Meteorological Applications

RESEARCH ARTICLE OPEN ACCESS

Balancing Informativity and Predictability in Circulation Type Forecasts: A Case Study of Energy Demand in Great Britain

Kristian Strommen[1] | Hannah M. Christensen[1] | Hannah C. Bloomfield[2]

[1]Department of Physics, University of Oxford, Oxford, UK | [2]School of Engineering, University of Newcastle, Newcastle upon Tyne, UK

Correspondence: Kristian Strommen (kristian.strommen@physics.ox.ac.uk)

Received: 11 March 2025 | **Revised:** 2 July 2025 | **Accepted:** 8 July 2025

Funding: This work was supported by the Natural Environment Research Council grant number (NE/P018238/1), the Grant Agreement (No. 101081383) funded by the European Union, the UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee, grant number 10049639 and the Newcastle University.

Keywords: circulation types | energy demand | forecasting | predictability | weather regimes

ABSTRACT

Weather regimes and weather patterns, here jointly referred to as circulation types, are used to generate forecasts for a variety of applications, such as energy demand and flood risk. However, there are usually many different choices available for precisely which circulation types to use. Ideally, one would like to use circulation types that are both highly informative for the application and also highly predictable, but in practice, there is often a tradeoff between informativity and predictability. We present a simple, general framework for how to construct a circulation type forecast that optimally balances these factors by segueing between different choices of circulation types at different lead times based on information-theoretic considerations. As an example, we apply this framework to the case of forecasting energy demand in Great British winters. We compare a set of 30 weather patterns produced by the UK Met Office with the much simpler two-state framework consisting of a positive and negative North Atlantic Oscillation (NAO) regime and show how to optimally combine the two across a winter season.

1 | Introduction

In many meteorological applications in the mid latitudes, it has become popular to make use of weather regimes or weather patterns. They are generally defined as a small number of patterns of atmospheric pressure over a particular region which explain a large percentage of the atmospheric variability and can be thought of as making up a low-dimensional truncation of the atmospheric flow. The atmosphere is thereby envisaged as transitioning discretely between these pressure patterns, which are assumed to be stationary (Michelangeli and Vautard 1995). There are many ways to construct these, ranging from the 'classical' approach based on clustering pressure data (Michelangeli and Vautard 1995; Neal et al. 2016), to more application-focused constructions (Bloomfield et al. 2019); see (Hannachi et al. 2017) for more constructions. Following past conventions (Neal et al. 2016), 'weather regimes' refer to situations where the patterns are continental scale and more persistent, whilst 'weather patterns' refer to more national scale patterns that vary more frequently. In this work, we will be constructing forecasts that combine the two, so for simplicity, we will use the term 'circulation type' (CT) to jointly refer to both kinds of patterns.

By assessing how the occurrence of each CT is related to one's application, one can carry out simple forecasts or classification analysis. One prominent example application is energy, since renewable energy availability and energy demand depend sensitively on weather (Grams et al. 2017; Bloomfield et al. 2019; Garrido-Perez et al. 2020; Mockert et al. 2023; Millin et al. 2024). For example, national electricity demand in the winter months

in Europe is strongly influenced by the temperature, which determines how much people heat their homes. Other examples include analysis of heatwaves (Rouges et al. 2023), drought (Richardson et al. 2017; Lavaysse et al. 2018), floods (Hendry et al. 2019) and other hydrological variables (Chang et al. 2023). These applications obviously rely on the existence of skilful CT forecasts, and there are now many studies showing that such skill does exist in many situations and across multiple timescales (Ferranti et al. 2015; Bloomfield et al. 2019; Grams et al. 2020; Osman et al. 2023).

The circulation type approach is appealing for at least two reasons. Firstly, at very extended lead times, weather forecasts struggle to capture predictable country level (or smaller) variability of surface variables important for climate impacts, such as temperature, wind speeds, and precipitation. Circulation types are often some form of frequently recurring synoptic scale patterns, and these are often the components of the midlatitude flow that are most predictable (Palmer 1999), both due to their intrinsically slower time scales and the fact that they are often responsive to tropical sea-surface temperature (SST) anomalies (Cassou 2008). Each CT also has its own 'signature' of surface impacts which can be estimated using observations. CT forecasts can thereby help isolate predictable signals in these surface variables on the extended sub-seasonal to seasonal time scales where regular weather forecasts become noisy. Secondly, they are a more accessible way for end users, who are often not trained meteorologists, to make use of weather forecasts. The small number of CTs allows users to focus on the part of the climate system that matters to them and build up intuition about the synoptic flow.

However, because there is no agreed upon definition of what a weather regime or weather pattern really is, a wide variety of different frameworks are present in the literature (Strommen et al. 2022). For example, for the Euro-Atlantic circulation relevant to the UK and Europe, there exist at least nine different CT frameworks, with the number of CTs in each ranging from 2 to 30; here the simplest case of two CTs corresponds to using the positive and negative phase of the North Atlantic Oscillation (NAO) (Hurrell et al. 2003; Wanner et al. 2001) as one's weather regimes, whilst the case of 30 CTs corresponds to the 30 weather patterns used in the UK Met Offices 'Decider' product (Neal et al. 2016, 2024), which we will refer to as the MO30 CTs (or just MO30) for short. This diversity presents an obvious practical problem for users: which regime framework is best suited for their application?

In a perfect world one would like to choose CTs which are both maximally informative for the application (e.g., national energy demand) and maximally predictable. However, we hypothesise that in practice a trade-off between informativity and predictability is often inevitable. Indeed, if one tries to identify a small number of large-scale pressure patterns that are in some sense intrinsic to the atmosphere, then there is no reason a priori for these to tell you much about electricity demand or flood risks in a particular country, as these often depend sensitively on very regional surface level conditions (Bloomfield et al. 2019). One way to try to increase the informativity of one's CTs is to construct them based on the target variable itself, as in the 'targeted circulation types' of (Bloomfield et al. 2019). However, because of the 'non-meteorological' origin of such CTs, there is no longer any reason to expect these to correspond to anything intrinsic to the atmosphere, and they may therefore not be notably predictable on longer timescales. Indeed, demand forecast skill was found to be higher in week 4 using more classical CT methods compared to the targeted circulation type method (Bloomfield et al. 2021). Another obvious way to get a more informative CT framework is to simply increase the number of CTs, since this allows more detailed variability to be captured. But on the other hand, the more CTs, the closer a CT forecast is to a basic forecast of the full midlatitude flow, which in the extratropics is generally thought to possess a predictability limit of around 2 weeks (Lorenz 1969; Judt 2020). Thus frameworks with too many types are unlikely to possess predictability beyond week 2, unlike what is observed in simpler frameworks, which can have predictability on sub-seasonal, seasonal and even interannual time-scales (Scaife et al. 2014; Strommen and Palmer 2018; Cassou 2008; Dunstone et al. 2016; Athanasiadis et al. 2017). A trade-off therefore appears likely either way.

If this hypothesis is correct, then it becomes natural to consider adopting a 'hybrid' CT forecast in order to capture the 'best of both worlds'. In other words, using a more informative framework on shorter timescales, where the predictability is inherently higher, and then switching to a less informative but more predictable framework on longer timescales. For example, this could involve using a framework with many CTs at short lead-times and a framework with few CTs at long lead-times. We remark here that the types of decisions stakeholders might make would likely differ between these two timescales. On shorter timescales the more informative forecasts could be used to make high-cost decisions, such as whether to evacuate an area at risk of flooding. On longer timescales, having a very early warning of potentially increased risk could allow stakeholders to draw up adaptation plans or adjust cost estimates.

The goal of this paper is to answer the following questions:

1. Is there indeed a trade-off between informativity and predictability in CT forecasts at a given lead time, and how would one measure such a trade-off?
2. What is a principled method of choosing either a single CT framework or combining two or more CT frameworks in such a way that the total value across a forecast window is maximised?
3. What is a robust way to measure how informative different CT frameworks are?

To illustrate our proposed answers, we consider the example of electricity demand and demand-net-wind (i.e., demand minus available wind energy) in Great British winters (November to February, abbreviated NDJF). Here demand-net-wind is an important quantity because wind is the dominant renewable energy source in Great Britain. We will examine two CT frameworks. On the one hand, the coarse 2-state framework of the positive (NAO+) and negative (NAO−), which we refer to as just 'NAO regimes', and on the other hand the detailed 30 UK Met Office patterns (MO30). This is motivated by the fact that the use of these CTs is being explored by the National Energy

System Operator (NESO), the independent, public corporation at the centre of Great Britain's (GB) energy system which, amongst its responsibilities, is in charge of balancing supply and demand on GB's national electricity transmission network. We will apply our methods to answer the following explicit question: what is the optimal way of combining the MO30 and NAO regimes in order to produce the most valuable circulation type forecast of GB demand and demand-net-wind?

We note that a novelty of our analysis is the explicit consideration of seasonal timescale CT forecasts. We will show that the CT method can provide value for forecasts of energy demand at lead times of up to 3 months ahead in winter, provided the correct framework is chosen. The potential for using seasonal forecasts to inform gas and electricity demand has been raised previously (Clark et al. 2017; Thornton et al. 2019), but not as far as we are aware in a CT context. The idea of trying to quantitatively compare different circulation type frameworks has been considered previously in some specific contexts (Beck and Philipp 2010; Schiemann and Frei 2010), but as far as we are aware never in a manner that considers the balance between informativity and predictability, nor in a manner that considers combining multiple frameworks. The idea that a framework with fewer CTs is more appropriate at longer timescales due to reduced predictability is present in (Neal et al. 2016). We go beyond this in two key ways. Firstly, by explicitly showing how to decide when such a reduction, or indeed when a switch from any one framework to another, should optimally take place. Secondly, by showing how a drastic reduction of complexity on seasonal timescales allows one to capture a valuable predictable signal (Scaife et al. 2014; Grams et al. 2020).

2 | Data and Methods

2.1 | Data

For our NWP forecasts of atmospheric pressure, we draw from a retrospective ensemble forecast of past winters using the Integrated Forecast System (IFS) produced by the European Centre for Medium-range Weather Forecasts (ECMWF). This so-called 'seasonal hindcast' consists of 35 individual ensemble forecasts, each initialised on November 1st for the years between 1980 and 2015, and producing daily data for every day in the winter season November through to the end of February. Each forecast uses 50 ensemble members. The IFS model cycle used is 48r1 (IFS documentation, n.d.) and is run using the Tco199 octahedral grid, which corresponds to a grid resolution of 50–80 km. These particular forecasts were run using sea-surface temperatures (SSTs) and sea ice prescribed from ERA5 reanalysis (Hersbach et al. 2020). This is done in order to enhance the seasonal forecast skill as much as possible, since most of the skill on longer lead times comes from the SSTs and sea ice. The skill this hindcast has at long lead times can be assessed by its skill at predicting the average winter NAO conditions (i.e., the DJF average NAO), as measured using the correlation between the ensemble mean winter NAO and the actually observed winter NAO. In our hindcast, this correlation is 0.55, which is statistically significant ($p < 0.05$) against a two-tailed t-test and comparable to the most skilful forecasts currently attained for the winter NAO in this period (Athanasiadis et al. 2017).

