interim. The envelope shows the 5th to 95th confidence interval, estimated from the individual ensemble members. The solid red line shows the non–biased corrected pooled ensemble whereas the yellow line shows the bias corrected pooled ensemble. The dashed red line shows the monthly mean maximum daily temperature in the large ensemble against the same quantity in CRUTS.

gion, comparing the values at the percentiles in the large ensemble of HadAM3P to the corresponding percentile values in ERA–Interim. The DJF season demonstrates the value of applying the bias correction to the ensemble. Prior to the bias correction, the ensemble is biased by up to 2.5°C in the quantiles, whereas after the mean in the ensemble is corrected to match the ERA–Interim mean, the bias in the ensemble is much reduced, with the yellow line lying much more closely to the 1:1 line than the solid red line. The correction in other seasons is less dramatic, but still has a noticeable affect on how close the ensemble data lies

to the 1:1 line. The dashed red line shows the quantiles of the monthly mean of $\mathsf{T}_{\mathsf{max}}$ in the ensemble plotted against the quantiles for the same quantity in CRUTS (Mitchell and Jones, 2005), the observed dataset used in section 5.2. This shows that the under–estimation of $\mathsf{T}_{\mathsf{max}}$ in the large ensemble is consistent across two data sets: CRUTS and ERA–Interim, and two time resolutions: daily and monthly.

		all EU	SREX NEU	SREX CEU	SREX MED	UK
DJF	*all*	0.40 (0.19)	2.63 (0.49)	0.80 (0.32)	1.31 (0.28)	0.47 (0.41)
Max.	*1–10th*	0.85 (0.49)	2.41 (0.47)	1.49 (0.72)	0.82 (0.56)	1.41 (1.16)
temp.	*90th+*	0.49 (0.16)	1.84 (0.88)	0.95 (0.26)	1.88 (0.53)	*0.08 (0.28)*
	KS	0.06 (0.04)	0.26 (0.06)	0.07 (0.04)	0.24 (0.04)	0.05 (0.04)
	p	0.99 (1.00)	**0.00 (0.99)**	0.96 (1.00)	**0.00 (1.00)**	1.00 (1.00)
MAM	*all*	0.44 (0.35)	1.39 (0.56)	0.57 (0.21)	1.95 (0.32)	0.24 (0.15)
Max.	*1–10th*	*0.32* (0.51)	2.21 (0.87)	0.56 (0.11)	1.48 (0.54)	0.41 (0.22)
temp.	*90th+*	1.04 (0.75)	*0.60 (0.95)*	*0.20 (0.52)*	2.28 (0.30)	0.42 (0.24)
	KS	0.05 (0.04)	0.12 (0.05)	0.06 (0.03)	0.19 (0.04)	0.04 (0.03)
	p	1.00 (1.00)	0.44 (1.00)	0.99 (1.00)	**0.05 (1.00)**	1.00 (1.00)
JJA	*all*	2.95 (0.86)	0.38 (0.29)	4.14 (1.44)	3.68 (0.82)	1.03 (0.66)
Max.	*1–10th*	1.59 (1.42)	*0.36 (0.60)*	*1.86 (2.25)*	2.41 (1.40)	0.57 (0.28)
temp.	*90th+*	4.67 (1.70)	*0.38 (0.49)*	7.01 (2.92)	5.27 (1.48)	2.74 (1.92)
	KS	0.53 (0.13)	0.09 (0.05)	0.45 (0.13)	0.65 (0.15)	0.10 (0.08)
	p	**0.00 (0.34)**	0.79 (1.00)	**0.00 (0.34)**	**0.00 (0.19)**	0.67 (0.89)
SON	*all*	1.38 (1.03)	1.09 (0.76)	2.02 (1.54)	1.78 (0.67)	0.78 (0.78)
Max.	*1–10th*	*0.11 (1.01)*	2.07 (1.25)	*0.40 (1.58)*	0.97 (0.75)	1.17 (1.12)
temp.	*90th+*	3.20 (2.24)	*0.52 (1.33)*	4.76 (3.38)	3.39 (1.67)	*2.03 (2.07)*
	KS	0.14 (0.09)	0.08 (0.06)	0.14 (0.09)	0.16 (0.07)	0.06 (0.06)
	p	0.26 (0.79)	0.89 (0.99)	0.26 (0.79)	0.14 (0.96)	0.99 (0.99)

Table 6.2.1: Root mean square deviations (RMSD) for all percentiles from 1 to 99 for the daily maximum temperature over the five regions (*all*), the 1st to 10th (*1–10th*) and the 90th to 99th percentiles (*90th+*). The deviations are from the values at the corresponding percentiles in ERA–Interim. Also shown are the results of the two-way Kolmogorov–Smirnoff test, showing the KS test statistic and the *p–value* for the pooled ensemble and ERA–Interim. Values in brackets are post bias correction. Bold values indicate whether applying the bias correction has increased the *p–value* of the KS test from below 0.05 to above 0.05. Italic values indicate that applying the bias correction has increased the error between the HadAM3P ensemble and ERA–Interim.

Table 6.2.1 quantifies the bias in the percentiles, by calculating the RMSD of the percentiles for each region and season, and evaluates the performance of the bias correction by calculating the RMSD after the correction has taken place. The two sample KS test is used to determine

how well the distribution of $\mathsf{T}_{\mathsf{max}}$ is represented in the ensemble, both before and after bias correction. For the KS test, the null hypothesis is that the distribution of $\mathsf{T}_{\mathsf{max}}$ in the ensemble and in ERA–Interim are the same, and the null hypothesis is rejected if the *p–value* is less than 0.05 (Wilks, 2011).

As can be seen in the Table, before bias correction is applied, there are 6 season–region combinations where the null hypothesis is rejected, i.e. the distribution of $\mathsf{T}_{\mathsf{max}}$ in the ensemble is distinct from the distribution of $\mathsf{T}_{\mathsf{max}}$ in ERA–Interim. These are highlighted in bold in the Table. After bias correction has been applied, all 20 season–region combinations have the same distribution of $\mathsf{T}_{\mathsf{max}}$ as ERA–Interim, as there are no *p–values* that are less than 0.05. In addition, the RMSD of the percentiles has decreased for *all* percentiles for all region–season combinations. However, for the low tails (*1–10th*) of $\mathsf{T}_{\mathsf{max}}$, applying the bias correction has caused the RMSD of the percentiles to increase in 5 season–region combinations and, for the high tails (*90th+*) to increase in 6 season–region combinations. An increase in RMSD which has occurred due to the bias correction is highlighted in italics.

6.2.2 Daily minimum temperature

Daily minimum temperature ($\mathsf{T}_{\mathsf{min}}$) is used in two different types of event attribution study. The first are investigations into whether anthropogenic greenhouse gas (GHG) emissions have changed the risk of occurrence of a certain circulation pattern which causes cold weather, such as a persistent high pressure system in winter (de Vries et al., 2013). Here it is necessary to look at changes in circulation patterns as, from a purely thermodynamic viewpoint, the risk of cold weather occurring will decrease under a warming climate. However, bringing atmospheric dynamics into the assessment, the warming due to an increase in GHG could alter circulation patterns and increase the frequency of cold air over particular regions. Here circulation patterns mean relatively stable states of the atmosphere that persist for days to weeks, such as a blocking high. This requirement to assess both dynamics and thermodynamics is why using a global climate model (GCM) is necessary.

The second type of attribution study requiring $\mathsf{T}_{\mathsf{min}}$ are heat waves. The minimum night

time temperature is one index which can define a heat wave, especially if public health concerns such as increased mortality are to be assessed. For example, D'Ippoliti et al. (2010) use both apparent temperature (taking into account relative humidity as well as maximum temperature) and minimum temperature to assess the effect on public health of heat waves across nine European cities.

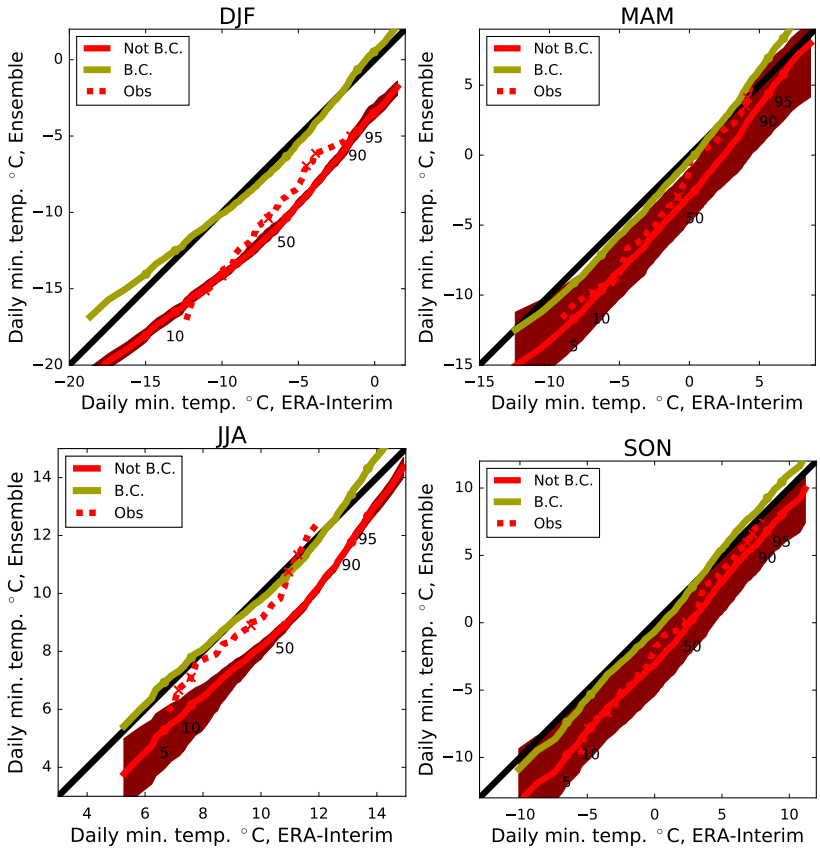


Figure 6.2.2: Quantile–Quantile plots of the daily minimum temperature, over the Northern European (SREX NEU) region, for each season. Quantiles from the large ensemble of HadAM3P for the pooled ensemble are plotted against the quantiles from ERA–Interim. The envelope shows the 5th to 95th confidence interval, estimated from the individual ensemble members. The solid red line shows the non–biased corrected pooled ensemble whereas the yellow line shows the bias corrected pooled ensemble. The dashed red line shows the monthly mean maximum daily temperature in the large ensemble against the same quantity in CRUTS.

Figure 6.2.2 shows the Q–Q plots for $\mathsf{T}_{\mathsf{min}}$ over the Northern European (SREX NEU) region, comparing the percentile values in the large ensemble of HadAM3P to those in ERA–Interim.

In the same manner as $\mathsf{T}_{\mathsf{max}}$, this Figure demonstrates the benefit of using a bias correction on $\mathsf{T}_{\mathsf{min}}$ in the SREX NEU region. Prior to bias correction, and shown by the thick red line, the large ensemble is consistently under–estimating $\mathsf{T}_{\mathsf{min}}$ for all seasons. After bias correction, shown by the yellow line, there is much more agreement between the large ensemble and ERA–Interim. This is confirmed by the *p–value* for the two sample KS test shown in Table 6.2.2. For SREX NEU, all seasons have a *p–value* of below 0.05, which are highlighted in bold in the Table. This indicates that the null hypothesis should be rejected and the two distributions of $\mathsf{T}_{\mathsf{min}}$ are distinct. After applying the bias correction, all *p–values* are now 0.99 or above, i.e. the distributions are the same. However, for MAM the RMSD has increased for the high percentiles (*90th+*), indicating that there is now more bias in the high percentiles of $\mathsf{T}_{\mathsf{min}}$.

Table 6.2.2 quantifies the bias in the percentiles of $\mathsf{T}_{\mathsf{min}}$ for all season and region combinations, in the same manner as Table 6.2.1 does for $\mathsf{T}_{\mathsf{max}}$. Before bias correction is applied, there are 15 out of the 20 season–region combinations where the *p–value* of the two sample KS test is less than 0.05, indicating that the two distributions are distinct. After bias correction there are no season–region combinations where this is still true. This shows that, as with $\mathsf{T}_{\mathsf{max}}$, correcting for the mean bias in the ensemble results in $\mathsf{T}_{\mathsf{min}}$ much more closely matching $\mathsf{T}_{\mathsf{min}}$ in ERA–Interim.

Applying the bias correction has also reduced the RMSD for *all* percentiles for each season–region combination. It has increased the RMSD in high percentiles (*90th+*) in 7 of the season–region combinations and increased the RMSD in low percentiles for just one season–region combination. This will be further commented upon in section 6.3.

6.2.3 Daily accumulated precipitation

Daily accumulated precipitation ($\mathsf{P}_{\mathsf{acc}}$) is the primary climate variable used in event attribution studies into high or persistent precipitation events, such as Tett et al. (2013), Sparrow et al. (2013) and Schaller et al. (2014). Other studies may attribute the contribution of anthropogenic GHG emissions to the impacts of high precipitation events, such as flooding

		all EU	SREX NEU	SREX CEU	SREX MED	UK
DJF	*all*	3.18 (0.15)	4.08 (0.54)	2.75 (0.23)	3.04 (0.52)	2.92 (1.02)
Min.	*1–10th*	3.50 (0.17)	3.12 (1.21)	2.87 (0.18)	3.82 (0.71)	4.82 (1.96)
temp.	*90th+*	3.14 (0.23)	3.65 (0.61)	2.33 (0.56)	1.96 (1.29)	*1.23 (1.67)*
	KS	0.49 (0.03)	0.36 (0.06)	0.27 (0.03)	0.55 (0.07)	0.32 (0.09)
	p	**0.00 (1.00)**	**0.00 (0.99)**	**0.00 (1.00)**	**0.00 (0.96)**	**0.00 (0.79)**
MAM	*all*	2.56 (0.49)	2.70 (0.79)	2.52 (0.57)	2.39 (0.31)	2.42 (0.68)
Min.	*1–10th*	3.34 (0.74)	3.51 (0.85)	3.46 (0.89)	3.20 (0.73)	4.07 (1.67)
temp.	*90th+*	1.73 (0.92)	*0.96 (1.78)*	1.45 (1.17)	2.28 (0.24)	1.68 (0.78)
	KS	0.24 (0.06)	0.22 (0.07)	0.20 (0.05)	0.23 (0.04)	0.24 (0.06)
	p	**0.00 (0.99)**	**0.01 (0.96)**	**0.03 (1.00)**	**0.01 (1.00)**	**0.00 (0.99)**
JJA	*all*	0.83 (0.53)	1.68 (0.27)	1.05 (0.92)	1.11 (0.54)	1.80 (0.32)
Min.	*1–10th*	1.51 (0.86)	1.53 (0.21)	*0.89 (1.42)*	2.06 (1.04)	2.23 (0.38)
temp.	*90th+*	*0.42 (1.05)*	1.09 (0.72)	2.51 (1.98)	*0.13 (0.95)*	1.15 (0.83)
	KS	0.18 (0.11)	0.34 (0.06)	0.17 (0.11)	0.23 (0.14)	0.33 (0.04)
	p	0.07 (0.55)	**0.00 (0.99)**	0.10 (0.55)	**0.01 (0.26)**	**0.00 (1.00)**
SON	*all*	2.11 (0.72)	2.52 (0.63)	1.74 (1.08)	2.19 (0.48)	2.67 (0.90)
Min.	*1–10th*	2.91 (0.83)	3.45 (0.90)	2.69 (1.34)	2.89 (0.65)	4.31 (1.68)
temp.	*90th+*	*0.61 (1.52)*	1.53 (1.06)	*0.96 (2.34)*	1.19 (1.08)	*1.06 (1.64)*
	KS	0.17 (0.08)	0.19 (0.05)	0.14 (0.09)	0.18 (0.07)	0.25 (0.09)
	p	0.10 (0.89)	**0.05 (1.00)**	0.26 (0.79)	0.07 (0.96)	**0.00 (0.79)**

Table 6.2.2: Root mean square deviations (RMSD) for all percentiles from 1 to 99 for the daily minimum temperature over the five regions for (*all*), the 1st to 10th (*1–10th*) and the 90th to 99th percentiles (*90th+*). The deviations are from the values at the corresponding percentiles in ERA–Interim. Also shown are the results of the two-way Kolmogorov–Smirnoff test, showing the KS test statistic and the *p–value* for the pooled ensemble and ERA–Interim. Values in brackets are post bias correction. Bold values indicate that the bias correction has improved the distribution of HadAM3P minimum temperature to the degree that the KS test cannot distinguish between the two distributions. Italic values indicate that applying the bias correction has increased the RMSD.

(Schaller et al. (2016), Pall et al. (2011)). As with $\mathsf{T}_{\mathsf{min}}$, it is important to examine changes in atmospheric circulation, as the mean increase in available precipitation due to a warming climate is known to be proportional to Clausius–Clapeyron (Allen and Ingram, 2002). However, it is not known where this extra available precipitation will fall or whether it will lead to an increase in the frequency of low–precipitation / high frequency events or an increase in the magnitude of high–precipitation / low frequency events. Huntingford et al. (2014) explore the potential causes of the high–precipitation and subsequent flooding events in the winter of 2013–2014 over the UK. On discussing whether the anthropogenic contribution to

GHG emissions could increase the risk of such events occurring, they concluded that any increase would be the resulting effect of anthropogenic GHG on a large number of connected processes in the climate system, which reiterates the importance of using GCMs to study the effect on the atmospheric circulation.

O'Gorman (2015) reports, from observations and modelling studies, that heavy precipitation events become more frequent in response to elevated GHG concentrations, and the warming associated with the increase. They also report that the extreme values of convective precipitation increase more rapidly than the relationship implied by Clausius-Clapeyron. Bengtsson et al. (2009) conducted a modelling study, using a high–resolution global climate model, to study the change in extratropical cyclones in a warmer climate. Two 32 year periods, one at the end of the 20th century and one at the end of the 21st century, were integrated under an IPCC A1B scenario. They found that cumulative precipitation along a storm track increased by 11% per track, which is approximately twice the increase found globally by Allen and Ingram (2002). Additionally, Bengtsson et al. (2009) report that the extreme precipitation increases by 27% and that regional changes in extreme precipitation can be even higher due to the change in position of the storm tracks. These two studies show that, even though global precipitation is predicted to increase by 6 to 7%, as predicted by Clausius–Clapeyron (Allen and Ingram, 2002), localised and extreme precipitation events are predicted to increase by a much greater amount, which has implications for both attribution studies and impact assessments.

Figure 6.2.3 shows the Q–Q plots for $\mathsf{P}_{\mathsf{acc}}$ over the Mediterranean (SREX MED) region, comparing the percentiles in the large ensemble to those in ERA–Interim. This Figure demonstrates how using a bias correction on $\mathsf{P}_{\mathsf{acc}}$ can enable the distribution in the large ensemble to more closely represent the distribution in ERA–Interim by just correcting the mean of the distribution. Before bias correction, shown by the thick black line, the large ensemble is consistently over–estimating $\mathsf{P}_{\mathsf{acc}}$ for the JJA season. For the other seasons (DJF, MAM, SON) the large ensemble closely resembles $\mathsf{P}_{\mathsf{acc}}$ in ERA–Interim. As shown by the purple line, after bias correction there is much more agreement between the large ensemble and ERA–Interim for $\mathsf{P}_{\mathsf{acc}}$ in JJA. This is confirmed by the *p–value* for the two sample KS

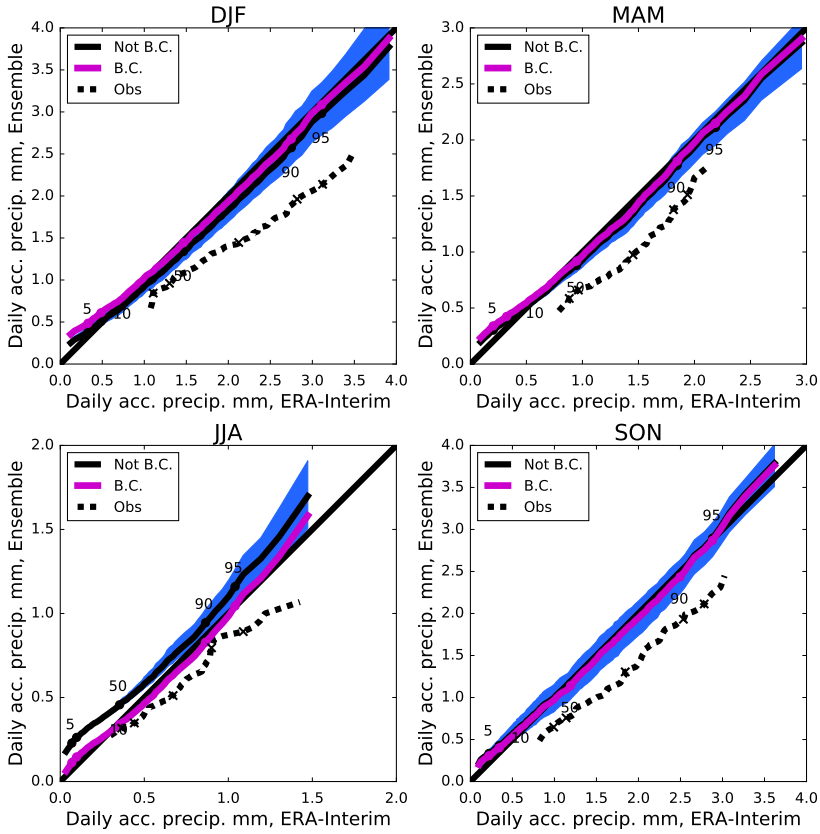


Figure 6.2.3: Quantile–Quantile plots of the daily accumulated precipitation, over the Mediterranean (SREX MED) region, for each season. Quantiles from the large ensemble of HadAM3P for the pooled ensemble are plotted against the quantiles from ERA–Interim. The envelope shows the 5th to 95th confidence interval, estimated from the individual ensemble members. The solid black line shows the non–biased corrected pooled ensemble whereas the magenta line shows the bias corrected pooled ensemble. The dashed black line shows the monthly mean maximum daily temperature in the large ensemble against the same quantity in GPCP.

test shown in Table 6.2.3. For SREX MED, the JJA all season has a *p–value* of below 0.05, indicating that the distribution of $\mathsf{P}_{\mathsf{acc}}$ is distinct between ERA–Interim and the large ensemble After the bias correction, the *p–values* is 0.79, indicating that the distributions are now indistinguishable, at the 95% confidence limit.

