Exploring the causes and consequences of phenological change in a wild bird population

A thesis submitted for the degree of Doctor of Philosophy

By Emily G. Simmonds

University of Oxford
St Cross College

Trinity Term 2017

Abstract

Exploring the causes and consequences of phenological change in a wild bird population

DPhil thesis by Emily G. Simmonds, St Cross College, submitted Trinity Term 2017

Changes in climate shape biological populations. They can alter spatial distributions, the timing of life history events, and even the species themselves. We are now experiencing a period of rapid directional climate change, alongside seasonal fluctuations. This thesis investigates temporal changes in life history events, phenology, as a climate response. I explore the causes and population level consequences of change in breeding phenology of two wild bird populations from Wytham Woods, UK.

I test how great tits (*Parus major*) and blue tits (*Cyanistes caeruleus*) achieve temporal synchrony between the peak demands of their breeding and the peak abundance of their prey species (winter moth caterpillars - *Operophtera brumata*) in an inter-annually variable environment. I demonstrate great tit (*Parus major*) incubation behaviour fine-tunes the timing of hatching in response to ambient temperatures right up until hatching (**Chapter two**). Temperatures within the nest box, however, appear to play little role in the breeding phenology of blue tits (*Cyanistes caeruleus*) (**Chapter two**). I discuss the merits and limitations of statistical approaches for cue identification (**Chapter six**), finding the method and time period of data used both affect the cue identified and predictive accuracy.

The second part of this thesis explores the influence of a directionally changing environment on great tit phenology. I use an integral projection model (IPM) to predict population dynamics over the 21st century, showing that if the cues used by both interacting species change sufficiently closely, temporal synchrony can be maintained through both phenotypic plasticity and micro-evolution (**Chapter 5**). However, if the cues diverge mismatch will arise (**Chapter 5**) causing population declines when certain thresholds are passed (**Chapter 4**).

This work contributes to understanding how phenological synchrony is achieved, how it might change in the future, and its population impacts. In Wytham Woods it appears that great tits have a great deal of flexibility in multiple components of their breeding cycle, allowing them to retain synchrony with their caterpillar prey in a fluctuating environment. These birds are relatively resilient to negative phenological impacts from climate change. Only if the cues used by the predator and prey completely diverge do we predict consistent declines in population size during this century.

Acknowledgements

I would like to take a chance to thank all of the people who have made this thesis possible. To my supervisors Ella Cole, Tim Coulson, and Ben Sheldon, thank you for your support and guidance throughout my DPhil and for challenging me to think as both a modeller and a field biologist. I would like to thank Keith McMahon, Will Hayward, and Andy Gosler for your patience and positivity in teaching me bird ringing, despite the early mornings. None of this work could have been possible without you. For their integral role in keeping an exceptional dataset going, I thank all of those who contributed to the Wytham Woods fieldwork over the years. I would like to give a special mention to those who overlapped with me; Keith McMahon for being a master at fixing faceplates and changing ice packs, Stephen Lang, Fraser Bell, Koosje Lamers, Nico Alioravainen and Will Hayward for your smiling faces that brightened my tired days, shared curiosities of breeding season life, and your help in my strange data collection ideas, Antica Culina for helping me to survive my first breeding season and for all of the lovely breakfasts that followed, and to Nicole Milligan for the beautiful maps, showing me the ropes, and for being a true inspiration for good fieldwork practice.

Next I would like to thank those that supported my indoor working life. For always ensuring I had the support and resources to conduct this work I would like to thank Heather Green, Andy Dixon, the whole of the Zoology IT team, John Hogg, and Tony Price. For organising our group meetings, social events, and pinning down my supervisors diaries I thank Lynne Bradley, Esme Taylor, and Sharon Cornwell. I also couldn’t have completed this thesis without the moral support and friendship provided

by both of my lab groups, E2D and the Sheldon Group, I cannot express enough my thanks to all those who I have shared my time at Oxford with. I would like to give a special mention to; Erik Sandvig, Freddy Hillemann, Josie Mahony, Leejiah Dorward, Rose Kentie, Claudio Quilodran, Marie-Claire Koschowitz, Jacques Deere, Severin Dreßen, Jeremy Cusack, Rocio Pozo, Jack Massey, and Marc Brouard, for all of the Thursday pub trips, weekday coffee breaks, weekend brunches, delicious homebrew, catch up teas, and generally keeping me going.

There are a few people that have put up with me more than most. Lara Veylit thank you for being there for me even when times were tough. You've been an endless source of support, always with an infectious smile and thirst for life. You are inspiring in more ways than you know, you were a truly wonderful housemate, and an amazing friend! Martin Cusack, you put up with me for more than two years, well done. Thank you, I know you wouldn't want me to say anything more. To Susanne Vogel, my kindred spirit, thank you for all of the times you've let me moan at you, for constantly helping me to see the good in things, for believing in me, and for the care and support you give me. It has all been appreciated more than you can know. Jenny Chapman, Luci Kirkpatrick, and Michelle Mclean, thank you all for keeping me laughing no matter what, for giving me excuses for fancy dress and holidays to exotic places.

Last but not least I'd like to thank my family for always being there for me, talking to me when I'm walking places, reminding me I am more than just an researcher, listening to me talk about birds and modelling, giving me places to retreat to for writing, running R scripts on my big computer, sending motivational texts, and generally helping me to believe in myself. To Jack Laverick, I cannot thank you

enough for all you have done for me during my thesis. Thank you for cooking me dinner, making my lunch, being my sounding post, collecting my helmet when I forget it, picking me up when I am down, apparently having unwavering faith that I can do this, for preventing me from rambling on forever or misplacing commas (shame you can't do that here), being overly picky with my figures, putting up with me working late, caring for me when I get ill, and generally for just being your wonderful self. You mean the world to me, thank you.

The work undertaken in this thesis was funded by the Natural Environment Research Council.

Author contributions

I declare that this thesis, *Exploring the causes and consequences of phenological change in a population of wild birds* is my own work. The contribution of other authors to any of the manuscripts is clearly stated below. None of the work included has been submitted, in whole or in part, in any previous application for a degree.

Chapter one

Authors: Emily G. Simmonds

Chapter two

Authors: Emily G. Simmonds (EGS), Ben C. Sheldon (BCS), Tim Coulson (TC), and Ella F. Cole (EFC)

EGS generated the research idea, designed and implemented data collection, analysed results, and wrote of the manuscript. EFC and BCS gave guidance on data collection. EFC, BCS, and TC contributed to the interpretation of results and commented on various stages of the manuscript.

Chapter three

Authors: Emily G. Simmonds (EGS), Ben C. Sheldon (BCS), Tim Coulson (TC), and Ella F. Cole (EFC)

EGS generated the research idea, designed and implemented data collection, analysed results, and wrote of the manuscript. EFC, BCS, and TC gave guidance on experimental design and contributed to interpretation of results. EFC and TC commented on various stages of the manuscript.

Chapter four

Authors: Emily G. Simmonds (EGS), Ella F. Cole (EFC), and Tim Coulson (TC)

EGS generated the research idea, constructed the model, analysed results, and wrote of the manuscript. TC gave guidance on methodological approach and model construction. EFC and TC contributed to interpretation of results and commented on various stages of the manuscript.

Chapter five

Authors: Emily G. Simmonds (EGS), Ella F. Cole (EFC), and Tim Coulson (TC)

EGS generated the research idea, constructed the model, analysed results, and wrote of the manuscript. TC gave guidance on methodological approach and model construction. EFC and TC contributed to interpretation of results and commented on various stages of the manuscript.

Chapter six

Authors: Emily G. Simmonds (EGS), Ben C. Sheldon (BCS), and Ella F. Cole (EFC)

EGS generated the research idea, conducted statistical analyses, analysed results, and wrote of the manuscript. EFC and BCS contributed to interpretation of results and commented on various stages of the manuscript.

Chapter seven

Authors: Emily G. Simmonds

Table of Contents

CHAPTER ONE

General introduction

Emily G. Simmonds

1.1. Living in a seasonal environment

Changes in climate shape biological populations. They can alter spatial distributions, the timing of life history events, and drive evolutionary change. Temperate regions are characterised by strong seasonality, where the climate changes throughout the year, presenting individuals in such environments with brief windows of optimum conditions for completing energetically demanding life history events i.e. flowering, breeding, moulting, hibernation or migration (Durant et al. 2007; Lack 1968; Perrins 1970b; Verhulst & Tinbergen 1991; Thackeray et al. 2010; Cleland et al. 2007). A failure to time these life history events with peak resource availability results in energetic demands which cannot be met, reducing individual fitness (Visser et al. 2011; Lane et al. 2012; van Asch 2007; Charmantier et al. 2008; Plard et al. 2014; Reed, Grøtan, et al. 2013; Reed, Jenouvrier, et al. 2013; Visser & Both 2005).

The majority of temperate species have some need to synchronise energetic demands with peak resource availability as resources fluctuate through the year. As the timing of peak resource availability shifts inter-annually with stochastic environmental variation (Saether et al. 2002; van Asch et al. 2007; van Asch & Visser 2007), this poses an adaptive challenge. Temperate species must therefore also shift the timing of their life history events, phenology, inter-annually. This is achieved through phenotypic plasticity (Charmantier et al. 2008; Gienapp et al. 2014; Charmantier & Gienapp 2014; Vedder et al. 2013).

Phenotypic plasticity, the same genotype producing different phenotypes in different environments (Pigliucci 2001), typically evolves where there is spatial or temporal variability in environmental conditions such that a different phenotype will have higher fitness in each environment (Phillimore et al. 2012). This is the case for phenology,

where the timing of peak resources varies in space and also between years. For species that live for multiple years, or disperse across different habitats, phenotypic plasticity can allow phenology to track optimum conditions over time and space during individual lifetimes (Tansey et al. 2017). The continuous relationship between a phenotype and the environment is termed the reaction norm (Woltereck 1909). The reaction norm for phenology can either take the form of a direct relationship between the environment and physiology, as occurs in plant and many invertebrate species (Bonhomme 2000; Kramer 1995; Chuine 2000; Rötzer et al. 2004), or indirectly (Visser et al. 2010). Endotherms, such as mammals and birds, have development that is buffered from direct environmental relationships (McNab 2012; Khaliq et al. 2014). These species must use proxy cues to predict when the period of peak resource availability will occur in a given year. For species occupying temperate environments, temperature changes appear to be the primary driver of plasticity in phenology (Menzel et al. 2006; Crick 2004; Parmesan 2006; Durant et al. 2007; Visser et al. 2004). The challenge of achieving synchrony with a resource peak in an inter-annually fluctuating environment is heightened for species in higher trophic levels. For these individuals, the resource they rely on is another species, which is itself responding to stochastic temperature variation, either directly or indirectly. The consumer species must therefore not just track the environmental changes, but track them in the same way as their prey species to achieve synchrony between peak abundance and peak requirement across years.

Phenotypic plasticity appears to ensure inter-annual synchrony for many pairs of species (Charmantier et al. 2008; Sparks & Yates 1997; Bourgault et al. 2010). However, exact synchrony is not always the optimal strategy. In some circumstances asynchrony between interacting species could be adaptive (Singer & Parmesan

2010; Johansson, Kristensen, et al. 2015). Life history trade-offs between survival and reproductive success or interactions between density dependence and competitive ability can all lead to scenarios when complete synchrony is not the optimal strategy (as detailed theoretically in (Johansson, Kristensen, et al. 2015)). In either a scenario of adaptive synchrony or adaptive asynchrony, individuals can stably achieve some level of matching to their resource peak in a stochastically fluctuating environment. This permits the energetic demands of reproduction to be satisfied, maximising reproductive success (Reed, Jenouvrier, et al. 2013; Visser et al. 2006; Simmonds et al. 2017).

1.2. Phenological synchrony and climate change

Stochastic environmental variation is not the only variation to which temperate species must adapt. Systematic environmental change can also alter the timing of resource peaks. We are currently experiencing a period of rapid and directional climate change accompanied by increases in variability (Mann et al. 2008), driven by anthropogenic greenhouse gas emissions (Mann et al. 2008; IPCC 2013). This change is projected to continue well into the next century (IPCC 2013) pushing species into novel climatic environments, which may not have been experienced for many thousands of generations. Whether phenotypic plasticity, which evolved to cope with stochastic environmental variation, is sufficient to keep pace with directional climate change is a key question.

Advances of phenology in response to rising temperatures have already been observed across taxa and geographic regions (Durant et al. 2007; Menzel et al. 2006; Sparks & Crick 1999; Thackeray et al. 2010; Thackeray et al. 2016; Dunn & Winkler 1999), resulting from phenotypic plasticity (Merilä & Hendry 2014). However,

the rates of change have not been uniform across groups, species, or populations; for example 22 % of plant and animal species in Europe showed no advance in timing of spring phenologies (Menzel et al. 2006). Even among species and populations that do advance, the rate at which this occurs can be highly variable (Both & Visser 2005; Parmesan & Yohe 2003). Often a lag exists in the advance of secondary consumers relative to primary producers and consumers, particularly in terrestrial systems (Thackeray et al. 2010). This could lead to temporal trophic mismatches under directional climatic change (Cushing 1969) and fitness consequences for the consumer species, potentially influencing population dynamics (Plard et al. 2014).

Whether or not a population can achieve and maintain synchrony with its resource under both stochastic and systematic (directional) climate change is dependent on three key factors:

1. The relationship between the predictive cue used to determine phenology of the focal species and the cue used to determine resource phenology.

If the cues used by a pair of interacting species and their environmental sensitivity are sufficiently close, then as long as plasticity does not reach limits (more detail on the limits of plasticity is discussed below), synchrony can be maintained under directional climate change. However, in reality, pairs of species are likely to have different cues or have different environmental sensitivities (Parmesan 2006; Thackeray et al. 2010; Thackeray et al. 2016). This can lead to mismatch when the relationship between the two cues is altered under climate change. Lower trophic levels often have direct relationships between development and temperature, resulting in a continuous and immediate temperature response right up to the

phenological event. For instance, winter moth caterpillars (*Operophtera brumata*) have temperature linked growth right up to pupation (Buse et al. 1999). In contrast, species in higher trophic levels do not have this direct temperature link, often initiating reproduction several weeks before the peak resource demand. As a result, they must use predictive cues prior to the reproductive event to indicate the timing of peak resources, with only minimal chance to fine-tune after this time (van Noordwijk et al. 1995). As climate change acts heterogeneously in time and space (on Climate Change 2007), the cues used by interacting species can be altered to different extents, potentially driving differential responses in phenology. This has been hypothesised as a driver of increasing mismatch between great tits (*Parus major*) and their winter moth caterpillar prey in the Netherlands (Visser et al. 2004; Visser et al. 1998). It has also caused yellow-bellied marmots in the USA to advance emergence from hibernation more rapidly than their food plants (Inouye et al. 2000). Mismatch can also arise when one of a pair of interacting species employs a temporally invariant cue, such as primarily relying on photoperiod, while the other uses a variable environmental cue. This has been the case for a roe deer (*Capreolus capreolus*) population in France, where lack of advance in parturition dates of roe deer led them to lag behind their food resource (Plard et al. 2014).