In order to test how sensitive our results are to the fixed initialisation date of November 1st, we compare against additional ECMWF forecasts drawn from the S2S project (Vitart et al. 2017), which consist of retrospective forecasts covering overlapping years but initialised at a range of dates in the winter months.

For our observations of two-metre air temperature (T2M) we use daily ERA5 reanalysis, regridded to a 1° degree regular grid (ERA5 Copernicus Climate Change Service (C3S) Climate Data Store (CDS) 2023).

When estimating historical electricity demand and demand-net-wind (i.e., demand minus available wind energy), we use the demand and wind energy models developed by Bloomfield, Brayshaw, and others (ERA5 derived time series of European country-aggregate electricity demand, wind power generation and solar power generation: hourly data from 1979–2019. [Dataset] 2020), as described in, for example, (Bloomfield et al. 2019). Note that this is not the true demand seen by electricity providers in the past, which is a complex function of a number of socio-economic drivers. Rather, this model estimates just the meteorological contribution to the demand, i.e., the part of demand driven by local weather conditions. We use daily values of these quantities.

2.2 | Computation of the NAO and MO30 Circulation Types

To define NAO regimes, we first compute an NAO pattern. We follow a standard method (Dawson et al. 2012) and compute the leading empirical orthogonal function (EOF) of daily NDJF 500 hPa geopotential height anomalies over the Euro-Atlantic sector 30°–90° N, 80° W–40° E, using data drawn from ERA5. The obtained NAO pattern is shown in the schematic Figure 1, Panel 3. For each IFS ensemble member and ERA5, the mean NDJF Z500 is then computed at each gridpoint in the domain and then subtracted in order to generate Z500 anomaly fields. These anomaly fields are then projected onto the ERA5 NAO pattern to generate NAO index timeseries. These NAO timeseries show no clear trends but a weak seasonal cycle. There is a discrepancy between the seasonal cycle found in the IFS and in ERA5, likely due to model drift, which results in a lead-time dependent bias in the IFS NAO forecasts. To not overly penalise the forecasts, we perform a calibration by removing this lead-time dependent bias. This is equivalent to removing the seasonal cycles from both IFS and ERA5 NAO timeseries, since the difference in the seasonal cycles is precisely the lead-time dependent bias. Finally, for each IFS ensemble member and ERA5, the circulation is said to be in the NAO+ regime (NAO− regime) if the NAO index is positive (negative) on a given day. We therefore avoid the use of a 'neutral regime', which is sometimes used to label days where the NAO index is close to 0. Note that all Z500 data here is regridded to a regular 1° grid prior to analysis.

For the MO30 types, the precise patterns of Neal et al. (2016) were provided to us by Robert Neal and the UK Met Office, in the form of 30 sea-level pressure (SLP) anomaly maps: they are shown by Neal et al. in their Figure 1. For each of the IFS

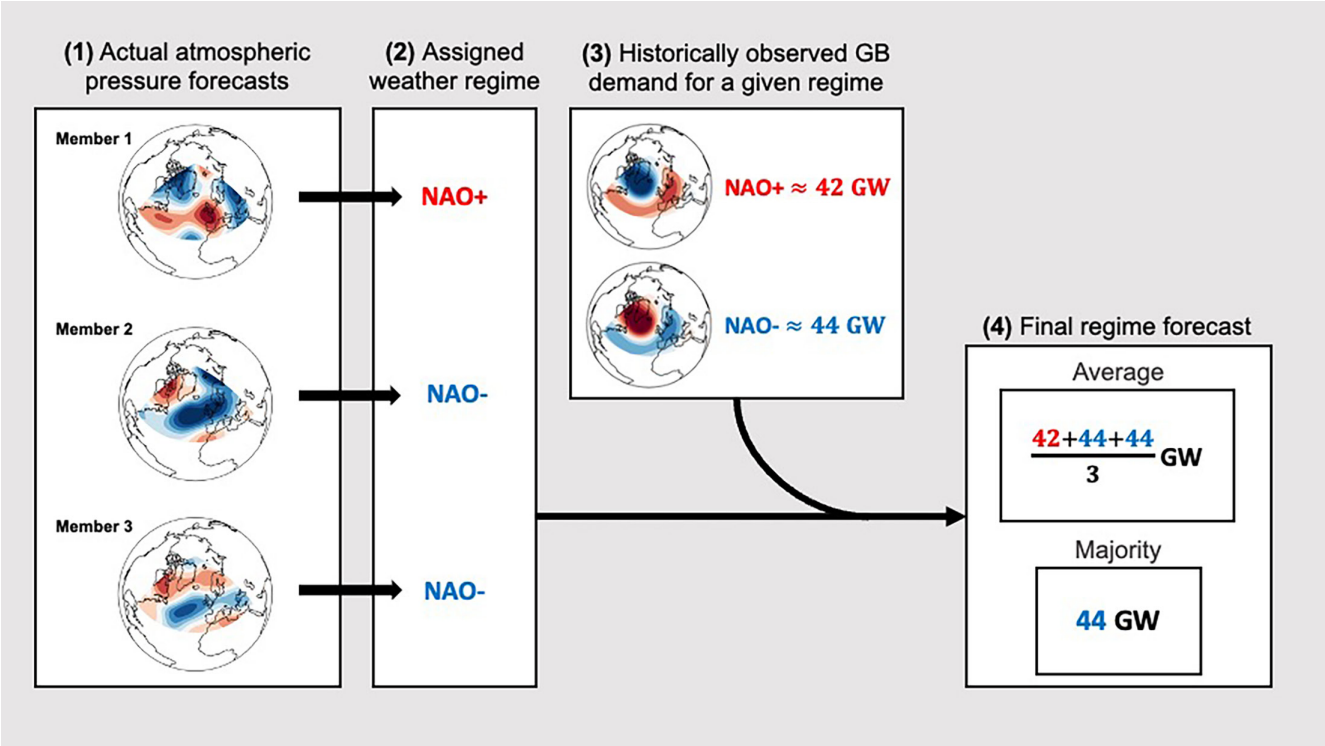

FIGURE 1 | Schematic description of the four steps involved in making a circulation type forecast of nationally averaged winter energy demand in Great Britain, starting from a 3-member NWP forecast of atmospheric pressure.

ensemble members and ERA5, we then consider their daily NDJF SLP over the region 35°–70° N, 30° W–20° E. Following Neal et al., the SLP data is regridded to a 5° degree grid. As with Z500, the mean NDJF SLP field is computed separately for each member and for ERA5, and subtracted to create NDJF SLP anomalies. Daily SLP anomaly forecasts at these gridpoints suffer from a lead-time dependent bias, like the NAO, so we calibrate the SLP fields by removing this lead-time dependent bias. To assign a MO30 CT to a given day, we follow the same methodology of Neal et al. (2016). Concretely, a 'distance' function is defined for gridded SLP anomaly fields as the area-weighted sum of squared differences. A given daily SLP field is then assigned to the MO30 CT for which this distance is minimised. Note that this methodology differs slightly from that of Neal et al. (2016), since they make systematic use of the EMULATE SLP climatology (Ansell et al. 2006) to compute anomalies. However, differences between the EMULATE SLP climatology and the SLP climatologies that we estimated directly using IFS and ERA5 data were found to be small. In particular, the CTs assigned to ERA5 are the same whether you use the EMULATE climatology or the climatology estimated from ERA5 itself. We thus expect this difference in methodology to be negligible.

A final point worth mentioning here is that the MO30 type contains patterns corresponding to the NAO+ and NAO−, so we could alternatively have used these patterns to define our NAO regimes. This would in practice allow direct traceability between the two sets of CTs when we transition from one framework to another. In this particular case, it turns out not to matter either way, because the NAO patterns deduced from ERA5 are virtually identical to those contained within the MO30 patterns, and so the CT assignment is the same whether one uses one set of patterns or the other. However, in other cases, such traceability may be an attractive feature worth pursuing.

2.3 | How to Make a Circulation Type Forecast for a Given Application

We assume there are four steps involved in making a CT forecast, from a given start date, of a variable X_t at a lead-time of $t = n$ days, i.e., n days into the future. Here X could, for example, be aggregated energy demand in Great Britain. Let $\rho_{obs}(X_t)$ be the pdf of X obtained from a historical record of observations, i.e., an observed climatological distribution of X. The pdf is assumed to be based on samples drawn from a particular period of each year (such as a particular season) across multiple years. The four steps are then:

1. Obtain an actual ensemble weather forecast of atmospheric pressure.
2. Assign each ensemble member to the closest matching circulation type.
3. Separately, condition $\rho_{obs}(X_t)$ on the occurrence each circulation type to obtain the conditioned pdfs

$$\rho_{obs}(X_t \mid \text{days } t \text{ for which that CT occurred}) \quad (1)$$

and extract the moments of interest (e.g., the mean of the conditioned pdf or the change in likelihood of extreme outcomes).

4. Identify the *mode* of the assigned CTs, i.e., the type that occurs most frequently across all ensemble members. The statistic obtained from the conditioned pdf for this *majority CT* is the final CT forecast.

We emphasise here that what we will from now on refer to as the 'majority CT' refers to a simple majority only, and we make no stipulations about how great the majority needs to be. In particular, for the MO30 CTs, it will in most cases be less than 50%. An alternative to using the majority CT in step 4 would be to compute a weighted average of the conditioned pdfs with weights given by the proportion of ensemble members predicting the given CT.

This procedure is described more intuitively in the schematic Figure 1, which gives an example of how to make a CT forecast of nationally averaged winter energy demand in Great Britain ('GB demand'), starting from a 3-member forecast of atmospheric pressure. The CTs being used in this example are the two NAO regimes. Starting in step 1 with actual meteorological forecasts, in step 2 one assigns each ensemble member to the appropriate CT. Separately, in step 3 one estimates the nationally averaged demand on winter days where the NAO+ regime prevailed (roughly 42 Gigawatts), and similarly for NAO− (roughly 44 Gigawatts). Note that step 3 uses data drawn entirely from observations, and in particular does not rely on the accuracy of an actual NWP forecast. Finally, in step 4, one combines the information from steps 2 and 3 to create one's desired CT forecast, either by taking the average or the majority (mode). This procedure can be used to create CT forecasts of any particular statistic one desires, including measures of 'extreme days'. For simplicity, we will focus on forecasts of changes to the mean.

For forecasts at longer lead-times, there can often be no genuine majority, with two or more CTs being tied for having the most members assigned to it. This is particularly the case for the MO30 types, due to the large number of patterns. In these cases, we just pick one of the options at random. The exact way of dealing with ties will turn out to be inconsequential to our results, since (a) the more frequently occurring MO30 ties occur at longer lead-times where the lack of MO30 forecast skill dominates any such choices, and (b) our analysis is ultimately based on metrics integrated across entire seasons and multiple years, which smooths away noise.

2.4 | Why We Use the Majority Circulation Type and Not the Average

In most circulation type applications, information from all ensemble members is combined by taking an average across distributions (Richardson et al. 2020; Bloomfield et al. 2021). However, in this paper we deliberately choose to use the majority CT instead. This choice is motivated by our interest in looking at seasonal timescale predictions.