Table 6.2.3 quantifies the errors in the distribution of $\mathsf{P}_{\mathsf{acc}}$ in the large ensemble, when compared to ERA–Interim. Unlike $\mathsf{T}_{\mathsf{max}}$ and $\mathsf{T}_{\mathsf{min}}$, there is only one season (JJA) which have *p–values* from the two sample KS test less than 0.05, and this is for only three of the five

		all EU	SREX NEU	SREX CEU	SREX MED	UK
DJF	*all*	0.03 (0.02)	0.15 (0.15)	0.29 (0.07)	0.13 (0.07)	0.24 (0.14)
Precip.	*1–10th*	0.02 (0.01)	0.24 (0.22)	0.21 (0.10)	*0.08 (0.18)*	0.18 (0.06)
	90th+	*0.02 (0.03)*	*0.22 (0.25)*	0.18 (0.13)	0.16 (0.05)	0.44 (0.23)
	KS	0.04 (0.02)	0.08 (0.07)	0.14 (0.04)	0.07 (0.06)	0.08 (0.04)
	p	1.00 (1.00)	0.89 (0.96)	0.26 (1.00)	0.96 (0.99)	0.89 (1.00)
MAM	*all*	0.02 (0.02)	0.05 (0.05)	0.09 (0.08)	0.07 (0.06)	0.16 (0.16)
Precip.	*1–10th*	0.01 (0.01)	0.12 (0.11)	0.09 (0.05)	*0.10 (0.14)*	*0.05 (0.09)*
	90th+	0.04 (0.04)	*0.04 (0.05)*	*0.14 (0.18)*	0.07 (0.04)	*0.38 (0.42)*
	KS	0.03 (0.02)	0.04 (0.04)	0.05 (0.04)	0.06 (0.07)	0.05 (0.05)
	p	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	0.99 (0.96)	1.00 (1.00)
JJA	*all*	0.21 (0.01)	0.16 (0.08)	0.92 (0.15)	0.12 (0.03)	0.51 (0.27)
Precip.	*1–10th*	0.19 (0.03)	*0.05 (0.19)*	0.58 (0.39)	0.17 (0.05)	*0.01 (0.45)*
	90th+	0.23 (0.02)	0.19 (0.05)	1.08 (0.14)	0.14 (0.05)	0.59 (0.22)
	KS	0.25 (0.02)	0.09 (0.05)	0.35 (0.09)	0.27 (0.09)	0.14 (0.18)
	p	**0.00 (1.00)**	0.79 (1.00)	**0.00 (0.79)**	**0.00 (0.79)**	0.26 (0.07)
SON	*all*	0.09 (0.04)	0.08 (0.07)	0.32 (0.08)	0.06 (0.05)	0.41 (0.28)
Precip.	*1–10th*	0.15 (0.07)	*0.03 (0.04)*	*0.14 (0.19)*	0.12 (0.09)	*0.02 (0.33)*
	90th+	*0.03 (0.08)*	0.18 (0.13)	0.40 (0.08)	0.09 (0.08)	*0.31 (0.56)*
	KS	0.09 (0.04)	0.04 (0.03)	0.12 (0.06)	0.05 (0.04)	0.09 (0.11)
	p	0.79 (1.00)	1.00 (1.00)	0.44 (0.99)	1.00 (1.00)	0.79 (0.55)

Table 6.2.3: Root mean square deviations (RMSD) for all percentiles from 1 to 99 for the daily accumulated precipitation over the five regions for (*all*), the 1st to 10th (*1–10th*) and the 90th to 99th percentiles (*90th+*). The deviations are from the values at the corresponding percentiles in ERA–Interim. Also shown are the results of the two-way Kolmogorov–Smirnoff test, showing the KS test statistic and the *p–value* for the pooled ensemble and ERA–Interim. Values in brackets are post bias correction. Bold values indicate that the bias correction has improved the distribution of HadAM3P minimum temperature to the degree that the KS test cannot distinguish between the two distributions. Italic values indicate that applying the bias correction has increased the RMSD.

regions. Applying the bias correction ensures that all region–season combinations have $\mathsf{P}_{\mathsf{acc}}$ distributions that are indistinguishable from ERA–Interim, in that their *p–values* are greater than 0.05. Applying the bias correction increases the RMSD in the low–percentiles (*1–10th*) for 8 out of the 20 season–region combinations and 7 out of 20 in the high–percentiles (*90th+*). Although some of these increases in RMSD are very small (0.02 to 0.03, for example), others are not (0.02 to 0.33). In fact, for the UK region, applying the bias correction to the JJA and SON seasons increases the KS test statistic and RMSD for one percentile range or more and decreases the *p–value*, indicating that the distribution after the bias correction is less similar

to ERA–Interim than before it was applied. Therefore, it may be prudent to only apply bias correction to season–region combinations where it has a positive impact on improving the distribution of the climate variable.

6.2.4 Daily mean sea level pressure

Although not a climate variable which in itself is subject to extreme event attribution, mean sea level pressure (MSLP) over Europe is associated with two types of events. The first, where there is a persistent high pressure system, is the blocking of cyclones exiting the Polar Jet. In summer this can lead to a heat wave forming over the area of the persistent block, as seen in the Moscow heat wave of 2010 (Otto et al., 2012). In winter the blocking can lead to a cold, dry snap, as warmer air from the Atlantic cannot advect to the area of the block (de Vries et al., 2013).

The second type of extreme event is associated with low MSLP systems advecting warm, moist air from the Atlantic, steered via the Polar Jetstream towards Europe. These systems are associated with three types of extreme weather event, which may occur separately but usually occur in concert with each other as an extra–tropical storm. The potentially sharp pressure gradient of a deep low pressure system produces high windspeeds, the advection of moist air from a warm climate to a cooler one causes high levels of precipitation and the low MSLP causes storm surges. Attribution studies of extra–tropical storms may use the MSLP as a proxy for the storm and its associated extreme events (von Storch et al., 2014; Feser et al., 2015).

Figure 6.2.4 shows the Q–Q plots for MSLP over the Mediterranean (SREX MED) region, comparing the percentile values in the large ensemble of HadAM3P to the corresponding percentiles in ERA–Interim. The DJF and MAM seasons show good agreement with ERA–Interim, whereas the JJA season shows the largest bias, with SON showing a smaller bias. Applying the bias correction improves the correspondence between the percentiles in the large ensemble and in ERA–Interim, especially for the JJA season and, to a lesser degree, the SON season. This is confirmed by Table 6.2.4, where the RMSD in *all* percentiles for JJA

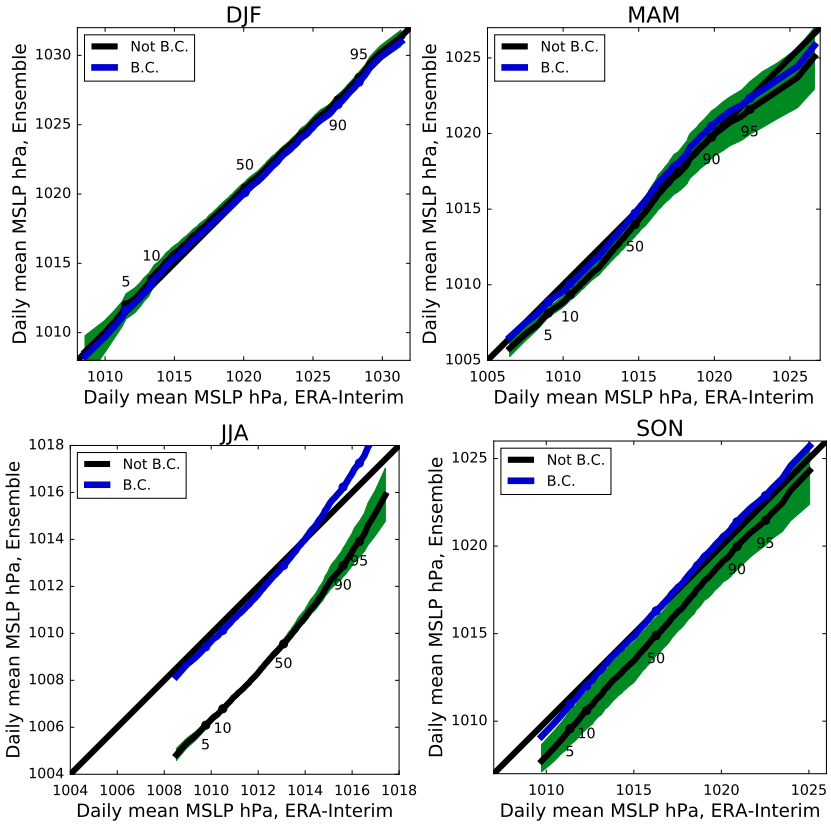


Figure 6.2.4: Quantile–Quantile plots of the daily mean sea level pressure (MSLP), over the Mediterranean (SREX MED) region, for each season. Quantiles from the large ensemble of HadAM3P for the pooled ensemble are plotted against the quantiles from ERA–Interim. The envelope shows the 5th to 95th confidence interval, estimated from the individual ensemble members. The solid black line shows the non–biased corrected pooled ensemble whereas the blue line shows the bias corrected pooled ensemble.

improves from 3.39hPa to 0.44hPa. Adjusting the mean of the MSLP in the large ensemble also makes the distribution indistinguishable from ERA–Interim MSLP, which is confirmed by the *p–value* of the two sample KS test changing from 0.0 before bias correction to 0.99 afterwards.

Table 6.2.4 quantifies the bias in the percentiles of MSLP for all season–region combinations, when compared to ERA–Interim. Before bias correcting, there are just 3 out of the 20 season–region combinations where the *p–value* of the two KS test is less than 0.05. This indicates that MSLP in the large ensemble is a good fit to the MSLP in ERA–Interim, at least over

		all EU	SREX NEU	SREX CEU	SREX MED	UK
DJF	*all*	0.81 (0.64)	2.31 (1.73)	1.20 (1.05)	0.39 (0.21)	2.20 (1.72)
MSLP	*1–10th*	1.65 (1.10)	4.16 (2.56)	*1.90 (2.45)*	0.34 (0.18)	*2.55 (3.92)*
	90th+	*0.40 (0.89)*	*1.09 (2.69)*	1.64 (1.07)	*0.18 (0.29)*	2.24 (0.92)
	KS	0.07 (0.05)	0.10 (0.06)	0.07 (0.05)	0.04 (0.02)	0.07 (0.05)
	p	0.96 (1.00)	0.67 (0.99)	0.96 (1.00)	1.00 (1.00)	0.96 (1.00)
MAM	*all*	0.68 (0.40)	2.62 (0.69)	0.56 (0.49)	0.84 (0.40)	0.40 (0.20)
MSLP	*1–10th*	1.58 (1.01)	4.07 (1.45)	1.50 (1.21)	1.04 (0.30)	0.42 (0.33)
	90th+	0.36 (0.35)	1.86 (0.83)	*0.75 (0.96)*	1.10 (0.63)	*0.25 (0.36)*
	KS	0.06 (0.04)	0.13 (0.03)	0.03 (0.02)	0.12 (0.05)	0.02 (0.02)
	p	0.99 (1.00)	0.34 (1.00)	1.00 (1.00)	0.44 (1.00)	1.00 (1.00)
JJA	*all*	1.87 (0.16)	0.37 (0.21)	1.98 (0.20)	3.39 (0.44)	0.52 (0.22)
MSLP	*1–10th*	1.91 (0.06)	*0.19 (0.42)*	1.94 (0.22)	3.89 (0.37)	0.69 (0.39)
	90th+	1.63 (0.34)	0.54 (0.24)	1.67 (0.56)	2.44 (1.19)	0.33 (0.31)
	KS	0.31 (0.04)	0.05 (0.03)	0.24 (0.03)	0.58 (0.06)	0.04 (0.02)
	p	**0.00 (1.00)**	1.00 (1.00)	**0.00 (1.00)**	**0.00 (0.99)**	1.00 (1.00)
SON	*all*	0.28 (0.15)	2.39 (0.27)	0.44 (0.34)	1.39 (0.28)	2.15 (0.98)
MSLP	*1–10th*	0.25 (0.08)	2.56 (0.29)	0.89 (0.67)	1.93 (0.46)	3.22 (1.29)
	90th+	0.51 (0.34)	1.99 (0.54)	*0.63 (0.79)*	0.98 (0.53)	*0.40 (1.67)*
	KS	0.03 (0.02)	0.12 (0.02)	0.04 (0.02)	0.17 (0.04)	0.08 (0.04)
	p	1.00 (1.00)	0.44 (1.00)	1.00 (1.00)	0.10 (1.00)	0.89 (1.00)

Table 6.2.4: Root mean square deviations (RMSD) for all percentiles from 1 to 99 for the daily MSLP over the five regions for (*all*), the 1st to 10th (*1–10th*) and the 90th to 99th percentiles (*90th+*). The deviations are from the values at the corresponding percentiles in ERA–Interim. Also shown are the results of the two-way Kolmogorov–Smirnoff test, showing the KS test statistic and the *p–value* for the pooled ensemble and ERA–Interim. Values in brackets are post bias correction. Bold values indicate that the bias correction has improved the distribution of HadAM3P minimum temperature to the degree that the KS test cannot distinguish between the two distributions. Italic values indicate that applying the bias correction has increased the RMSD.

the European domains analysed here. Applying the bias correction results in no season–region combinations having a *p–value* less than 0.05. However, there are 3 season–region combinations in the low percentile range (*1–10th*) and 7 in the high percentile range (*90th+*) where applying the bias correction increases the RMSD, even though the *p–value* of the KS test is above 0.05 and bias correction is not really necessary. Therefore, it may be prudent, as with precipitation to only apply bias correction to season–region combinations that require it.

6.2.5 Daily maximum windspeed

Although there are, at the time of writing, currently no attribution studies which attempt to attribute an increase or decrease in windspeed to the anthropogenic increase in GHG, there is scope to do such a study in the future using a large ensemble of the HadAM3P model run via volunteer distributed computing. Additionally, the daily maximum windspeed ($\mathsf{U}_{\mathsf{max}}$) is a good measure of how the boundary layer processes are represented in the model.

It should be noted that the quantity validated in this section is the daily maximum windspeed. That is, for the HadAM3P model, the maximum windspeed encountered within a grid box over a 24 hour period. As HadAM3P has a 15 minute timestep, this equates to the maximum 15 minute average windspeed within 24 hours. In ERA–Interim the 10m windspeed is not available as either an analysis or forecast variable. Instead, the 10m U and V components of the wind are available at 6 hourly intervals as analysis variables. For this thesis, the U and V components of the wind are used to calculate the windspeed at every 6 hour interval and the maximum of these 4 values in a 24 hour period is taken as ($\mathsf{U}_{\mathsf{max}}$). Although these two quantities are not identical, they are close enough to be comparable. Furthermore, the 3 second gust wind speed is considered to be the most damaging to buildings and infrastructure (Haylock, 2011). An attribution study would, by preference, use the 3 second gust as the variable, so as to measure the impacts of any change in risk of the wind storm event occurring. However, this diagnostic is not available for either the large ensemble of the HadAM3P model, nor for ERA–Interim. Therefore, $\mathsf{U}_{\mathsf{max}}$ is used as a next best diagnostic variable.

Figure 6.2.5 shows the Q–Q plots for $\mathsf{U}_{\mathsf{max}}$ over the UK (SREX UK) region, comparing the percentile values in the large ensemble of HadAM3P to those in ERA–Interim. The large ensemble of HadAM3P is consistently overestimating $\mathsf{U}_{\mathsf{max}}$ for all seasons, with DJF having the largest error and MAM having the smallest error. After bias correction, the percentiles of $\mathsf{U}_{\mathsf{max}}$ match very closely between the large ensemble and ERA–Interim.

Table 6.2.5 quantifies the bias in the percentiles of $\mathsf{U}_{\mathsf{max}}$ for all season–region combinations,

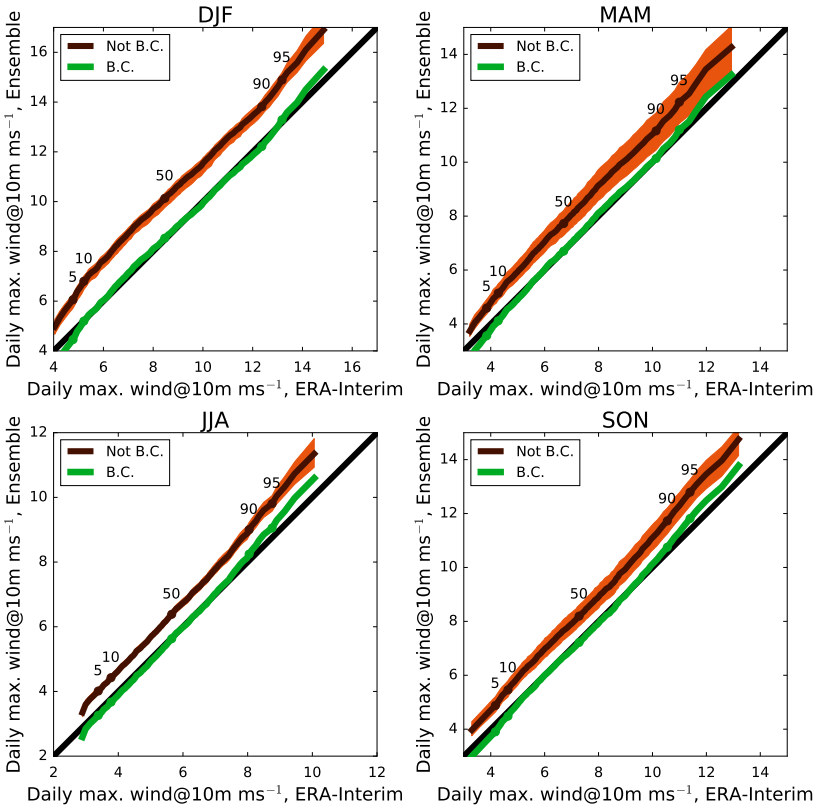


Figure 6.2.5: Quantile–Quantile plots of the daily maximum windspeed, over the United Kingdom (SREX UK) region, for each season. Quantiles from the large ensemble of HadAM3P for the pooled ensemble are plotted against the quantiles from ERA–Interim. The envelope shows the 5th to 95th confidence interval, estimated from the individual ensemble members. The solid brown line shows the non–biased corrected pooled ensemble whereas the green line shows the bias corrected pooled ensemble.

when compared to ERA–Interim and confirms that bias correction produces a distribution of $\mathsf{U}_{\mathsf{max}}$ that more closely matches ERA–Interim. Prior to bias correction, there are only 2 season–region combinations where the *p–value* of the two sample KS test is greater than 0.05. Even for the 2 season–region combinations, the *p–value* is low, closer to 0.0 than 1.0. However, after bias correction, the majority of season–region combinations have very high *p–values* ($p \geq 0.89$) with only MAM for SREX MED having a relatively low *p–value* of 0.34. Furthermore, there are no season–region combinations where the RMSD of the percentiles is greater after bias correction than before bias correction, even for the low and high percentiles. This shows that the offset for $\mathsf{U}_{\mathsf{max}}$ between the large ensemble and ERA–Interim is linear

		all EU	SREX NEU	SREX CEU	SREX MED	UK
DJF	*all*	1.43 (0.13)	1.46 (0.24)	1.46 (0.21)	1.37 (0.17)	1.62 (0.15)
Wind	*1–10th*	1.25 (0.24)	1.14 (0.38)	1.14 (0.40)	1.13 (0.31)	1.40 (0.37)
speed	*90th+*	1.74 (0.25)	1.93 (0.41)	1.90 (0.37)	1.69 (0.26)	1.84 (0.24)
@10m	KS	0.71 (0.06)	0.38 (0.07)	0.51 (0.07)	0.52 (0.06)	0.24 (0.03)
	p	**0.00 (0.99)**	**0.00 (0.96)**	**0.00 (0.96)**	**0.00 (0.99)**	**0.00 (1.00)**
JJA	*all*	1.61 (0.10)	1.13 (0.22)	1.73 (0.19)	1.89 (0.16)	1.04 (0.13)
Wind	*1–10th*	1.47 (0.22)	0.78 (0.40)	1.43 (0.38)	1.64 (0.34)	0.79 (0.32)
speed	*90th+*	1.87 (0.18)	1.59 (0.42)	2.10 (0.29)	2.21 (0.23)	1.29 (0.24)
@10m	KS	0.75 (0.05)	0.33 (0.06)	0.69 (0.08)	0.79 (0.08)	0.19 (0.03)
	p	**0.00 (1.00)**	**0.00 (0.99)**	**0.00 (0.89)**	**0.00 (0.89)**	**0.05 (1.00)**
MAM	*all*	1.58 (0.11)	1.09 (0.22)	2.02 (0.18)	1.70 (0.18)	0.77 (0.14)
Wind	*1–10th*	1.44 (0.21)	0.76 (0.38)	1.76 (0.36)	1.50 (0.28)	0.65 (0.16)
speed	*90th+*	1.86 (0.20)	1.60 (0.47)	2.43 (0.31)	2.15 (0.37)	1.20 (0.41)
@10m	KS	0.94 (0.08)	0.43 (0.08)	0.88 (0.08)	0.95 (0.13)	0.17 (0.03)
	p	**0.00 (0.89)**	**0.00 (0.89)**	**0.00 (0.89)**	**0.00 (0.34)**	0.10 (1.00)
SON	*all*	1.41 (0.08)	1.34 (0.19)	1.47 (0.18)	1.42 (0.11)	0.99 (0.16)
Wind	*1–10th*	1.32 (0.16)	1.04 (0.36)	1.19 (0.34)	1.28 (0.21)	0.79 (0.25)
speed	*90th+*	1.64 (0.17)	1.77 (0.39)	1.84 (0.31)	1.64 (0.17)	1.46 (0.44)
@10m	KS	0.74 (0.04)	0.38 (0.05)	0.60 (0.07)	0.72 (0.06)	0.16 (0.03)
	p	**0.00 (1.00)**	**0.00 (1.00)**	**0.00 (0.96)**	**0.00 (0.99)**	0.14 (1.00)

Table 6.2.5: Root mean square deviations (RMSD) for all percentiles from 1 to 99 for the daily maximum windspeed over the five regions for all percentiles (*all*), the 1st to 10th (*1–10th*) and for 90th to 99th percentiles (*90th+*). The deviations are from the values at the corresponding percentiles in ERA–Interim. Also shown are the results of the two-way Kolmogorov–Smirnoff test, showing the KS test statistic and the *p–value* for the pooled ensemble and ERA–Interim. Values in brackets are post bias correction. Bold values indicate that the bias correction has improved the distribution of HadAM3P minimum temperature to the degree that the KS test cannot distinguish between the two distributions. Italic values indicate that applying the bias correction has increased the RMSD.

and can be successfully corrected by adding a single offset value.