2. The ability of individuals in a population to exhibit plasticity in the timing of life history events.

Even if the cues used by interacting species are shared, or change to the same degree under climate change, mismatch can still occur if responses to the cue differ. If one species exhibits a stronger response to the same cue, disruptions to synchrony will occur similar to that observed for great tits and yellow-bellied marmots detailed above (Visser et al. 2004; Visser et al. 1998; Inouye et al. 2000). The slower advance

of secondary consumers in terrestrial systems could result from weaker reaction norms in this trophic level, relative to lower trophic levels (Thackeray et al. 2010; Thackeray et al. 2016). Such a difference will always produce mismatch, as shown theoretically by Gienapp et al (Gienapp et al. 2014).

The ability of phenotypic plasticity in phenology to track optimal environmental conditions can also reach limits when pushed into new environments. Plasticity and reaction norms that are adaptive in one environment are not always adaptive in a new one, potentially moving the phenotype further from the new optimum rather than closer (Ghalambor et al. 2007; Visser 2008). Under these circumstances evolution on the reaction norm is required to maintain adaptive plasticity (Gienapp et al. 2008). This could occur for great tit species who have evolved plasticity to achieve temporal synchrony in a fluctuating environment. Under climate change the advance of breeding in warmer temperatures may be maladaptive if species in lower trophic levels change differently. This could be the case for the pedunculate oak (*Quercus robur*) which might delay its phenology as winters warm and cooling requirements for budburst are no longer achieved (van Asch & Visser 2007). Further to this, there are also limits and costs to phenotypic plasticity (DeWitt et al. 1998; Auld et al. 2009; Murren et al. 2015) which can be reached through environmental change. Physiological or developmental limits (DeWitt et al. 1998) occur when particular processes are not able to operate under certain conditions, such as body size cannot increase infinitely. Furthermore, plasticity in one aspect of life history could be bounded by trade-offs with other life history events (Johansson, Kristensen, et al. 2015) e.g. breeding timing cannot occur before spring migration in migratory birds. Both of these limits result in phenotypic plasticity occurring within certain bounds but then phenotypes ceasing to change beyond a certain point. Changing environmental

conditions could also limit plastic responses by altering the balance between costs and benefits of plasticity (Pigliucci 2001; DeWitt et al. 1998). If the phenological cues used are no longer a reliable indicator of the timing of peak resources, maintaining the sensory equipment to perceive them may be a fitness disadvantage (DeWitt et al. 1998). All of these processes act to limit the ability of phenotypic plasticity to be persistently adaptive across novel environments.

As climate change progresses, increasing both mean temperatures and temperature variability (IPCC 2013; MET Office 2009), the environmental conditions experienced by species will be pushed into ranges they have not experienced for many thousands of years, or ever. This current change is also occurring more rapidly than has been seen since at least 1880 (Mann et al. 2008). It is not known under these conditions whether phenotypic plasticity will continue to keep pace with resources or whether limits will be reached (DeWitt et al. 1998; Auld et al. 2009).

3. The ability of a population to undergo microevolution if the limit of plasticity is reached.

If the limits of plasticity are reached as novel conditions are encountered, synchrony could be maintained if micro-evolution can occur, either in phenology directly or on plasticity in phenology (Ghalambor et al. 2007). The ability for species to undergo evolution in their phenology at a pace fast enough to respond to anthropogenic climate change is contested. Several studies suggest that a strong selective pressure exists on the timing of reproduction (Charmantier et al. 2008; Gienapp et al. 2006). Others, indicate low heritability of phenological traits causing slow evolutionary responses (Gienapp et al. 2014; Chevin et al. 2010; Vedder et al. 2013). The role of phenotypic plasticity in evolutionary responses is also contentious (Ghalambor et al.

2007; Chevin & Lande 2015). Plastic responses could buffer a population against new environmental change, allowing them to persist in the environment and evolve (Robinson & Dukas 1999; Badyaev 2005; Chevin & Lande 2015) or plasticity could reduce evolution through selection on a non-heritable phenotypic trait (Price et al. 2003; Falconer 1990). If directional climate change continues for sufficiently long then some form of evolutionary phenological response may be required to shift reaction norms to suit the new environment (Visser 2008; Ghalambor et al. 2007). However, the extent to which this will be possible and how this will be influenced by phenotypic plasticity, is still unknown.

1.3. Predicting phenological responses

Responses to climate change that have already been observed differ by ecosystem (Thackeray et al. 2010), trophic levels (Thackeray et al. 2016), species (Crick 2004; Menzel et al. 2006), and even between populations of the same species (Husby et al. 2010; Gienapp et al. 2005). Predicting how phenology will change, where mismatches may occur, and what the population level consequences of such changes may be, are essential for building a picture of population dynamics under climate change. Such studies can highlight populations that may be at risk of future decline and help to direct management initiatives. To achieve these predictions, several steps are required; firstly understanding how current synchrony has been achieved and what drives phenology. Secondly, using this knowledge to predict the population level consequences of changes in phenology.

1.3.1. Understanding how phenological synchrony has been achieved in an inter-annually variable environment

In order to gain a full understanding of how plasticity aids synchrony we must identify the environmental cues that drive phenological change, while quantifying flexibility.

This will provide insights into the drivers of reproductive timing and allow predictions of future phenology.

To quantify plasticity in breeding phenology we must look beyond just easily observable events i.e the onset or completion of breeding (clutch initiation dates - e.g. (Charmantier et al. 2008; Visser et al. 1998; Nussey et al. 2005; Perrins 1965b), birth dates - e.g. (Plard et al. 2014), flowering dates - e.g. (Menzel et al. 2006)). Consideration of flexibility in different elements of the breeding cycle is needed to build a full picture of the mechanisms through which species can adjust to environmental changes. There is increasing evidence of high variability in post-initiation reproductive components such as incubation behaviour (Stenning 2008; Álvarez & Barba 2014; Lord et al. 2011; Cresswell & McCleery 2003; García-Navas & Sanz 2011; Matthysen et al. 2010; Hepp et al. 2006; Ardia et al. 2010), gestation periods (Asher et al. 2005; Moyes et al. 2011; Scott et al. 2008; Racey & Swift 1981), and clutch size (von Haartman 1969; Kluijver 1950; Balen & Balen van 1973; Lack 1955; Perrins 1965b; Perrins 1979; Lack 1958a; Haftorn 1981). Such variability, if linked to environmental drivers, could allow species to fine-tune their phenology. Plastic changes in one component of the breeding cycle may act to emphasise or depress changes in other components, all combining to influence the final phenology. Understanding how flexible different elements are while identifying their drivers are necessary steps to be able to predict how phenology will change into the future. A focus on only one element may miss the full picture and generate inaccurate predictions.

Understanding the limits of plasticity may be as important as understanding the extent of flexibility. As discussed above, plasticity that has evolved to deal with inter-

annually variable conditions will not always be optimal or operating to the same level in novel environmental conditions (Section 1.2). Deducing the limits of phenotypic plasticity is difficult to achieve. To accurately determine these limits experimental manipulations must be conducted, pushing individuals into novel environmental conditions. For the majority of animal species this is both ethically and logistically problematic, particularly with wild living vertebrates. These constraints make identifying the cues driving phenology a challenging endeavour. Temperature manipulations can be conducted in natural systems for plants and invertebrates, which tend to be more sedentary than vertebrate species. Phenology has been successfully manipulated in the wild for subalpine plant species (Dunne et al. 2003), aphids (*Acyrthosiphon svalbardicum*) and their host plant (*Dryas octopetala*) (Strathdee et al. 1995), and a tundra perennial herb (*Ranunculus nivalis*) (Henry & Molau 1997), to name a few examples.

Wild manipulations of phenology in vertebrate species have been attempted with varying success. While nocturnal incubation and incubation intensity were successfully altered through nest box heating in blue tits (*Cyanistes caeruleus*) and tree swallows (*Tachycineta bicolor*) (Vedder 2012; Ardia et al. 2009), heating and cooling of nest boxes failed to alter clutch initiation dates in great tits (Nager & van Noordwijk 1992). Captive experimentation offers an alternative to natural manipulations for hard-to-study species. Such studies have yielded valuable insights into cue-phenology relationships. Simulated warming of flowering plants (*Cardamine hirsuta*) resulted in an average delay to flowering of 3.6 days (Cao et al. 2016). In contrast, whole tree heating of the Norway spruce (*Picea abies*) lead to a two to three week advance of bud burst (Hänninen et al. 2007), demonstrating that the response to temperature changes are not uniform across plant species. For vertebrate species,

experimental work in temperature controlled aviaries found increasing temperature, not mean values or vegetative cues, prompts laying in great tits (Schaper et al. 2011; Schaper et al. 2012). However, the results found in experimental manipulations (captive or natural) do not always correspond with observed trends. A comparison of results generated from warming experiments on 1634 plant species and field observations indicated that experiments under predict plant responses to warming in leafing and flowering (Wolkovich et al. 2012). A captive blue tit experiment also found that populations with close wild breeding phenology had non-overlapping breeding times when held in outdoor aviaries (Lambrechts et al. 1999). In order to accurately identify the cues driving phenological change experimental manipulations must be combined with statistical analyses of observational data. An extensive statistical toolkit for phenological cue identification exists, but with little current standardisation in approach.

1.3.2. Predicting population level consequences of mismatch

Phenological change in response to climate change is widely documented (Visser et al. 2004; Visser & Both 2005; Walther et al. 2002; Parmesan 2006; Menzel et al. 2006; Cleland et al. 2007). Whether these changes will translate into temporal mismatches and/or population declines is less certain (Miller-Rushing et al. 2010; Bennett et al. 2015). Not all populations experiencing environmental change are mismatching with interacting species. Shared cues and/or phenotypic plasticity can maintain synchrony at present for some populations (Buse et al. 1999; Buse & Good 1996; Cresswell & McCleery 2003; Charmantier et al. 2008; Gienapp et al. 2005; Visser & Both 2005; Vedder et al. 2013). However maintenance of synchrony is not uniform even within species, with different populations of the same pairs of species having differing responses to climate change. An illustrative example is great tit reproductive timing and peak abundance of their winter moth caterpillar prey in the

Netherlands and the UK. In the UK great tits appear to maintain synchrony with their caterpillar prey, whereas Dutch great tits have advanced breeding at a slow rate in recent decades so lag behind caterpillar peak abundance (Visser et al. 2004; Visser et al. 2006; Charmantier et al. 2008). Predicting the circumstances under which mismatches will occur is a key component of generating projections/predictions of future phenological change.

Further to the variability in phenological responses, there is also substantial variability in the population level impacts of temporal mismatch. Despite reported strong links between phenological mismatch and individual fitness (Reed, Jenouvrier, et al. 2013; Reed, Grøtan, et al. 2013), several populations experiencing mismatch are not suffering associated population declines (Visser & Both 2005; Reed, Grøtan, et al. 2013; Dunn & Møller 2014). A lack of decline is not always the case, some species of European breeding birds (Møller et al. 2008; Both 2010) and ungulates (Plard et al. 2014) have seen population declines at least in part associated with mismatch, population reductions will not always be the result.

Several studies have proposed possible mechanisms for a lack of population level consequences of phenological mismatch (a detailed theoretical exploration was conducted by (Johansson, Kristensen, et al. 2015)). These include:

- Compensatory mechanisms, whereby any fitness costs of phenological mismatch are balanced by increased fitness elsewhere through processes such as; density dependence (Reed, Grøtan, et al. 2013; Grøtan et al. 2009; Johansson, Kristensen, et al. 2015), life history trade-offs (Reed, Jenouvrier, et al. 2013; Reed, Grøtan, et al. 2013; Johansson, Kristensen, et al. 2015), connected life stages (Johansson, Kristensen, et al. 2015; Visser & Both 2005), and

improvements in conditions at other times of year (Perrins 1965b; Van Balen 1980). This could buffer the population from some of the negative impacts of climate change.

- Reduced importance of the temporal synchrony. The strength of the phenological interaction could be weak due to a wide temporal abundance of resources, resulting in small amounts of asynchrony having limited impacts on the ability to meet energetic requirements (Johansson, Kristensen, et al. 2015; Dunn & Møller 2014). Additionally, if other potential resource or interacting species are available, mismatch with any one may not result in population declines if the focal species can replace the mismatched species with another (Johansson, Kristensen, et al. 2015; Cholewa & Wesołowski 2011). Spatial heterogeneity in timing could allow some members of a population to retain synchrony even if other spatial patches mismatch (Johansson, Kristensen, et al. 2015; Phillimore et al. 2012; Phillimore et al. 2010).
- Fitness benefits of asynchrony. It has been proposed that phenological asynchrony may not always be maladaptive (Johansson, Kristensen, et al. 2015; Singer & Parmesan 2010). Adaptive asynchrony (Johansson, Kristensen, et al. 2015; Singer & Parmesan 2010) can arise from several sources: (i) frequency dependence - where earlier phenology can bring competitive advantages (Kokko et al. 2006; Cleland et al. 2015; Johansson, Kristensen, et al. 2015), (ii) where bet hedging leads to sub-optimal phenology in particular years being adaptive on average, (iii) life history trade-offs and connected life stages can also lead to conditions where mismatching produces the highest fitness (Johansson, Kristensen, et al. 2015; Singer & Parmesan 2010).

To accurately predict population dynamics of seasonally breeding species under climate change a predictive framework is needed that includes not only phenological change, but also changes in demographic rates, the influence of population density, and key environmental drivers. It is likely mismatch and associated population declines are spatially and taxonomically heterogeneous across multiple scales.