It is well established that it is possible to make skilful forecasts of some aspects of the Euro-Atlantic winter circulation initialised months ahead. Most relevantly, ensemble forecasts initialised on November 1st can skilfully forecast the DJF mean NAO (Scaife et al. 2014; Dunstone et al. 2016; Athanasiadis et al. 2017). However, these skilful forecasts exhibit low signal-to-noise ratios, meaning that the predictable signal being detected by the forecast is larger than the ensemble suggests it is *prima facie* (Eade et al. 2014; Siegert et al. 2016; Scaife and Smith 2018). In the present context, where we will be considering CT forecasts of the NAO+ and NAO− patterns, these low signal-to-noise ratios practically mean that at long lead-times t

$$\mathbb{P}(\text{NAO} + \text{ at time t}) \approx \mathbb{P}(\text{NAO} - \text{ at time t}) \approx 1/2 \qquad (2)$$

Since the NAO+ and NAO− events partition the whole space, the average across the conditioned pdfs in such a situation is close to climatology, meaning one would gain next to no information. In other words, this averaging procedure would effectively remove any benefit from seasonal NAO forecast skill. Note that this is in principle the case whether one is interested in the 'central forecast' (i.e., the change in the mean of the distribution) or a measure of changes to the extremes/tails, since the limiting factor is either way that the forecasted distribution is very close to climatology. However, it should be emphasised here that even small changes to the probabilities of an extreme outcome can have outsized impacts on cost-loss analysis, so it is possible that this limitation is less severe for users focused on extreme outcomes. Indeed, the discussion above hinges on the notion of what it means to be 'close to' the climatological pdf, which could be measured in different ways, including ways which focus on the tails of the distribution. Further thoughts on this theme are included in the Discussion.

To overcome this limitation and regain some benefit, one must take the ensemble mean signal 'more seriously'. It is tempting here to consider using the CT assigned to the ensemble mean as the basis for a CT forecast. Unfortunately, this fails badly in cases with a lot more than 2 types (e.g., MO30), since the CT assigned to the highly smoothed out ensemble mean pressure field can easily have no overlap with the CT assigned to the sharper pressure field of any individual ensemble member. The better way is to use the majority CT. This works well at short lead-times, since there is more certainty in the forecast and hence the majority CT dominates the ensemble. It also works for longer lead-time forecasts of the NAO regimes, since given a sufficiently large ensemble, the ensemble mean NAO regime will match the majority regime NAO, for two reasons. Firstly, the presence of only two CTs means the ensemble mean field will definitely be assigned to the NAO+ or NAO−. Secondly, in a large ensemble, the members exhibiting an NAO+ (NAO−) regime will occupy a wide range of positive (negative) NAO index values (due to chaotic variability), and hence the mean NAO index is determined by the overall number of such positive (negative) index members and not the index of any particular member (e.g., outliers). Thus, skilful seasonal forecasts of the NAO index translate into skilful forecasts of the majority NAO regime, something we will confirm explicitly in Section 3.

In summary, using only the majority CT is a principled strategy that is uniform across lead times and has the potential to extract more information. This approach also has the benefit that analysis of skill and optimisation becomes much easier, as we show in the next section. The clear downside is that some risk is introduced: weighted averaging across all ensemble

members effectively hedges for the possibility that the majority CT forecast will be wrong sometimes. More broadly, due to Equation (2), restricting ourselves to the majority CT at longer lead times can discount as much as 50% of the information nominally contained in the full ensemble. In the Discussion, we will comment on how hedging for risk can be reintroduced in our framework, as well as the use of scenario-based approaches to complement our methodology.

3 | How to Optimally Choose One or More Circulation Type Frameworks

In this section we outline a general method for how to answer the following question: given different sets of CTs with differing levels of informativity/predictability, how does one pick a single 'optimal' one, or create an 'optimal' hybrid of several of them? We only give explicit details for how to pick a single one and how to combine two frameworks. Extending the method to more than two sets of CTs is simple and we explain how at the end. We will exemplify our method throughout using the example of how to combine MO30 and NAO frameworks for GB demand and demand-net-wind forecasts.

In all that follows we assume we are working within a fixed 'season' consisting of K days, and that the climatological pdf being made use of is the distribution of all values attained by the target variable X over these K days, as sampled from a large number of years. For example, the season might be extended winter (NDJF), with $K = 120$ days.

Before proceeding, we emphasise that we will deliberately make some simplifying assumptions to streamline the analysis. The main goal is to show how one can answer these kinds of questions in a flexible manner, so that interested readers can tweak the precise assumptions according to their interest.

3.1 | Determining the Expected Value of a Single Circulation Type Forecast

Given a CT forecast as defined in Section 2.3, value is obtained from the forecast of a given day when two conditions are met: (1) the majority CT forecast needs to be correct (i.e., match the actually observed CT on that day), and (2) the conditioned pdf needs to be sufficiently different from the climatological pdf: i.e., the CTs need to carry meaningful information about the target variable X. One standard way to compute how much information the CTs carry is to use mutual information content, as we explain in Section 4. For now, we will just assume that, given a particular choice of CTs R and forecast target X, one can assign a *constant* value $V = V_{R,X}$ which summarises how much value the user gains on average from a single correct CT forecast. Conversely, we assume that if the CT forecast is incorrect then the user loses the same amount of value V. This is a reasonable assumption for CT frameworks with very few CTs, since the CTs will in this case all differ notably from each other by construction, meaning the conditioned pdfs will typically also differ notably. However, in frameworks with many CTs, the majority CT might be 'wrong', but nevertheless look fairly similar to the observed CT, meaning not all value will be lost.

Finally, we assume that our majority CT forecast at lead-time t is correct with some *fixed* probability p_t which is a non-increasing function of t. In other words, we assume that the forecast skill is independent of the particular year it was made and that the skill is always going down over time. For forecasts of meteorological variables one generically expects skill to decline exponentially with time (Krishnamurthy 2019). These probabilities of success, or 'hit-rates', will at least to some extent depend on the initialisation date, but for simplicity we will suppress this from the notation. Let p_{clim} be the probability of guessing correctly by guessing climatology, i.e., guessing the climatologically most frequently occurring CT for any given day. Intuitively one expects $p_t \gg p_{clim}$ (i.e., high skill) for short lead-times and $p_t \approx p_{clim}$ (i.e., no skill beyond climatology) for long lead-times.

Example: MO30 vs NAO for GB Demand.

Figure 2 shows estimates of the daily hit-rate, p_t, using MO30 in (a) and NAO regimes in (c), based on the ECMWF seasonal forecasts. Stipled lines show estimates of p_{clim}. For the NAO regimes, $p_{clim} = \frac{1}{2}$, whilst for MO30, $p_{clim} \approx \frac{1}{16}$.

We note that these climatological hit-rates are the same hitrates you expect from guessing by chance. Indeed, for the NAO this is clear: the climatological value of our NAO timeseries at each day is zero, since we have removed the seasonal cycle. However, since there is no neutral regime in our NAO framework, the climatological guess amounts to randomly picking NAO+ or NAO− with equal probability every day, i.e., a random guess. For the MO30 CTs, we need the additional information that from around November 15th onwards the MO30 types occurring are almost entirely restricted to MO30 patterns 15–30 (see Figure S1a). We remind the reader that the MO30 patterns are defined using year-round data, and that patterns 15–30 are precisely the high-amplitude patterns associated with boreal winter. Of these 16 patterns, the most likely one to occur on all subsequent days in NDJF is pattern 21 (see Figure S1a). However, this is a weak constraint, since the actually occurring pattern can in practice be any of the other 15 remaining patterns with almost equal probability (not shown). Thus the climatological hit-rate from November 15th onwards is the same as a random guess with probability of success $\frac{1}{16}$. Prior to November 15th the climatological guess is only fractionally better than this random guess, so the climatological MO30 hit-rate is approximately $\frac{1}{16}$ for all lead-times. Because the climatological hit-rates are in both cases the same as random guesses, a two-sigma range around p_{clim} is computed analytically by assuming the climatological guesses are binomial trials. This range is shown with grey shading.

Figure 2a,c show the expected behaviour, with hit-rates much higher than chance at early lead-times which converge towards p_{clim} at longer lead-times. For MO30 the hit-rates have lost skill at around day 12. Figure 2b shows the number of ensemble members making up the majority CT, and one can see that the size of the majority drops at a similar rate, converging to around 6 (out of a total of 50) at day 12. This loss of MO30 skill is consistent with related estimates found in (Neal et al. 2016). We also verified that it is not an artefact of the fixed November 1st start-date: the hit-rates diminish at the same rate in S2S forecasts initialised at a variety of different dates in NDJF (see Figure S1b).

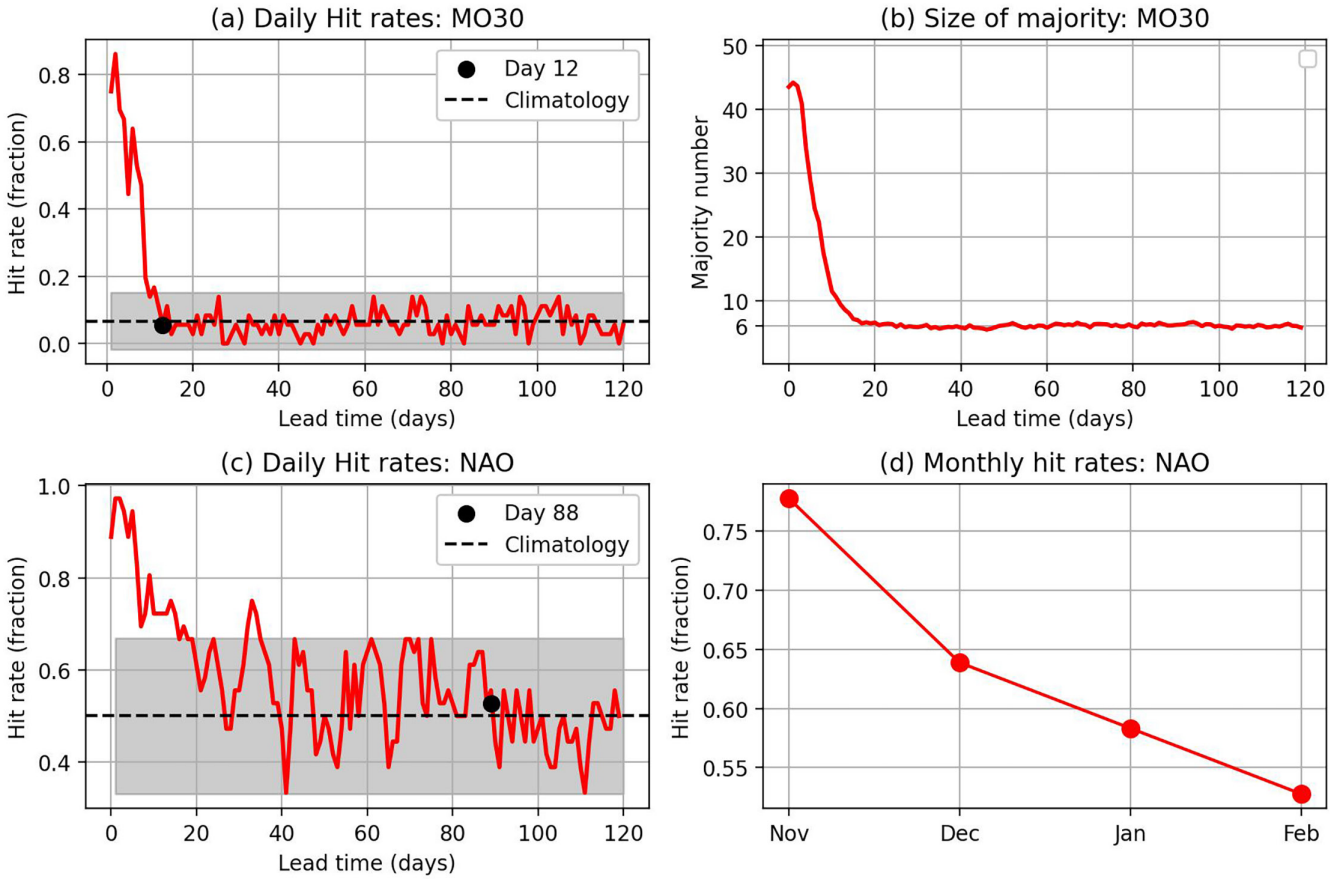

FIGURE 2 | (a) The daily ECMWF forecast hit-rates (= probabilities of success) using MO30 CTs at different lead times (days after Nov 1st). (b) The number of ensemble members making up the majority MO30 CT at different lead times. (c) The daily ECMWF forecast hit-rates using NAO regimes. (d) The monthly ECMWF forecast hit-rates using NAO regimes. (a and c), the stipled line is the probability of getting it right by chance, and shading is the 95% confidence interval around this probability.