6.3 Summary and conclusions

This chapter has presented a single forecast verification technique (Q–Q plots, Jolliffe and Stephenson (2003)), over five regions in Europe, which can be used to test the validity of a very large ensemble of a climate model computed using volunteer distributed computing.

By applying this technique prior to a probabilistic event attribution study, it is possible to understand the biases in the model, when used in a large ensemble, and to determine whether these biases can be corrected. These techniques are applied to the very large ensemble of the HadAM3P climate model computed for the period between 1989 and 2009.

Section 6.2.1 shows that, even though the distribution of near–surface daily maximum temperature over the UK and Ireland of the ensemble may not initially match the distribution from the reference dataset (in this chapter ERA–Interim is used), after applying a bias correction, which ensures that the ensemble distribution and reference distribution have the same mean, there is good correspondence between the ensemble and reference datasets. Section 6.2.2 shows that the same is true for the daily minimum temperature. Section 6.2.3 shows that the daily accumulated precipitation is already well represented in the model, with only 3 out of 20 season–region combinations requiring bias correction. Similarly for daily mean MSLP (section 6.2.4), only 3 season–region combinations require bias correction, the same 3 as require it for precipitation. Finally, for daily maximum windspeed nearly all season–region combinations require bias correction as the model consistently represents windspeed at too high a value for each percentile, when compared to ERA–Interim. This is remedied by correcting the mean of the windspeed to the mean of ERA–Interim windspeed.

Applying the bias correction, by correcting the mean of the large ensemble to the mean of the reference dataset (ERA–Interim) reduces the bias in all variables for all region–season combinations to the point that the *p–value* in all variables and region–season combinations is increased to be above the threshold where the null hypothesis of the two sample KS test is rejected. In many cases, the *p–value* is increased to ≥ 0.9. For some variables this has the unwanted consequence of increasing the RMSD for either the low (*1–10th*) percentiles or the high (*90th+*) percentiles. However, due to the low sample size of the reference dataset, these high and low percentiles may not be well defined, whereas for the large ensemble of HadAM3P, the very large sample size ensures that the true values of these extreme percentiles are known. A different method of bias correction, such as percentile matching, or ensuring that the standard deviation of the ensemble matches the standard deviation of the reference data could be used. However, this would be correcting a well defined quantity (high / low

percentiles in the ensemble) to a poorly defined quantity in the reference dataset. This is the reason why the bias correction used in this thesis only corrects for the mean of the ensemble.

The ability to apply such a simple bias correction to all five climate variables analysed in this section and for the distribution of the variable to closely match the distribution from the reference dataset gives confidence that the model can be used to conduct attribution studies into extreme weather events involving temperature, precipitation, windspeed and MSLP.

7 Validation of weather systems in a very large ensemble of climate models

Chapters 5 and 6 perform verifications of the very large ensemble of the HadAM3P climate model by comparing the long term mean of climate variables from the ensemble against observations for the entire globe, and by testing the distributions of daily climate variables in the ensemble against the observed distributions over Europe.

The very large ensemble, and by extension, the model itself performs well. This is especially true when some identified biases are adjusted for by correcting the mean of the climate variables to match the mean of the observations, as described in sections 6.2.1 to 6.2.5. However, if the model is to be used for probabilistic event attribution over Europe, then it is important to verify that the good performance of the model in this region is due to the correct modelling of weather events.

This chapter tests the ability of the model to reproduce weather systems that are similar to those weather systems in the observed record. This is done by applying the regridding and feature identification algorithms of chapter 2 to the mean sea level pressure variable from the very large ensemble, and then using the feature tracking algorithm of chapter 2 to track the progress of the identified features as they move through the field of climate variables. The same algorithms (hereafter referred to as MAS) are then applied to the ERA–Interim dataset (Dee et al., 2011) and the tracks of features present in the very large ensemble are compared to those present in the ERA–Interim dataset. Here ERA–Interim is treated as the observations, although it is actually a reanalysis product of a data assimilation which combines observed quantities of climate variables, from both satellite and ground–based measurements, with a short forecast from a global circulation model every six hours to produce a dataset of more

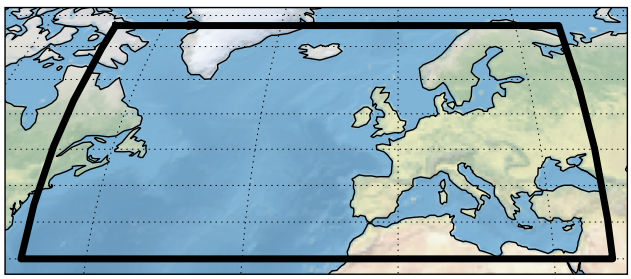
climate variables than are available in the observed datasets. Although ERA–Interim cannot be considered to be observations, it provides a consistent dataset which closely matches observations, especially for mean sea level pressure (Dee et al., 2011).

7.1 Applying the algorithms to the Europe–Atlantic region

Much of the weather in Western Europe is dominated by cyclones (low pressure systems) originating in the Atlantic Ocean and moving across the ocean into Northern Europe. These systems bring warm wet air from the Atlantic Ocean, leading to high levels of precipitation as well as strong winds, resulting from the pressure gradient. The jet stream in the Northern Hemisphere can steer these cyclones, positioning them either further North or further South. A key advantage of using an objective identification and tracking system is that the statistics of individual tracks of cyclones can be recovered. This is in contrast to a band-pass filtering method, where only spatial patterns which relate to weather systems with a particular lifetime can be recovered, and that do not identify which systems are high pressure systems and which are low pressure (Anderson et al., 2003; Blackmon et al., 1977).

The MAS feature identification and tracking algorithms are applied in order to identify low–pressure features and track them across the Europe–Atlantic region. This region encompasses most of the North Atlantic, all of Western Europe, the Mediterranean and part of North Africa. Figure 7.1.1 gives the longitude and latitude bounds of the region. This region was chosen as this is where there is a high level of cyclone activity which particularly affects Europe, as can be seen in Figure 3.1.1. Additionally, the MAS algorithms and subsequent analysis of the tracks produced are only applied to the period 1989 to 2009, so as to reflect the time period analysed in Neu et al. (2013) and enable ERA–Interim to be used as the dataset to compare to.

In light of the result of section 3.3.2, which determined that 6 hourly instantaneous data is required by the tracking algorithm to produce tracks that are comparable to those in Neu et al. (2013), both the tracking analysis of ERA–Interim and the large ensemble of HadAM3P



	Longitude	Latitude
Top left	71.25W	70N
Top right	41.25E	70N
Bottom right	41.25E	30N
Bottom left	71.25W	30N

Figure 7.1.1: Longitude and latitude coordinates of the four corners of the Europe–Atlantic region.

is performed on 6 hourly MSLP data, with 6 hourly geopotential height data being used as the steering vector in the tracking phase of the analysis. Returning 6 hourly global data from the volunteer distributing computing of global climate model runs, as described in chapter 4, would require a large amount of data to be returned from the volunteer's computers. This would place a large burden on the storage and bandwith requirements of the users and also the storage of data in the project's servers. To alleviate this, the MSLP and geopotential height (at 700 hPa) data required by the tracking algorithm at 6 hourly intervals is subset, on the volunteer's computer, to the Europe–Atlantic region as the model run progresses. Additionally, only the part of the ensemble which models the years 1989 to 2009 returns 6 hourly data. To ensure that comparable tracks in ERA–Interim are being identified, the data in ERA–Interim is also subset to the Europe–Atlantic region before the MAS feature identification and tracking algorithm is applied. For the large ensemble 1000 ensemble members per year, out of a possible 2500, are used for the tracking analysis. These 1000 members were chosen at random, with no selection criteria being applied.

To detect low–pressure features in ERA–Interim and the large ensemble, the same procedure as section 3.1 is performed. Firstly, the data is regridded to the hierarchical triangular mesh for all seasons. Low pressure features are identified via the algorithm in section 2.3. Data at level 3 in the mesh is subtracted as the background field, features are identified at level 4 in the mesh and then refined to level 5, which is the highest level in the mesh for both ERA–Interim and HadAM3P. ERA–Interim has a resolution of $1.5 \times 1.5^{\circ}$ and HadAM3P has a resolution of $1.875 \times 1.25^{\circ}$, producing a grid box area at 45°N of $\sim 200\text{km}^2$ for both sets of data.

As described in chapter 2, the MAS algorithm recognises a low–pressure feature when a local minimum is detected within the MSLP field at the identification level, after the background field has been removed. This local minimum is then refined to the highest level in the mesh and grown to encompass a region as a feature object. In the following sections, a local minimum is defined to have a pressure that is at least 2hPa lower than the background field (has a -2hPa anomaly) and all data has been contoured to the nearest 1hPa. Triangles added to the object at the refinement level during the object growing process must have a value that is less than or equal to the anomaly of the local minimum. This object growing process enables the algorithm to pick out low-pressure systems, as opposed to single low-pressure points within the field.

Although the choice of anomaly of -2hPa is somewhat arbitrary, Anderson et al. (2003) shows that the number and position of storms identified after a number of different filters have been applied are comparable to each other, but that using a mean hemispherical value of MSLP reduces the density of features. Using a two–dimensional climatological mean value improves the density at the expense of weakening the systems, in terms of depths of low and pressure gradient. Anderson et al. (2003) show that using a spectral spacial filter to remove large scale background flow results in the most satisfactory identification of low pressure features. Using the hierarchical mesh allows the background field to be defined as data lower in the grid, and subtracting this is akin to the spectral spacial filter method, in that large scale features are removed while allowing smaller scale perturbations to the field to be recognised as cyclones.

The MAS feature identification and tracking algorithms have been verified in section 3.1 and have been shown to be comparable to the 15 methods investigated in Neu et al. (2013), in that the total cyclone centre density, track length distribution, low pressure centre distribution and lifetime distribution of the new method is within the spread of the 15 methods. All feature identification and tracking algorithms make some mistakes, but it is still relevant to use a tracking algorithm to validate a model as applying the same algorithm to the model and the reanalysis data will result in the same biases.

Once the detection of the low–pressure systems has been carried out for each ensemble member, the tracks that these systems produce over time are determined by applying the tracking algorithm of chapter 2 to each ensemble member. These tracks are produced on a season by season basis, so, for example, the identified low-pressure systems that occur within an ensemble member within the months September, October and November (SON) are fed into the tracking algorithm to produce a track set for SON for that ensemble member. This is repeated for 1000 ensemble members per year and so, over the years between 1989 and 2009, 21,000 ensemble members are analysed to produce 84,000 track sets: one track set is produced per season per ensemble member.

As detailed in section 2.4, the tracking algorithm relies on two distinct phases. The first phase builds a locally optimal set of tracks by evaluating candidate track points present at a time step against the tracks which end in the previous time step. For a candidate point to be added to a track, it has to pass at the distance rule and one other of the four remaining hierarchical rules and also minimise the cost function of the rule, when compared to the other candidate points. The five rules are a distance rule, overlapping rule, intensity matching rule, steering wind rule and curvature rule. A second phase ensures that a globally optimal set of tracks is obtained by applying interchanging segments of track between the locally optimal tracks and selecting the tracks which minimise the curvature rule.

Unlike Massey (2012), where a number of free parameters required tuning to the data to which the tracking algorithm is applied, the new tracking method presented here only requires one parameter, and that is the search radius. This radius is set to 950km, as this is found to be the 98th percentile of the distance between track points when the MAS algorithms are applied to 6 hourly ERA–Interim data in section 3.3.1. Using this value, the curvature cost function of Equation 2.4.10 will actually allow the distance between track points to reach 1900km in the global optimisation phase, whereas in the local optimisation phase, the maximum distance between points remains at 950km. The increase in permitted distance in the global optimisation phase could lead to track aliasing, where two tracks are erroneously concatenated into one track. However, as shown in Table 3.3.1 this rarely happens and is above the 99th percentile of track point displacement when MAS is applied to ERA–Interim.

Future work could investigate the sensitivity of the global optimisation algorithm to the permitted distance and determine which value would lead to an optimal set of tracks with the least track aliasing.

The feature detection and tracking algorithms are applied to the large ensemble of HadAM3P using the same definition of a low–pressure system and the search radius as for the ERA–Interim dataset. In this way a direct comparison can be made between the tracks in the very large ensemble and the tracks in the ERA–Interim dataset. ERA–Interim is a deterministic dataset, which can also be thought of as a single realisation of the possible weather under a climate scenario. With only one realisation available over a 21 year period, there are only 84 seasons of tracks available to analyse in ERA–Interim. The testing protocol of Neu et al. (2013) is followed so, as in section 3.1 the tracks are filtered so only those that last for 24 hours (5 timesteps) or more are included.

7.2 Counts of tracks over the Europe–Atlantic region

After applying the MAS feature identification and tracking algorithm to the ERA–Interim reanalysis data and the large ensemble of HadAM3P over the Europe–Atlantic region, a number of metrics for the tracks are calculated. The first of these is the count of the number of tracks which occur within each season within a year, from 1989 to 2009.

Figure 7.2.1 shows the time series of track counts for each season from 1989 to 2009 produced by applying the MAS feature identification and tracking algorithms to both ERA–Interim and the large ensemble. Track counts are given for each season and, for the large ensemble the ensemble mean is produced and the 5th and 95th percentiles of track counts. ERA–Interim and HadAM3P are both driven at the lower boundary by observed SSTs. ERA–Interim uses a series of NCEP produced SST products (see Dee et al. (2011), Table I for details) whereas the large ensemble of HadAM3P uses HadISST v1.1 (Rayner et al., 2003) (see section 4.2.1 for more details). Probabilistic event attribution (PEA) studies using large ensembles of a single climate model take great care in producing SSTs that closely represent the observations, reckoning that the SSTs have a large effect on the frequency of

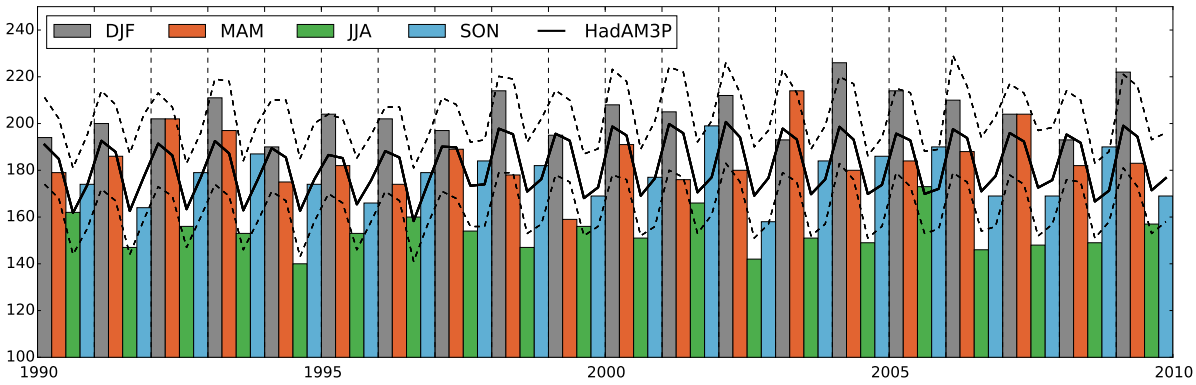


Figure 7.2.1: Counts of the number of tracks in the Europe–Atlantic domain per season over the period 1989 to 2009 produced by the MAS tracking algorithm applied to both ERA–Interim and the large ensemble of HadAM3P. The solid bars show the track counts from ERA–Interim, the solid black line shows the mean track counts for the large ensemble and the dashed lines show the 5th to 95th percentile of the track counts for the large ensemble.

occurrence of the extreme event. The non-anthropogenic counter–factual ensemble removes the estimated anthropogenic contribution to the SSTs by subtracting the SST warming estimated from the CMIP5 ensemble (Schaller et al., 2016). This indicates that there is an assumption in PEA studies that the SSTs have a large influence on regional weather, and the extremes of that weather. If the temperature gradient of the sea–surface is to have an influence on the formation, position and evolution of extratropical cyclones (low pressure systems), as Woollings et al. (2010) suggests, then the large ensemble should have a similar count of tracks, for a season and year, as the corresponding season and year in ERA–Interim. However, Figure 7.2.1 shows this not to be the case, at least in the HadAM3P model.

The ensemble mean of the track counts over the Europe–Atlantic domain in the large ensemble of HadAM3P capture, to some degree, the seasonal cycle seen in the track counts in ERA–Interim. In both ERA–Interim and HadAM3P there are more tracks in DJF and MAM and fewer in JJA and SON, with DJF having the highest track count. However, the amplitude of the seasonal cycle is not as pronounced as it is in ERA–Interim, with an approximately 30 track difference between DJF and JJA in HadAM3P, as opposed to approximately 60 in ERA–Interim. Furthermore, there are years in ERA–Interim with particularly large storm track count in DJF (1993, 1998, 2004 and 2009) or low storm track count in JJA (1992,

1994, 2002). The timings of these high / low track counts are not represented in the large ensemble of HadAM3P, with those years having very similar ensemble mean track counts as any other years.

However, examining the 95th percentile of the track counts in the large ensemble shows that there are some ensemble members which can equal or exceed the DJF track counts in ERA–Interim. Similarly, examining the 5th percentile shows that there are some ensemble members which can match or be lower than the JJA track counts in ERA–Interim. The correlation between the track counts in ERAInterim and the ensemble mean track count in the large ensemble is 0.81. The correlation coefficient here is the rvalue obtained by fitting a linear regression model (Wilks, 2011) to the track counts in ERAInterim and in the large ensemble. The correlation between the 95th percentile of track counts and the ERA–Interim track counts is 0.79, whereas the correlation between the 5th percentile and ERA–Interim track counts is 0.82, both showing equivalent correlation to the ensemble mean. This result shows that the model is capable of producing seasons with low and high storm activity, at least for some ensemble members, and that the range of track counts produced in the ensemble is comparable to the range seen in the ERA–Interim reanalysis. However, the model is not able to correctly predict the timings of high and low storm activity seasons from the SSTs.

7.3 Spatial density of tracks over the Europe–Atlantic region

Figure 7.3.1 shows the density of cyclone centres per timestep and 1000km^2 for the ERA–Interim reanalysis data after applying MAS. This Figure applies the same method of calculating the cyclone density as Figure 1 in Neu et al. (2013) and Figures 3.1.1 and 3.1.3 in this thesis. The value calculated is the percentage of timesteps at which a cyclone occurs within a grid box and per unit area of 1000km^2. For ERA–Interim this is calculated by iterating through all the storm tracks identified between 1989 and 2009 for each season and, for each point in the track, identifying which grid box in the Europe–Atlantic region the point fits

into, for a grid which has a resolution of $1.5 \times 1.5°$. Once the track point counts per grid box have been calculated, these values are divided by the number of timesteps over the 1989 to 2009 season. This equates to the number of days in a season (90, 91, or 92, depending on season) multiplied by the number of years (21) and the number of timesteps per day (4). The resulting track counts per timestep is then normalised to the area of the grid box and multiplied by 1000 to get the quantity of cyclone occurrences per 1000km^2 per timestep.