1.4. The study system

The study system used in this thesis consists of the pedunculate oak (*Quercus robur*), winter moth caterpillar (*Operophtera brumata*), blue tit (*Cyanistes caeruleus*) and great tit (*Parus major*) in Wytham Woods, UK. Wytham Wood is an approximately 400 hectare mixed deciduous woodland (Perrins 1965b), with small patches of conifer plantation. Within the woodland there are 1203 tit nest boxes, of which 187 are blue tit only boxes (with entrance hole diameters too small for great tits). Four species of tit use the nest boxes for breeding; great tits, blue tits, marsh tits (*Poecile palustris*), and coal tits (*Periparus ater*). The larger holed boxes were erected in 1960, with blue tit boxes added later in 2002 and 2008. These boxes are split into nine 'sections/rounds' within the woodland; the sections differ in habitat type and history of forest management, as well as in the density of nestboxes. There is a long history of phenological study in this population ((Lack 1958a; Perrins & McCleery 1989; Husby et al. 2010; Charmantier et al. 2008; van Noordwijk et al. 1995; Cresswell & McCleery 2003; Vedder 2012)
as just a few examples) mainly focusing on breeding phenology. A portion of the breeding population of great tits have been studied since 1947 (Lack 1947), and across the whole woodland in a standardised way since 1960. Annual censuses of the breeding population, reproductive timing, and success provide an exceptional dataset on which to conduct studies of phenological change.

The timing of spring phenology is important for all components of this system. Oak trees release their leaves from dormancy at a time when the risk of frost has passed to avoid damage to new leaves, through reaching a chilling requirement over winter. This is achieved through autumn/winter chilling accumulation coupled with spring thermal accumulation (van Asch & Visser 2007; Powell 1987; Sparks & Carey 1995). Winter moth caterpillars time their hatching to match with the bud burst of the oak trees so that they can feed on the new leaves before tannins build up (van Asch & Visser 2007; Feeny 1970). Hatching too early means that they starve. Hatching too late slows growth, causing pupation at a smaller size and increased mortality as a result of high levels of tannins in their food source (van Asch & Visser 2007). Winter moth hatching appears to be linked to ambient temperatures (Buse & Good 1996), which also influence their development and growth (Buse et al. 1999). In turn, the great tits preferentially feed their young on Lepidoptera larvae (Perrins 1991) and therefore need to time the peak energetic requirement of chick rearing to the peak abundance of the winter moth caterpillar. How great tits time their reproductive effort so that this synchrony is achieved is not yet known. Clutch initiation dates have frequently been linked to spring temperatures (Perrins 1965b; Perrins & McCleery 1989; van Noordwijk et al. 1995; Charmantier et al. 2008; Husby et al. 2010). However, the exact cue used has not been established, and the potential for phenological flexibility in other elements of the great tit breeding cycle has not been examined in detail.

1.5. Thesis outline

Overall aim:

“ *Identify the impact of climate change on phenological synchrony between great tits and their caterpillar prey and predicting the population level consequences of this change.”*

This aim is addressed through several key steps. Firstly, I explore the causes of phenological change. In **Chapters two** and **three** I investigate some of the mechanisms that have permitted great tits and blue tits to achieve synchrony with winter moth caterpillars in an inter-annually fluctuating environment. **Chapter two** focuses on quantifying plasticity in breeding behaviour after reproductive initiation. This helps to assess the full flexibility of the great tit breeding cycle, identifying areas of plasticity and constraints. I quantify individual variation in different aspects of incubation behaviour (onset, duration, and daily intensity), and conduct a comprehensive assessment of the causes and consequences of this variation. I use detailed in-nest temperature recordings, to show that birds can improve their synchrony with resources post-laying, primarily by varying the onset of incubation, with duration changes playing only a small role. I then explore which spring temperature cues drive variance in each aspect of incubation behaviour, with both mean (for onset) and maximum (for intensity) temperatures showing a relationship to incubation behaviour. The results suggest that multiple aspects of the breeding cycle influence the final timing of peak energetic demand. Such adjustments could offset the fitness impacts of poor initial timing. In **Chapter three** I conduct an experimental test of a potential temperature cue as a driver of breeding phenology. While identifying cues statistically is useful to provide insights into which cues birds might be using, the only way to establish a causal link is through experimental manipulations. Here I experimentally manipulated in-nest temperatures in early spring for 89 blue tit boxes. Blue tits were used here so as not to disturb the long-term great tit monitoring. Nests were split into three treatments; heated, cooled and control. In-nest temperature in the heated and cooled boxes was manipulated by an average of +/- 0.6 ºC from control temperatures using heating devices and ice packs respectively. This was conducted to test the hypothesis that in-nest nocturnal

temperatures can act as a cue to the roosting female and potentially delay or advance breeding, and impact nest box choice. The results showed a slight trend towards earlier phenology in heated nest boxes in addition to a higher occupancy rate in cooled boxes, however neither of these trends was found to be statistically significant. The ability to distinguish statistical signals was hampered by unexpectedly low occupancy rates across all experimental treatments, highlighting the difficulties of conducting temperature manipulations in the wild.

For the second half of this thesis, I switch from exploring the causes of phenological change to predicting the consequences of phenological change under novel climate conditions. In **Chapters four** and **five** I construct an integral projection model (IPM), incorporating quantitative genetics, to tease apart the ecological and evolutionary processes driving phenological change and predict future great tit population dynamics. In **Chapter four** I introduce the model framework and parameterise it for the Wytham Woods great tits. I test different levels of model complexity (different numbers of explanatory drivers) and cross validate results against observed data. I find that a fully parameterised model generated from exploratory statistical analyses of drivers of fundamental demographic rates is needed to capture the combined trait and population dynamics of the study system. Environmental drivers in the parameterised model are then perturbed to identify the relative contribution of phenology and phenological synchrony to population dynamics. The results showed that while phenotypic plasticity dominates inter-annual changes in phenology, micro-evolution plays a directional role in shifting phenology, potentially visible across five decades. Spring temperature and changes to phenological synchrony were identified as the key drivers of alterations to population size. **Chapter five** builds on these analyses, using the same model to generate directed predictions of the future

population size and phenological synchrony of the study population to the end of the 21st Century. This is achieved by combining the population model with climate projections from the UKCP09 estimates under low, medium, and high greenhouse gas emissions scenarios. I test explicitly how different cue usage by interacting species can influence the likelihood of temporal mismatch under climate change. The results show that if phenological cues are shared the great tit population will be able to maintain synchrony with their prey species, in all but the highest emissions scenario, even showing population increases. However, if phenological cues are not shared, mismatch occurs, accompanied by population plateaus or declines.

The conclusion of this thesis occurs in the final two chapters. **Chapter six** returns to the concepts of cue identification in phenology through a critical review of our current toolkit. I focus on the use of the cues identified by commonly applied methods for predictive purposes. I compare and contrast the predictive capabilities of five commonly used methods. The accuracy, statistical robustness, and biological relevance of the cues identified through these methods is crucial to the interpretation of the results of this thesis, and any predictive or explanatory phenological work. As a consequence, the results of this chapter have significant relevance across the whole body of work, and to the wider field of phenology. I explore how the timing and aggregate statistic of the identified cue changes based on the method used, the number of years of data included, and the timing of this data relative to present. I apply all methods to the same model dataset of lay date timing of the Wytham Woods great tits, assessing the predictive capacity of each method and their sensitivity to sample size and time period of data. The five methods inconsistently and inaccurately predicted lay dates, with predictions generally lagging one to fifteen days behind observed timings, suggesting that the current toolkit does not accurately

capture the biological cue used by great tits to time their breeding. **Chapter seven** completes the thesis with a synthesis and discussion of some outstanding questions that remain in phenology and suggestions of how future research directions could begin to address them

CHAPTER TWO

Incubation behaviour adjustments improve synchrony between hatch dates and caterpillar peak in a wild bird population

Emily G. Simmonds, Ben C. Sheldon, Tim Coulson, and Ella F. Cole

Ecology and Evolution 2017
DOI: 10.1002/ece3.3446

2.1. Abstract

For organisms living in seasonal environments, synchronising the peak energetic demands of reproduction with peak food availability is a key challenge. Understanding the extent to which animals can adjust behaviour to optimise reproductive timing, and the cues they use to do this, is essential for predicting how they will respond to future climate change. In birds, the timing of peak energetic demand is largely determined by the timing of clutch initiation, however, considerable alterations can still occur once egg laying has begun. Here, we use a wild population of great tits (*Parus major*) to quantify individual variation in different aspects of incubation behaviour (onset, duration and daily intensity), and conduct a comprehensive assessment of the causes and consequences of this variation. Using a 54-year dataset, we demonstrate that timing of hatching relative to peak prey abundance (synchrony) is a better predictor of reproductive success than clutch initiation or clutch completion timing, suggesting adjustments to reproductive timing via incubation are adaptive in this species. Using detailed in-nest temperature recordings, we found that post-laying, birds improved their synchrony with the food peak primarily by varying the onset of incubation, with duration changes playing a lesser role. We then used a sliding time window approach to explore which spring temperature cues best predict variance in each aspect of incubation behaviour. Variation in the onset of incubation correlated with mean temperatures just prior to laying, however incubation duration could not be explained by any of our temperature variables. Daily incubation intensity varied in response to daily maximum temperatures throughout incubation, suggesting female great tits respond to temperature cues even in late stages of incubation. Our results suggest that multiple aspects of the breeding cycle influence the final timing of peak energetic demand.

Such adjustments could compensate, in part, for poor initial timing, which has significant fitness impacts.

2.2. Introduction

For species living in seasonal environments, reproductive success can be maximised by timing reproduction to coincide with annual peaks in resource abundance (Lack 1968; Perrins & McCleery 1989; van Noordwijk et al. 1995; Parmesan 2007; Nussey et al. 2005). The timing of these resource peaks can vary considerably between years in response to environmental variation. To achieve synchrony across years, iteroparous animals must exhibit plasticity in their reproductive phenology. If the resource tracked is another species that themselves track environmental conditions, this can be particularly challenging. Synchrony between trophic levels can be achieved and maintained simply by a shared sensitivity to temperature cues, however interacting species may differ in their sensitivity to environmental variation. Plants and insects respond more directly to temperature changes than homeotherms (e.g. temperate mammals and birds), consequently the ability of the homeotherm to track the resource species may be constrained (van Noordwijk et al. 1995; Plard et al. 2014; Gaillard et al. 1993; Visser et al. 1998). Successful matching is dependent on whether animals can perceive ambient temperature and utilise this as a cue to predict optimal reproductive timing in a given year. Such mechanisms, which ensure matching in an inter-annually varying system, could potentially be disrupted by novel climatic change, leading to mismatches between predators and resources (Gienapp et al. 2005; Gienapp et al. 2014). For instance, cues which previously predicted the timing of peak prey abundance may no longer do so if climate patterns are altered. Understanding the factors that constrain the extent to which animals can track their resource is important for predicting future levels of mismatch. Determining which

elements of the reproductive cycle are flexible and their sensitivity to temperature is a key component of this.

The most commonly studied aspects of reproductive phenology have been those which are easily observable; clutch initiation date (Charmantier et al. 2008; Visser et al. 1998; Schaper et al. 2012; Nussey et al. 2005; Lack 1958b), clutch size (von Haartman 1969; Kluijver 1950; Balen & Balen van 1973; Lack 1955; Lack 1958b; Perrins 1965a; Perrins 1979; Haftorn 1981), birth date (Plard et al. 2014), flowering date (Menzel et al. 2006), and hatch date (Cresswell & McCleery 2003; Tomás 2015). However, it is well established that there is also considerable variation in other aspects of the reproductive cycle such as; incubation behaviour (Stenning 2008; Álvarez & Barba 2014; Lord et al. 2011; Cresswell & McCleery 2003; García-Navas & Sanz 2011; Matthysen et al. 2010; Hepp et al. 2006; Ardia et al. 2010), conception date (Scott et al. 2008), and gestation length (Asher et al. 2005; Moyes et al. 2011; Scott et al. 2008; Racey & Swift 1981). The phenology of many of these reproductive behaviours cannot be observed directly, nonetheless, they could have a significant role in determining the timing of the peak energetic demands of reproduction, usually during offspring rearing. Mammals primarily have one method of flexibility after a reproductive event has been initiated, gestation period. This has been shown to vary in response to food availability and temperature in several species (Asher et al. 2005; Moyes et al. 2011; Scott et al. 2008; Racey & Swift 1981) . In contrast, for bird species, there are several different mechanisms which can alter phenology (hatch date and consequently timing of peak food demand), after a reproductive event has begun and right up until hatching.

The beginning of the reproductive effort for birds is the building of a nest and the onset of egg laying. The lay date of the first egg of a clutch has been well studied and is highly variable (Lack 1955; Lack 1958b; Perrins 1965a; Perrins 1979; Visser et al. 1998; van Noordwijk et al. 1995; Charmantier et al. 2008), with annual shifts of up to several weeks in some species. Changes in lay date have been extensively linked to changes in early spring temperatures (Lack 1955; Lack 1958b; Perrins 1965a; Perrins 1979; Visser et al. 1998; van Noordwijk et al. 1995; Charmantier et al. 2008; Schaper et al. 2012) and represent a plastic response to the environment (Charmantier et al. 2008) . Once egg laying has commenced, the majority of birds lay a maximum of one egg per day until their clutch is complete. Therefore the size of a clutch and the rate of egg laying can delay or advance hatch date. Clutch size and egg laying rate have been shown to vary based on timing of laying, with clutch sizes decreasing as clutch initiation dates become later (von Haartman 1969; Kluijver 1950; Balen & Balen van 1973; Lack 1955; Lack 1958b; Perrins 1965a; Perrins 1979; Haftorn 1981; Matthysen et al. 2010).

Incubation behaviour can also impact the timing of hatching in birds species, acting via two main mechanisms. First, the onset of incubation can be advanced or delayed, relative to when a clutch is completed. Second, the duration of the incubation period can be adjusted based on the intensity of incubation effort. Variability in the relative onset of incubation has been demonstrated across a diverse range of bird species (e.g. *Paridae* and *Anatidae*) and can vary by up to a week either side of clutch completion (Stenning 2008; Alvarez & Barba 2014; Lord et al. 2011; Cresswell & McCleery 2003; García-Navas & Sanz 2011; Matthysen et al. 2010; McClintock et al. 2014; Hepp 2004; Loos & Rohwer 2004). Such changes are also known to have knock on impacts on reproductive success. Beginning incubation prior to completion

of a clutch can increase hatching asynchrony and lead to rapid brood reduction in years of poor resources (Lord et al. 2011; Stenning 2008; García-Navas & Sanz 2011; Álvarez & Barba 2014; Ardia et al. 2010). Variation in incubation onset has been linked to changes in spring temperatures both experimentally (Vedder 2012; Álvarez & Barba 2014; Bryan & Bryant 1999) and through observational studies (Matthysen et al. 2010; Cresswell & McCleery 2003). Incubation duration and intensity have been less extensively studied but have also been shown to vary with temperatures and potentially with individual condition (Conway & Martin 2000; McClintock et al. 2014; Ardia et al. 2010).