For the NAO regimes, there is a suggestion of marginal skill out to around day 90, as the daily hit-rates are more often than not greater than 50% despite being consistently within the grey confidence band. This existence of skill is clearer when using monthly hit-rates, shown in Figure 2d. Here we define the majority CT for a given month as the most frequently recurring CT across all days in the given month (and across all ensemble members in the case of the forecast), and compute hit-rates by comparing to ERA5. The fact that the monthly hit-rates do not hit 50% until February is consistent with the existence of winter NAO forecast skill in this hindcast, as measured by a statistically significant correlation of 0.55 between the ensemble-mean and ERA5 DJF mean NAO.

Given this set-up, the expected value from a CT forecast at lead-time t is

$$\mathbb{E}(\text{value}_t) = p_t V + (1 - p_t)(-V) = (2p_t - 1)V \tag{3}$$

where we have used our assumption that there is a fixed amount of value V obtained when the forecast is correct and the equivalent loss of value $-V$ if the forecast is wrong. This is positive if and only if $p_t > 1/2$, reflecting the fact that if one is wrong more often than right then one necessarily experiences a net loss of value. The expected value integrated across a winter season ('integrated seasonal value' for short) is then simply

$$\mathbb{E}(\text{seasonal value}) = \sum_t (2p_t - 1)V = 2V\left(\sum_t p_t\right) - KV \tag{4}$$

where the sum is over all the days making up the season, and K is the length of the season. The goal now is to make this quantity as large as possible, relative to some baseline forecast. In our case, the baseline forecast is one which guesses climatology; in practice this will often amount to the same as guessing randomly, as shown earlier with MO30 and NAO regimes. Then the expression for the integrated seasonal value from this baseline is identical to that of Equation (4) with p_t replaced by p_{clim}. Consequently, the integrated seasonal value from a CT forecast, relative to the baseline, is

$$2V \sum_t (p_t - p_{clim}) \tag{5}$$

This sum is growing as long as $p_t > p_{clim}$, i.e., as long as there is skill in the forecast. It is thus beneficial to use of the forecast at all lead-times for which $p_t > p_{clim}$ and not thereafter. Furthermore, if one wants to use just a single set of CTs at all lead-times, then one should pick the CTs which maximise the expression in Equation (5). It can also be seen that the contribution to overall value from a given lead-time is determined by the product $V(p_t - p_{clim})$, where the first factor measures informativity and

the second factor measures predictability. Thus any trade-off between informativity and predictability is measured with this product. One wants to make this product as large as possible, but a high value of one factor can easily be outweighed by a low value of the other factor. We will see an example of this in action in the next section.

3.2 | Optimal Combinations of Two Circulation Type Forecasts

Consider now the situation where one has two CT forecasts to compare. To make the link to our example application more simple, let us assume one of these forecasts is based on MO30 and the other on NAO regimes. Let p_t and p_{clim} (q_t and q_{clim}) be the MO30 (NAO) forecast and climatological probabilities respectively. Suppose one knows already that $V_{\mathrm{MO30}} > V_{\mathrm{NAO}}$ (i.e., that MO30 patterns are more informative than NAO patterns) but that $q_t > p_t$ for large lead-times (i.e., that NAO patterns are more predictable on longer timescales). Then a natural approach is to construct a hybrid forecast which uses MO30 for M days and swaps to NAO regimes thereafter. The integrated seasonal value of such a forecast is

$$2V_{\mathrm{MO30}} \sum_{t \le M} (p_t - p_{clim}) \; + 2V_{\mathrm{NAO}} \sum_{t > M} (q_t - q_{clim}) \tag{6}$$

and the goal is to choose M such that this expression is maximised. Whilst this can easily be done given explicit values of all the variables, a more intuitive approach is as follows. The added value from the CT forecast at a particular lead-time t is $2V_{\mathrm{MO30}}(p_t - p_{clim})$ for MO30 and $2V_{\mathrm{NAO}}(q_t - q_{clim})$ for the NAO regimes. The integrated seasonal value is therefore maximised by choosing MO30 whenever

$$2V_{\mathrm{MO30}}(p_t - p_{clim}) > 2V_{\mathrm{NAO}}(q_t - q_{clim}) \tag{7}$$

and NAO regimes otherwise. Thus the optimal choice of M can be obtained by identifying the point of intersection between the curves $V_{\mathrm{MO30}}(p_t - p_{clim})$ and $V_{\mathrm{NAO}}(q_t - q_{clim})$, both viewed as functions of t.

In practice, estimates of the hit-rates p_t and q_t are noisy, so identifying the intersection numerically is dubious. On the other hand, the sums over many hit-rates occurring in Equation (6) are relatively smooth, making it possible to robustly identify the value of M which maximises (6), as the following example shows.

Example: MO30 vs NAO for GB Demand.

Figure 3a shows the curves $V_{\mathrm{MO30}}(p_t - p_{clim})$ and $V_{\mathrm{NAO}}(q_t - q_{clim})$ in a hypothetical scenario where we assume that $V_{\mathrm{NAO}} = 1$ and $V_{\mathrm{MO30}} = 2$. The raw estimates of these curves are shown in thinner colours in the background, whilst the thick colours of the same colour show smooth fits based on a Savitsky-Golay filter (Savitzky and Golay 1964), a standard method for smoothing timeseries via convolutions. The noisiness of the raw curves reflects the noisiness in the hit-rate estimates shown in Figure 1a,c. The smoothed curves intersect at day 9, suggesting that in this case one should swap from

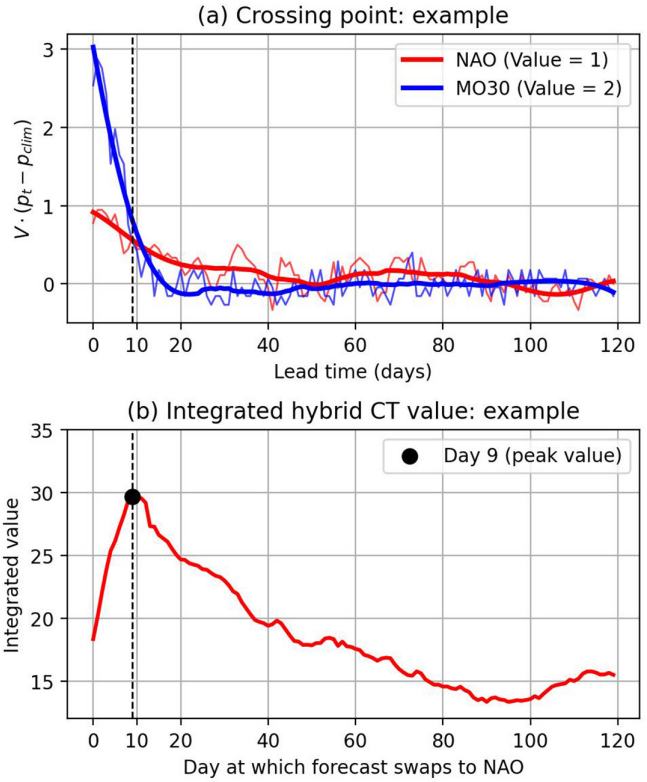

FIGURE 3 | (a) Estimates of daily added value from the MO30 (blue) and NAO (red) CT forecasts, assuming that MO30 patterns are twice as valuable as NAO patterns. The raw estimates are shown in thin lines, whilst smoothed estimates are shown in thick lines. The black stipled line highlights day 9, where the smooth curves intersect. (b) Estimates the integrated seasonal value for a MO30/NAO hybrid forecast, as a function of the day at which one swaps from MO30 to NAO. The black dot highlights the maximum value (day 9). The black stipled line is again highlighting day 9. Smoothing is performed using a Savitsky-Golay filter with a window-length of 50 and polynomial order 5. This is roughly comparable to applying a 14-day running mean.

MO30 to NAO regimes from day 9 onwards. Figure 3b shows, by contrast, raw estimates of the curve defined by Equation (6) as a function of M, given the same assumption that MO30 patterns are twice as valuable as NAO regimes. In this case a fairly clear maximum can be estimated with no additional smoothing required, and it turns out that this maximum occurs at $M = 9$. This agrees with the estimate from Figure 3a, but we emphasise that different choices of how to smooth the curves in Figure 3a can easily change the intersection point by several days. The estimate from Figure 3b is therefore more principled.

If one now defines the units of V to be such that $V_{\mathrm{NAO}} = 1$, then one can estimate the value of M for a wide range of possible values of V_{MO30}. This is shown in Figure 4. This suggests that if MO30 patterns and NAO regimes are equally valuable, one should swap at day 8, whilst if MO30 patterns are at least 2.9 times more valuable then one should only swap at day 12, when the MO30 skill has vanished. Further increases to V_{MO30} at this point have no effect on the day to swap, due to this vanishing of skill.

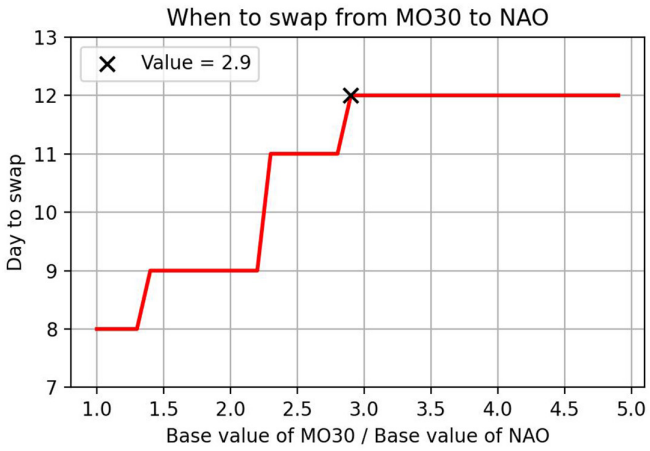

FIGURE 4 | Estimates of the day to swap from MO30 to NAO as a function of the ratio V_{MO30} / V_{NAO}. Estimates are obtained by picking, for a given value of the ratio, the choice of M which maximises the expression in Equation (6). The value of 2.9 has been highlighted: At this point the red curve stabilises at day 12.