The total cyclone centre densities for the large ensemble are calculated in the same way, but each ensemble member is iterated over and, as there are 1000 ensemble members per year used in the tracking analysis, this equates to 21,000 ensemble members over the 1989 to 2009 period. After the track counts on the $1.5 \times 1.5°$ grid have been calculated, the track counts per timestep are calculated by dividing by the number of years (21) multiplied by the number of days in a season (90), the number of timesteps per day (4) and the number of ensemble members (1000). The resulting track count per timestep is then normalised to 1000km^2. 90 is used as the number of days per season as the HadAM3P model uses a 360 day calendar, so all seasons have 90 days. The same resolution grid ($1.5 \times 1.5°$) as ERA–Interim is used so that the cyclone densities are directly comparable between ERA–Interim and the large ensemble. The method which generates the central point of feature objects (Equation 2.3.1) produces a point that is off grid and so tracks produced from MAS do not have any underlying grid.

Figure 7.3.1 shows the total cyclone centre densities, per 1000km^2 and per 6 hour timestep, per season for both the tracks produced by applying MAS to the ERA–Interim data (in the left column) and the large ensemble (in the right column). The spatial distribution of the cyclone density in the large ensemble of HadAM3P matches that of ERA–Interim, in that large scale features and known areas of high cyclone activity correspond between the two datasets.

In DJF, the HadAM3P track set shows high activity (a location with a high percentage of tracks per 1000km^2 per timestep) off the southern coast of Greenland, a path of activity across the Atlantic which skirts the top of Scotland, high activity in the Mediterranean Sea

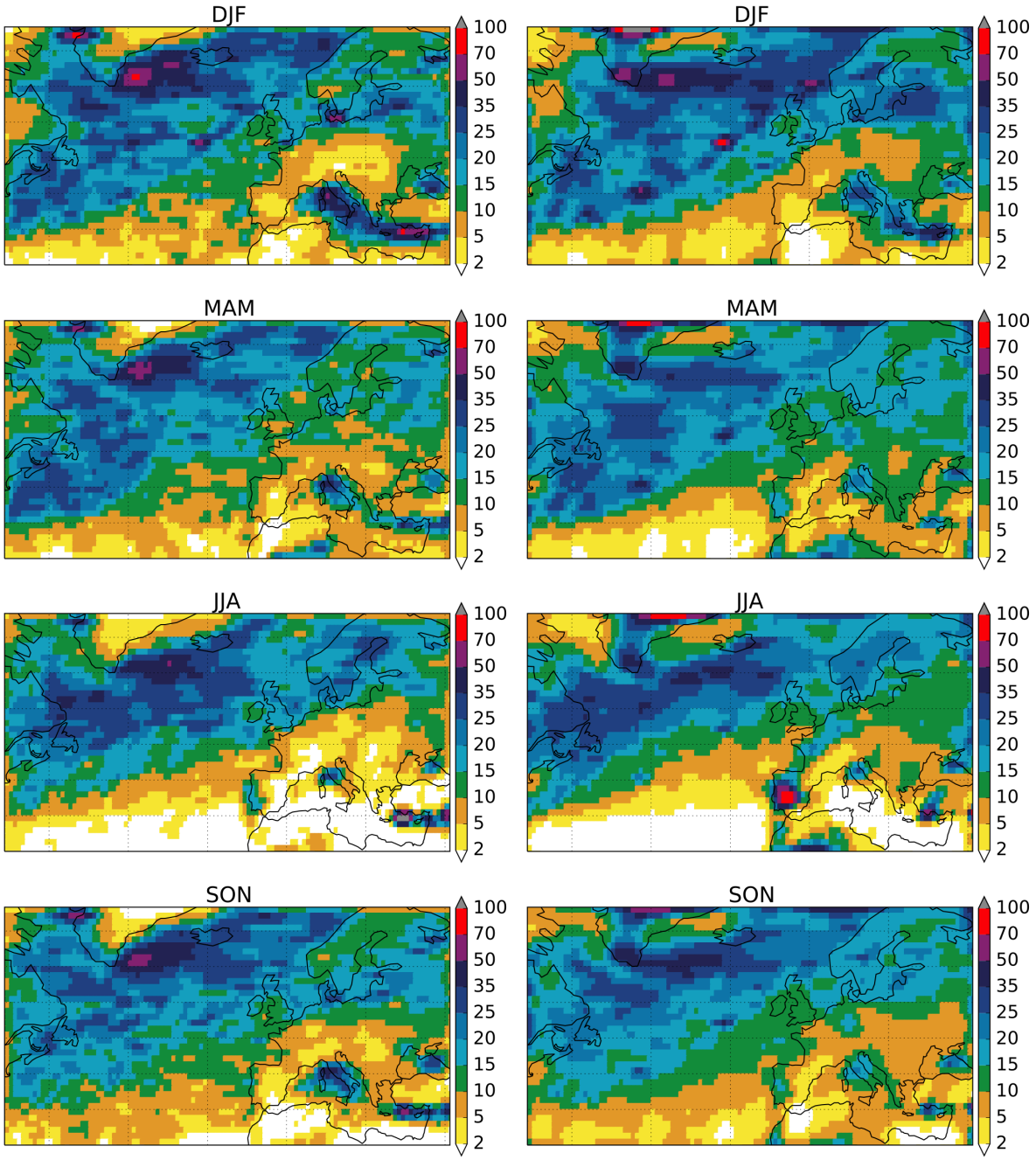


Figure 7.3.1: Seasonal total cyclone centre density (percentage of cyclone occurrence per 1000km^2 per 6 hour timestep) for the Europe–Atlantic region for tracks derived from ERA–Interim (left column) and from the large ensemble of HadAM3P (right column).

and increased activity to the east of the Baltic Sea. These areas of activity are also present in the ERA–Interim track set and in the track sets from the 15 algorithms detailed in Neu

et al. (2013). However, although there is relatively high activity in the Mediterranean Sea in the HadAM3P track set, it is not as high as in the ERA–Interim track set, with total cyclone centre density per timestep being up to 20% lower in HadAM3P. This corresponds to the area of positive MSLP bias in Figure 5.2.3 and so the model's high MSLP in the Mediterranean in DJF is reducing the frequency of low pressure systems in that area.

In MAM, the HadAM3P track set shows the same structure in total cyclone centre density over the Atlantic as the ERA–Interim track set. However, there is a reduced level of activity off the southern coast of Greenland and, again, a reduced cyclone density over the Mediterranean, although the track set does show higher cyclone density in the Mediterranean when compared to surrounding areas. There is an increased activity over the western edge of the Iberian peninsula.

In JJA, the HadAM3P track set shows the same spatial characteristics as the ERA–Interim track set. However, there is an area of high activity over the Iberian peninsula in HadAM3P that is not present in ERA–Interim. Figure 5.2.1 shows that this is an area with a positive temperature bias and so this erroneously high track density could be due to heat lows (Rácz and Smith, 1999). As heat lows can be persistent but are shallow systems, they could be filtered out by requiring features to have a higher than 2hPa anomaly between the background field and the identification level.

In SON, the HadAM3P track set again shows a similar spatial distribution of track densities when compared to ERA–Interim. The two areas with the greatest discrepancy are the Mediterranean and the southern coast of Greenland. Although low pressure systems do occur in these regions, the cyclone densities are much reduced from those in ERA–Interim.

Massey (2012) has an error where Mediterranean storms are not identified due to using a threshold value for pressure depth that is too high for Mediterranean storms to be successfully identified. However, the revised feature identification and tracking algorithm presented in this thesis (MAS) takes into account the background flow of the model, and identifies features in relation to this flow. Section 3.1.1 shows that MAS is able to successfully identify cyclones over the Mediterranean, and so the result here is due to the deficiencies in the HadAM3P

	Europe–Atlantic	EU	SREX NEU	SREX CEU	SREX MED	UK
DJF	3.44	2.08	1.47	0.81	1.23	0.48
MAM	3.08	1.72	1.31	0.72	0.84	0.43
JJA	3.02	1.75	1.38	0.55	0.92	0.51
SON	3.11	1.83	1.42	0.76	0.86	0.48

Table 7.3.1: Root mean square errors (RMSE) of the seasonal total cyclone centre densities in the large ensemble of HadAM3P when compared to the total cyclone centre densities in ERA–Interim.

model and not in the tracking algorithm.

Table 7.3.1 shows the root mean square errors (RMSE) of the large ensemble of HadAM3P track set when compared to the ERA–Interim track set, calculated over the five regions detailed in section 1.7 and for the entire Europe–Atlantic region. This Table shows that there are comparable errors across the seasons, with the DJF season having the highest RMSE for all regions, excepting over the UK. DJF is the season with the highest level of cyclonic activity in the North Atlantic and Europe and so it is reasonable that this season also has the highest error. This Table also confirms that, for all seasons, the Mediterranean (SREX MED) and Northern Europe (SREX NEU) have the highest errors of all the regions. For SREX NEU this error is likely due to the high number of cyclones present in the region, with small differences in the cyclone density per grid box cumulatively adding to the error. There is also a discrepancy over Norway for each season. For SREX MED this error is due to the differences in the cyclone densities between the track sets over the Mediterranean Sea.

7.4 Track metrics over the Europe–Atlantic region

Several more track metrics can be computed and compared between track sets. For the purpose of this thesis these are track lifetime, track length, track intensity and track propagation speed. These track metrics are the same as those computed for the Northern and Southern Hemisphere in sections 3.1.1 and 3.1.2 and in Figure 3 in Neu et al. (2013). However, the distributions of the track metrics presented in this section are different to those in sections 3.1.1 and 3.1.2 as the metrics in those sections were calculated over an entire

hemisphere whereas here they are calculated on a sub–section of the Northern Hemisphere over Europe and the North Atlantic.

7.4.1 Track lifetime

The first track metric to compare between ERA–Interim and the large ensemble is the track lifetime. This is the number of days for which a track persists. Tracks which last less than one day (4 or fewer timesteps in the 6 hourly data) have already been filtered out prior to analysis and so the minimum track lifetime is one day. The track lifetimes are calculated per track and then binned into one day wide bins. The frequency of these bins is then normalised so as to obtain a relative frequency. ERA–Interim has only one set of tracks per season and year, and so the normalisation factor is just the total number of tracks for that season for all years. The large ensemble of HadAM3P has 1000 ensemble members per year and so the normalisation factor is the total number of tracks for that season for all ensemble members in all years.

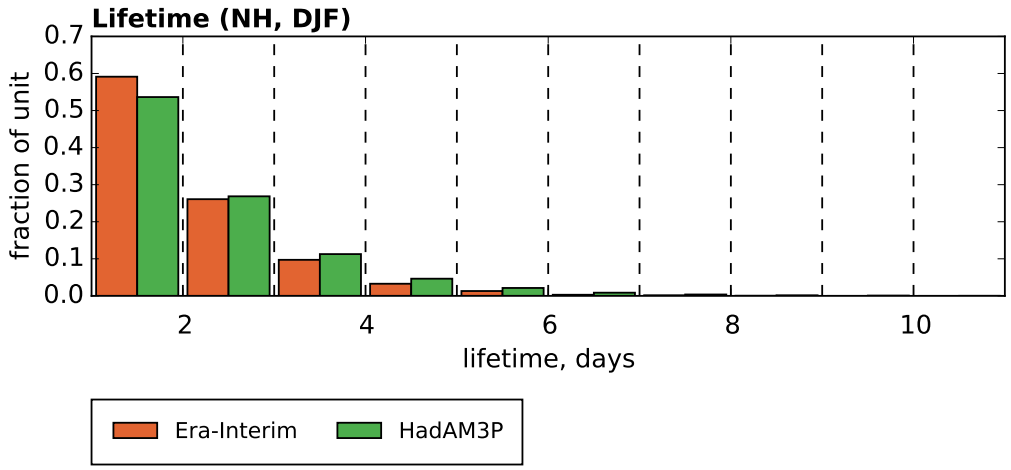


Figure 7.4.1: Normalised distribution of cyclone lifetime (the number of days the cyclone persists for) for the cyclones identified in DJF by applying the MAS feature identification and tracking algorithm to the Europe–Atlantic region of ERA–Interim and the large ensemble of HadAM3P for the period 1989 to 2009.

Figure 7.4.1 shows the normalised distribution of track lifetimes for the ERA–Interim track set and the track set from the large ensemble of HadAM3P, for DJF only. The distributions are very similar with ERA–Interim having slightly more tracks that persist for one day, and

Track lifetime (days)	KS	p	Mean		Std. dev.	
DJF	0.06	1.0	1.60	1.77	0.86	1.07
MAM	0.02	1.0	1.69	1.75	1.04	1.07
JJA	0.01	1.0	1.72	1.77	1.07	1.17
SON	0.03	1.0	1.7	1.77	1.04	1.07

Table 7.4.1: Results of the two-way Kolmogorov–Smirnoff test, showing the KS test statistic and *p-values* for the distributions of lifetimes for the cyclones identified by applying the MAS feature identification and tracking algorithm to the Europe–Atlantic region of ERA–Interim and the large ensemble of HadAM3P for the period 1989 to 2009. Also shown are the track lifetime mean and standard deviation for the ERA–Interim track set (in the grey column) and the large ensemble (in the white column).

HadAM3P having slightly more tracks that persist for two days or more. This is confirmed by Table 7.4.1 which shows the results of performing the two sample Kolmogorov–Smirnoff (KS) test on the distributions of track lifetime for both track sets. For all seasons, the KS test statistic (KS in the table) is low and the *p–value* is 1.0. This indicates that the null hypothesis (that the two distributions are drawn from the same population) cannot be rejected and so the distributions of track lifetimes is equivalent between the two track sets.

Also shown in Table 7.4.1 are the means and standard deviations of the track lifetimes for both the ERA–Interim track set and the HadAM3P track set. The means and standard deviations and very similar between track sets, for all seasons, which adds confidence that the distributions are equivalent.

Considering only a small domain, such as the Europe–Atlantic domain, will affect the lifetime of tracks, reducing, on average, the lifetime of tracks when compared to performing the tracking analysis over the entire Northern Hemisphere, as in section 3.1.1. However, the analysis for ERA–Interim has been performed on the same domain as the HadAM3P model, by subsetting the ERA–Interim data to the same Europe–Atlantic domain, and so the results presented here for the model are directly comparable to those for the reanalysis.

7.4.2 Track length

The track length is the sum of the distances between subsequent points in a track. The distance between two track points, which are defined in latitude–longitude coordinates, is calculated using the haversine distance, which is defined in Equation 2.4.3. Therefore the track length is the sum of these point to point distances taken over the surface of a sphere that has the radius of the Earth. The track lengths are calculated for each track, binned into 500km wide bins for each season and the frequency of each bin is normalised by the total number of tracks for the season in the track set. For ERA–Interim, this is the total number of tracks in the season for 1989 to 2009, whereas for the large ensemble of HadAM3P, it is the total number of tracks in the season for all ensemble members in all years between 1989 and 2009.

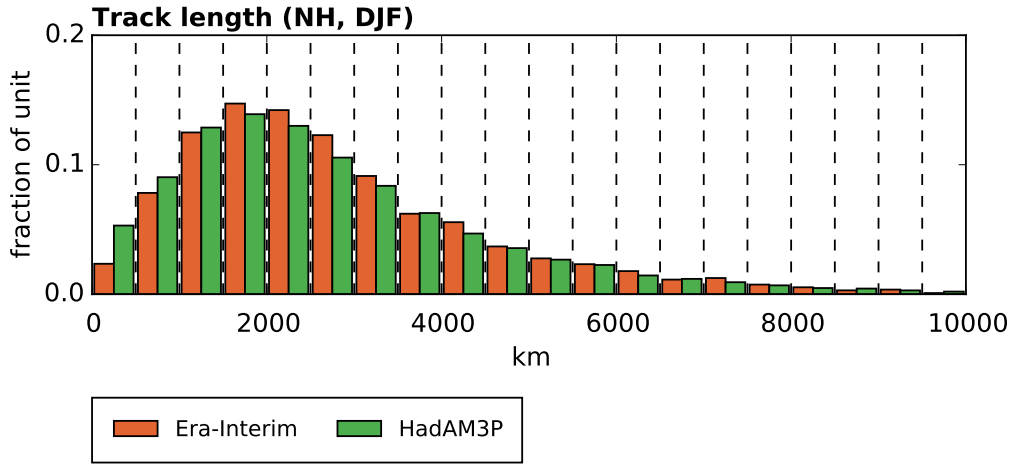


Figure 7.4.2: Normalised distribution of track length for the cyclone tracks identified in DJF by applying the MAS algorithm to the Europe–Atlantic region of ERA–Interim and the large ensemble of HadAM3P for the period 1989 to 2009.

Figure 7.4.2 shows the normalised distribution of track lengths for the ERA–Interim track set and the large ensemble of HadAM3P track set, for DJF only. The distributions are very similar, with the HadAM3P track set having more tracks in the 0 to 500km, 500 to 1000km and 1000 to 1500km ranges and fewer in the 1500 to 2000km, 2000 to 2500km, 2500 to 3000km and 3000 to 3500km ranges. For longer track lengths the distributions are also very similar. This is confirmed in Table 7.4.2 which shows the results of a KS test between the

Track Length (km)	KS	p	Mean		Std. dev.	
DJF	0.06	1.0	2560.61	2350.52	1682.16	1627.98
MAM	0.03	1.0	2065.66	2000.0	1424.26	1490.52
JJA	0.04	1.0	1803.03	1788.66	1327.29	1413.91
SON	0.05	1.0	2040.82	2010.42	1408.2	1529.19

Table 7.4.2: Results of the two-way Kolmogorov–Smirnoff test, showing the KS test statistic and *p-values* for the cyclone track lengths identified by applying the MAS algorithm to the Europe–Atlantic region of ERA–Interim and the large ensemble of HadAM3P for the period 1989 to 2009. Also shown are the track length mean and standard deviation for the ERA–Interim track set (in the grey column) and the large ensemble (in the white column).

track lengths in the ERA–Interim track set and the HadAM3P track set. The *p–value* of the KS test for all seasons is 1.0 and the KS test statistic is low, indicating that the hypothesis that the distributions of track lengths in the ERA–Interim track set and the HadAM3P track set are the same cannot be rejected.

Table 7.4.2 also shows the means and standard deviations of the track lengths for each track set in each season. Across the seasons, the track lengths in the ERA–Interim track set and the HadAM3P track set have similar (within 10%) means and standard deviations. This adds confidence that the track lengths are drawn from the same distribution and that the HadAM3P model is representing the lengths of cyclone tracks well. As noted in section 7.4.1 the size of the Europe–Atlantic domain will have an effect on the mean and maximum length of the tracks.

7.4.3 Track intensity

The track intensity is defined as the lowest MSLP value that occurs at any point during the cyclone track's lifetime. This is calculated, per season, for each track in both the ERA–Interim track set and the HadAM3P track set and the intensity values are binned into 10hPa wide bins. The frequencies of the bins are normalised by dividing by the total number of tracks in the track set. Figure 7.4.3 shows the normalised distribution of track intensity for both the ERA–Interim track set and the track set derived from the large ensemble of HadAM3P, for DJF. There are relatively fewer very deep cyclones in HadAM3P than in ERA–

Interim, with intensities in the range 950 to 960hPa having about half the relative frequency than in ERA–Interim. All intensities below 1010hPa have a lower relative frequency of occurrence in the large ensemble of HadAM3P than in ERA–Interim.

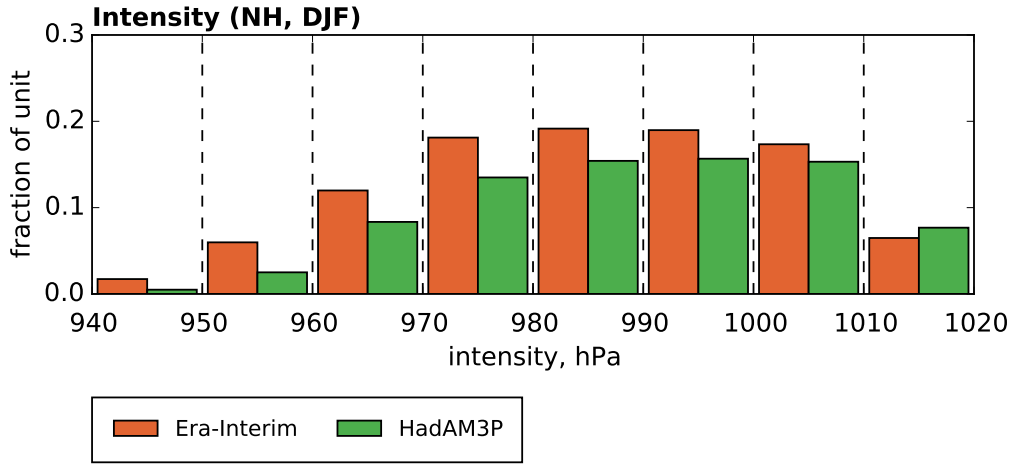


Figure 7.4.3: Normalised distribution of track intensity (minimum MSLP value along the track) for cyclone tracks identified in DJF by applying the MAS algorithm to the Europe–Atlantic region of ERA–Interim and the large ensemble of HadAM3P for the period 1989 to 2009.