Although each aspect of incubation behaviour has been shown to vary and have some relationship to temperature, the precise temperature cues that trigger variation in these traits are, as yet, unknown. Identifying the temperature metric (mean, maximum, minimum, or temperature range) driving variability in each aspect of incubation behaviour, in addition to the temporal window during which these cues are important, is necessary for understanding how incubation behaviour could be used to improve hatching synchrony. Alterations to incubation behaviour are likely to be most important when temperatures fluctuate throughout the spring (for instance when initial warming suddenly turns cold). Consequently, it is important to understand the cues used and the limits of plasticity in different elements of incubation behaviour in order to accurately predict how hatching timing might change under different climate scenarios. If different aspects of incubation behaviour respond to different temperature cues, the different aspects could change at varying rates in the future. In order to determine the role of onset, duration and intensity of incubation in the final timing of hatching a detailed assessment of all aspects of incubation behaviour is required.

We seek to conduct a detailed study of the incubation behaviour of wild great tits (*Parus major*), exploring the extent to which they adjust incubation to improve timing of chick hatching in relation to the peak abundance of their prey species, winter moth caterpillars (*Operophtera brumata*). Passerine songbirds are a good study system to address questions of plasticity and constraints in multiple aspects of the breeding cycle because their reproductive phenology has been extensively studied (Charmantier et al. 2008; Lack 1955; Lack 1958b; Perrins 1965a; Perrins 1979; Visser et al. 1998; van Noordwijk et al. 1995) and, particularly for nest box breeding populations, phenology can be easily monitored. Great tits start incubation gradually, beginning at a few hours each night and increasing to cover the whole night several days prior to clutch completion (Haftorn 1981). There is then a transition to begin incubating during daylight hours, which again gradually increases up to a point when almost the entire day is spent incubating (Haftorn 1981). The start of daytime incubation is generally recognised to be the start of true incubation (Haftorn 1981). Identifying the precise onset of full incubation is necessary to characterise different components of incubation behaviour, however, it is challenging to achieve as nest observations alone are insufficient (Stenning 2008).

Here we carry out a thorough exploration of the causes and consequences of within-population variance in the various components of incubation behaviour. First we explore whether great tits improve their reproductive success, and synchrony with their food source, through adjustments to incubation behaviour. We then investigate the mechanisms behind these patterns by quantifying the extent to which different aspects of incubation behaviour vary, exploring whether this variation can be explained by temperature cues, and ultimately determining which aspects of

incubation behaviour are important in improving synchrony with the caterpillar peak. We thus address five key questions:

1. Is reproductive success better explained by timing of hatching relative to the caterpillar peak than timing of clutch initiation relative to the caterpillar peak?

2. Do adjustments to timing of hatching made after clutch initiation improve synchrony between chick hatching and the caterpillar peak?

3. How much within-population variation exists in (a) onset of incubation relative to clutch completion, (b) incubation duration, and (c) incubation intensity?

4. Can this observed variability be explained by ambient temperature cues, and if so, which temperature measures best capture variation?

5. To what extent do these three aspects of incubation behaviour (relative onset, duration, and intensity) contribute to improving synchrony between timing of chick hatching and the caterpillar peak?

2.3. Methodology

2.3.1. Data collection

2.3.1.1. Long-term breeding timing

The nest box breeding great tit population of Wytham woods has been studied since 1960 using a standardised procedure (Perrins 1965a; Perrins & McCleery 1989). Nest boxes (N = 1203 with an average of 225 occupied by great tits each year) are visited weekly from early April. During these weekly checks nest stage and number of eggs are recorded, and the date that a female initiated her clutch is then inferred by assuming a laying rate of one egg per day and counting back from the number of eggs observed on the weekly check. When at least three eggs are present, they are

weighed so that species can be assigned (blue tits, coal tits (*Periparus ater*), and marsh tits (*Poecile palustris*) also use the Wytham nest boxes). Clutch size is defined as the maximum number of eggs observed in the nest. Date of hatching is established by visiting the nest on the estimated hatch date (date of clutch completion plus 11 days) and then every other day until the eggs hatch. If eggs have hatched prior to the hatch check, the largest chicks are weighed and assigned age based on their weight. Parents are identified at the nest either remotely using RFID (Radio-frequency identification) scanners (all previously trapped birds are fitted with RFID tags) or by catching using spring-loaded nest box traps. All surviving nestlings are tagged with uniquely identifiable metal leg rings and RFID tags at two weeks old. Nests are then checked post-fledging to determine the number of chicks that left the nest.

The timing of caterpillar peak abundance is taken to be the median date on which caterpillars descend to the ground to pupate. This data has been collected as part of a long term study in Wytham woods, supplied by Dr L. Cole.

2.3.1.2. In-nest temperature data collection and incubation onset identification

During the 2014 breeding season, in-nest temperature was recorded at 163 great tit nests using iButton thermometers (DS1921G-F5, accurate to ± 1 ºC, HomeChip Ltd) set to record temperature every 20 minutes. These iButtons were secured in the nest cup by wrapping blunted garden wire around the iButton and using the protruding ends to anchor the iButton in the nest material. The iButtons and wire were then sealed into small cotton pouches. Pouch colour was matched as closely as possible to the nesting material to minimise visibility to the female great tit (for further details see Figure S2.1, Supporting Information). iButtons were placed in every second great

tit nest discovered across the woodland, throughout the season, ensuring a spatially and temporally even spread of sampling (163 of the 337 great tit nests). iButtons were placed in nests prior to the start of incubation (eggs were cold to the touch). Incubation in great tits occurs gradually (Haftorn 1981) and therefore eggs can feel cold even after daytime incubation has begun if a fieldworker visits when the female is not incubating. Consequently only nests which showed, from in-nest temperatures, at least one day of non-incubation after placement of the iButton were included in our analyses. Four nests that did not meet this criteria were removed from the analysis.

Of the 163 iButtons placed, 109 were retrieved (54 disappeared from the nest and were assumed to have been removed by the resident great tit). There was potential for bias to be created if females who removed iButtons had a tendency for a certain incubation behaviour. However, there was no statistical difference between the clutch completion to hatch period for those females who removed iButtons and those that did not (full analysis reported in 2.7.3, Supporting Information). Of the retrieved iButtons a further six nests were removed from analyses because they were abandoned, three prior to incubation onset and three after incubation but prior to hatching. There was no significant association between abandonment and whether a nest had an iButton or not ($\chi^2 = 3.06$, DF = 1, P = 0.08, N = 337). A further six iButtons were removed from analyses due to indistinct readings, probably due to deep burial in the nest by the resident female. We therefore present data from the remaining 93 nests where onset of daytime incubation was clearly identifiable.

Date of onset of daytime incubation was determined by combining in-nest temperature measures with hourly local ambient temperature measures (see below for details of ambient temperature data collection) and calculating the difference

between the two. Every in-nest temperature was paired with a local ambient reading from the same hour. Each in-nest iButton was matched to the closest ambient temperature iButton using GPS coordinates in ArcGIS (ESRI 2010). In order to identify the onset of daytime incubation it was necessary to distinguish when a female was incubating in the daytime. This posed a methodological challenge because the temperature readings from the iButtons did not represent the exact egg temperature due to different conductive properties of the egg and the iButton. iButtons were also prone to slight burial and movement within the nest cup based on female behaviour, potentially altering the temperature readings. As a result it was necessary to calibrate each iButton daily. The in-nest iButton was calibrated to its daily position by selecting a period each day, when the female is known to be incubating (the period just after dark, c. 7 pm, to midnight (Haftorn 1981)) and using the temperatures recorded at this point as a threshold for what can be considered an incubating temperature. In-nest temperature was calculated for this every night and temperatures over 4 ºC greater than ambient were taken to indicate incubation. This cut off was chosen in order to avoid classing small deviations (1 or 2 ºC) caused by ambient temperature differences, as incubation. All in-nest iButtons showed differences of < 1 ºC from local ambient temperature during the active day (7 am to 7 pm) prior to the onset of daytime incubation, therefore we are confident a 4 ºC cut off is sufficient to indicate incubation. The minimum of these incubating temperatures was then taken as the minimum incubation temperature for the focal nest and current iButton position. Recorded temperatures during the following active day (7 am to 7 pm), which exceeded the defined threshold for a given nest and day were classed as showing incubation is taking place. The onset of daytime incubation was consequently defined as the day when at least 50 % of the active day recordings

were classed as "incubating" (see Figure S2.2, Supporting Information, for further details of how thresholds were defined).

2.3.1.3. Ambient temperature data collection

Local ambient temperature was collected via a grid of ambient temperature iButtons (DS1923-F5, accurate to ± 0.5 ºC, HomeChip Ltd) set to measure absolute temperature every 30 minutes. 200 of these ambient temperature iButtons were distributed in a grid system across Wytham Woods with positions chosen to reflect the density of nest boxes. For further details see (Cole & Sheldon 2017).

2.3.2. Statistical analyses

2.3.2.1. Is reproductive success better explained by timing of hatching relative to the caterpillar peak than timing of clutch initiation relative to the caterpillar peak?

Whether hatch timing relative to winter moth caterpillar peak abundance (taken as the median date on which caterpillars were observed descending to pupate) is important for reproductive success was tested using a Poisson generalised linear model (GLM) with number of fledglings as the response variable. Fixed effects were hatching synchrony (observed hatch date minus median caterpillar date of that year) and its quadratic. Clutch size, clutch initiation synchrony (observed clutch initiation date minus median caterpillar date of that year), its quadratic, and section of the woodland were also included as explanatory variables to take account of number of eggs laid, individual condition and local habitat quality.

2.3.2.2. Do adjustments to timing of hatching made after clutch initiation improve synchrony between chick hatching and the caterpillar peak?

In order to address the overarching question of whether adjustments after clutch initiation improve synchrony between chick hatching and the caterpillar peak, we address two sub-questions.

A. Is annual population-level variance in hatching timing reduced through incubation adjustments?

The annual variance of hatch dates and clutch initiation dates were calculated from 1960 to 2014. To take account of the potential influence of changes in clutch size, we also calculated the variance of clutch completion dates. The variances of clutch initiation date, clutch completion date, and observed hatch date were compared using an ANOVA including a fixed effect of year.

B. Is synchrony with the food source improved by adjustments made between clutch initiation and hatching?

To distinguish whether the observed synchrony between hatch dates and caterpillar peak abundance was a significant improvement to the null expectation of this synchrony, without incubation alterations, a paired T-Test was conducted. Observed synchrony was calculated as the difference between each nest's hatch date and the woodland annual caterpillar timing. All synchrony values had 13 added to them as 13 days prior to the caterpillar peak was indicated to be the optimal timing of hatching (see above for the analysis of reproductive success). This gives an index with negative values indicating hatch timing earlier than optimum, positive values indicating hatch dates later, and 0 indicating optimal synchrony. The null expectation, assumed no incubation behaviour alterations and was calculated as the clutch completion date plus 14 days (to represent duration of incubation) minus annual caterpillar timing. Synchrony index measures were then squared in order to remove negative values.

2.3.2.3. How much within-population variation exists in (a) onset of incubation relative to clutch completion, (b) incubation duration and (c) incubation intensity?

Using the identified onset of full incubation for the 2014 breeding season we calculated two primary aspects of incubation behaviour; the relative onset (interval between clutch completion and start of daytime incubation) and duration (interval between onset of incubation and observed hatch date). In addition, we also quantified the daily intensity of incubation effort. Incubation effort for each day was determined by calculating the number of 20 minute periods during the active day (7 am to 7 pm - the active day) that exceeded the threshold temperature for incubation, as a proportion of the total number of active day recordings. The range, mean, and variance of each of these aspects of incubation behaviour were calculated.

Relationships between each of the aspects of incubation behaviour were also tested. Linear models (LM) and GLMs were run for each combination of behaviours. Due to sequential timing, relative incubation onset could not be causally influenced by either intensity or duration. However, the onset itself could influence the intensity of incubation effort and consequently impact the duration of daytime incubation. Furthermore, other factors, such as individual condition and clutch size, could also impact incubation effort. These associations were tested in two analyses. The first was a Binomial GLM, with mean intensity of incubation as a response variable and relative incubation onset as an explanatory variable, accounting for clutch size and clutch initiation date (a proxy for individual condition (Rowe et al. 1994)). The second was a LM with incubation duration as the response variable and mean intensity, relative incubation onset, and clutch size as explanatory variables. The duration of daytime incubation should be a result of the amount of incubation required for an embryo to develop and hatch, scaled by the intensity at which it was incubated. As a

result, we would expect incubation duration to show a relationship with mean incubation intensity. However, this relationship could also be modulated by other processes, which influence the amount of incubation required. This could be altered by clutch size (Haftorn 1981) and the amount of prior incubation, through relative incubation onset. A nest that began incubation prior to clutch completion could require a greater intensity of incubation effort for the same duration as a nest that began incubation after clutch completion. This occurs due to accumulated hours of incubation during nocturnal incubation reducing the amount required from full incubation for nests which delay onset.

2.3.2.4. Can this observed variability be explained by ambient temperature cues, and if so which temperature measures best capture variation?

For both relative incubation onset and incubation duration, exploratory analyses were conducted using sliding time window methods in order to identify the temporal temperature window and temperature measure, which best explain variance in different components of incubation behaviour. This method does not establish a causal link between the behaviour of interest and temperature, but instead attempts to identify the time window in which temperature may be most important for determining the behaviour in question.

12 different length windows were tested for these analyses (see Figure 2.1). The minimum time window (shown in red) was unique to each nest and spanned from the lay date to the date of onset of full incubation, hereafter termed the "laying period". The time windows then increase in one day increments spanning backwards from the start of laying, up to the maximum window of 10 days prior to laying, up until incubation onset (see Figure 2.1). Window 12 is only used in the analysis of incubation duration and spans from the onset of incubation up until hatching.