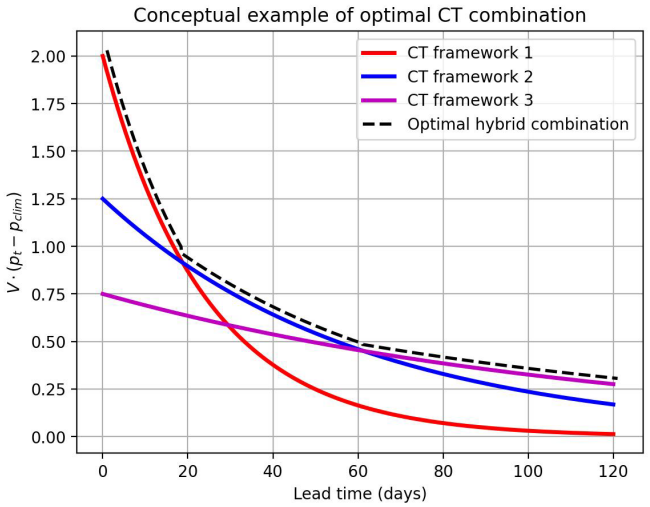

FIGURE 5 | Daily added value for three sets of artificially constructed circulation type frameworks, assumed to have differing levels of informativity and predictability. The black stipled line shows the curve corresponding to the optimal hybrid combination.

We therefore see, both in the example case of Figure 3 or the range of cases explored in Figure 4, a confirmation of the intuitively expected trade-off between informativity and predictability. Even if the MO30 patterns are assumed to be more than five times as informative as the NAO regimes, by day 12 this extra informativity is completely outweighed by the lower predictability.

3.3 | Optimal Combination of Multiple Circulation Type Forecasts

It is now clear how to extend to combining more than two CT frameworks. The conceptual idea is shown in Figure 5. Here we have created three artificial CT frameworks by generating three smooth, exponentially decreasing hit-rate curves and assigning differing baseline values to each set of CTs. The optimal combination is obtained by following a path that always stays at the highest value curve; thus in this example, one would swap from framework 1 to 2 at around day 20, and from framework 2 to 3 at around day 60.

In practice, one would proceed by first estimating the baseline value for each framework, along with the hit rates. Visual assessment of the raw curves of daily added value then suggests which order to place the CT frameworks in, allowing one to write down an expression like that in Equation (6). A direct estimate of the days to swap at can then be obtained numerically.

4 | How to Estimate the Informativity of a Circulation Type Framework

Section 3 explains how to optimally combine multiple CT frameworks, given prior knowledge about how informative each framework is. The next question is then how to numerically estimate the level of informativity. This will also allow us to assess if our assumption that MO30 patterns are more informative than NAO patterns is correct.

In a CT forecast as we define it, information about the variable of interest is obtained by conditioning the climatological pdf based on the forecasted CT. Thus a CT forecast is only informative if the conditioned pdfs tend to look different from the climatological pdf: if the conditioned pdfs look very different, there is a lot of information, and if they look the same as climatology then there is no information.

What we are looking for is therefore a precise quantitative answer to a question of the form "how much information about a variable X do I gain by knowing the value of a different variable Y?". The branch of mathematics dealing with questions like this is called information theory, and the answer it gives is the *mutual information content* of X and Y, which we denote here by $MI(X, Y)$. In the following Section 4.1, we give a brief and informal overview of this concept with no proofs. There are numerous resources available on information theory should the reader want details, such as the textbook Elements of Information Theory (Cover and Thomas 2005). We also recommend the excellent and informally written book by John Baez (Baez 2024).

Readers who are happy to take the details as a black box can safely skip to the example application in Section 4.2. The key point for the application is that, given a climatological pdf and some CT framework, the mutual information between the climatology and the CTs measures in a robust and quantitative way how different the conditioned pdfs look from climatology. High mutual information means the conditioned pdfs look very different, whilst low mutual information means the conditioned pdfs are very close to climatology. Mutual information code is implemented in many standard software packages (e.g., python).

4.1 | A Brief Introduction to Mutual Information Content

One first defines the *entropy* of a random variable X, denoted $H(X)$, as the sum

$$H(X) = -\sum_x \rho_X(x)\log\big(\rho_X(x)\big) \tag{8}$$

where ρ_X is the pdf of X. The logarithm is typically with respect to base 2, in which case the units are called 'bits', but any base will do. This formula can be read as an expected value:

$$H(X) = \mathbb{E}\big(-\log\big(\rho_X(x)\big)\big) \tag{9}$$

Informally, $-\log\big(\rho_X(x)\big)$ measures the amount of information gained about X by observing the value $x \in X$, and the expected value of this quantity thereby measures the average amount of information contained in a single random observation of X. It can be thought of as measuring the amount of uncertainty present in X: high entropy means that observations of X generally carry a lot of information, implying that one needs to make many observations in order to accurately estimate ρ_X. Conversely, low entropy means ρ_X can be well understood with relatively few observations. In accordance with this informal description, one can show that the entropy is maximised when ρ_X is a uniform distribution, as one would in such a case need information from the entire range of possible values in order to describe ρ_X. It is minimised when X is a delta distribution (i.e., only takes a single value), as then only one observation is needed in order to fully describe X. In general, entropy is high for a wide distribution and low for a narrow distribution.

One can similarly define the conditional entropy $H(X|\,Y)$, which informally measures the uncertainty present in the conditional pdf $\rho_{X|Y}$. It can be shown that

$$H(X) \geq H(X|\,Y) \tag{10}$$

for any choice of Y, which reflects the fact that knowledge of Y can never *decrease* one's knowledge of X: at worst it gives you no information about X, in which case the inequality is an equality.

Finally, one defines

$$MI(X,Y) = H(X) - H(X|\,Y) \tag{11}$$

It is a strictly positive quantity due to Equation (10), and measures how much the entropy decreases by knowing Y. Put differently, it measures how much bigger the entropy of ρ_X is compared to $\rho_{X|Y}$. Thus the mutual information is high if one needs many fewer observations to describe the conditional pdf compared to the original pdf, whilst it is low if one needs a similar number of observations in both cases. In particular, $MI(X,Y) = 0$ if and only if X and Y are independent. In general, the mutual information is high if the conditional pdf is narrower than the original pdf, and low if it is not. One can show that

$$MI(X,Y) = MI(Y,X) \tag{12}$$

justifying the word 'mutual': the information Y has about X is the same as the information X has about Y.

It is convenient to normalise the mutual information to facilitate comparison between differing data sets. This can be done by dividing by the entropy:

$$MI_{norm}(X,Y) = MI(X,Y)\,/\,H(X) \tag{13}$$

(One can alternatively divide by the entropy of Y, due to the symmetry from Equation (12)). This quantity is strictly between 0 and 1. As before, it is 0 if and only if the variables are independent, and 1 if and only if the variables are completely determined by each other, i.e., if and only if there is a deterministic and invertible function f such that $X = f(Y)$. This can be compared to Pearson correlation, a normalised variant of the covariance which is 1 if and only if $X = g(Y)$ with g a *linear* function. Mutual information is in this sense a generalisation of correlation which allows for non-linearities.

Finally, we note that one can show that the formula for entropy is the *unique* formula satisfying intuitive properties about what 'information' should mean, and so the expression for mutual information content is in some sense unique.

4.2 | Applying Mutual Information Content to Circulation Type Forecasts

We now exemplify the theory described above in the context of GB demand forecasts.

Example: MO30 vs NAO for GB Demand.

In this case, $X = X_t$ is either a timeseries of observed daily GB electricity demand (GB demand) or the electricity demand minus available wind-energy (GB DNW). The climatological pdf of X is denoted ρ_X, whilst $Y = Y_t$ is an integer-valued timeseries keeping track of which CT (labelled by integer indices) is occurring at time t. The mutual information $MI(X,Y)$ measures how much one constrains the possible values of GB demand by knowing which CT is occurring. There are many standard packages for computing it, e.g., in python. Figure 6a shows the climatological pdf of GB demand (black) as well as the pdf when conditioned on the NAO+ and NAO− regimes. The (normalised) mutual information was estimated as around 0.02, implying near-independence. Consistent with this, visual inspection shows that the conditioned pdfs are not substantially different from climatology, with just a mild shift in the mean. Figure 6b shows the joint pdf of GB demand and the MO30 patterns. Thus the sum of all entries in Figure 6b is 1, and the fact that the high amplitude patterns 15–30 have higher values reflects the relative scarcity of patterns 1–14 during boreal winter. Note that the distribution of demand and DNW across all days is exactly the black histogram highlighted in Figure 6a,c, respectively. The (normalised) mutual information between demand and MO30 was estimated as around 0.10, around 4.65 times greater than that between the NAO regimes and demand. Figure 6c,d show similar plots but for DNW. Here the MO30 patterns are even more informative than NAO regimes: we estimated the ratio of mutual