To quantify this difference, Table 7.4.3 shows the results of performing a two sample KS test on the track sets produced by applying MAS to the ERA–Interim reanalysis and the large ensemble of HadAM3P. Although, inspecting Figure 7.4.3 appears to show that the distribution of intensities is different between ERA–Interim and the large ensemble, in DJF, this is not apparent in the KS test statistic. The *p–value* of 0.95 does not indicate that the hypothesis that the distributions are drawn from the same population should be rejected. The KS test statistic of 0.08 is also not particularly high. For the other seasons, this is also true, that none of the *p–values* are lower than 0.05. However, the value for SON (0.26) is much lower than the values for the other seasons.

Table 7.4.3 also shows the mean and standard deviation of the track intensities in the ERA–Interim and large ensemble of HadAM3P track sets. The means and standard deviations of the intensities for the DJF, MAM and JJA seasons are close to those in ERA–Interim (being within 10%), whereas, for SON, there is a greater discrepancy in both the mean and the standard deviation. This, at least partly, explains the higher KS test statistic and lower

Intensity (hPa)	KS	p	Mean		Std. dev.	
DJF	0.08	0.95	978.39	981.14	15.95	14.50
MAM	0.09	0.83	987.02	989.62	11.83	10.37
JJA	0.08	0.92	992.02	993.66	7.66	6.90
SON	0.15	0.26	984.13	989.05	13.12	10.93

Table 7.4.3: Results of the two-way Kolmogorov–Smirnoff test, showing the KS test statistic and *p-values* for the track intensity (minimum MSLP value along a track) identified by applying the MAS algorithm to the Europe–Atlantic region of ERA–Interim and the large ensemble of HadAM3P for the period 1989 to 2009. Also shown are the intensity mean and standard deviation for the ERA–Interim track set (in the grey column) and the large ensemble (in the white column).

p–value.

The difference in frequencies for low pressure cyclones between the large ensemble and ERA–Interim reflects the findings in section 5.2.3 and sections 6.2.4: that the model exhibits a positive bias in MSLP over certain areas of Europe, which affects the model's ability to represent very low pressure systems.

7.4.4 Track propagation speed

The track propagation speed is defined as the mean speed at which the central low pressure point in the cyclone moves over the lifetime of the cyclone. Both the ERA–Interim and HadAM3P track sets are derived from data with a 6 hour timestep. Therefore, the mean propagation speed is the total track length (in km) divided by the track lifetime, in hours. The propagation speed is calculated for all tracks in the ERA–Interim track set and all tracks in the track set derived from the large ensemble of HadAM3P. These propagation speeds are then binned into 10 km/hr wide bins and normalised by the total number of tracks.

Figure 7.4.4 shows the distribution of propagation speed for the ERA–Interim and HadAM3P track sets. The distributions are quite similar, except for a negative shift in the mean for the HadAM3P track set. This can most clearly be seen in the frequency of propagation speed in the 30 to 40 km/hr range and 40 to 50 km/hr range, where frequency of propagation speed in the HadAM3P tracks in a bin more closely matches the frequency in ERA–Interim in the

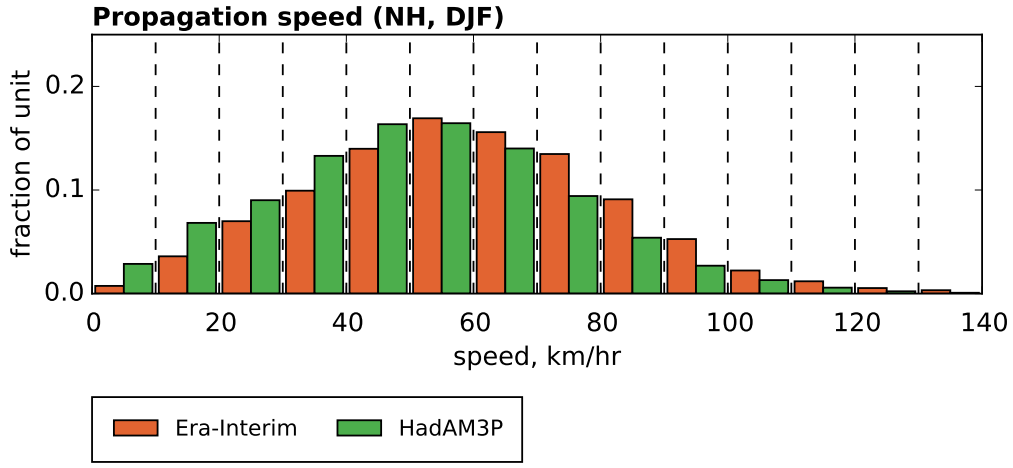


Figure 7.4.4: Normalised distribution of the mean propagation speed for the cyclone tracks identified in DJF by applying the MAS algorithm to the Europe–Atlantic region of ERA–Interim and the large ensemble of HadAM3P for the period 1989 to 2009.

Propagation speed (km/hr)	KS	p	Mean		Std. dev.	
DJF	0.13	0.31	53.6	45.67	23.60	23.19
MAM	0.09	0.81	43.0	38.88	19.82	20.99
JJA	0.05	1.0	35.66	34.54	17.30	19.37
SON	0.09	0.83	43.13	38.37	20.82	21.51

Table 7.4.4: Results of the two-way Kolmogorov–Smirnoff test, showing the KS test statistic and *p-values* for the track propagation speed (mean speed of the cyclone) identified by applying the MAS algorithm to the Europe–Atlantic region of ERA–Interim and the large ensemble of HadAM3P for the period 1989 to 2009. Also shown are the intensity mean and standard deviation for the ERA–Interim track set (in the grey column) and the large ensemble (in the white column).

next bin.

The shift in the mean of the distribution of track propagation speed in DJF, between the ERA–Interim track set and HadAM3P track set is confirmed by Table 7.4.4, which shows the results of applying a KS test to the distributions of the two track sets, as well as the mean and standard deviation of the propagation speed. For DJF propagation speed, the HadAM3P track set has a very similar standard deviation as the ERA–Interim track set, but the mean is $\sim 15\%$ less. This has an effect on the *p–value* from the KS test which,

although it is greater than the 0.05 threshold needed to reject the null hypothesis, is lower than the *p–value* for other seasons. For example, JJA has a *p–value* of 1.0 and a mean and standard deviation that are close to the ERA–Interim values. SON and MAM also have a mean propagation speed that is lower in the HadAM3P track set than in the ERA–Interim track set. Overall, in the three seasons with the higher storm activity (DJF, MAM, SON) and especially the season with the highest (DJF), the HadAM3P model is representing the cyclones as moving too slowly, when compared to ERA–Interim. The model can represent fast moving cyclone systems but not at the same frequency as in ERA–Interim.

7.5 Distribution of variables associated with cyclones

Extratropical cyclones are the dominant driver of weather in the mid–latitudes and produce much of the day to day variability in that region (Chang et al., 2002). They are also responsible for many extreme weather events, such as high precipitation, high windspeeds and storm surges caused by low MSLP (Leckebusch et al., 2006). Chapter 6 examines the daily mean values of MSLP, precipitation and windspeed when averaged over one of five regions of Europe. Although this analysis produces metrics which can be used to validate the large ensemble of a climate model, it does not allow the investigation of the values that are within the extreme tails of the distributions of climate variables which occur within a single gridbox. Due to the large data volume produced by the large ensemble of HadAM3P, it is not practical to analyse the climate variables for all the grid boxes which occur in the European regions. Using the tracks produced from the MAS allows for the variables associated with the cyclones to be extracted from the ERA–Interim and HadAM3P climate variable fields and then compared to each other.

To extract the MSLP, precipitation and 10m windspeed variables associated with the track of a cyclone identified by the MAS algorithms, the track is read in from the track file, and the MSLP, precipitation and windspeed are also read in from the ensemble member seasonal data (or ERA–Interim season) from which the track was generated. At each track point, a 1000km radius is searched to find the maximum 10m windspeed, maximum precipitation rate

and minimum MSLP value. Using a 1000km radius with 6 hourly climate data is the same method as used in the XWS storm catalogue (Roberts et al., 2014). After applying this method, a maximum accumulated precipitation, maximum 10m windspeed and minimum MSLP value that is associated with the cyclone centre is recorded for every point in the track. The values of these variables are in the tails of the distributions of the variables, when the distribution is taken over all timesteps and the whole region. Here, the values are associated with a cyclone which occurs between two timesteps and in a region that has been identified as having potentially extreme weather due to the presence of an extratropical cyclone. It is the distribution of these associated variables that are analysed in the rest of this section. The variables have not been bias corrected using the method presented in section 6.1.1.

7.5.1 Mean sea level pressure

Figure 7.5.1 shows the quantile–quantile plot (Q–Q plot) for the minimum MSLPs occurring within 1000km of the cyclone centres, comparing the percentile values in the large ensemble of HadAM3P to the corresponding percentiles in ERA–Interim. As can be seen in the Figure, the only season that matches the very low MSLP values in ERA–Interim is DJF, with MAM, JJA and SON having a positive bias for all percentiles below the 90th. These values, below the 90th percentile ($\sim$ 1000hPa), represent the depth of most cyclones, as shown in Figure 7.4.3. This confirms the findings in sections 5.2.3,6.2.4 and 7.4.3: that there are biases in the daily and seasonal MSLP and that these biases are represented in the cyclone tracks as a positive bias in the minimum MSLP associated with a cyclone centre. This is only the case for MAM, JJA and SON, as DJF produces a good representation of the minimum MSLP associated with a cyclone centre.

Table 7.5.1 quantifies the bias in the percentiles of MSLP, when comparing the minimum MSLP associated with tracks in the HadAM3P ensemble and the minimum MSLP associated with tracks in ERA–Interim. The Table shows that DJF has the lowest RMSD of all the seasons and that most of the error, for all seasons, occurs in the lowest percentiles, i.e. the strongest cyclones. Table 7.5.1 also shows the results of applying a two sample KS test

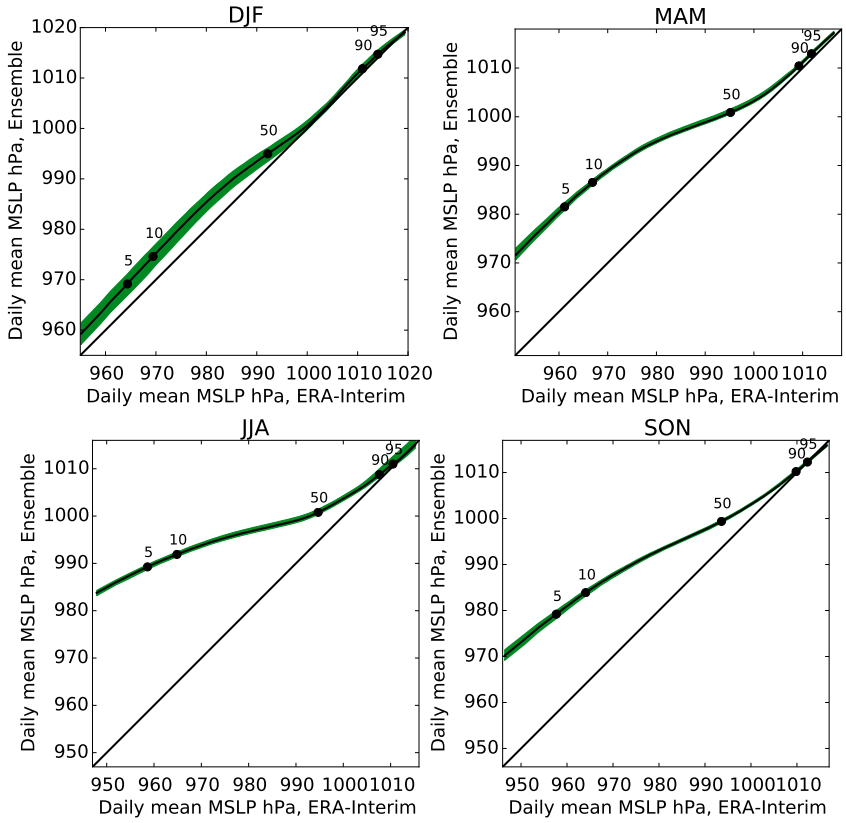


Figure 7.5.1: Quantile–Quantile plot of the minimum MSLP occurring within 1000km radius of a cyclone centre in the tracks produced by applying the MAS algorithms to the ERA–Interim reanalysis between 1989 to 2009 and the large ensemble of HadAM3P for the same period. The envelope shows the 5th to 95th confidence interval, estimated from the individual ensemble members.

to the distribution of minimum MSLPs from the ERA–Interim tracks and the minimum MSLPs from the large ensemble of HadAM3P. This confirms that the only season where the hypothesis is not rejected, and that the distributions are the same between ERA–Interim and HadAM3P, is DJF, with a *p–value* of 0.55, whereas the other seasons have a *p–value* of 0.0.

7.5.2 Precipitation

Figure 7.5.2 shows the Q–Q plot of the maximum daily accumulated precipitation ($\mathsf{P_{acc}}$) associated with a cyclone, that is occurring within 1000km radius of the cyclone centre. $\mathsf{P_{acc}}$ is calculated for all tracks in the ERA–Interim track set and all tracks in the HadAM3P track

MSLP (hPa)	DJF	MAM	JJA	SON
all	3.53	11.05	14.52	11.01
1–10th	4.82	20.23	30.80	21.56
90th+	0.65	1.08	0.61	0.29
KS	0.11	0.27	0.34	0.25
p	0.55	0.00	0.00	0.00

Table 7.5.1: Root mean square deviations (RMSD) for all percentiles from 1 to 99 for the minimum MSLP associated with a cyclone centre (*all*), the 1st to 10th (*1–10th*) and the 90th to 99th percentiles (*90th+*). The deviations are from the values at the corresponding percentiles in ERA–Interim. Also shown are the results of the two-way Kolmogorov–Smirnoff test, showing the KS test statistic and the *p–value* for the HadAM3P ensemble and ERA–Interim.

set. The Q–Q plot presents the values of the percentiles plotted against each other. As can be seen in the Figure, the track set derived from the large ensemble of HadAM3P has associated $\mathsf{P}_{\mathsf{acc}}$ values that resemble those associated with the ERA–Interim track set for percentiles below the 90th, across all seasons. Percentiles above the 90th have large discrepancies from those in the ERA–Interim, with the large ensemble of HadAM3P producing too little rainfall associated with a cyclone for the high percentiles. The percentiles here are in the extreme tails of the distribution of precipitation over the entire European region, and so this shows that the HadAM3P cannot represent very heavy precipitation events, such as the greater than 300mm/day that occurred over Cumbria in December 2015 due to Storm Desmond (van Oldenborgh et al., 2016).

The inability of the HadAM3P model to capture the high $\mathsf{P}_{\mathsf{acc}}$ events is confirmed in Table 7.5.2. This shows that the RMSD for the high percentiles (*90th+*) is much greater than for all of the percentiles or for the lower (*1–10th*), indicating that it is the higher percentiles which contribute the most error to the RMSD of all percentiles. Applying the KS test shows that only the SON distribution of $\mathsf{P}_{\mathsf{acc}}$ associated with the cyclone tracks in the HadAM3P ensemble can be considered a distinct distribution when compared to $\mathsf{P}_{\mathsf{acc}}$ in ERA–Interim, as the *p–value* is less than 0.05. JJA and MAM have quite high *p–values* and so the distributions can be considered to be non distinct. However, although the *p–value* in DJF is not low enough to reject the null hypothesis of the KS test, it is sufficiently low (0.07) to warrant further investigation as to whether the distributions are drawn from the

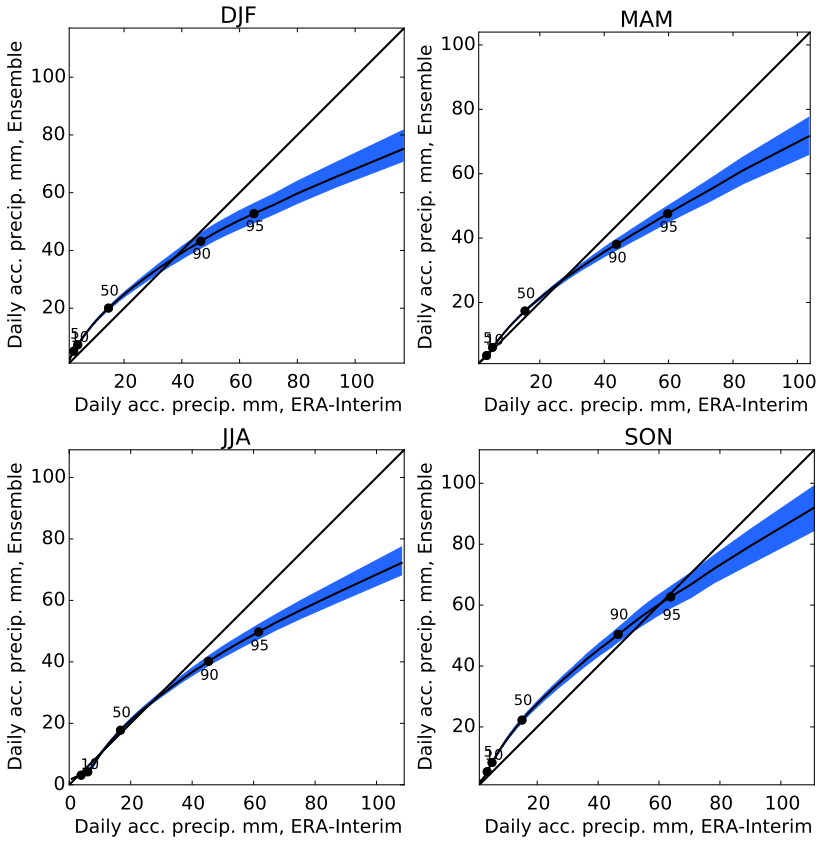


Figure 7.5.2: Root mean square deviations (RMSD) for all percentiles from 1 to 99 for the maximum precipitation (in mm/day) associated with a cyclone centre (*all*), the 1st to 10th (*1–10th*) and the 90th to 99th percentiles (*90th+*). The deviations are from the values at the corresponding percentiles in ERA–Interim. Also shown are the results of the two-way Kolmogorov–Smirnoff test, showing the KS test statistic and the *p–value* for the HadAM3P ensemble and ERA–Interim.

same population.

For all seasons, the Q–Q plots give enough evidence to suggest that the equivalence of the distributions of $\mathsf{P}_{\mathsf{acc}}$ only holds for percentiles that are lower than the 90th, even though the KS test indicates the distributions are equivalent, for all percentiles, for three of the four seasons.

Precipitation (mm/day)	DJF	MAM	JJA	SON
all	7.21	5.29	5.53	6.21
1–10th	2.59	0.39	1.18	2.09
90th+	19.75	16.59	17.80	7.75
KS	0.18	0.08	0.06	0.22
p	0.07	0.89	0.99	0.01

Table 7.5.2: Root mean square deviations (RMSD) for all percentiles from 1 to 99 for the maximum precipitation associated with a cyclone centre (*all*), the 1st to 10th (*1–10th*) and the 90th to 99th percentiles (*90th+*). The deviations are from the values at the corresponding percentiles in ERA–Interim. Also shown are the results of the two-way Kolmogorov–Smirnoff test, showing the KS test statistic and the *p–value* for the HadAM3P ensemble and ERA–Interim.

7.5.3 Windspeed at 10m

As mentioned in section 6.2.5 an attribution study would, by preference, use the 3 second maximum gust windspeed at 10m, rather than the 6 hourly maximum windspeed at 10m, which is the maximum windspeed at any timestep over the 6 hour period. However, the gust windspeed variable is not available in the output from the large ensemble of HadAM3P and so the maximum windspeed over 6 hours ($\mathsf{U}_{\mathsf{max}}$) is used instead. As HadAM3P has a 15 minute timestep, this is akin to the maximum of the 15 minute mean windspeed. The maximum windspeed over 6 hours in ERA–Interim is used to compare to.

Figure 7.5.3 shows the Q–Q plot of $\mathsf{U}_{\mathsf{max}}$ associated with a cyclone, comparing the percentile values in the ERA–Interim track set with those in the large ensemble of HadAM3P. In the DJF season, $\mathsf{U}_{\mathsf{max}}$ in HadAM3P is greater at every percentile than $\mathsf{U}_{\mathsf{max}}$ in ERA–Interim. For MAM and SON there is a reasonably close match between $\mathsf{U}_{\mathsf{max}}$ in the two track sets. For JJA, $\mathsf{U}_{\mathsf{max}}$ in the HadAM3P ensemble is much lower than in ERA–Interim, for all percentiles above the 10th.