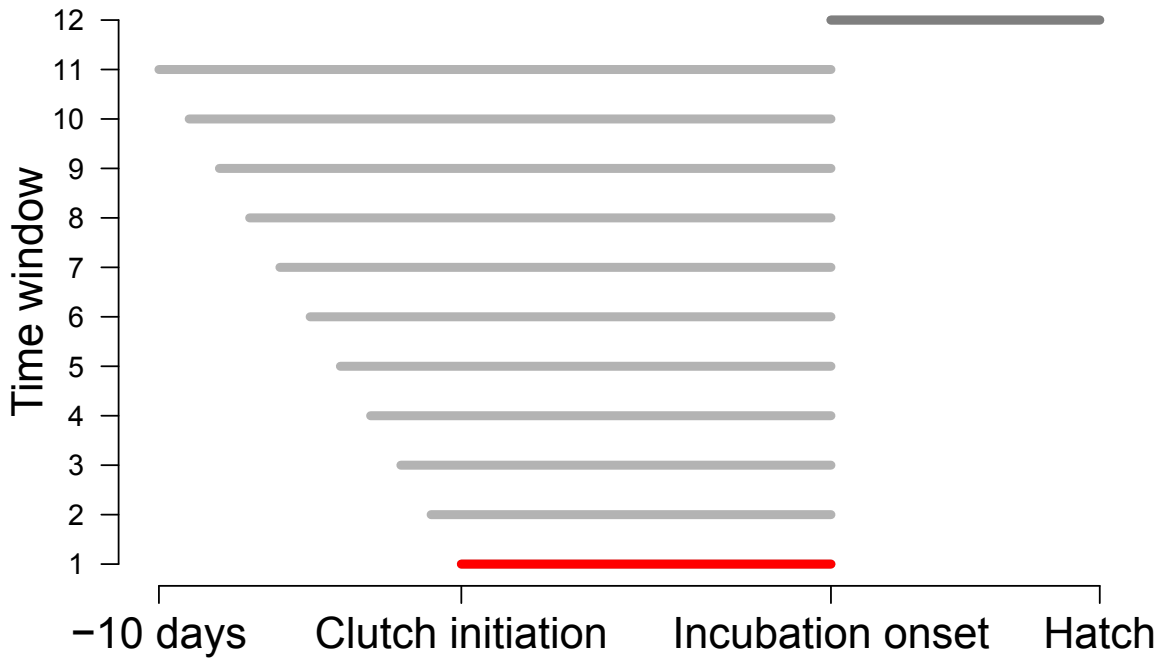

Figure 2.1: Plot of 12 sliding windows of temperature.
Time through the breeding cycle is shown on the x axis and name of the window on the y axis. Window 12, dark grey, was used only for analyses of incubation duration and covers the incubation period. The minimum window, from clutch initiation date to the onset of incubation, is highlighted in red

Candidate models for model selection were LMs and took the form of two configurations of null model (models with no temperature variables included), which were compared with four configurations of temperature model, each containing a single temperature variable (see Table 2.1a). Four different temperature measures were tested; mean temperature, mean daily minimum temperature (MMin), mean daily maximum temperature (MMax) and mean daily temperature range (Trange). In total this gave 44 different explanatory temperature variables (four measures across 11 windows). 176 candidate temperature models were compared with relative incubation onset as the response variable, consisting of 44 different temperature variables in each of the four model configurations and the two null models. For models with incubation duration as a response variable each candidate model also included a fixed effect of relative incubation onset. As incubation occurs at night prior to the onset of full incubation, any delay or advance in full incubation onset will alter

the number of accumulated "incubation hours" and therefore will likely influence the final duration of full incubation. An additional four candidate models were also run due to the additional time window for this variable, totalling 180 candidate models.

Table 2.1: Candidate model configurations for the analysis of incubation behaviour

a) Candidate models for the analysis of relative onset of incubation and incubation duration. Models for incubation duration also include the relative onset of incubation as a fixed variable. b) Candidate models for the analysis of daily incubation intensity

a)	
Model name	Explanatory fixed variables
Null 1	Clutch initiation date
Null 2	Clutch initiation date + Clutch size
Temperature 1	Temperature variable
Temperature 2	Temperature variable + Clutch initiation date
Temperature 3	Temperature variable * Clutch initiation date
Temperature 4	Temperature variable + Clutch size
b)	
Model name	Explanatory fixed variables
Null 1	Clutch initiation date
Temperature 1	Incubation day + Temperature variable
Temperature 2	Incubation day + Temperature variable + Clutch initiation date
Temperature 3	Incubation day + Temperature variable * Clutch initiation date
Temperature 4	Incubation day + Temperature variable + Clutch size

For incubation intensity, Binomial generalised linear mixed effects models (GLMM) were used with daily incubation effort as the response variable and temperature and stage of incubation (incubation day) as fixed effects. The random effect of individual was also included to take account of individual differences in incubation effort. Incubation days ran from the onset of full incubation (day 1) to the day prior to hatch day. Hatch date itself was excluded as this is when iButtons were removed. Candidate models each included a single temperature variable from daily mean temperature, daily minimum temperature, daily maximum temperature, and daily temperature range. Five different configurations of model were trialled (see Table 2.1b). Therefore a total of 20 candidate models were trialled.

Candidate models were compared using the ΔAIC. The preferred model was defined as the model with the lowest AIC. Models are considered as not significantly different to the preferred model if the ΔAIC is less than 2. It should be noted that as each model including a temperature variable contains at least a partially overlapping or correlated variable (as all temperature measures and windows are likely to be correlated) we would not expect these analyses to always produce any single clearly preferred model.

2.3.2.5. To what extent do these three aspects of incubation behaviour (relative onset, duration and intensity) contribute to improving synchrony between timing of chick hatching and the caterpillar peak?

The influence of relative incubation onset and incubation duration on the observed synchrony was tested by comparing the observed hatch date synchrony (hatch date minus caterpillar peak timing) to a null expectation of hatch synchrony, if either aspect of incubation behaviour had not varied. This is similar to the technique applied to the long term data, however in this instance we are able to distinguish between different aspects of incubation behaviour and quantify their impacts independently. A null expectation was created for both relative incubation onset (calculated as the clutch completion date plus the duration of full incubation minus caterpillar timing) and incubation duration (calculated as the clutch completion date plus the relative incubation onset and the mean duration of full incubation in 2014 - 12 days minus caterpillar timing). The variance of the observed hatching synchrony was compared to the two null estimations using pairwise F-Tests for equality of variance. As there was no logical null estimate for the duration of daytime incubation, therefore isolation of the influence of this element of incubation behaviour on mean synchrony was not possible. Therefore only the influence of relative incubation onset on mean

synchrony was tested. This was conducted using a paired T-Test between the observed synchrony and a null estimate with no onset changes.

2.4. Results

2.4.1. Hatching timing relative to the food peak abundance influences reproductive success and is a better predictor than relative clutch initiation timing

The number of chicks fledged showed a significant relationship with observed hatching synchrony (difference between hatch date and caterpillar peak date) (EST = -0.074, SE = 0.0037, P < 0.01) and its quadratic (EST = -0.0017, SE = 0.00013, P < 0.01) (see Figure 2.2). Those hatching too early (more than 13 days prior to the caterpillar peak) or too late (less than 13 days prior to the caterpillar peak) fledged fewer young. The number of chicks fledged also showed a significant positive relationship with clutch initiation synchrony, but not the quadratic (EST = -0.00012, SE = 0.000099, P = 0.24). Earlier layers fledged more young, however the effect size was almost half that of hatch timing (EST = 0.036, SE = 0.0069, P < 0.01).

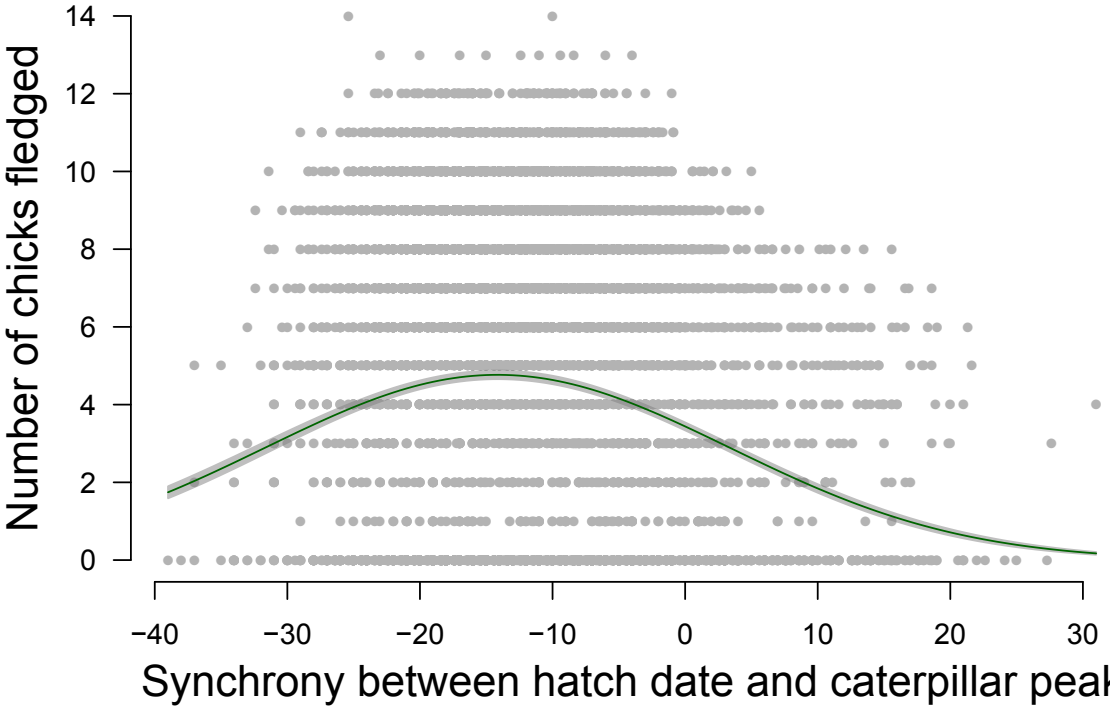

Figure 2.2: The number of chicks fledged against hatching synchrony, from 1960 to 2014.
The plotted line is generated from a Poisson GLM with fixed effects of clutch initiation date relative to caterpillar timing, its quadratic, clutch size, hatch date relative to caterpillar timing, its quadratic, and section of woodland. The shaded area represents the 95 % confidence interval of the plotted relationship

2.4.2. Incubation alterations lead to hatch dates having lower variance than clutch initiation or completion dates

From 1960 to 2014 hatch dates in the Wytham Woods great tit population showed significantly lower annual variance than clutch initiation dates. Hatch dates had an average variance of 3.7 days less than clutch initiation dates (variance difference = -13.6, SE = 0.97, $P < 0.01$). This difference was partly driven by changes in clutch size, but not exclusively. Variance in clutch completion date, was significantly lower than in clutch initiation date (variance difference = 5.9, SE = 0.97, $P < 0.01$) and significantly greater than observed hatch date variance (variance difference = 7.6, SE = 0.97, $P < 0.001$).

2.4.3. Incubation alterations improve synchrony between hatch dates and the caterpillar food peak

Results of a paired T-Test show that the asynchrony between observed hatch dates and caterpillar timing is 3.35 days lower (closer to optimal synchrony) than for the null estimate of hatching asynchrony, if no incubation alterations occurred. This is statistically significant ($T = -26.6$, $DF = 11438$, $P < 0.01$).

2.4.4. Relative incubation onset, incubation duration and intensity are highly variable

For the 2014 breeding season, relative incubation onset ranged from three days prior to clutch completion up to 12 days after, with an average of four and a half days delay. Incubation duration ranged from seven to 19 days with an average duration of 11.5 days. Incubation intensity, once full incubation had begun, ranged from 5 % to 100 % of the active day, with a mean incubation effort of 70 %. The maximum alteration to hatch date, assuming independence of different incubation behaviours and an expected incubation duration of 11.5 days, is an advance of seven and a half days or a delay of 19.5 days.

The mean intensity of incubation showed no significant relationship with either clutch size ($EST = 0.04$, $SE = 0.03$, $P = 0.14$) or clutch initiation date ($EST = 0.01$, $SE = 0.01$, $P = 0.42$), but did show a significant negative relationship to relative incubation onset ($EST = -0.04$, $SE = 0.01$, $P < 0.001$). Incubation duration also showed a significant negative relationship with the relative onset of incubation ($EST = -0.45$, $SE = 0.05$, $P < 0.001$), but did not show any significant relationship with either clutch size ($EST = 0.03$, $SE = 0.12$, $P = 0.79$) or mean intensity ($EST = -3.92$, $SE = 2.09$, $P = 0.06$).

2.4.5. Relative incubation onset varies in response to mean local temperatures around the laying period

10 candidate models had ΔAIC values of within two of the lowest AIC model, giving 11 candidate models with equal support. All of these models included mean temperature or mean maximum temperature. The majority (nine of 11) included an additive effect of clutch initiation date or an interaction between clutch initiation date and temperature. Effect sizes for the relationship between temperature and relative incubation onset varies from -5.4 for mean temperature to -0.7 for mean maximum temperature. As all of these models are similar in composition, only the model with the lowest AIC is discussed further (but a full list of model parameters can be found in 2.7.4, Supporting Information). The model with the lowest AIC included explanatory variables of mean temperature for time window eight, clutch initiation date and an interaction between the two.

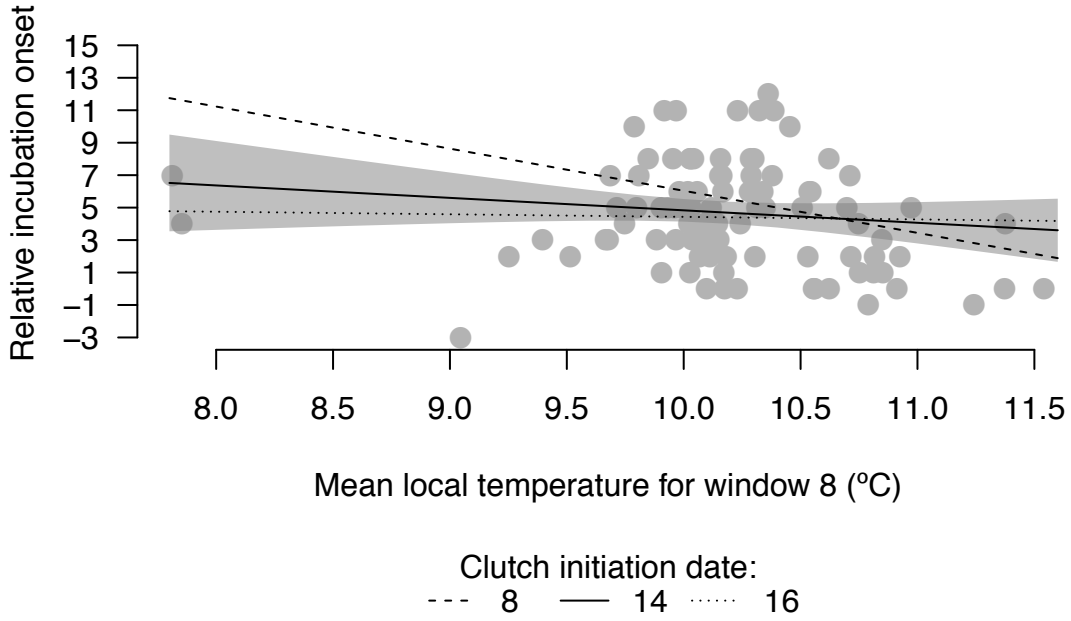

Figure 2.3: Relative onset of incubation against mean temperature for window 8.