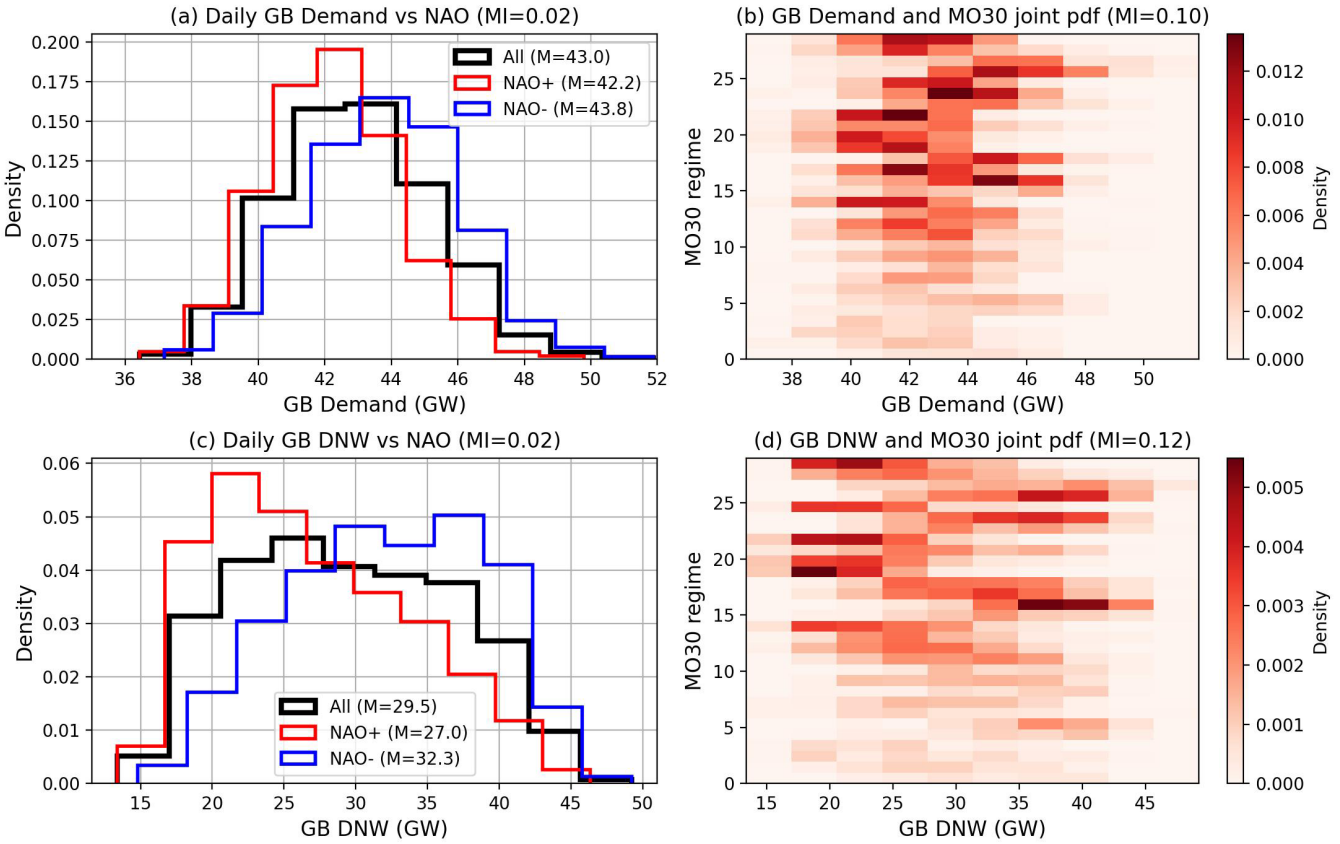

FIGURE 6 | (a) Histogram of daily NDJF GB demand (black), and demand conditioned on the NAO+ (red) and NAO− regime (blue). The values M in the legend are the means of the distributions. (b) The joint histogram of GB demand and MO30 regimes (thus the sum across all entries is 1). (c and d) The same but for demand-net-wind (DNW). The value MI in each title is the mutual information content between demand/DNW and the NAO/MO30 CTs. Histograms are estimated over the period 1979–2015.

information contents here to be 6.15. Consistent with this, the joint pdf in Figure 6d shows that some of the MO30 patterns (e.g., patterns 19 and 17) are particularly strongly associated with large DNW anomalies.

The fact that MO30 patterns are less informative for demand than for DNW likely reflects the fact that demand is mostly a function of the GB mean T2M: the extra granularity of the MO30 patterns over NAO regimes is not as helpful at constraining such a large-scale mean. On the other hand, DNW depends sensitively on the energy produced at wind farms in particular locations: the extra granularity allows for variability in these regions to be more accurately captured. Additionally, wind is a lot more variable than temperature in general, meaning any constraints on winds are likely to be more valuable.

Referring back to Figure 4, these estimates imply that one should use MO30 for days 1–12 (for both demand and DNW), and then swap to NAO regimes afterwards.

5 | Discussion

We discuss some ways to extend or augment our framework in various directions, in the hope of encouraging future work.

5.1 | Using the Gridpoint Forecast

Our Example application to GB demand led to the conclusion that one should use the MO30 framework only up to day 12 of the forecast window, and swap to the NAO regimes after that. That is, the extra informativity coming from having 30 patterns vs. only two is lost after day 12, due to the loss of skill at that point in time. However, within this time horizon, it is natural to consider whether one would be better off using the gridpoint forecast instead. Indeed, GB demand is a strong function of GB mean T2M, and this explicit quantity might be deterministically predictable at such short lead-times, bypassing any need to condition indirectly on historical pdfs.

To test this, in Figure 7a we compare the root mean-square error (RMSE) of two forecasts of daily GB T2M, here defined as the population-weighted T2M average across major GB cities in order to be consistent with the convention of NESO. The blue curve is the forecast based on MO30 patterns: each day, one conditions the climatological pdf of daily winter GB T2M by the predicted majority CT, and uses the mean of the conditioned pdf as one's forecast for that day's T2M. The orange curve is that based on the IFS gridpoint forecast: one simply computes the GB T2M of all ensemble members and takes the ensemble mean as one's basic forecast. To make a direct comparison with the

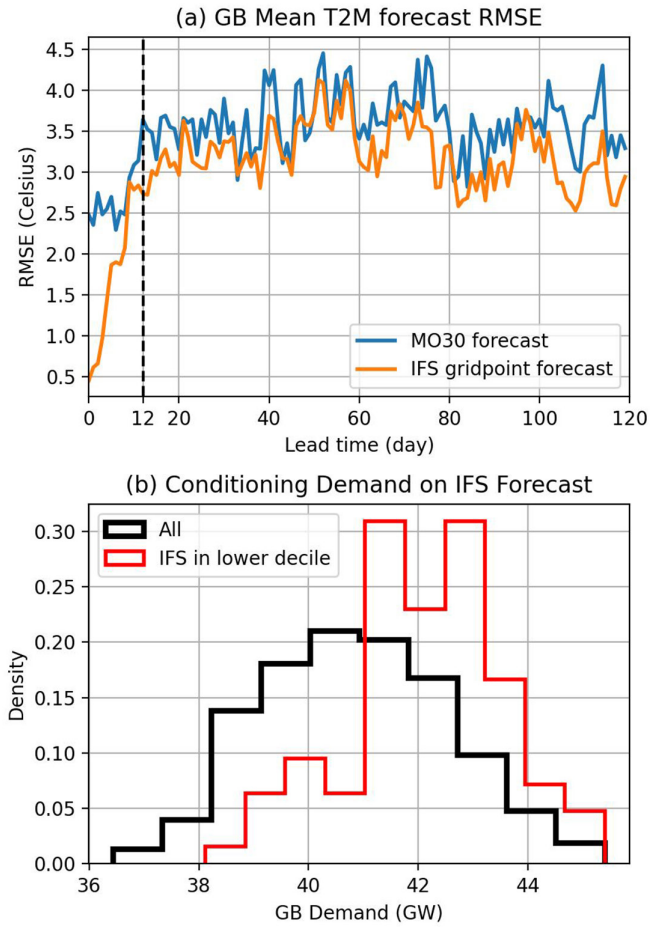

FIGURE 7 | (a) The root-mean-square error (RMSE) of two forecasts of daily GB T2M. The blue line is a forecast based on conditioning the climatological pdf using MO30 CT forecasts, and the orange line is based on a calibrated IFS gridpoint forecast. The stipled line marks day 12, when MO30 skill is lost. (b) The climatological histogram of GB winter demand (black), and the histogram obtained when conditioning on days where the IFS gridpoint forecast of GB T2M is in the lower decile of its climatology.

CT forecasts, a lead-time dependent bias is removed from the raw forecast: it is this calibrated gridpoint forecast that is used when computing the RMSE. It can be seen that the IFS gridpoint forecast in the first 12 days clearly outperforms the MO30 forecast, showing particularly low RMSE values in the first week. This agrees with the results of past work, such as (Bloomfield et al. 2021) and (Rouges et al. 2023). In fact, the gridpoint forecast appears to outperform the MO30 forecast for all 120 days, but this is almost certainly due to the fact that the IFS forecast will automatically contain a realistic seasonal cycle, whilst our conditioned MO30 forecast does not. A seasonal cycle can be introduced to a CT forecast by constructing a climatological pdf for, e.g., each successive week of the season and using these to condition, though care would be required to ensure sufficient samples are available to constrain these more granular pdfs.

The superiority of the gridpoint forecast in the first 12 days is corroborated also using information theory. The mutual information between the GB mean T2M forecasted by the IFS and the observed GB demand is 0.27 when measured across days 1–12 only, compared to 0.14 between the conditioned MO30 forecast and demand. If we replace the forecasted MO30 CTs with the actually observed MO30 CTs in ERA5 (i.e., assume that the forecast was in fact perfect), the mutual information with GB demand is still only 0.17 across days 1–12, implying that even given perfect CT predictability it is still better to use the gridpoint forecast. A similar conclusion was also made in (Bloomfield et al. 2021). An example of the informativity of the gridpoint forecast is shown in Figure 7b, showing how the pdf of demand changes when you condition on days where the gridpoint forecast of GB T2M is in the lower decile of its climatology. A substantial increase in the likelihood of high demand days is clearly visible. Interestingly, the mutual information between the IFS GB mean T2M forecast and demand/DNW is only 0.06 to 0.07 when measured across all days in winter, compared to 0.10 to 0.12 for MO30 patterns. This suggests that MO30 patterns are in principle superior at longer lead times, though the lack of skill beyond day 12 means this information is not currently accessible.

We conclude that it is almost certainly better to use a calibrated gridpoint forecast for days 1–7. This is due to the high skill the gridpoint forecast has in this period of predicting the actual intensity of T2M anomalies, in contrast to the CT-based T2M forecast which always utilises the same historical average. For days 7–12, it has been shown that a forecast obtained by conditioning the gridpoint forecast on the majority CT can sometimes outperform both a gridpoint forecast and a CT-based forecast (Bloomfield et al. 2021). This is closely related to the notion of flow-dependent predictability, the phenomenon whereby the atmosphere is more predictable when starting from particular flow states (Ferranti et al. 2015). Conditioning on the majority CT thus acts as a way of allowing the increased predictability associated with certain CTs to boost the forecast skill. It therefore seems plausible that for days 7–12 the optimal forecast is a blend between the gridpoint and MO30 forecasts, but we do not test this here.

5.2 | Hedging for Risk and Windows of Opportunity

Using the majority CT as one's forecast is sensible for the purpose of conditioning the climatological pdf of interest, but it has an obvious drawback in that one loses all information about how certain or uncertain the forecast is, i.e., the ensemble spread. If the user is interested in hedging risk, it is necessary to reintroduce some of this information.

A simple and pragmatic approach here is to introduce a time-varying weight w_t to the CT forecast. For example, if the forecast for the variable of interest X at time t is obtained as the mean of a climatology conditioned by the majority CT forecast at time t, multiply this forecast by a weight w_t which depends on the proportion of ensemble members actually making up the majority CT. The exact value of this weight could be influenced by several factors, including how risk-averse the user is. The weight could also be made state-dependent. Such state-dependence is appealing because forecast reliability can vary depending on the CT. For example, seasonal forecasts of the NAO+ are in general underconfident, whilst forecasts of the NAO− are overconfident (Strommen et al. 2023). Here an underconfident forecast of some event is one where the forecasted probability of the event

occurring is systematically lower than it ought to be, whilst an overconfident forecast is one where the forecasted probability of occurrence is systematically higher than it ought to be. Thus one might wish to weight NAO− forecasts lower than NAO+ forecasts in order to avoid too many false positives. It will in general be important for users to understand the reliability of any given CT forecast.

A related point here is the existence of so-called windows of opportunity, meaning situations where the forecast uncertainty is lower than usual and predictability is higher (Spaeth et al. 2024). This could in theory allow users to amplify the weights given to a forecast at particular lead-times depending on whether the conditions of the window of opportunity are met.

Finally, we note that a scenario-based approach would be highly complementary to our methodology, especially for risk-averse users. In such an approach, rather than simply presenting the most likely CT forecast (i.e., the majority CT), one presents the estimated probability for any of the available CTs occurring. For example, in the case of the 2-state NAO regimes, one might report that there is a 55% chance of an NAO+ regime with less than average demand, and a 45% possibility of an NAO− regime with higher than average demand. If the impacts of any of the lower probability outcomes are deemed significant, the user could thereby still decide to take action. Such an approach is, for example, used in the UK Met Office 'Decider' product.