The errors in $\mathsf{U}_{\mathsf{max}}$ are enumerated in Table 7.5.3, which shows the RMSD for each season and for all percentiles, the lower percentiles (*1–10th*) and the higher percentiles (*90th+*). The results of the two sample KS test is also shown. For DJF, although there are some errors in the lower percentiles, there are more in the higher percentiles, indicating that most of the

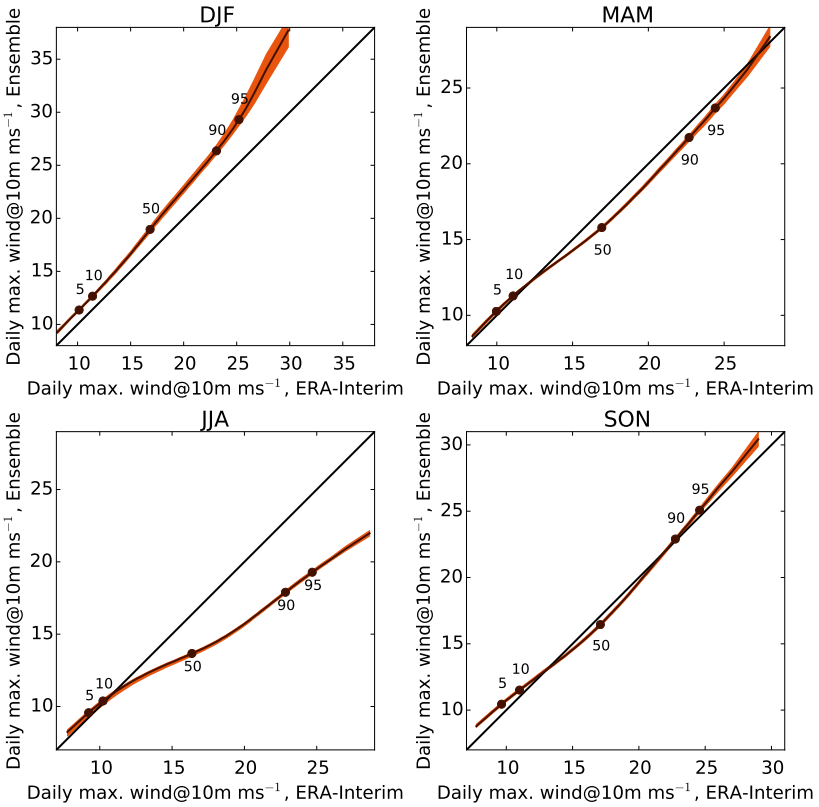


Figure 7.5.3: Q–Q plot of the maximum windspeed at 10m occurring within 1000km radius of a cyclone centre in the tracks produced by applying the MAS algorithms to the ERA–Interim reanalysis between 1989 to 2009 and the large ensemble of HadAM3P for the same period. The envelope shows the 5th to 95th confidence interval, estimated from the individual ensemble members.

error is from these higher percentiles. The *p–value* of the two sample KS test is low enough in DJF to consider the distribution of $\mathsf{U}_{\mathsf{max}}$ in the HadAM3P ensemble and in ERA–Interim to be distinct, even though it is not lower than the formal limit of 0.05. For MAM and SON, the *p–value* of the KS test indicates that the null hypothesis cannot be rejected and that the distributions of $\mathsf{U}_{\mathsf{max}}$ in the individual seasons should be considered to be drawn from the same population. JJA has a *p–value* of 0.0 indicating that the two distributions are distinct. There are also large errors in the high (*90th+*) percentiles.

It is surprising that the windspeeds associated with cyclone tracks in the HadAM3P ensemble in MAM and SON have such a similar distribution to the windspeeds associated with the ERA–Interim tracks, as the MSLP associated with the tracks in those seasons show no

Windspeed@10m (ms^{-1})	DJF	MAM	JJA	SON
all	2.50	0.89	3.21	0.52
1–10th	1.22	0.27	0.33	0.79
90th+	4.89	0.72	5.59	0.71
KS	0.19	0.11	0.32	0.06
p	0.05	0.55	0.00	0.99

Table 7.5.3: Root mean square deviations (RMSD) for all percentiles from 1 to 99 for the maximum windspeed at 10m associated with a cyclone centre (*all*), the 1st to 10th (*1–10th*) and the 90th to 99th percentiles (*90th+*). The deviations are from the values at the corresponding percentiles in ERA–Interim. Also shown are the results of the two-way Kolmogorov–Smirnoff test, showing the KS test statistic and the *p–value* for the HadAM3P ensemble and ERA–Interim.

correlation (Figure 7.5.1), and the two sample KS test returns a *p–value* of 0.0 (Table 7.5.1), indicating that the distributions of MSLP are drawn from different populations. As $\mathsf{U}_{\mathsf{max}}$ is proportional to depression depth in cyclones, it would be expected that the errors in MSLP in MAM and SON would cause larger errors in $\mathsf{U}_{\mathsf{max}}$ in those seasons. This is true for JJA, where the large error in the MSLP is linked to the large error in the windspeed. The opposite is the case for DJF, where there is little error in the MSLP (*p–value* of 0.55) and a larger error in the windspeed (*p–value* of 0.05).

7.6 Summary and conclusions

This chapter has presented the results of applying the feature detection and tracking algorithms of chapter 2 to both the very large ensemble of the HadAM3P model and the ERA–Interim reanalysis. By comparing the statistics and location of tracks from the HadAM3P ensemble with those from ERA–Interim, the ability of the model to replicate the weather systems seen in the reanalysis dataset is assessed. The ability of the model to represent weather systems is important for the validation exercise as it ensures that any good correlation between weather statistics in the observational record and the large ensemble is due to the model being able to correctly capture the physical processes involved in the atmospheric flow.

Low pressure systems that track from the Atlantic Ocean over Northern Europe are examined

in terms of their statistics and their spatial location. Figure 7.4.2 shows that there is good correspondence in the length of tracks of low-pressure systems between the large ensemble and ERA–Interim, and this is confirmed with a two way Kolmogorov-Smirnov (K-S) test (Table 7.4.2). The model is also able to replicate the distribution of track persistence, as shown in Figure 7.4.1 and Table 7.4.1, where the *p–values* of the KS tests indicate that the distribution of track lifetimes in the HadAM3P ensemble is non distinct from the ERA–Interim reanalysis.

Figure 7.4.3 shows that the model is unable to replicate the exact distribution of the minimum MSLP values of the ERA–Interim storm tracks. The distributions do have some resemblance to each other, and the mean and standard deviation values are close (except for SON) indicating that the model is at least capturing some of the properties of the distribution. The K-S test indicates that, for all seasons, the *p–value* of the KS test is not low enough to reject the null hypothesis that the distributions are drawn from the same population, even though the frequency of occurrence of very low MSLP values is half that of ERA–Interim.

The model fares much better in correctly positioning the locations of the cyclone tracks, as shown in Figure 7.3.1. There is very good spatial correspondence between the most frequent positions of tracks in the very large ensemble and in the ERA–Interim dataset.

The model has mixed results when comparing the climate variables associated with the tracks in HadAM3P with those in ERA–Interim. The associated variables are the maximum or minimum values of a variable occurring within 1000km of a cyclone centre. For MSLP, only the DJF season has good representation of minimum MSLP associated with a track (Figure 7.5.1, *p–value* of 0.55 in Table 7.5.1) whereas the other seasons have *p–values* of 0.0 and distributions that cannot represent the low values of MSLP associated with a storm track. Precipitation is well represented, except for the very high percentiles (90th+), as shown in Figure 7.5.2 and Table 7.5.2. Finally, windspeeds associated with a cyclone are well represented in MAM and SON, poorly represented in JJA and have a positive bias in DJF, which may be resolved by a correction to the mean windspeed, as seen in Table 7.5.3 and Figure 7.5.3. However, it should be noted that the MSLP, windspeed and precipita-

tion variables have not been bias corrected. A correction of the variable in the HadAM3P ensemble to the mean of the variable in ERA–Interim could be carried out, calculating the ERA–Interim mean from the 6 hourly data over either an area average over the EU region or individual points within the region. This could improve the correlation between the variables associated with a cyclone in the HadAM3P ensemble and those in the ERA–Interim ensemble.

In conclusion, it is deemed that the model can represent low-pressure systems to a reasonable degree, enough to have confidence in using the model in a probabilistic event attribution study examining weather events that are related to low-pressure systems over Europe. However, some caveats must be made and they are that very high precipitation events are not well represented, the very low MSLP events are not well represented for seasons other than DJF and that the windspeeds are not well represented in JJA.

8 Conclusions and future work

8.1 Summary, conclusions and discussion

At the beginning of this thesis, the question is posed of how to validate a climate model so as to have confidence in its representation of the distribution of weather variables so that it can be used in probabilistic event attribution (PEA) studies. To answer the question, this thesis concentrates on determining which methods and techniques are best suited to validating the statistics of weather variables produced by a climate model, develops and implements a new set of algorithms which detect and track features in meteorological data, devises and performs an experiment of a very large ensemble of a climate model and applies the validation methods and techniques to the output from this experiment.

In **chapter 2**, the development of a new method to detect storm like features in meteorological data is documented. In order to overcome the problems encountered when detecting features in rectangularly gridded data, a method of regridding data to an equal–area hierarchical triangular mesh is developed. This mesh has the advantage that there is no singularity at the pole and that the resolution is constant throughout the mesh, unlike a rectangular grid. Climate data is regridded to the hierarchical mesh and the background field is removed. This is achieved by smoothing a lower resolution version of the data, from a lower level in the mesh, to the higher level at which feature identification is performed and subtracting the smoothed data from the higher resolution data. The resulting data is then contoured to 1hPa levels, which simplifies the search for minima in the data. Feature identification locates low pressure features by examining adjacent triangles at an *identification level* and refining these to a higher level. From these triangles a feature object is derived by expanding the triangles via a concentric shell so as to encapsulate a cyclone (low pressure system) as a

feature object. Central points are obtained for these feature objects by a weighted average of the positions of the individual triangles in the object.

The second part of chapter 2 details a new method of tracking the feature points derived above, which uses a two part process. Firstly a set of locally optimal tracks is derived by using a hybrid prediction–optimisation scheme. To add feature objects to tracks a number of hierarchical rules need to be passed and the tracks are optimised locally by finding the object which passes the highest rule with the lowest cost. To remedy the occasional wrong assignment of a feature–object at the beginning and end of a track, a global optimisation scheme is employed which exchanges segments between tracks so as to maximise track lengths and minimise track mean curvature across all tracks.

Chapter 3 validates the tracking algorithm by applying it to DJF Northern Hemisphere MSLP data and Southern Hemisphere JJA MSLP from ERA–Interim. This follows the same testing protocol as Neu et al. (2013) and so direct comparison between the new identification and tracking algorithm and the 15 methods documented in Neu et al. (2013) is made. The new method is found to perform well, being consistent with the 15 other methods. Taking the output of the 15 methods as an ensemble, the output from the new method is statistically indistinguishable from the other ensemble members.

The new method is also applied to a higher resolution version of the ERA–Interim data from 1979 to 2009, over Europe and the North Atlantic only, and compared to the XWS windstorm catalogue (Roberts et al., 2014). Comparing the tracks produced by the new method shows that there is a good match, both spatially and in terms of the MSLP depth along the track, for all but 2 of the 20 named tracks in the XWS catalogue. The sensitivity of the tracking algorithm to the temporal resolution of the input data is also investigated by degrading the ERA–Interim reanalysis to 12 hourly and 24 hourly data. However, due to the reduction of the performance of the algorithm on degraded data, the large ensemble of a climate model is designed to return 6 hourly data over a domain covering Europe and the North Atlantic, so as to reduce the storage and bandwidth burden.

To sample all potential weather systems produced by a climate model acting under a climate

scenario, a very large ensemble of a single climate model is required. **Chapter 4** outlines the construction of an experiment using volunteer distributed computing (VDC) to obtain this very large ensemble. The HadAM3P model from the UK Met. Office is used and a climate scenario is constructed for the 1960 to 2000 period using observed datasets for the sea surface temperatures, sea ice fraction and atmospheric composition, including CO_2, CH_4, N_2O, the halocarbons, and other atmospheric gases such as ozone, natural and anthropogenic sulphate concentrations, volcanic eruptions and dimethyl sulphide. The natural variability of the Sun is also modelled by providing an offset to the solar constant.

In order to produce a large population of ensemble members that will fully sample the possible weather events that can occur, an initial condition perturbation in potential temperature is used. Within a year, each ensemble member will have a similar climatology for the year, but the slightly different starting values for the potential temperature will lead to different weather patterns. The temperature at a single gridpoint in the ensemble members starts to diverge after 3 days and, by day 7, has a large degree of variability. The large ensemble of a climate model is computed via volunteer distributed computing (VDC) and 100,000 ensemble members spanning a 50 year period are analysed in Chapters 5 to 7. In order to model the tails of the distribution with confidence, a large ensemble like this is required. Only by relying on VDC can such a large ensemble be computed, certainly for a DPhil. student, as to do it via a supercomputer would be prohibitively expensive.

Chapter 5 analyses the 100,000 ensemble members by comparing the seasonal means of the ensemble means to the seasonal means of the ERA–40 dataset for surface temperature, precipitation, MSLP and upper–level wind. However, it is first noted that there are differences between the observed datasets (CRUTS for temperature, GPCP for precipitation) and the ERA–40 data. Either of these datasets could be used to compare the model to, with the possibility that the model validity or non–validity could in fact be attributed to bias in the observed or reanalysis dataset. This leads to the concept of using the maximum reanalysis bias, as a non–parametric indicator of a substantial deviation from the mean, and the ensemble is itself treated in the same manner as the observed datasets. If the bias between the ensemble mean and the observation is larger than the bias between ERA–40 and the

observed data, then the bias is treated as significant. This method allows for variability and error within the observed datasets, but does not link the validity of the ensemble to the validity of one single dataset. The ensemble is considered valid if it is indistinguishable from the observations or the reanalysis, in that its bias is within the biases between the datasets.

Using this method for surface temperature shows that there are large, and significant biases in the temperature, especially over Europe in JJA. However, by considering the difference between high or low percentiles (95th or 5th) and the 50th percentile, much of this bias disappears and indicates that a probabilistic attribution study can still be relevant even if the mean state is itself severely biased. This is important as PEA studies are essentially examining the difference between two percentiles in the model, as the extreme event is expressed in terms of an anomaly to the climatological value. Therefore, when measuring a change in return time of surface temperature, the result is valid as the model can represent the differences in percentiles in surface temperature.

The same method is used to compare the ensemble mean seasonal mean precipitation bias, using ERA–40 and GPCP to construct the maximum bias. There are biases in the accumulated precipitation for a season, particularly a dry bias over Central Europe and North America and a wet bias over North Africa in JJA and a dry bias over North Africa in DJF. That the dry bias in JJA over North America and Central Europe corresponds to the areas with a high warm bias is not surprising, as increased evaporation due to the heat will lead to a corresponding lack of available precipitation. This indicates that the increase in the available precipitation is closely related to the increase in the temperature. Taking the difference between percentiles in the tails of the distribution and the median reduces these biases in all season and region combinations.

MSLP in the large ensemble is compared to ERA–40 and it is found that there is some bias, mostly a shallower Equator to Northern Pole gradient in DJF and a positive bias in the Mediterranean in SON and DJF which are periods of storminess. Chapter 7 shows that this affects the total cyclone centre density in the Mediterranean, reducing it when compared to ERA–Interim. Examining the MSLP in terms of percentiles does not remove this bias.

Biases in the upper–level wind are shown to be small over Europe but quite large around the tropics.

To determine how the spread in the model is related to the bias, the RMSE is compared to the standard deviation for temperature and precipitation. For temperature, the RMSE and standard deviation have a good correlation, indicating that the model can capture the variability in temperature that is necessary to capture the high and low temperature events that are of interest in extreme event attribution. However for precipitation, there are instances where the RMSE and standard deviation do not correlate closely, indicating that the model may not be able to capture the full spread of the ensemble leading to an underrepresentation of very wet events or very dry events.

Chapter 5 deals with testing the model's validity on a seasonal timescale, using a well known method to construct bias maps over the 40 year period. Bias is the mean absolute error over a time period, and does not give an insight into how well the model represents the tails of the distribution of climate variables which manifest as extreme events in the observations. **Chapter 6** uses a technique from forecast validation to test whether the model can represent these extreme values on a daily timescale.

Chapter 6 calculates Q–Q plots for the very large ensemble of the HadAM3P climate between 1989 and 2009, comparing the ensemble to ERA–Interim. Even though the distribution of near–surface daily maximum temperature in the ensemble may not initially match the distribution from the ERA–Interim, after applying a bias correction there is good correspondence. The bias correction ensures that the ensemble distribution and reference distribution have the same mean The same also applies to daily minimum temperature. Daily accumulated precipitation is already well represented in the model, with only 3 out of 20 season–region combinations requiring bias correction. Similarly for daily mean MSLP only 3 season–region combinations require bias correction, the same 3 as require it for precipitation. Finally, for daily maximum windspeed the model consistently represents windspeed at too high a value for each percentile, when compared to ERA–Interim. However, this is remedied by correcting the mean of the windspeed to the mean of ERA–Interim windspeed. Applying the bias

correction, by adjusting the mean of the large ensemble to the mean of the ERA–Interim dataset reduces the bias in all variables for all region–season combinations to the point that there is no statistical distinction between ERA–Interim and the ensemble.

The ability to apply such a simple bias correction to all five daily climate variables analysed in this thesis and for the distribution of the variable to closely match the distribution from the reference dataset gives confidence that the model can be used to conduct attribution studies into extreme weather events involving temperature, precipitation, windspeed and MSLP.

Whereas Chapter 6 analysed the climate variable outputs from the large ensemble in an essentially one-dimensional sense, **Chapter 7** uses the feature identification and tracking software developed in Chapter 2 to analyse the weather systems in a two–dimensional manner. These weather systems are largely responsible for the variability seen in the near–surface temperature and precipitation variables over the UK and Ireland and so it is important to ensure that the model is matching the distribution of the observed variables for the right reasons, by ensuring that it can represent the position, frequency and magnitude of these systems correctly. To do this, the feature identification and tracking software is used to detect low–pressure systems in both the very large ensemble of HadAM3P and the ERA–Interim data. The track lengths in the large ensemble have a very similar distribution to those in the ERA–Interim data, as does the persistence. However, the intensity of the cyclones in the HadAM3P tracks show some differences. Here, the inability to match the maximum depression depth is possibly the most important shortcoming of the modelling of the low pressure systems, as it will affect the model's ability to produce realistic wind speeds. Therefore, an attribution study examining any increase in destructive wind storms will have to carefully check the wind speeds against observed wind speeds. However, there is a very good spatial correspondence between the ERA–Interim low-pressure system tracks and those in the large ensemble. Along with the good match between the distributions of track lengths, this is likely to be one of the main reasons that there is such a good correspondence between the distribution of precipitation in the ensemble and in the ERA–Interim data.

However, the model has mixed results when comparing the climate variables associated with the tracks in HadAM3P with those in ERA–Interim. Only the DJF season has good representation of minimum MSLP associated with a track whereas the other seasons are statistically distinguishable from ERA–Interim and have distributions that cannot represent the low values of MSLP associated with a storm track. Precipitation is well represented, except for the very high percentiles (90th+). Windspeeds associated with a cyclone are well represented in MAM and SON, poorly represented in JJA and have a positive bias in DJF, which may be resolved by a correction to the mean windspeed.

In conclusion, it is deemed that the model can represent low-pressure systems to a reasonable degree, enough to have confidence in using the model in a probabilistic event attribution study examining weather events that are related to low-pressure systems over Europe. However, some caveats must be made and they are that very high precipitation events are not well represented, the very low MSLP events are not well represented for seasons other than DJF and that the windspeeds are not well represented in JJA.

Overall, this thesis has shown that, with some bias correction, a large ensemble of a climate model can represent the distribution of observed climate variables that have been used for PEA studies previously, namely near-surface temperature and precipitation. The timescale on which an extreme event is predicted is very important and considering events within a seasonal timescale gives the most reliable results. Finally, it is important to validate the ability of a model to represent the atmospheric flow, in order to have confidence that the model is replicating the distributions of climate variables for the correct reasons. In this study, the low-pressure systems are well represented, with the exception of the maximum depression depth.

8.2 Future work : the Weather-@-Home experiments

8.2.1 Improvements to the model

One area of future work would be to improve the model used in the validation, and subsequent attribution studies. The model used in this thesis has biases in UK temperature, UK summer rainfall and in the maximum depression depth of low-pressure systems over the North Atlantic. In numerical weather prediction, it has been found that improving the model has improved the weather forecast, and so it seems likely that improving the climate model used in attribution studies would improve the representation of the weather-like events.

One way to improve the model is to increase its resolution. Stratton (2004) and Berckmans et al. (2013) discovered that, in the HadAM3 model (a precursor to HadAM3P), increasing the horizontal resolution improves the position and strength (in terms of maximum depth) of the low-pressure systems, when compared to ERA-40. However, increasing the resolution of a climate model increases the amount of computational time required, which can cause problems with volunteer distributed computing. Increasing the computational time can lead to an attrition of volunteers, as found by the climate*prediction*.net team (Christensen et al., 2005). The Weather@Home (W@H) project (Massey et al., 2015) circumnavigates this problem by coupling a free running global climate model (GCM), at N96 resolution, to a 50km regional climate model (RCM) that is relaxed at its perimeter to boundary conditions from the GCM. This setup, with a RCM over Europe and run with a climate scenario matching observations from 1960 to 2012, has been validated using the same methods as in Chapters 6 and 7 and published in Massey et al. (2015).

One potential problem with the W@H setup, in terms of model resolution, is that the high resolution regional model is driven at the boundaries by the same resolution model as used in this thesis (it is, in fact, the same model). The improvements to representation of low-pressure systems and high-pressure blocks gained by increasing the resolution of the model globally, as reported by Stratton (2004) and Berckmans et al. (2013) may be due to an improvement in modelling the atmospheric flow outside of the domain of the regional

model. If this is the case, then there will be no improvement within the RCM as the boundaries are still relaxed to a model which represents the flow in these areas erroneously. A comparison of the weather statistics and low-pressure systems between the driving GCM and the RCM should indicate whether the improved resolution within the domain improves the representation of these phenomena, or whether the resolution of the model needs to be increased outside the domain.

8.2.2 Improvements to the experimental design

Using the W@H setup, a number of experiments are currently being conducted, or are planned, which address some of the issues surrounding PEA. One of these problems is how to take into account uncertainty, both in model formulation and in climate scenario forcings.