The plotted lines are generated from predictions from linear models. Clutch initiation dates are held at the 1st quartile value, the median value, and the 3rd quartile value, to illustrate the significant interaction between clutch initiation date and temperature. Shaded area represents the 95 % confidence interval for predictions with median clutch initiation date only to improve the readability

The relative incubation onset had a significant negative correlation to mean temperature for window eight, with individuals starting incubation earlier when mean temperature was higher (5 days advance in the onset of incubation per 1 ºC increase in mean temperature), see Figure 2.3 (EST = -5.03, SE = 1.72, P = 0.004). The interaction between clutch initiation date and temperature took the form of early layers having the strongest negative relationship between temperature and their incubation onset and later layers showing almost no relationship with temperature, as illustrated in Figure 2.3 (EST = 0.3, SE = 0.13, P = 0.02). Figure 2.3, shows the relationship including outliers; the relationship was also tested without these data points (mean temperature values < 8.5 ºC). Parameter values did not change considerably on removal of the outliers, consequently we opted to retain all data. Full output of both analyses can be found in the 2.7.4, Supporting Information.

2.4.6. Variability in incubation duration shows no significant relationship with temperature

The duration of full incubation showed no significant relationship with any temperature measure or temporal window once the relative onset of incubation was taken into account. The candidate model with the lowest AIC included the temperature range for window 11 and the relative incubation onset, however the effect of temperature was not statistically significant at P = 0.05 (EST = 0.26, SE = 0.15, P = 0.09).

Full model selection results and model parameters can be found in 2.7.5, Supporting Information.

2.4.7. Daily intensity of incubation effort increases with higher daily maximum temperatures

The proportion of the active day spent incubating was significantly positively correlated with time through incubation (~ 5 % increase per day of incubation) and local maximum daily temperature (~ 2 % increase per 1ºC), according to the lowest AIC model. There was also a weak but significant negative interaction between temperature and day of incubation (EST = -0.008, SE = 0.001, P < 0.001); with the relationship between temperature and intensity being weaker the further through incubation a bird is. All other candidate model configurations had ΔAIC values of > 2. These results show that the further through incubation a female is, and the warmer the daily temperature, the higher the proportion of the day is spent incubating (see Figure 2.4).

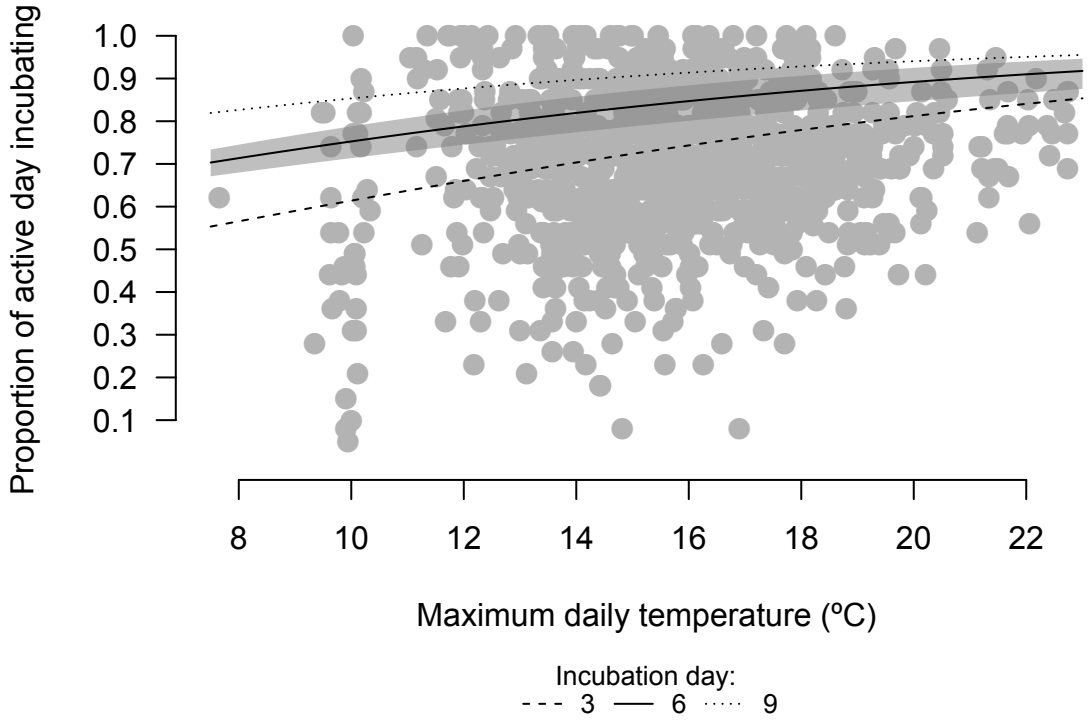

Figure 2.4: The proportion of the active day spent incubating against maximum daily temperature.

The plotted lines are generated from predictions from a binomial GLMM using mean incubation day and one standard deviation above and below the mean incubation day(start of full incubation = incubation day 1). For the purposes of prediction, random effects were ignored. Shaded area represents the 95 % confidence interval for the relationship between temperature and intensity when incubation day is 6 only in order to improve the readability

Full model selection results and model parameters can be found in 2.7.6, Supporting Information.

2.4.8. Changes to relative incubation onset alter the mean, but not the variance, of synchrony between hatch dates and food peak abundance

Null estimations of hatching synchrony, assuming no changes in relative incubation onset or no changes to incubation duration, and the observed hatching synchrony were compared. This allowed us to tease apart the contributions of different aspects of incubation behaviour to the observed hatching synchrony for the 2014 breeding season. The inclusion of alterations to incubation duration altered the variance of synchrony among nests, from 44.8 without duration changes to 37.8 for observed hatching synchrony. The inclusion of changes to the relative onset of incubation also showed a difference in variance, but to a much lower extent (1.1 higher than observed variance). None of these differences in variance were statistically significant. F-Test of variances for a baseline with no duration changes compared to observed hatching synchrony showed a ratio of variances of 1.18 (DF = 92/92, P = 0.21). The F-Test for no onset changes compared to observed hatching synchrony showed a ratio of variances of 1.02 (DF = 92/92, P = 0.44).

A paired T-test testing the difference in mean synchrony with and without incubation onset changes showed a significant difference. Mean synchrony was 4.5 days lower, closer to optimum, for observed compared to null synchrony (difference = -4.5, T = 13.29, DF = 92, P < 0.01).

2.5. Discussion

This study examined whether incubation behaviour is used to improve the timing of hatching relative to peak resource abundance (winter moth caterpillars), and which

temperature cues best predict within year variation in incubation. We quantified variability in different aspects of incubation behaviour for a single breeding season (the onset relative to clutch completion, duration and intensity) and explored how such changes are linked to temperature and how they can influence hatch timing. Using a population of wild great tits we show that across 54 years, individual hatching dates show lower within-year variance than clutch initiation dates, suggesting that alterations are occurring post clutch initiation that are bringing hatching dates closer to the mean. These alterations were found to be driven by both clutch size changes and incubation behaviour. Synchrony between hatch dates and caterpillar peak abundance was shown to be significantly improved for observed hatching compared to a null estimation (where incubation behaviour was assumed to be the same for all individuals). This suggests that incubation behaviour is being used to improve synchrony with peak food availability.

In this study we were interested in how different aspects of incubation behaviour drive the observed post clutch initiation adjustments and to identify the temperature cues which best explain this variability. In order to distinguish different elements of incubation behaviour we used a passive method of in-nest temperature recording. Through this we were able to accurately distinguish the onset of full incubation and consequently quantify the relative onset, duration and intensity of incubation. We demonstrate that incubation adjustments are driven by variability in multiple aspects of incubation behaviour. The onset of incubation relative to clutch completion, the duration of full incubation and intensity of incubation effort are all highly variable. Cumulatively alteration to these behaviours could lead to almost eight days advance or 20 days delay in hatch timing. At least some of the variation observed in these behaviours can be explained by temperature changes from just prior to the laying

period, right up until hatching. An interplay between the energetic costs of incubation (Visser & Holleman 2001) and the fitness costs of timing hatching poorly in relation to peak food abundance could also play a role in determining optimal incubation behaviour. However this study focuses on the role incubation behaviour plays in the timing of hatching.

Different aspects of incubation behaviour correlated with different temperature variables. Sliding time window analyses identified a critical temporal window of mean temperatures from seven days prior to laying up until incubation onset (see Figure 2.1) as the best predictor of variability in relative incubation onset. Warmer mean temperatures during this period corresponded to earlier relative onsets with an advance of approximately five days per ºC but mediated by an interaction with clutch initiation date. Earlier layers displayed the strongest negative temperature-onset relationship, supporting previous findings that early layers are most able to alter their timing (Cresswell & McCleery 2003). This suggests that temperatures around laying do act as a cue for the onset of incubation, supporting several previous findings (Stenning 2008; Cresswell & McCleery 2003; Álvarez & Barba 2014). However, the cues for incubation behaviour were not found to be uniform across different aspects. While some of the findings from this study support previous work suggesting that mean temperature is an important phenological cue (Cresswell & McCleery 2003; García-Navas & Sanz 2011; Álvarez & Barba 2014) we also show that, for some aspects of incubation behaviour, temperature extremes are more important. This demonstrates a need to test multiple temperature measures when considering phenological cues. It should also be noted that temperatures do auto-correlate throughout the year and between different measures (mean, maximum, minimum and range), this can make distinguishing a definitive temperature cue difficult.

The different aspects of incubation behaviour studied here, were not completely independent. Both the intensity and duration of incubation appeared to be partly constrained by the onset of incubation. The daily intensity of incubation effort showed a significant positive relationship with daily maximum temperatures during the incubation period, even when accounting for stage of incubation. In contrast, incubation duration showed no significant relationship to any temporal window or temperature measure tested here. This could arise because incubation duration is constrained due to an interaction between the number of hours of incubation required for development and the relative onset of incubation. Consequently it may not exhibit plasticity in response to temperature variation, despite changes to intensity. Incubation durations are highly variable, but a significant portion of this variability can be explained by the relative onset of incubation. Mean incubation intensity showed no relationship with clutch size or lay date (proxy for individual condition) but a significant negative relationship with the relative incubation onset. Higher mean intensities corresponded with earlier incubation onsets. This could indicate an attempt by females to advance their hatch date via both onset and duration of incubation. However, this relationship is not shown for duration itself. Incubation duration showed no significant relationship to either clutch size or mean intensity of incubation. It did however, show a significant negative relationship with relative incubation onset, earlier onsets corresponded to longer durations. The lack of a relationship between intensity of incubation effort and the duration of incubation could be explained by the role of night time incubation. As great tits incubate at night from several days prior to clutch completion (Haftorn 1981) those individuals initiating full incubation prior to clutch completion will have accumulated fewer active incubation hours than an individual who initiates after the clutch is completed. Consequently

females incubating early will need to input more hours of full incubation than those who delayed. This would lead to a modulation of any intensity-duration relationship with females with earlier onset requiring a greater intensity of incubation in order to achieve the same total duration as a female with later onset.

Here, multiple aspects of the breeding cycle have been shown to play a role in achieving synchrony between hatching timing and caterpillar peak abundance. We have demonstrated that elements of incubation behaviour alter hatch dates in different ways. The relative onset of incubation significantly alters the mean hatching synchrony to create better matching with the caterpillar peak abundance. The duration of incubation altered the variance of hatching timing, although not significantly in the focal year. The relative onset of incubation had no influence on the variance of hatching timing, therefore all variance changes are likely attributable to alterations of incubation duration. Although not statistically significant in 2014, we did find that incubation behaviour significantly altered variance in hatching timing across long term data, which could be driven by duration changes. As a result, incubation duration changes do appear to play a notable role in eventual synchrony between hatching and caterpillar peak abundance across years.

Hatching timing, relative to the caterpillar peak, has a significant impact on reproductive success and consequently, all alterations discussed here play a key role in fitness. Further to direct effects on the relative hatching timing, incubation alterations could also have other knock on fitness impacts. The shifting of incubation onset earlier could influence reproductive success through increases in asynchrony in the hatching of individual eggs within a clutch (Stenning 2008; Lord et al. 2011; Johnson et al. 2013; Cresswell & McCleery 2003) . This can drive rapid brood

reduction in years with little food or when matching is poor, therefore reducing recruitment and having impacts on population dynamics.

When assessing plasticity in different elements of the breeding cycle, it is important to consider the limits and constraints on this flexibility. There is an inherent asymmetry in adjustments to incubation behaviour. As demonstrated with females in our study, it is much easier to delay incubation onset than to advance. An advance in incubation onset is bounded by initial lay date but eggs remain viable many weeks after being laid (Perrins 1970a) so females have more flexibility to delay incubation than they do to advance it (van Noordwijk et al. 1995). Therefore, if temperatures after laying advance faster than prior to laying, as is occurring for great tit populations in the Netherlands (Visser et al. 1998; Visser et al. 2003), plasticity in incubation behaviour alone may not be sufficient to compensate for a late lay date in comparison to subsequently early caterpillar peak. This is demonstrated by reduced reproductive success for this population in years when temperatures change drastically after laying (van Noordwijk et al. 1995). Additionally, females may be energetically constrained from advancing incubation very early in the spring when resources are scarce (van Noordwijk et al. 1995).