5.3 | Focusing on Extreme Outcomes

Many users may be more concerned about flagging an increased risk of 'extreme days'. For example, a big concern in forecasts of energy demand is to provide early warning for days where demand might be very high and renewable energy availability very low. Circulation type methods have proven themselves able to shed insight on such events (Mockert et al. 2023), meaning the methodology we have described here could also be used for such events. However, the full mutual information content may be a less helpful diagnostic here, since the contribution from changes to the tails of the distribution is given low weight by construction.

The key notion from information theory is that the quantity $-\log\left(\rho_X(x)\right)$ is the 'correct' way to measure the information present in a particular outcome. Therefore, if one's main concern are events above a particular threshold u say, then in place of $MI(X,Y)$ one could consider how the quantity $-\log(\mathbb{P}(X>u))$ changes when one conditions X on a set of CTs. CTs which produce a large change in this quantity should be ones which are most effective at flagging changes to the risk of exceeding the threshold, though we have not explored this quantitatively here.

5.4 | Constructing Maximally Informative Circulation Types

A typical limitation of CT forecasts is that the CTs are simply not all that informative. That is, conditioning climatology on a CT often leads to a pdf which is only marginally different. Our case study of GB demand shows an example of this. The conditioned pdfs are not much sharper than climatology (Figure 6), and this is reflected in the fact that the magnitudes of the mutual information contents are small, with the largest value being only 0.12 for DNW using MO30 patterns (a value 0 is no information and a value of 1 is perfect information).

It has previously been shown that if one constructs circulation types by clustering the target variables themselves (as opposed to circulation variables like pressure), then one can obtain circulation types that are more closely linked to the target (Bloomfield et al. 2019). It is natural to ask if one can do even better by constructing types that explicitly attempt to maximise the mutual information content between the types and one's target. Such circulation types would by construction highlight the circulation types that are associated with the strongest deviations from climatology. One potential way to approach this construction is to perform the clustering using k-medoids (Kaufman and Rousseeuw 1990), as opposed to k-means. This is because k-medoids allows one to use an arbitrary distance function to create the clusters, and it may thus be possible to use a distance function based on mutual information content.

6 | Conclusions and Outlook

The goal of this study was to answer three questions. Firstly, how does one measure possible trade-offs between informativity and predictability in circulation type forecasts, and do such trade-offs actually occur? Secondly, what is a principled method of selecting the best circulation type framework, or combining multiple frameworks? Thirdly, how should one measure the informativity of a circulation type framework?

To answer these questions, we have presented a general framework for combining multiple circulation type frameworks in order to create a hybrid forecast which optimally balances informativity and predictability. Specifically, the expected value of a circulation type forecast at lead-time t is maximised when the product $V_{\mathrm{R}}\left(p_t - p_{clim}\right)$ is maximised, where V_R measures how informative circulation type framework R is, p_t is the hit-rate of the forecast at lead-time t and p_{clim} is the climatological hit-rate. The optimal combination of frameworks is the one which swaps from one framework to the next according to which framework maximises this product at a given t, as demonstrated conceptually in Figure 5. This product therefore gives a precise way to measure the trade-off between informativity and predictability. Our application to GB demand suggests that such trade-offs do occur in practice. This therefore answers the first question, whilst our proposed methodology for optimal combinations answers the second question. We also explained why, for a circulation type forecast based on conditioning a climatological pdf, mutual information content is a natural way to estimate V_R, which answers the third question. Finally, we highlighted several ways in which the method could be modified to better suit users needs.

Our application to the case of forecasting energy demand in Great British winters suggests that the MO30 patterns used in

the UK Met Office's 'Decider' product are around 4.5 times more informative than the simpler NAO regimes, but that the more extended predictability of the NAO means that MO30 should only be used out to day 12, after which it is better to swap to NAO regimes. There is some marginal gain from using NAO forecasts out to around day 90, consistent with earlier studies (Clark et al. 2017). We argue that stakeholders exposed to a changing climate may derive particular benefit from utilising this extended range information, which can allow for an early assessment of risk exposure and consequently better forward planning. However, we also emphasised that for days 1–7, the gridpoint forecast is very likely superior to a CT forecast, and for days 7–12, one is likely better off using a gridpoint forecast conditioned on the MO30 patterns, similar to the approach taken in (Bloomfield et al. 2021). We point out here that the conclusion to swap from MO30 to something else at day 12 is independent of the particular application, because it is purely a consequence of the loss of forecast skill.

It might be argued that swapping from the 30 MO30 patterns to the 2 NAO regimes is a somewhat brutal transition. We tested whether the intermediate set of 8 MO30 patterns considered in (Neal et al. 2016) is an improvement in this regard. However, more than 90% of the time in boreal winter the actually occurring patterns of these 8 intermediates are the two corresponding to the NAO+ and NAO−. Thus these are functionally equivalent to just using NAO regimes. Other plausible options for an intermediate set of CTs could be the classical four circulation regimes (Vautard 1990) or the more stable set of three geopotential jet-regimes (Dorrington and Strommen 2020). The targeted circulation types of (Bloomfield et al. 2019) are also an obvious possibility. As mentioned in the discussion, it is also of clear interest to assess whether a new bespoke set of CTs can be constructed which maximises the mutual information content with GB demand or DNW, or indeed with an arbitrary target variable.

Our quantitative results depended on our choice of forecast model. In particular, the fact we found it is worthwhilst to swap to NAOs at all depends entirely on the existence of seasonal forecast skill in the ECMWF model: this skill is not shared by all models (Athanasiadis et al. 2017). More skillful forecasts will be able to access more information for a wider range of lead times, so it is always best to identify and use the NWP forecast which exhibits the highest skill for one's area of interest, or even to pool multiple forecasts together into a larger multi-model forecast. It is also important to understand biases in any forecast model used. For example, understanding when the model is overconfident is crucial for avoiding false positives. In the case of circulation types, another relevant bias is that models sometimes underestimate how persistent certain CTs tend to be (Strommen and Palmer 2018), which is important because the biggest risks can be associated with particularly persistent CT events (e.g., particularly long periods of high demand and low winds).

Finally, the use of NWP forecasts for a given application can sometimes be limited by the horizontal resolution of the forecast. For example, UK floods are tightly linked to the precipitation happening at individual catchments, and lower resolution models will not be able to accurately resolve what happens at the catchment level. A future avenue of work is therefore to study the value of high-resolution models for meteorological applications, something we intend to pursue as part of the EERIE Project (doi: 10.3030/101081383).

Author Contributions

Kristian Strommen: conceptualization, investigation, writing – original draft, methodology, visualization, writing – review and editing. **Hannah M. Christensen:** conceptualization, funding acquisition, writing – review and editing, visualization, project administration. **Hannah C. Bloomfield:** investigation, methodology, visualization, writing – review and editing, data curation.

Acknowledgements

The authors express their gratitude to NESO, in particular Daniel Drew and Benjamin Sloman, for their expert input and encouragement. We also thank Robert Neal and the UK Met Office for sharing data and participating in discussions on the results. Two anonymous reviewers helped improved the manuscript greatly. The ECMWF forecasts used were carried out as part of an ECMWF Special Project, which the lead author gratefully acknowledges. This work also makes use of S2S data. S2S is a joint initiative of the World Weather Research Programme (WWRP) and the World Climate Research Programme (WCRP). The original S2S database is hosted at ECMWF as an extension of the TIGGE database. This publication is part of the EERIE project (Grant Agreement No. 101081383) funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. University of Oxford's contribution to EERIE is funded by UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee, grant number 10049639. Kristian Strommen and Hannah M. Christensen acknowledge funding from this grant. Hannah M. Christensen was further funded by Natural Environment Research Council grant number NE/P018238/1. Hannah C. Bloomfield was funded by a Newcastle University Academic Track Fellowship.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are openly available in Zenodo (Decider vs NAO Regime Forecasts (Version 1). [Dataset] 2025) at https://doi.org/10.5281/zenodo.15001098.

References

Ansell, T. J., P. D. Jones, R. J. Allan, et al. 2006. "Daily Mean Sea Level Pressure Reconstructions for the European–North Atlantic Region for the Period 1850–2003." *Journal of Climate* 19, no. 12: 2717–2742. https://doi.org/10.1175/jcli3775.1.

Athanasiadis, P. J., A. Bellucci, A. A. Scaife, et al. 2017. "A Multisystem View of Wintertime NAO Seasonal Predictions." *Journal of Climate* 30, no. 4: 1461–1475. https://doi.org/10.1175/jcli-d-16-0153.1.

Baez, J. 2024. What is Entropy. https://arxiv.org/abs/2409.09232.

Beck, C., and A. Philipp. 2010. "Evaluation and Comparison of Circulation Type Classifications for the European Domain." *Physics and Chemistry of the Earth, Parts A, B, and C* 35, no. 9–12: 374–387. https://doi.org/10.1016/j.pce.2010.01.001.

Bloomfield, H. C., D. J. Brayshaw, and A. J. Charlton-Perez. 2019. "Characterizing the Winter Meteorological Drivers of the European

Electricity System Using Targeted Circulation Types." *Meteorological Applications* 27, no. 1: e1858. https://doi.org/10.1002/met.1858.

Bloomfield, H. C., D. J. Brayshaw, P. L. M. Gonzalez, and A. Charlton-Perez. 2021. "Pattern-Based Conditioning Enhances Sub-Seasonal Prediction Skill of European National Energy Variables." *Meteorological Applications* 28, no. 4: e2018. https://doi.org/10.1002/met.2018.

Cassou, C. 2008. "Intraseasonal Interaction Between the Madden–Julian Oscillation and the North Atlantic Oscillation." *Nature* 455, no. 7212: 523–527. https://doi.org/10.1038/nature07286.

Chang, A. Y.-Y., K. Bogner, C. M. Grams, S. Monhart, D. I. V. Domeisen, and M. Zappa. 2023. "Exploring the Use of European Weather Regimes for Improving User-Relevant Hydrological Forecasts at the Subseasonal Scale in Switzerland." *Journal of Hydrometeorology* 24, no. 10: 1597–1617. https://doi.org/10.1175/jhm-d-21-0245.1.

Clark, R. T., P. E. Bett, H. E. Thornton, and A. A. Scaife. 2017. "Skilful Seasonal Predictions for the European Energy Industry." *Environmental Research Letters* 12, no. 2: 024002. https://doi.org/10.1088/1748-9326/aa57ab.

Cover, T. M., and J. A. Thomas. 2005. *Elements of Information Theory*. Wiley. https://doi.org/10.1002/047174882x.

Dawson, A., T. N. Palmer, and S. Corti. 2012. "Simulating Regime Structures in Weather and Climate Prediction Models." *Geophysical Research Letters* 39, no. 21: L21805. https://doi.org/10.1029/2012gl053284.

Decider vs NAO Regime Forecasts (Version 1). [Dataset]. 2025. Zenodo. 10.5281/zenodo.15001098.

Dorrington, J., and K. J. Strommen. 2020. "Jet Speed Variability Obscures Euro-Atlantic Regime Structure." *Geophysical Research Letters* 47, no. 15: e2020GL087907. https://doi.org/10.1029/2020gl087907.