Accounting for uncertainty in sea surface temperatures

Although there is an initial condition perturbation in the HadAM3P ensemble, there is no accounting for the uncertainty in the the forcing at the lower boundary by the sea-surface temperatures (SST). Observed SST datasets, such as HadISST (Rayner et al., 2003), will inevitably have some errors in the processes used to observe and assimilate the data, and this will lead to uncertainty in the values of the SSTs. Applying a perturbation to the SSTs, to form a large ensemble of climate model experiments, takes this uncertainty into account. HadISST v2 will present the SSTs as a 100 member ensemble (Dr. N. Rayner - personal comms.), in the same manner as the surface temperature is presented in CRUTEM4 (Jones et al., 2012), and so such an experiment will be conducted as soon as HadISST v2 is available.

Increasing the variability in sea surface temperatures

Woollings et al. (2010) show that by increasing the temporal resolution of the SSTs, the positioning of the North Atlantic storm track in high resolution climate models is improved, more closely matching the reanalysis data of ERA–40. The large ensemble experiment in this thesis uses the HadISST dataset for SSTs and sea ice, which has monthly temporal resolution (Rayner et al., 2003). In section 4.2.1, an attempt to recover some of the variability of

the SSTs was made, using the temporal interpolation method of Sheng and Zwiers (1998). However, although this algorithm did restore some of the inter-month variability, it could not reproduce the intra-month variability seen in satellite based observations of the SSTs, such as OSTIA (Donlon et al., 2012). The lack of this intra-month variability could be having a detrimental effect on the model's ability to represent the North Atlantic storm track.

One of the current active projects using the W@H setup is producing an initial condition ensemble forced at the lower boundary by the OSTIA SSTs. OSTIA provides daily resolution SSTs which, for the sake of reducing the size of download data to a client on the VDC network, are averaged to five-day mean values. Using the output from this ensemble, a storm tracking exercise could be performed with a view as to seeing whether the increase in temporal resolution of the SSTs improves the storm track statistics and position, relative to the HadISST forced ensemble computed in this thesis. Alternatively, a coupled atmosphere-ocean seasonal forecast model could also be used, although this would require an amount of model development.

Accounting for uncertainty in model parameterisations

Both the HadAM3P model described in chapter 4 and the W@H modelling setup described in section 8.2.1 use models with "standard physics", that is, the schemes which parameterise sub grid scale processes are not altered in any way. These parameterisations often have constant values associated with them, for example the speed of ice falling through a cloud, which are obtained from a combination of observations, experimental analysis and model tuning. In the climate*prediction*.net project it was found that by varying these parameter values, large ranges of future global mean temperatures are produced by the coupled atmosphere-ocean model (Stainforth et al., 2002; Rowlands et al., 2012).

An experiment to produce a large ensemble of perturbed parameter models can be conducted along the same lines as the initial condition ensemble experiment described in chapter 4. Once the parameters to be perturbed have been identified, and the ranges of their perturbations decided, a large ensemble is formed consisting of model runs with combinations of these perturbed parameters, the SSTs, atmospheric composition, an initial condition for a

particular year, and an initial condition perturbation. W@H has the ability to perturb the parameters in both the global model and regional model, in unison, so that a parameter value is equal between the global model and regional model, although they are corrected for length scales if necessary. A perturbed parameter ensemble (PPE) experiment of this kind has already been conducted using the W@H setup

It is not likely that the large temperature ranges that Stainforth et al. (2005) found will be produced by the W@H ensemble, as the lower boundary is always fixed. However, there may be some influence on both the low-pressure systems and on the surface variable. Stratton (2004) found that the improvement of the circulation in the eastern Atlantic on increasing the model resolution was not solely due to the increase in the resolution of the dynamical core of the model, but also an improvement in the parameterisation of the hydrological cycle. Knight et al. (2007) discovered that the greatest contribution to variability in a large perturbed parameter ensemble of climate models is due to perturbing the parameter values associated with clouds. By conducting this PPE experiment using the W@H, the influence of the parameters on the weather patterns and weather statistics will be determined.

8.2.3 Future attribution studies and extreme event sets

This thesis has performed an extensive appraisal of methods used to validate a climate model in order to use it in a probabilistic event attribution study. These methods, particularly the feature identification and tracking methods, could be used in the event attribution studies themselves. Attribution studies such as Pall et al. (2011), Otto et al. (2012) and Massey et al. (2012) have so far concentrated on examining surface variables over a particular domain in an essentially one dimensional manner. This has involved examining river runoff (Pall et al., 2011), area-averaging a region (Otto et al., 2012) or producing an analogue to an observed one-dimensional time series such as the Central England Temperature (Massey et al., 2012). Although useful studies, using the output from climate models in a one-dimensional way only indicates how the climate variables are changing and give no clue as to why they are changing.

Using a two-dimensional system, such as the storm tracking algorithms, would allow an investigation into how the atmospheric flow is changing under a warming climate, and provide some insight into potential changes in the mean state, and the variability around the mean state, of the atmospheric circulation. For example, an attribution experiment to determine whether the large rainfall over the UK and Ireland in Autumn and Winter 2012 could be conducted using two very large ensembles of climate models, one forced with an observed climate scenario, and the other with a "natural" climate scenario, constructed using the same method as in Pall (2006). Upon completion of the ensembles, via VDC, the feature identification and tracking algorithms would be performed upon every member of each ensemble. By constructing total cyclone centre density maps it can be determined whether there is a shift in the mean state of the North Atlantic storm track and whether this can be attributed to an increase in global temperature, due to anthropogenic greenhouse gas emissions,

Finally, the W@H setup provides a valuable framework to explore loss and damage to insurance portfolios, via catastrophe modelling. By using the large ensembles of climate models, the potential weather systems under a climate scenario can be more fully sampled than by relying on observations. This work is currently ongoing and a large ensemble is being produced for the years 1985 to 2012, forced using the OSTIA SSTs, and verified against a regional model run over the same period forced at the lateral boundary conditions (LBC) from the ERA Interim reanalysis dataset.

The feature identification and tracking algorithm will be applied to the members of the ensemble and, along with the tracks of features recovered, values of the field in the proximity of the feature will also be recovered. The same analysis will be performed on the regional model forced at the LBCs with ERA-Interim data. This will form an "event-set", containing a population of winter storms over Western Europe, and the variables associated with those storms. The frequency and magnitude of the systems associated with the features, in relation to the systems identified in the ERA-Interim forced model, will form the risk component of the dataset and the loss will be calculated by using the output of the event-set as the input to catastrophe models.

This project, and work conducted in this thesis, indicates one way forward for extreme event attribution studies, in using very large ensembles of climate models and an objective event identification system to both validate a climate model and to make attribution statements on how weather is being altered under our changing climate.

Bibliography

Añel, J. A., J. I. López-Moreno, F. E. L. Otto, S. Vicente-Serrano, N. Schaller, N. Massey, S. T. Buisán, and M. R. Allen, 2014: The extreme snow accumulation in the western Spanish Pyrenees during winter and spring 2013. *Bulletin of the American Meteorological Society*, **95**, S73–S76.

Adler, R. F., G. J. Huffman, A. Chang, R. Ferraro, P. Xie, J. Janowiak, B. Rudolf, U. Schneider, S. Curtis, D. Bolvin, A. Gruber, J. Susskind, and P. Arkin, 2003: The version 2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979-present). *Journal of Hydrometeorology*, **4**, 1147–1167.

Allen, M. R., 2003: Liability for climate change. *Nature*, **401**, 891–892.

Allen, M. R. and W. J. Ingram, 2002: Constraints on future changes in climate and the hydrologic cycle. *Nature*, **419**, 224–232.

Anderson, D., K. I. Hodges, and B. J. Hoskins, 2003: Sensitivity of feature-based analysis methods of storm tracks to the form of background field removal. *Monthly Weather Review*, **131**, 565–573.

Anderson, D. P., 2004: BOINC: A system for public-resource computing and storage. *5th IEEE / ACM International Workshop on Grid Computing*.

Anderson, D. P., C. Christensen, and B. Allen, 2006: Designing a runtime system for volunteer computing. *Supercomputing '06 (The International Conference for High Performance Computing, Networking, Storage and Analysis)*.

Anderson, D. P., J. Cobb, E. Korpela, M. Lebofsky, and D. Werthimer, 2002: SETI@home: an experiment in public-resource computing. *Communications of the ACM*, **45**(11), 56–61.

Andrews, D. G., 2000: *An Introduction to Atmospheric Physics*, Cambridge University Press.

Bardin, M. Y. and A. Polonsky, 2005: North atlantic oscillation and synoptic variability in the european-atlantic region in winter. *Izvestiya Atmospheric and Oceanic Physics*, **41**(2), 127–136.

Benestad, R. and D. Chen, 2006: The use of a calculus-based cyclone identification method for generating storm statistics. *Tellus A*, **58**(4), 473–486.

Bengtsson, L., K. I. Hodges, and N. Keenlyside, 2009: Will extratropical storms intensify in a warmer climate? *Journal of Climate*, **22**(9), 2276–2301.

Bengtsson, L., K. I. Hodges, and E. Roeckner, 2005: Storm tracks and climate change. *Journal of Climate*, **19**, 3518–3543.

Berckmans, J., T. Woollings, M.-E. Demory, P.-L. Vidale, and M. Roberts, 2013: Atmospheric blocking in a high resolution climate model: influences of mean state, orography and eddy forcing. *Atmospheric Science Letters*, **14**(1), 34–40.

Blackmon, M. L., 1976: A climatological spectral study of the 500mb geopotential height of the northern hemisphere. *Journal of the Atmospheric Sciences*, **33**, 1607–1623.

Blackmon, M. L., J. M. Wallace, N.-C. Lau, and S. L. Mullen, 1977: An observational study of the Northern Hemisphere wintertime circulation. *Journal of the Atmospheric Sciences*, **34**, 1040–1053.

Blake, E. S. and E. J. Gibney, 2011: The deadliest, costliest and most intense United States tropical cyclones from 1851 to 2010 (and other frequently requested hurricane facts). *Technical Memorandum NWS NHC–6 NOAA, Natural Hurricane Centre, Miami*.

Blender, R., K. Fraedrich, and F. Lunkeit, 1997: Identification of cyclone-track regimes in the North Atlantic. *Quarterly Journal of the Royal Meteorological Society*, **123**(727–741).

Blender, R. and M. Schubert, 2000: Cyclone tracking in different spatial and temporal resolutions. *Monthly Weather Review*, **128**, 377–384.

Box, G. E. P., 1976: Science and statistics. *Journal of the American Statistical Association*, **71**(356), 791–799.

Brohan, P., J. J. Kennedy, I. Harris, S. F. B. Tett, and P. D. Jones, 2006: Uncertainty estimates in regional and global observed temperature changes: a new dataset from 1850. *Journal of Geophysical Research: Atmospheres (1984–2012)*, **111**(D12).

Burden, R. L. and J. D. Faires, 2001: *Numerical Analysis*, 7th edition, Brooks / Cole.

Cattiaux, J. and P. Yiou, 2012: Contribution of atmospheric circulation to remarkable European temperatures of 2011. *Bulletin of the American Meteorological Society*, **93**(7), 1054–1057.

Chang, E. K. M., S. Lee, and K. L. Swanson, 2002: Storm track dynamics. *Journal of Climate*, **15**, 2163–2183.

Charlson, R. J., J. E. Lovelock, M. O. Andreae, and S. G. Warren, 1987: Oceanic phytoplankton, atmospheric sulphur, cloud albedo and climate. *Nature*, **326**, 655–661.

Christensen, C., T. Aina, and D. Stainforth, 2005: The challenge of volunteer computing with lengthy climate modelling simulations. *First International Conference on e-Science and Grid Technologies (e-Science 2005), 5-8 December 2005, Melbourne, Australia*.

Christidis, N. and P. Stott, 2012: Lengthened odds of the cold UK winter of 2010/11 attributable to human influence. *Bulletin of the American Meteorological Society*, **93**, 1060–1062.

Christidis, N. and P. A. Stott, 2015: Extreme rainfall in the united kingdom during winter 2013/14: The role of atmospheric circulation and climate change. *Bulletin of the American Meteorological Society*, **96**(12), S46–S50.

Christidis, N., P. A. Stott, and A. Ciavarella, 2014: The effect of anthropogenic climate change on the cold spring of 2013 in the United Kingdom. *Bulletin of the American Meteorological Society*, **95**, S79–S82.

Christy, J. R., W. B. Norris, R. W. Spencer, , and J. J. Hnilo, 2007: Tropospheric temperature change since 1979 from tropical radiosonde and satellite measurements. *Journal of Geophysical Research: Atmospheres (1984–2012)*, **112**(D6).

Collins, W., D. Stevenson, C. Johnson, and R. Derwent, 1997: Tropospheric ozone in a global-scale three-dimensional Lagrangian model and its response to NOx emission controls. *Journal of Atmospheric Chemistry*, **26**(3), 223–274.

Compo, G. P., J. S. Whitaker, P. D. Sardeshmukh, N. Matsui, R. J. Allan, X. Yin, B. E. Gleason, R. Vose, G. Rutledge, P. Bessemoulin, et al., 2011: The twentieth century reanalysis project. *Quarterly Journal of the Royal Meteorological Society*, **137**(654), 1–28.

Das, R., B. Qian, S. Raman, R. Vernon, J. Thompson, P. Bradley, S. Khare, M. Tyka, D. Bhat, D. Chivian, D. Kim, W. Sheffler, L. Malmstrm, A. Wollacott, C. Wang, I. Andre, and D. Baker, 2007: Structure prediction for CASP7 targets using extensive all-atom refinement with Rosetta@home. *Proteins*, **69 Suppl 8**, 118–28.

de Vries, H., R. van Westrhenen, and G. J. van Oldenborgh, 2013: The February 2012 European cold spell that didn't bring the Dutch another 11-City Tour. *Bulletin of the American Meteorological Society*, **94**, S26–S28.

Dee, D. P., S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. J. Geer, L. Haimberger, S. B. Healy, H. Hersbach, E. V. Hlm, L. Isaksen, P. Kllberg, M. Khler, M. Matricardi, A. P. McNally, B. M. Monge-Sanz, J.-J. Morcrette, B.-K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.-N. Thpaut, and F. Vitart, 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, **137**(656), 553–597.

D'Ippoliti, D., P. Michelozzi, C. Marino, F. DeDonato, B. Menne, K. Katsouyanni, U. Kirchmayer, A. Analitis, M. Medina-Ramón, A. Paldy, et al., 2010: The impact of heat waves on mortality in 9 European cities: results from the EuroHEAT project. *Environ Health*, **9**, 37.

Dong, B., R. Sutton, and L. Shaffrey, 2014: The 2013 hot, dry, summer in western Europe. *Bulletin of the American Meteorological Society*, **95**, S62–S66.

Dong, B., R. Sutton, and T. Woollings, 2013: The extreme European summer 2012. *Bulletin of the American Meteorological Society*, **94**, S28–S32.

Donlon, C. J., M. Martin, J. Stark, J. Roberts-Jones, E. Fiedler, and W. Wimmer, 2012: The operational sea surface temperature and sea ice analysis (OSTIA) system. *Remote Sensing of Environment*, **116**, 140–158.

Feser, F., M. Barcikowska, S. Haeseler, C. Lefebvre, M. Schubert-Frisius, M. Stendel, H. V. Storch, and M. Zahn, 2015: Hurricane gonzalo and its extratropical transition to a strong European storm. *Bulletin of the American Meteorological Society*, **96**, S51–S54.

Field, C., V. Barros, T. Stocker, D. Qin, D. Dokken, K. Ebi, M. Mastrandrea, K. Mach, G. Plattner, S. Allen, M. Tignor, and P. Midgley, 2012: *IPCC 2012: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change*, Cambridge University Press.

Finkel, R. A. and J. L. Bentley, 1974: Quad trees: A data structure for retrieval on composite keys. *Acta Informatica*, **4**, 1–9.

Fischer, E. M. and C. Schar, 2010: Consistent geographical patterns of changes in high-impact european heatwaves. *Nature Geoscience*, **3**(6), 398–403.

Foley, J. D., A. van Dam, S. K. Feiner, and J. F. Hughes, 1990: *Computer Graphics, Principles and Practice*, Addison Wesley.

Frame, D. J., T. Aina, C. M. Christensen, N. E. Fall, S. H. E. Knight, C. Piani, S. M. Rosier, K. Yamazaki, Y. Yamazaki, and M. Allen, 2009: The climate*prediction*.net BBC climate change experiment: Design of the coupled model ensemble. *Philosophical Transactions of The Royal Society*, **367**, 855–870.

Frich, P., L. Alexander, P. Della-Marta, B. Gleason, M. Haylock, A. K. Tank, and T. Peterson, 2002: Observed coherent changes in climatic extremes during the second half of the twentieth century. *Climate Research*, **19**, 193–212.

Gibson, J. K., P. Kllberg, S. Uppala, A. Hernandez, A. Nomura, and E. Serrano, 1997: ERA description. *ECMWF Re-Analysis Project Report Series No. 1*, 71pp.

Greeves, C. Z., V. D. Pope, R. A. Stratton, and G. M. Martin, 2007: Representation of Northern Hemisphere winter storm tracks in climate models. *Climate Dynamics*, **28**, 683–702.

Hanley, J. and R. Caballero, 2012: Objective identification and tracking of multicentre cyclones in the era-interim reanalysis dataset. *Quarterly Journal of the Royal Meteorological Society*, **138**(664), 612–625.

Hansen, J., R. Ruedy, M. Sato, and K. Lo, 2010: Global surface temperature change. *Reviews of Geophysics*, **48**(4).

Haylock, M. R., 2011: European extra-tropical storm damage risk from a multi-model ensemble of dynamically-downscaled global climate models. *Natural Hazards and Earth System Science*, **11**(10), 2847–2857.

Hewson, T. D. and H. A. Titley, 2010: Objective identification, typing and tracking of the complete life-cycles of cyclonic features at high spatial resolution. *Meteorological Applications*, **17**(3), 355–381.

Hodges, K., 2008: Confidence intervals and significance tests for spherical data derived from feature tracking. *Monthly Weather Review*, **136**(5), 1758–1777.

Hodges, K. I., 1994: A general method for tracking analysis and its application to meteorological data. *Monthly Weather Review*, **122**, 2573–2586.

—, 1995: Feature tracking on the unit sphere. *Monthly Weather Review*, **123**, 3458–3465.

—, 1996: Spherical nonparameteric estimators applied to the UGAMP model integration for AMIP. *Monthly Weather Review*, **124**, 2914–2932.

—, 1999: Adaptive constraints for feature tracking. *Monthly Weather Review*, **127**, 1362–1373.

Hodges, K. I., B. J. Hoskins, J. Boyle, and C. Thorncroft, 2003: A comparison of recent reanalysis datasets using objective feature tracking: Storm tracks and tropical easterly waves. *Monthly Weather Review*, **131**, 2012–2037.

Hodges, K. I., R. Lee, and L. Bengtsson, 2011: A comparison of extratropical cyclones in recent reanalyses era-interim, nasa merra, ncep cfsr, and jra-25. *Journal of Climate*, **24**(18), 4888–4906.

Holton, J. R., 2004: *An Introduction to Dynamic Meteorology*, 4th edition, Elsevier Academic Press.

Hoskins, B. J. and K. I. Hodges, 2002: New perspectives on the Northern Hemisphere winter storm tracks. *Journal of Atmospheric Sciences*, **59**, 1041–1061.

—, 2005: A new perspective on Southern Hemisphere storm tracks. *Journal of Climate*, **18**, 4108–4129.

Houghton, J., Y. Ding, D. Griggs, M. Noguer, P. van der Linden, X. Dai, K. Maskell, and C. Johnson, 2001: *IPCC: 2001 Third Assessment Report: Climate Change 2001: The Scientific Basis.*, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Huntingford, C., T. Marsh, A. A. Scaife, E. J. Kendon, J. Hannaford, A. L. Kay, M. Lockwood, C. Prudhomme, N. S. Reynard, S. Parry, J. A. Lowe, J. A. Screen, H. C. Ward, M. Roberts, P. A. Stott, V. A. Bell, M. Bailey, A. Jenkins, T. Legg, F. E. L. Otto, N. Massey, N. Schaller, J. Slingo, and M. R. Allen, 2014: Potential influences on the United Kingdom's floods of winter 2013/14. *Nature Climate Change*, **4**(9), 769–777.

IPCC, 1990: *Climate Change: The IPCC Scientific Assessment*, Cambridge University Press, Cambridge, Great Britain, New York, NY, USA and Melbourne, Australia.

Jacob, D., J. Petersen, B. Eggert, A. Alias, O. Christensen, L. Bouwer, A. Braun, A. Colette, M. Dqu, G. Georgievski, E. Georgopoulou, A. Gobiet, L. Menut, G. Nikulin, A. Haensler, N. Hempelmann, C. Jones, K. Keuler, S. Kovats, N. Krner, S. Kotlarski, A. Kriegsmann, E. Martin, E. van Meijgaard, C. Moseley, S. Pfeifer, S. Preuschmann, C. Radermacher, K. Radtke, D. Rechid, M. Rounsevell, P. Samuelsson, S. Somot, J.-F. Soussana, C. Teichmann, R. Valentini, R. Vautard, B. Weber, and P. Yiou, 2014: Euro-cordex: new high-resolution climate change projections for european impact research. *Regional Environmental Change*, **14**(2), 563–578.