The patterns shown here are not unique to great tits. Variation in timing after reproduction has been initiated has been shown in blue tits (Stenning 2008; García-Navas & Sanz 2011; Matthysen et al. 2010), tree swallows (Ardia et al. 2010), ducks (Hepp et al. 2006), cervids (Asher et al. 2005; Asher 2007; Moyes et al. 2011), and bats (Racey & Swift 1981) . In order to predict how populations of temperate species will be impacted by further climate changes it is essential to consider the mechanisms behind phenological synchrony, the responsiveness to temperature, as

well as their constraints. Many species appear to display considerable flexibility in multiple aspects of their breeding phenology. Therefore they might have greater potential to adapt to further climatic variability than would be assumed from looking at only single aspects of reproduction. In this study, we demonstrated that great tits do alter incubation behaviour in response to ambient temperatures from prior to laying right up until hatching, even adjusting to daily maximum temperature changes during incubation. These adjustments improve the synchrony between hatch timing and caterpillar peak abundance and have knock on impacts to reproductive success through this improved matching.

2.6. Acknowledgements

We are grateful to all of the Wytham fieldworkers who collected population census data on the Wytham great tits. This work was supported by NERC grant NE/K006274/1 to Ben Sheldon. We also thank Will Cresswell and another anonymous reviewer for their helpful comments and suggestions.

2.7. Supporting Information

2.7.1. iButton construction and placement

Illustrative photographs of how iButton thermometers were placed in great tit nests.

Figure S2.1: Photographs of the construction of in-nest iButtons. From left to right: the iButton alone, the iButton with its garden wire anchor, the iButton secured in cloth pouch, the finished iButton placed in a nest amongst the eggs

2.7.2. Defining the onset of daytime incubation

The definition of the onset of incubation can be a challenging subject. As incubation is a gradual processes, usually beginning at night and then gradually extending into the active day (daylight hours) creating a clear cut off for when this incubation began is somewhat arbitrary. Haftorn (Haftorn 1981) discusses this dilemma in some length and defines two definitions of incubation onset; onset 'senso stricto', which would refer to the onset of the first incubation (usually at night) and onset of steady incubation (what we term full incubation) when a female's attentiveness reaches optimum. It was the onset of full incubation that we are interested in here and (Haftorn 1981) defined this as when 60-85 % of the active day was spent incubating (however this was specific to Norway). In a paper by Haftorn (Cresswell & McCleery 2003) on great tits in the UK (Wytham Woods), they discussed a threshold of 75 % of the day spent incubating as being important. If a day is taken to be 24 hours then this 75 % corresponds to 50 % of the active day, so comparable but just below the threshold of 60 % used for Haftorn (Haftorn 1981) in Norway. Following on from this pervious work, in this study we chose a definition of full incubation of 50 % of the active day spent incubating to be comparable with these previous studies. The cut-off of 50 % also takes account of the uneven start of incubation. Gibb, Kluijver, Hinde, and Haftorn (Gibb 1950; Kluijver 1950; Hinde 1952; Haftorn 1981) all note that incubation during the active day begins in the afternoon and then gradually extends to morning. We observe this same pattern in every one of the nests in this study, therefore the cut off of 50 % of the active day also represents the time when incubation extends beyond midday and into the morning, defining a difference between partial daytime incubation and more steady incubation.

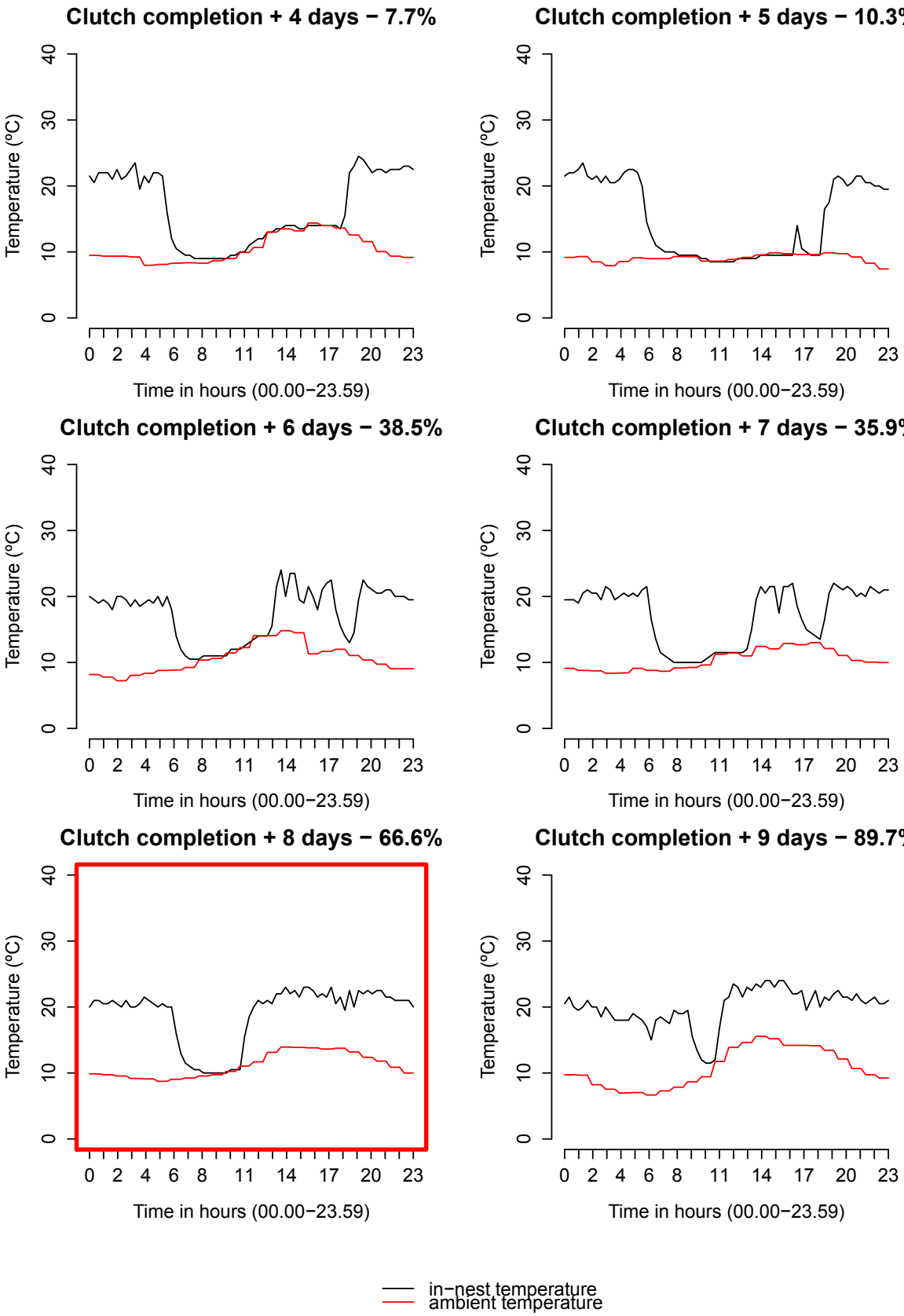

Figure S2.2: Graphs demonstrating the identification of full incubation onset.

Each window is a single day (00:00-23:59) labelled with the day after clutch completion and the % of the day spent incubating. The red outline of the graph box indicates the day identified as the onset of full incubation

2.7.3. Testing for potential bias in the incubation behaviour of individuals who removed iButtons

Of the 163 iButton placed in great tit nests in the 2014 breeding season, 54 were removed by females and therefore no incubation data could be obtained from these nests. It is possible that the females who rejected iButtons had a tendency towards a certain incubation behaviour, such as advancing or delaying. If this were the case, this could create a bias in the remaining sample from nests where individuals did not remove iButtons. As a result we tested whether the clutch completion to hatch interval varied significantly between nests where iButtons were removed and nests where iButtons remained for the duration of incubation. The clutch completion to hatch interval was used as a proxy for incubation behaviour as exact incubation onset was not known for nests where iButtons were removed. This is an aggregate measure that should capture both changes to the relative onset of incubation and duration but does not distinguish between the two.

An unpaired T-Test was performed to determine if the mean clutch completion to hatch interval varied between the groups and variance differences were assessed with an F-Test of variance. Both tests showed no significant differences (T = -0.57, DF = 60.3, P = 0.57 and F = 0.73, DF = 103/38, P = 0.22). The lack of a difference between the distribution of clutch completion to hatch intervals between those nests where iButtons were removed and those where they remained can also be seen visually in Figure S2.3

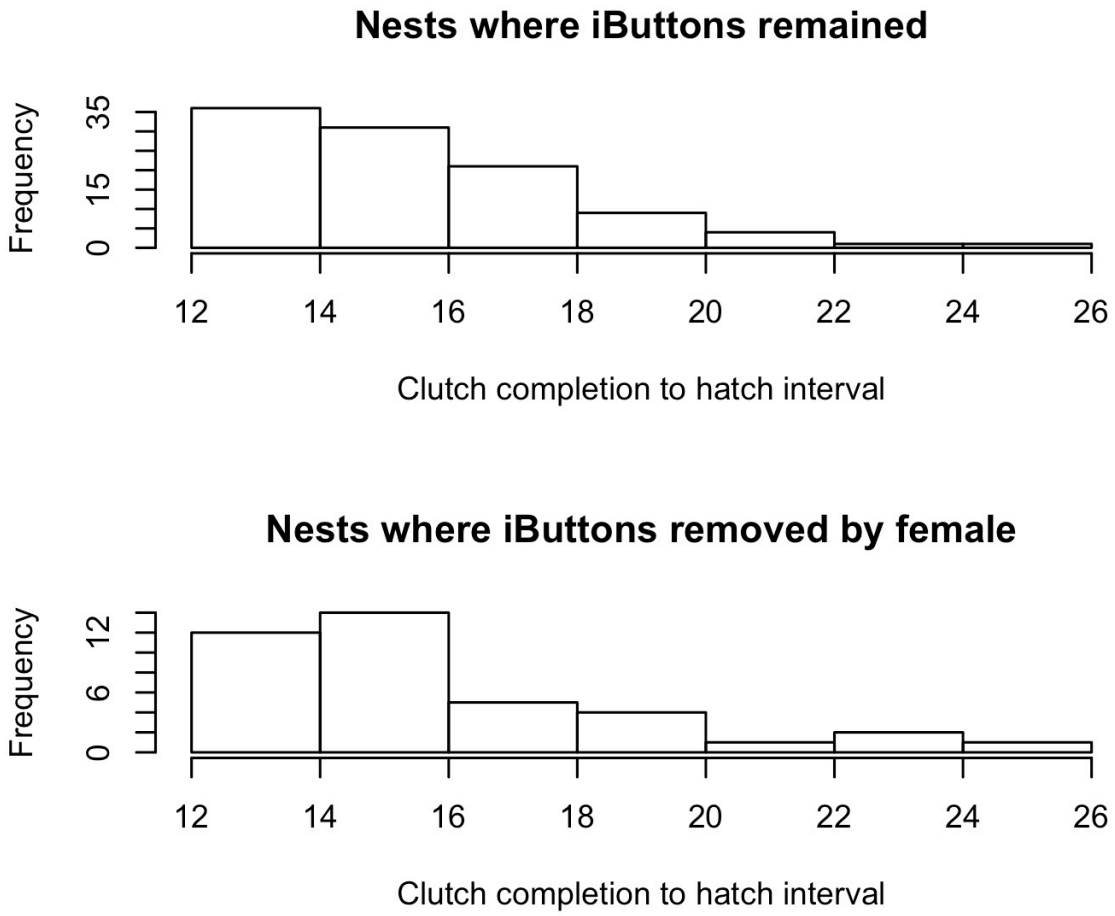

Figure S2.3: Histograms of the distribution of clutch completion to hatch intervals in nests where iButtons were removed and nests where iButtons remained for the duration of incubation

2.7.4. Which temperature window and measure best explains variance in the relative onset of incubation?

Below we present the full results of the model selection procedure for the sliding time window analysis on the relative onset of incubation. We tested 176 candidate models with different fixed effect configurations and different temporal temperature windows and temperature measures and two null models. Shown here are the AIC results of all models with a ΔAIC of < 2 (11 models) and the first model with a ΔAIC of > 2. We also present the null models, which have ΔAIC values considerably above 2.

Table S2.1: Results of model selection from the sliding time window analysis for the relative onset of incubation.

Models with ΔAIC values of less than two are included in addition to the first model with a ΔAIC of greater than two and both null models

Model configuration	Temperature variable	AIC	Delta AIC	K
Clutch initiation * temperature	Mean - window 8	481.86	0	5
Clutch initiation * temperature	Mean - window 9	481.9	0.05	5
Clutch initiation + temperature	Mean maximum - window 8	482.86	1.01	4
Clutch initiation * temperature	Mean - window 7	482.89	1.03	5
Clutch initiation * temperature	Mean - window 10	483.26	1.4	5
Clutch initiation + temperature	Mean maximum - window 1	483.36	1.51	4
Clutch initiation + temperature	Mean maximum - window 10	483.41	1.55	4
Clutch initiation + temperature	Mean maximum - window 9	483.41	1.55	4
Clutch initiation + temperature	Mean maximum - window 7	483.69	1.83	4
temperature	Mean maximum - window 8	483.75	1.89	3
temperature	Mean maximum - window 1	483.8	1.94	3
temperature	Mean maximum - window 10	483.91	2.06	3
Clutch initiation	Null	487.47	5.62	3
Clutch initiation + clutch size	Null	489.44	7.58	4

We also present the parameter estimates for all 11 'preferred' models (ΔAIC < 2). Dependent on the model configuration some table elements are missing because these variables were not included in the model itself.

Table S2.2: Parameter estimates for the temperature variable, clutch initiation date and their interaction for all models with ΔAIC <2.
Models appear in the same order as Table 1.

Temperature measure	Temperature			Clutch initiation			Interaction		
	Est	SE	P	Est	SE	P	Est	SE	P
Mean - window 8	-5.03	1.72	0.004	-3.25	1.35	0.02	0.3	0.13	0.02
Mean - window 9	-5.4	1.83	0.004	-3.53	1.43	0.02	0.33	0.14	0.02
Mean max- window 8	-1.01	0.39	0.01	-0.09	0.06	0.1	NA	NA	NA
Mean - window 7	-4.51	1.68	0.01	-2.78	1.3	0.04	0.26	0.12	0.04
Mean - window 10	-4.94	1.84	0.01	-3.13	1.43	0.03	0.3	0.14	0.03
Mean max - window 1	-0.93	0.37	0.02	-0.09	0.06	0.13	NA	NA	NA
Mean max - window 10	-0.97	0.39	0.02	-0.09	0.06	0.12	NA	NA	NA
Mean max - window 9	-0.97	0.4	0.02	-0.09	0.06	0.11	NA	NA	NA
Mean max - window 7	-0.93	0.39	0.02	-0.09	0.06	0.12	NA	NA	NA
Mean max - window 8	-0.73	0.36	0.04	NA	NA	NA	NA	NA	NA
Mean max - window 1	-0.71	0.35	0.05	NA	NA	NA	NA	NA	NA

Results of this sliding window analysis showed two outlier values where mean temperatures for window 8 (the temperature variable of the lowest AIC model) were below 8.5 ºC. To ascertain the effect of these outliers on our parameter estimates, we removed them and re-ran the model with the lowest AIC. The parameter values produced were largely the same as the model with all data, therefore, we opted to retain the maximum amount of data and present the full model in the main article. Parameter values shown in Table S2.3.