Dunstone, N., D. Smith, A. Scaife, et al. 2016. "Skilful Predictions of the Winter North Atlantic Oscillation One Year Ahead." *Nature Geoscience* 9, no. 11: 809–814. https://doi.org/10.1038/ngeo2824.

Eade, R., D. Smith, A. Scaife, et al. 2014. "Do Seasonal-To-Decadal Climate Predictions Underestimate the Predictability of the Real World?" *Geophysical Research Letters* 41, no. 15: 5620–5628. https://doi.org/10.1002/2014gl061146.

ERA5 Copernicus Climate Change Service (C3S) Climate Data Store (CDS). 2023. Monthly Averaged Data on Single Levels From 1940 to Present. 10.24381/cds.f17050d7.

ERA5 derived time series of European country-aggregate electricity demand, wind power generation and solar power generation: hourly data from 1979–2019. [Dataset]. 2020. University of Reading. 10.17864/1947.272.

Ferranti, L., S. Corti, and M. Janousek. 2015. "Flow-Dependent Verification of the ECMWF Ensemble Over the Euro-Atlantic Sector." *Quarterly Journal of the Royal Meteorological Society* 141, no. 688: 916–924. https://doi.org/10.1002/qj.2411.

Garrido-Perez, J. M., C. Ordóñez, D. Barriopedro, R. Garca-Herrera, and D. Paredes. 2020. "Impact of Weather Regimes on Wind Power Variability in Western Europe." *Applied Energy* 264: 114731. https://doi.org/10.1016/j.apenergy.2020.114731.

Grams, C. M., R. Beerli, S. Pfenninger, I. Staffell, and H. Wernli. 2017. "Balancing Europe's Wind-Power Output Through Spatial Deployment Informed by Weather Regimes." *Nature Climate Change* 7, no. 8: 557–562. https://doi.org/10.1038/nclimate3338.

Grams, C. M., L. Magnusson, and L. Ferranti. 2020. "How to Make Use of Weather Regimes in Extended-Range Predictions for Europe." *ECMWF Newsletter* 165: 14–19. https://www.ecmwf.int/en/newsletter/165/meteorology/how-make-use-weather-regimes-extended-range-predictions-europe.

Hannachi, A., D. M. Straus, C. L. E. Franzke, S. Corti, and T. Woollings. 2017. "Low-Frequency Nonlinearity and Regime Behavior in the Northern Hemisphere Extratropical Atmosphere." *Reviews of Geophysics* 55, no. 1: 199–234. https://doi.org/10.1002/2015rg000509.

Hendry, A., I. D. Haigh, R. J. Nicholls, et al. 2019. "Assessing the Characteristics and Drivers of Compound Flooding Events Around the UK Coast." *Hydrology and Earth System Sciences* 23, no. 7: 3117–3139. https://doi.org/10.5194/hess-23-3117-2019.

Hersbach, H., B. Bell, P. Berrisford, et al. 2020. "The ERA5 Global Reanalysis." *Quarterly Journal of the Royal Meteorological Society* 146, no. 730: 1999–2049. https://doi.org/10.1002/qj.3803.

Hurrell, J. W., Y. Kushnir, G. Ottersen, and M. Visbeck. 2003. "An Overview of the North Atlantic Oscillation." In *The North Atlantic Oscillation: Climatic Significance and Environmental Impact*, 1–35. American Geophysical Union. https://doi.org/10.1029/134gm01.

IFS Documentation. n.d. IFS Documentation. https://www.ecmwf.int/en/publications/ifs-documentation.

Judt, F. 2020. "Atmospheric Predictability of the Tropics, Middle Latitudes, and Polar Regions Explored Through Global Storm-Resolving Simulations." *Journal of the Atmospheric Sciences* 77, no. 1: 257–276. https://doi.org/10.1175/jas-d-19-0116.1.

Kaufman, L., and P. J. Rousseeuw. 1990. *Finding Groups in Data*. Wiley. https://doi.org/10.1002/9780470316801.ch2.

Krishnamurthy, V. 2019. "Predictability of Weather and Climate." *Earth and Space Science* 6, no. 7: 1043–1056. https://doi.org/10.1029/2019ea000586.

Lavaysse, C., J. Vogt, A. Toreti, M. L. Carrera, and F. Pappenberger. 2018. "On the Use of Weather Regimes to Forecast Meteorological Drought Over Europe." *Natural Hazards and Earth System Sciences* 18, no. 12: 3297–3309. https://doi.org/10.5194/nhess-18-3297-2018.

Lorenz, E. N. 1969. "The Predictability of a Flow Which Possesses Many Scales of Motion." *Tellus A: Dynamic Meteorology and Oceanography* 21, no. 3: 289. https://doi.org/10.3402/tellusa.v21i3.10086.

Michelangeli, P.-A., and R. Vautard. 1995. "Bernard Legras. Weather Regimes: Recurrence and Quasi Stationarity." *Journal of the Atmospheric Sciences* 52, no. 8: 1237–1256. https://doi.org/10.1175/1520-0469(1995)052<1237:wrraqs>2.0.co;2.

Millin, O. T., J. C. Furtado, and C. Malloy. 2024. "The Impact of North American Winter Weather Regimes on Electricity Load in the Central United States." *npj Climate and Atmospheric Science* 7, no. 1: 254. https://doi.org/10.1038/s41612-024-00803-1.

Mockert, F., C. M. Grams, T. Brown, and F. Neumann. 2023. "Meteorological Conditions During Periods of Low Wind Speed and Insolation in Germany: The Role of Weather Regimes." *Meteorological Applications* 30, no. 4: e2141. https://doi.org/10.1002/met.2141.

Neal, R., D. Fereday, R. Crocker, and R. E. Comer. 2016. "A Flexible Approach to Defining Weather Patterns and Their Application in Weather Forecasting Over Europe." *Meteorological Applications* 23, no. 3: 389–400. https://doi.org/10.1002/met.1563.

Neal, R., J. Robbins, R. Crocker, et al. 2024. "A Seamless Blended Multi-Model Ensemble Approach to Probabilistic Medium-Range Weather Pattern Forecasts Over the UK." *Meteorological Applications* 31, no. 1: e2179. https://doi.org/10.1002/met.2179.

Osman, M., R. Beerli, D. Büeler, and C. M. Grams. 2023. "Multi-Model Assessment of Sub-Seasonal Predictive Skill for Year-Round Atlantic–European Weather Regimes." *Quarterly Journal of the Royal Meteorological Society* 149, no. 755: 2386–2408. https://doi.org/10.1002/qj.4512.

Palmer, T. N. 1999. "A Nonlinear Dynamical Perspective on Climate Prediction." *Journal of Climate* 12, no. 2: 575–591. https://doi.org/10.1175/1520-0442(1999)012<0575:andpoc>2.0.co;2.

Richardson, D., H. J. Fowler, C. G. Kilsby, and R. Neal. 2017. "A New Precipitation and Drought Climatology Based on Weather Patterns."

International Journal of Climatology 38, no. 2: 630–648. https://doi.org/10.1002/joc.5199.

Richardson, D., R. Neal, R. Dankers, et al. 2020. "Linking Weather Patterns to Regional Extreme Precipitation for Highlighting Potential Flood Events in Medium: To Long-Range Forecasts." *Meteorological Applications* 27, no. 4: e1931. https://doi.org/10.1002/met.1931.

Rouges, E., L. Ferranti, H. Kantz, and F. Pappenberger. 2023. "European Heatwaves: Link to Large-Scale Circulation Patterns and Intraseasonal Drivers." *International Journal of Climatology* 43, no. 7: 3189–3209. https://doi.org/10.1002/joc.8024.

Savitzky, A., and M. J. E. Golay. 1964. "Smoothing and Differentiation of Data by Simplified Least Squares Procedures." *Analytical Chemistry* 36, no. 8: 1627–1639. https://doi.org/10.1021/ac60214a047.

Scaife, A. A., A. Arribas, E. Blockley, et al. 2014. "Skillful Long-Range Prediction of European and North American Winters." *Geophysical Research Letters* 41, no. 7: 2514–2519. https://doi.org/10.1002/2014gl059637.

Scaife, A. A., and D. Smith. 2018. "A Signal-to-Noise Paradox in Climate Science." *npj Climate and Atmospheric Science* 1, no. 1: 28. https://doi.org/10.1038/s41612-018-0038-4.

Schiemann, R., and C. Frei. 2010. "How to Quantify the Resolution of Surface Climate by Circulation Types: An Example for Alpine Precipitation." *Physics and Chemistry of the Earth, Parts A/B/C* 35, no. 9–12: 403–410. https://doi.org/10.1016/j.pce.2009.09.005.

Siegert, S., D. B. Stephenson, P. G. Sansom, A. A. Scaife, R. Eade, and A. Arribas. 2016. "A Bayesian Framework for Verification and Recalibration of Ensemble Forecasts: How Uncertain Is NAO Predictability?" *Journal of Climate* 29, no. 3: 995–1012. https://doi.org/10.1175/jcli-d-15-0196.1.

Spaeth, J., P. Rupp, M. Osman, C. M. Grams, and T. Birner. 2024. "Flow-Dependence of Ensemble Spread of Subseasonal Forecasts Explored via North Atlantic-European Weather Regimes." *Geophysical Research Letters* 51, no. 14: e2024GL109733. https://doi.org/10.1029/2024GL109733.

Strommen, K., M. Chantry, J. Dorrington, and N. Otter. 2022. "A Topological Perspective on Weather Regimes." *Climate Dynamics* 60, no. 5–6: 1415–1445. https://doi.org/10.1007/s00382-022-06395-x.

Strommen, K., M. MacRae, and H. Christensen. 2023. "On the Relationship Between Reliability Diagrams and the "Signal-To-Noise Paradox"." *Geophysical Research Letters* 50, no. 14: e2023GL103710. https://doi.org/10.1029/2023gl103710.

Strommen, K., and T. N. Palmer. 2018. "Signal and Noise in Regime Systems: A Hypothesis on the Predictability of the North Atlantic Oscillation." *Quarterly Journal of the Royal Meteorological Society* 145, no. 718: 147–163. https://doi.org/10.1002/qj.3414.

Thornton, H. E., A. A. Scaife, B. J. Hoskins, et al. 2019. "Skilful Seasonal Prediction of Winter Gas Demand." *Environmental Research Letters* 14, no. 2: 024009. https://doi.org/10.1088/1748-9326/aaf338.

Vautard, R. 1990. "Multiple Weather Regimes Over the North Atlantic: Analysis of Precursors and Successors." *Monthly Weather Review* 118, no. 10: 2056–2081. https://doi.org/10.1175/1520-0493(1990)118<2056:mwrotn>2.0.co;2.

Vitart, F., C. Ardilouze, A. Bonet, et al. 2017. "The Subseasonal to Seasonal (S2S) Prediction Project Database." *Bulletin of the American Meteorological Society* 98, no. 1: 163–173. https://doi.org/10.1175/bams-d-16-0017.1.

Wanner, H., S. Brönnimann, C. Casty, et al. 2001. "North Atlantic Oscillation—Concepts and Studies." *Surveys in Geophysics* 22, no. 4: 321–381. https://doi.org/10.1023/a:1014217317898.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1:** Supporting Information.