Jennings, G. A., 1994: *Modern Geometry With Applications*, Springer.

Jolliffe, I. T. and D. B. Stephenson, eds., 2003: *Forecast Verification: a Practitioner's Guide in Atmospheric Science*, John Wiley & Sons.

Jones, P. D., D. H. Lister, T. J. Osborn, C. Harpham, M. Salmon, and C. P. Morice, 2012: Hemispheric and large-scale land surface air temperature variations: An extensive revision and an update to 2010. *Journal of Geophysical Research: Atmospheres (1984–2012)*, **117**(D5).

Jones, R., M. Noguer, D. Hassell, D. Hudson, S. Wilson, G. Jenkins, and J. Mitchell, 2004: *Generating high resolution climate change scenarios using PRECIS*, Met Office Hadley Centre, Exeter, UK, 40pp pp.

Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, A. Leetmaa, R. Reynolds, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Ropelewski, J. Wang, R. Jenne, and D. Joseph, 1996: The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, **77**, 1963–1710.

Kam, J., T. R. Knutson, F. Zeng, and A. T. Wittenberg, 2015: Record annual mean warmth over europe, the northeast pacific, and the northwest atlantic during 2014: Assessment of anthropogenic influence. *Bulletin of the American Meteorological Society*, **96**(12), S61–S65.

Kerr, Y., J. Lagouarde, and J. Imbernon, 1992: Accurate land surface temperature retrieval from AVHRR data with use of an improved split window algorithm. *Remote Sensing of Environment*, **41**(2), 197–209.

Kew, S. F., M. Sprenger, and H. C. Davies, 2010: Potential vorticity anomalies of the lowermost stratosphere: A 10-yr winter climatology. *Monthly Weather Review*, **138**(4), 1234–1249.

King, A. D., S. C. Lewis, S. E. Perkins, L. V. Alexander, M. G. Donat, D. J. Karoly, and M. T. Black, 2013: Limited evidence of anthropogenic influence on the 2011-12 extreme rainfall over southeast Australia. *Bulletin of the American Meteorological Society*, **94**, S55–S58.

Knight, C. G., S. H. E. Knight, N. Massey, T. Aina, C. Christensen, D. J. Frame, J. A. Kettleborough, A. Martin, S. Pascoe, B. Sanderson, D. A. Stainforth, , and M. R. Allen, 2007: Association of parameter, software and hardware variation with large scale behavior across 57,000 climate models. *PNAS*, **104**(30), 12259–12264.

Kreyszig, E., 1993: *Advanced Engineering Mathematics*, seventh edition, John Wiley & Sons.

Krivova, N., L. Balmaceda, and S. Solanki, 2007: Reconstruction of solar total irradiance since 1700 from the surface magnetic flux. *Astronomy & Astrophysics*, **467**, 335–346.

Kunszt, P. Z., A. S. Szalay, and A. R. Thakar, 2001: The hierarchical triangular mesh, *Mining the sky*, Springer, 631–637.

Lawrimore, J. H., M. J. Menne, B. E. Gleason, C. N. Williams, D. B. Wuertz, R. S. Vose, and J. Rennie, 2011: An overview of the Global Historical Climatology Network monthly mean temperature data set, version 3. *Journal of Geophysical Research: Atmospheres (1984–2012)*, **116**(D19).

Leckebusch, G. C., B. Koffi, U. Ulbrich, J. G. Pinto, T. Spangehl, and S. Zacharias, 2006: Analysis of frequency and intensity of european winter storm events from a multi-model perspective, at synoptic and regional scales. *Climate Research*, **31**(1), 59–74.

Lee, M. and H. Samet, 2000: Navigating through triangle meshes implemented as linear quadtrees. *ACM Transactions on Graphics (TOG)*, **19**(2), 79–121.

Lionello, P., F. Dalan, and E. Elvini, 2002: Cyclones in the mediterranean region: the present and the doubled co˜ 2 climate scenarios. *Climate Research*, **22**(2), 147–159.

Loomis, E., 1885: Areas of low pressure, their form, magnitude, direction and velocity of movement, *Contributions to Meteorology*, Tuttle, Moorehouse, and Taylor, eds., New Haven CT, 67.

Martin, J. E., 2006: *Mid–Latitude Atmospheric Dynamics: A First Course*, John Wiley & Sons.

Massey, N., 2012: Feature tracking on the hierarchical equal area triangular mesh. *Computers and Geosciences*, **44**, 42–51.

Massey, N., T. Aina, C. Rye, F. E. L. Otto, S. Wilson, R. G. Jones, and M. R. Allen, 2012: Have the odds of warm November temperatures and of cold December temperatures in Central England changed? *Bulletin of the American Meteorological Society*, **93**, 1041–1067.

Massey, N., R. Jones, F. E. L. Otto, T. Aina, S. Wilson, J. M. Murphy, D. Hassell, Y. H. Yamazaki, and M. R. Allen, 2015: weather@home-development and validation of a very large ensemble modelling system for probabilistic event attribution. *Quarterly Journal of the Royal Meteorological Society*, **141**(690), 1528–1545.

Mera, R., N. Massey, D. E. Rupp, P. Mote, M. Allen, and P. C. Frumhoff, 2015: Climate change, climate justice and the application of probabilistic event attribution to summer heat extremes in the california central valley. *Climatic Change*, **133**(3), 427–438.

Mitchell, T. D. and P. D. Jones, 2005: An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *International Journal of Climatology*, **25**, 693–712.

Murphy, J. M., D. M. H. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, M. Collins, and D. A. Stainforth, 2004: Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, **430**, 768–772.

Murray, R. J. and I. Simmonds, 1991: A numerical scheme for tracking cyclone centres from digital data. Part I:development and operation of the scheme. *Australian Meteorological Magazine*, **39**, 155–166.

Nakicenovic, N. and R. Swart, 2000: *IPCC 2000: Emissions Scenarios*, Cambridge University Press, UK.

Neu, U., M. G. Akperov, N. Bellenbaum, R. Benestad, R. Blender, R. Caballero, A. Cocozza, H. F. Dacre, Y. Feng, K. Fraedrich, et al., 2013: IMILAST: A community effort to intercompare extratropical cyclone detection and tracking algorithms. *Bulletin of the American Meteorological Society*, **94**(4), 529–547.

New, M., M. Hulme, and P. D. Jones, 1999: Representing twentieth century space-time climate variability. Part 1: Development of a 1961-90 mean monthly terrestrial climatology. *Journal of Climate*, **12**, 829–856.

O'Gorman, P. A., 2015: Precipitation extremes under climate change. *Current Climate Change Reports*, **1**(2), 49–59.

Otto, F. E., N. Massey, G. Oldenborgh, R. Jones, and M. Allen, 2012: Reconciling two approaches to attribution of the 2010 Russian heat wave. *Geophysical Research Letters*, **39**(4).

Pall, P., 2006: *Constraints on, and attribution of, changes in extreme precipitation under climate change*, Ph.D. thesis, University of Oxford.

Pall, P., T. Aina, D. A. Stone, P. A. Stott, T. Nozawa, A. G. Hilberts, D. Lohmann, and M. R. Allen, 2011: Anthropogenic greenhouse gas contribution to flood risk in England and Wales in autumn 2000. *Nature*, **470**(7334), 382–385.

Pall, P., M. Allen, and D. Stone, 2006: climate*prediction*.net seasonal attribution project. *http://attribution.cpdn.org/about*.

Parker, H. R., R. J. Cornforth, E. Boyd, R. James, F. E. Otto, and M. R. Allen, 2015: Implications of event attribution for loss and damage policy. *Weather*, **70**(9), 268–273.

Perkins, S. E., S. C. Lewis, A. D. King, and L. V. Alexander, 2014: Increased simulated risk of the hot Australian summer of 2012/13 due to anthropogenic activity as measured by heat wave frequency and intensity. *Bulletin of the American Meteorological Society*, **95**, S34–S37.

Pinto, J. G., T. Spangehl, U. Ulbrich, and P. Speth, 2005: Sensitivities of a cyclone detection and tracking algorithm: individual tracks and climatology. *Meteorologische Zeitschrift*, **14**(6), 823–838.

Pope, V., M. Gallani, P. Rowntree, and R. Stratton, 2000: The impact of new physical parameterizations in the Hadley Centre climate model: HadAM3. *Climate Dynamics*, **16**, 12–146.

Pope, V., J. Pamment, D. Jackson, and A. Slingo, 2001: The representation of water vapor and its dependence on vertical resolution in the Hadley Centre climate model. *Journal of Climate*, **14**, 30653085.

Pope, V. and R. Stratton, 2002: The processes governing horizontal resolution sensitivity in a climate model. *Climate Dynamics*, **19**, 211–236.

Post, F. H., B. Vrolijk, H. Hauser, R. S. Laramee, and H. Doleisch, 2003: The state of the art in flow visualisation: Feature extraction and tracking. *Computer Graphics Forum*, **22**(4), 775–792.

Prior, J. and M. Beswick, 2008: The exceptional rainfall of 20 July 2007. *Weather*, **63**(9), 261–267.

Rácz, Z. and R. K. Smith, 1999: The dynamics of heat lows. *Quarterly Journal of the Royal Meteorological Society*, **125**(553), 225–252.

Raible, C., P. Della-Marta, C. Schwierz, H. Wernli, and R. Blender, 2008: Northern hemisphere extratropical cyclones: A comparison of detection and tracking methods and different reanalyses. *Monthly Weather Review*, **136**(3), 880–897.

Rayner, N. A., P. Brohan, D. E. Parker, C. K. Folland, J. J. Kennedy, M. Vanicek, T. J. Ansell, and S. F. B. Tett, 2006: Improved analyses of changes and uncertainties in sea surface temperature measured in situ since the mid-nineteenth century: The HadSST2 dataset. *Journal of Climate*, (19), 446–469.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research: Atmospheres (1984–2012)*, **108**(D14).

Renka, R. J., 1997: Algorithm 772: STRIPACK: Delaunay triangulation and Voronoi diagram on the surface of a sphere. *ACM Transactions on Mathematical Software*, **23**(3), 416–434.

Roberts, J. F., A. J. Champion, L. C. Dawkins, K. I. Hodges, L. C. Shaffrey, D. B. Stephenson, M. A. Stringer, H. E. Thornton, and B. D. Youngman, 2014: The XWS open access catalogue of extreme European windstorms from 1979 to 2012. *Natural Hazards and Earth System Science*, **14**(9), 2487–2501.

Rowlands, D., D. Frame, D. Ackerley, T. Aina, B. Booth, C. Christensen, M. Collins, N. Faull, C. E. Forest, B. Grandey, E. Gryspeerdt, E. Highwood, W. Ingram, S. Knight, A. Lopez, N. Massey, F. McNamara, N. Meinshausen, C. Piani, S. Rosier, B. Sanderson, L. Smith, D. Stone, M. Thurston, K. Yamazaki, H. Yamazaki, and M. Allen, 2012: Broad range of 2050 warming from an observationally constrained large climate model ensemble. *Nature Geoscience*, **5**(4), 256–260.

Rudeva, I. and S. K. Gulev, 2007: Climatology of cyclone size characteristics and their changes during the cyclone life cycle. *Monthly Weather Review*, **135**(7), 2568–2587.

Russo, S., A. Dosio, R. G. Graversen, J. Sillmann, H. Carrao, M. B. Dunbar, A. Singleton, P. Montagna, P. Barbola, and J. V. Vogt, 2014: Magnitude of extreme heat waves in present climate and their projection in a warming world. *Journal of Geophysical Research: Atmospheres*, **119**(22), 12,500–12,512.

Salari, V. and I. Sethi, 1990: Feature point correspondence in the presence of occlusion. *IEPM*, **12**, 87–91.

Samet, H., 1990: Applications of spatial data structures: Computer graphics, image processing, and gis. *Addison-Wesley, R eading, MA*, **1**, 99.

Sato, M., J. Hansen, M. McCormick, and J. Pollack, 1993: Stratospheric aerosol optical depth, 1850-1990. *Journal of Geophysical Research: Atmospheres (1984–2012)*, **98**, 22987–22994.

—, 2011: Forcings in GISS climate model : Stratospheric aerosol optical thickness, `http://data.giss.nasa.gov/modelforce/strataer//`.

Scaife, A. A., T. Woollings, J. Knight, G. Martin, and T. Hinton, 2010: Atmospheric blocking and mean biases in climate models. *Journal of Climate*, **23**(23), 6143–6152.

Schaller, N., A. L. Kay, R. Lamb, N. R. Massey, G. J. van Oldenborgh, F. E. L. Otto, S. N. Sparrow, R. Vautard, P. Yiou, I. Ashpole, A. Bowery, S. M. Crooks, K. Haustein, C. Huntingford, W. J. Ingram, R. G. Jones, T. Legg, J. Miller, J. Skeggs, D. Wallom, A. Weisheimer, S. Wilson, P. A. Stott, and M. R. Allen, 2016: Human influence on climate in the 2014 Southern England winter floods and their impacts. *Accepted for publication in Nature Climate Change*.

Schaller, N., F. E. L. Otto, G. J. van Oldenborgh, N. R. Massey, S. Sparrow, and M. R. Allen, 2014: The heavy precipitation event of May-June 2013 in the upper Danube and Elbe basins. *Bulletin of the American Meteorological Society*, **95**, S69–S72.

Sedgewick, R., 1992: *Algorithms in C++*, Addison-Wesley.

Sheng, J. and F. Zwiers, 1998: An improved scheme for time-dependent boundary conditions in atmospheric general circulation models. *Climate Dynamics*, **14**, 609–613.

Simmons, A. J., P. D. Jones, V. da Costa Bechtold, A. C. M. Beljaars, P. W. Kallberg, S. Saarinen, S. M. Uppala, P. Viterbo, and N. Wedi, 2004: Comparison of trends and low-frequency variability in CRU, ERA-40, and NCEP/NCAR analyses of surface air temperature. *Journal of Geophysical Research: Atmospheres (1984–2012)*, **109**(D24).

Sinclair, M. R., 1994: An objective cyclone climatology for the Southern Hemisphere. *Monthly Weather Review*, **122**, 2239–2256.

—, 1997: Objective identification of cyclones and their circulation intensity, and climatology. *Weather and Forecasting*, **12**, 595–612.

Sinclair, M. R. and I. G. Watterson, 1999: Objective assessment of extratropical weather systems in simulated climates. *Journal of Climate*, **12**, 3467–3485.

Sinnott, C. W., 1984: Virtues of the haversine. *Sky and Telescope*, **68**(2), 159.

Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K. Averyt, M. Tignor, and H. Miller, 2007: *IPCC 2007: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Sparrow, S., C. Huntingford, N. Massey, and M. Allen, 2013: The use of very large atmospheric model ensemble to assess potential anthropogenic influence on the UK Summer 2012 high rainfall totals. *Bulletin of the American Meteorology Society*, **94**, S36–S41.

Spencer, R. W., J. R. Christy, and N. C. Grody, 1990: Global atmospheric temperature monitoring with satellite microwave measurements: Method and results 1979-84. *Journal of Climate*, **3**, 1111–1128.

Stainforth, D., J. Kettleborough, M. Allen, M. Collins, and A. Heaps, 2002: Climateprediction.com: Distributed computing for public interest modelling research. *Computing in Science and Engineering*, **4**(3).

Stainforth, D. A., T. Aina, C. Christensen, M. Collins, N. Faull, D. J. Frame, J. A. Kettleborough, S. Knight, A. Martin, J. M. Murphy, C. Piani, D. Sexton, L. A. Smith, R. A. Spicer, A. J. Thorpe, and M. R. Allen, 2005: Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature*, **433**, 403–406.

Stott, P. A., M. Allen, N. Christidis, R. M. Dole, M. Hoerling, C. Huntingford, P. Pall, J. Perlwitz, and D. Stone, 2013: Attribution of weather and climate-related events, *Climate Science for Serving Society*, Springer, 307–337.

Stott, P. A., N. Christidis, F. E. Otto, Y. Sun, J.-P. Vanderlinden, G. J. van Oldenborgh, R. Vautard, H. von Storch, P. Walton, P. Yiou, and F. W. Zwiers, 2016: Attribution

of extreme weather and climate-related events. *Wiley Interdisciplinary Reviews: Climate Change*, **7**(1), 23–41.

Stott, P. A., D. A. Stone, and M. R. Allen, 2004: Human contribution to the European heatwave of 2003. *Nature*, **432**, 610–614.

Stratton, R. A., 2004: Report on aspects of variability on high-resolution versions of HadAM3, `http://www.metoffice.com/research/hadleycentre/pubs/HCTN/HCTN\_53.pdf`.

Tans, P. and R. Keeling, 2011: Trends in carbon dioxide, `http://www.esrl.noaa.gov/gmd/ccgg/trends/mlo.html#mlo_full`.

Taylor, K. E., 1986: An analysis of the biases in traditional cyclone frequency maps. *Monthly weather review*, **114**(8), 1481–1490.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, **93**(4), 485–498.

Teanby, N. A., 2006: An icosahedron-based method for even binning of globally distributed remote sensing data. *Computers and Geosciences*, **32**(9), 1442 – 1450.

Tett, S. F. B., K. Deans, E. Mazza, and J. Mollard, 2013: Are recent wet northwestern European summers a response to sea ice retreat? *Bulletin of the American Meteorological Society*, **94**, S32–S35.

Trigo, I. F., 2006: Climatology and interannual variability of storm-tracks in the euro-atlantic sector: a comparison between era-40 and NCEP/NCAR reanalyses. *Climate Dynamics*, **26**, 127–143.

Trigo, R., J. A. nel, D. Barriopedro, R. Garcia-Herrera, L. Gimeno, R. Nieto, R. Castillo, M. Allen, and N. Massey, 2013: The record winter drought of 2011-2012 in the iberian peninsula. *Bulletin of the American Meteorological Society*, **94**(9), S41–S45.

Twomey, S., 1974: Pollution and the planetary albedo. *Atmospheric Environment (1967)*, **8**(12), 1251–6.

Uppala, S. M., P. W. Kållberg, A. J. Simmons, U. Andrae, V. D. C. Bechtold, M. Fiorino, J. K. Gibson, J. Haseler, A. Hernandez, G. A. Kelly, X. Li, K. Onogi, S. Saarinen, N. Sokka, R. P. Allan, E. Andersson, K. Arpe, M. A. Balmaseda, A. C. M. Beljaars, L. V. D. Berg, J. Bidlot, N. Bormann, S. Caires, F. Chevallier, A. Dethof, M. Dragosavac, M. Fisher, M. Fuentes, S. Hagemann, E. Hólm, B. J. Hoskins, L. Isaksen, P. A. E. M. Janssen, R. Jenne, A. P. McNally, J. Mahfouf, J. Morcrette, N. A. Rayner, R. W. Saunders, P. Simon, A. Sterl, K. E. Trenberth, A. Untch, D. Vasiljevic, P. Viterbo, and J. Woollen, 2005: The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*, **131**, 2961–3012.

van Oldenborgh, G. J., F. E. L. Otto, K. Haustein, and H. Cullen, 2016: Climate change increases the probability of heavy rains like those of Storm Desmond in the UK – an event attribution study in near-real time. *Submitted for publication in Hydrology and Earth System Sciences*.

van Oldenborgh, G. J., A. van Urk, and M. Allen, 2012: The absence of a role of climate change in the 2011 Thailand floods. *Bulletin of the American Meteorological Society*, **93**, 1047–1049.

Vautard, R., G. V. Oldenborgh, S. Thao, B. Dubuisson, G. Lenderink, A. Ribes, S. Planton, J. M. Soubeyroux, and P. Yiou, 2015: Extreme fall 2014 precipitation in the cevennes mountains. *Bulletin of the American Meteorological Society*, **96**(12), S56–S60.

von Storch, H., F. Feser, S. Haeseler, C. Lefebvre, and M. Stendel, 2014: A violent midlatitude storm in northern Germany and Denmark, 28 October 2013. *Bulletin of the American Meteorological Society*, **95**, S76–S78.

Wang, X. L., V. R. Swail, and F. W. Zwiers, 2006: Climatology and changes of extratropical cyclone activity: Comparison of ERA–40 with NCEP–NCAR reanalysis for 1958-2001. *Journal of Climate*, **19**(13), 3145–3166.

Webster, S., A. R. Brown, D. R. Cameron, and C. P. Jones, 2003: Improvements to the representation of orography in the met office unified model. *Quarterly Journal of the Royal Meteorological Society*, **129**(591), 1989–2010.

Wernli, H. and C. Schwierz, 2006: Surface cyclones in the era40 data set (1958-2001). part i: novel identification method and global climatology. *Journal of the Atmospheric Sciences*, **63**(10), 2486—2507.

Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*, third edition, Academic Press.

Woollings, T., B. Hoskins, M. Blackburn, D. Hassell, and K. Hodges, 2010: Storm track sensitivity to sea surface temperature resolution in a regional atmosphere model. *Climate Dynamics*, **35**, 341–354.

Yiou, P. and J. Cattiaux, 2013: Contribution of atmospheric circulation to wet North European summer precipitation of 2012. *Bulletin of the American Meteorological Society*, **94**, S39–S41.

—, 2014: Contribution of atmospheric circulation to wet southern European winter of 2013. *Bulletin of the American Meteorological Society*, **95**, S66–S69.

Zwiers, F. W., X. Zhang, and Y. Feng, 2011: Anthropogenic influence on long return period daily temperature extremes at regional scales. *Journal of Climate*, **24**(3), 881–892.