Table S2.3: Parameter values for the model with the lowest AIC with outliers removed.

Temperature - mean window 8			Clutch initiation			Interaction		
Est	SE	P	Est	SE	P	Est	SE	P
-5.80	1.79	0.001	-3.55	1.39	0.01	0.33	0.13	0.01

2.7.5. Which temperature window and measure best explains variance in the duration of incubation?

Here we present the full results of the model selection procedure for the sliding time window analysis on incubation duration. We tested 180 candidate models with different fixed effect configurations and different temporal temperature windows and temperature measures and two null models. 67 models had a ΔAIC of < 2, including one of the null models (with clutch initiation date only). None of these models showed a significant relationship between temperature and incubation duration. Consequently here we present only the three models with the lowest AIC values (the 3rd of which is a null model) and the other null model.

Table S2.4: Results of model selection from the sliding time window analysis for incubation duration.
Models with the three lowest ΔAIC values and the remaining null model are shown. All models also included a fixed effect of relative incubation onset

Model configuration	Temperature variable	AIC	Delta AIC	K
Temperature	Range - window 11	347.98	0	5
Temperature	Range - window 9	348.03	0.05	5
Clutch initiation	Null	348.08	0.1	4
Clutch initiation + clutch size	Null	351.03	3.05	5

Table S2.5: Parameter estimates for linear model of incubation duration with the lowest AIC.

Temperature - range window 11			Relative incubation onset		
Est	SE	P	Est	SE	P
0.26	0.15	0.08	-0.41	0.05	<0.001

2.7.6. Which temperature window and measure best explains variance in the daily intensity of incubation effort?

Here we present the full results of the model selection for the temperature cues driving variation in the intensity of incubation effort. We tested 20 candidate models which included the effects of various temperature measures, day through incubation, clutch size and clutch initiation date and a null model. Models with a ΔAIC of < 2 are presented below, in addition to the null model.

Table S2.6: Results of model selection for GLMMs for daily incubation intensity.
Models with ΔAIC values of less than two are included in addition to the first model with a ΔAIC of greater than two and the null model

Model configuration	Temperature variable	AIC	Delta AIC	K
Incubation day * temperature	Daily maximum	7538.66	0	5
Incubation day + temperature	Daily maximum	7570.38	31.72	4
Incubation day + temperature	Temperature range	7600.07	61.41	4
Incubation day	Null	8249.51	710.85	3

Parameter estimates for binomial GLMM model of proportion of the active day incubated as a function of temperature and incubation day. Random effect of individual had 0 variance

Table S2.7: Parameter estimates for binomial GLMM model of proportion of the active day incubated as a function of temperature and incubation day.
Random effect of individual had 0 variance

Variable	Estimate	SE	P
Daily max temperature	0.1	0.01	<0.001
Incubation day	0.22	0.02	<0.001
Daily max temperature:Incubation day	-0.008	0.001	<0.001

CHAPTER THREE

Experimental manipulation of nocturnal nest cavity temperature in wild blue tits

Emily G. Simmonds, Ben C. Sheldon, Tim Coulson, and Ella F. Cole

3.1. Abstract

Advances in the timing of reproduction in temperate species are some of the most well documented biotic responses to increasing global temperatures. However, the magnitude and rate of these advances in timing are not equal across all taxonomic groups. These differences can lead to disruption of interspecific relationships if species respond differently to temperature changes. Understanding the relationship between temperature and phenology is a key step in predicting future population trends for species living in seasonal environments. However, experimentally manipulating temperature in the wild is logistically challenging and has consequently rarely been attempted. In this study we experimentally test whether in-nest temperatures in early spring act as a cue for breeding phenology in a population of wild blue tits (*Cyanistes caeruleus*). We split nests into three treatments; heated, cooled, and control. In-nest temperature in the heated and cooled boxes was manipulated by an average of ± 0.6 ºC from control temperatures using heating devices and ice packs respectively. We assessed the impact of our experimental manipulation on box occupancy and reproductive timing. We found trends towards earlier phenology in heated nest boxes in addition to a higher occupancy rate in cooled boxes, however neither of these trends was found to be statistically significant. Our ability to distinguish statistical signals was hampered by unexpectedly low occupancy rates across all experimental treatments. Based on the results we cannot say if nocturnal in-nest temperature is an important cue for nest box choice or the timing of laying.

3.2. Introduction

Advances in phenology, the timing of life history events, have been widely documented amongst temperate species over recent decades, with examples spanning multiple ecosystems and almost all taxonomic groups (Thackeray et al. 2010). These advances have coincided with increases in global temperatures as a result of anthropogenic climate change (Menzel et al. 2006; Parmesan 2006; Singer & Parmesan 2010). Failure to time breeding to match with the peak abundance of resources can result in energetic demands that cannot be met and a consequent reduction in fitness (Reed, Grøtan, et al. 2013; Reed, Jenouvrier, et al. 2013). Therefore understanding how temperature influences breeding phenology is crucial for predicting the impact of climatic change on reproductive success. Despite this importance, the precise role that temperature plays in altering breeding phenology has not yet been determined.

Matching peak energetic demands to a shifting resource peak is particularly challenging for consumer species, which rely on species in lower trophic levels as their resource. The consumer species must respond to temperature in the same way as the resource in order to maintain synchrony, this is challenging for species with temperature independent development, where embryos have a fixed developmental period. In contrast to species with temperature-dependent development, which respond directly to temperature changes (Perrins 1979), temperature-independent species must use predictive cues in order to determine the optimal breeding time (van Noordwijk et al. 1995; Perrins 1979). Not all species respond to the environment in the same way, this can have significant impacts on interspecific interactions (Thackeray et al. 2010; Parmesan 2007; Visser & Holleman 2001; Plard et al. 2014). If interacting species differ in their response to environmental cues or the specific

cues they use to time their phenology, they will shift their timing to differing degrees, which can disrupt synchrony between species (Cushing 1969). For instance, if a prey species advances their phenology at a faster rate than a predator species, due to greater responsiveness to an environmental cue this could result in temporal mismatch. Peak abundance of resources would occur prior to the peak energetic demand, reducing survival and reproductive success of the predator (Reed, Grøtan, et al. 2013). This has been the case for great tits (*Parus major*) in the Netherlands. Great tits in the Hoge Veluwe population have not advanced their phenology as much as resources despite increases in spring temperatures, whereas peak abundance of their prey species, winter moth caterpillars (*Operapthera brumata*) has advanced by over one week in two decades (Visser et al. 1998). This has resulted in fitness reductions for these great tits (Reed, Grøtan, et al. 2013). In order to understand how and where these mismatches may occur, it is important to identify the precise cues that individuals use to time their reproductive behaviour. Only through understanding the mechanisms that control reproductive timing can we predict how timing will change into the future and whether temporal mismatch will be likely.

Attempts to identify the cues that drive reproductive phenology consist of several approaches. The most commonly used approach is to conduct regression based analyses on observational data (van Noordwijk et al. 1995; Charmantier et al. 2008; Menzel et al. 2006; Parmesan & Yohe 2003; Torti & Dunn 2005). This approach typically consists of regressing candidate cues (often temperature) against a phenological event of interest. Either testing different temporal windows or days of temperature

(Roberts 2008; Thackeray et al. 2016; van de Pol et al. 2016; Phillimore et al. 2013). The results of studies of this type, have shown that the breeding phenology of northern temperate species is highly correlated with various measures of spring temperature (Visser et al. 1998; van Noordwijk et al. 1995; Charmantier et al. 2008; Cleland et al. 2007; Menzel et al. 2006; Parmesan & Yohe 2003). April temperatures were shown to correlate with clutch initiation date in gulls (Brommer et al. 2008) and temperatures during March and April explained 70 % of the variation in clutch initiation date for great tits and blue tits (*Cyanistes caeruleus*) in the UK (Perrins & McCleery 1989). However, it can be difficult to pinpoint exact temperature cues through these analyses given that temporal windows and the temperature variables tested are all at least partially correlated. Furthermore, regression based studies do not allow the determination of causality, they can only assess if two variables correlate. In order to establish causal links between environmental cues and phenological events we must therefore use experimental manipulations are required.

Experimental manipulations of temperature cues pose a logistical challenge, particularly for large vertebrate species. While sessile species, such as plants, can be manipulated in situ, mobile species such as most animals, cannot. Animal species tend to cover considerable distances, daily, weekly or seasonally, experiencing temperature cues across a large spatial and temporal range. Consequently, manipulating temperature across these scales in the wild is often impossible. As a result, captive experiments provide one alternative (Schaper et al. 2011; Schaper et al. 2012). These captive experiments provide a good environment to determine the role of different cues and test causality. For instance, captive experiments on great tits used temperature controlled aviaries to explore the influence of temperature on clutch initiation date. This experiment showed that while an increasing temperature

caused clutch initiation date to advance, changes to mean temperature or temperature variation alone did not impact phenology (Schaper et al. 2012). However, experiments in captivity do not tell us whether these relationships would hold in the wild. There have been examples of experimental manipulations failing to capture observed patterns (Lambrechts et al. 1999; Wolkovich et al. 2012). Captive environments have highly controlled conditions, often investigating a single focal cue at a time (Schaper et al. 2011; Schaper et al. 2012) and consequently failing to take account of the influence of other variables. It is unlikely in reality that a single driver determines all phenological change. Furthermore, captive environments create artificial conditions, which are potentially stressful for the animals and can lead to adjustment of behaviour away from what it would be in the wild (Lambrechts et al. 1999).

To identify whether relationships identified during captive experiments or regression analyses also hold in the wild we need to perform experiments on natural systems. Cavity nesting birds, such as great tits and blue tits, provide a good study system for such experiments. Reproductive timing plays an important role in reproductive success in these species, as a failure to time the peak energetic requirement of reproduction with the peak availability of food resources can result in reduced reproductive success (Reed, Grøtan, et al. 2013). There have also been many statistical assessments of cues on these populations (Husby et al. 2010; van Noordwijk et al. 1995; Perrins & McCleery 1989; Perrins 1965b) and some captive experiments (Schaper et al. 2012), which can be used to inform field experiments. Additionally nest cavities provide an enclosed environment that can be manipulated. Previous studies that have experimentally heated and cooled nest cavities, have shown that cavity temperature influences nocturnal incubation (Vedder 2012) and the

incubation intensity (Ardia et al. 2010) in blue tits and tree swallows (*Tachycineta bicolor*), respectively. In contrast, in spite of observational evidence that suggests nest temperatures influence clutch initiation date (Dhondt & Eyckerman 1979; O'Connor 1978) similar experimental manipulations have been unable to demonstrate this effect. A previous experiment on great tits attempting to manipulate clutch initiation date (Nager & van Noordwijk 1992), through heating and cooling nest boxes in two consecutive years, failed to identify a significant signal of temperature on clutch initiation date. This experiment heated and cooled groups (N = 40 boxes per year) of nest boxes at night by ± 0.75 ºC from two to four weeks prior to the median clutch initiation date of each year; however, no difference in phenology was detected between treatments. Here we repeat and extend this study using a population of wild blue tits. We heated and cooled 87 (N = 29 per treatment) nest boxes by an average of ± 0.6 ºC from 6 pm to midnight from 17 days before the first egg was laid, and testing whether this manipulation influences nest box choice, commencement of nest building and clutch initiation date. We hypothesise that female blue tits will preferentially choose to nest in warmer nest boxes in order to minimise their nocturnal energy expenditure (Ardia et al. 2009; Vedder 2012; Dhondt & Eyckerman 1979). Furthermore, we predict that the temperature of the nest cavity will also act as a cue for clutch initiation date (Dhondt & Eyckerman 1979; O'Connor 1978), therefore the warmer nests will also experience the earliest clutch initiation dates and the coolest boxes, the latest clutch initiation date.

3.3. Methodology

3.3.1. The Study system

This study was conducted in spring 2015 using a population of wild blue tits in Wytham Woods, UK. The nest box breeding population of blue tits has been studied in detail using standardised procedures (Perrins 1965a; McCleery & Perrins 1985)

since 2002. Annually 250-450 pairs of blue tits breed in the woodland (Perrins 1979). 147 blue tit nest boxes were erected across the woodland in 2002, with a further 41 boxes added in 2008. These blue tit boxes have entrance holes with a diameter too small to allow larger species, such as great tits, to enter the boxes. By using these boxes rather than those with larger entrance holes, which are also present in the study area, we were able to exclude great tits from our experimental manipulations. We use 87 of these blue tit nest boxes as the focus of our study (Figure 3.1). These 87 were chosen because they are located close together in a topographically similar area of the woodland, consequently limiting habitat differences. Proximity of the boxes was essential for this experiment to ensure all experimental treatments could be set running within a two hour window each evening.

3.3.2. Environmental data collection

Altitude data were extracted by Wilkin et al (Wilkin et al. 2006) from an Inverse Distance Weighting interpolation of a 50 m resolution Land Form profile Digital-Terrain-Model data set provided by Ordnance Survey.

Local ambient temperature was collected via a grid of ambient temperature iButtons (DS1923-F5, accurate to ± 0.5 ºC, HomeChip Ltd) set to measure absolute temperature every 30 minutes. 200 of these ambient temperature iButtons were distributed in a grid system across Wytham woods with positions chosen to reflect the density of nest boxes. QGIS (Team 2016) was used to match each experimental nest box to a nearest ambient temperature iButton, therefore allowing the calculation of local ambient temperature for each box. The measure of local ambient temperature used in this study is mean temperature for the experimental period, from 1st April to 6th May or until the first egg is laid, whichever is first.

Long-term temperature data is taken from the Radcliffe Meteorological Station (located 5 km East of Wytham Woods). This station records absolute temperature during the day and calculates daily minimum and maximum temperatures from 1815 to the present day (Radcliffe Meterological Station n.d.).

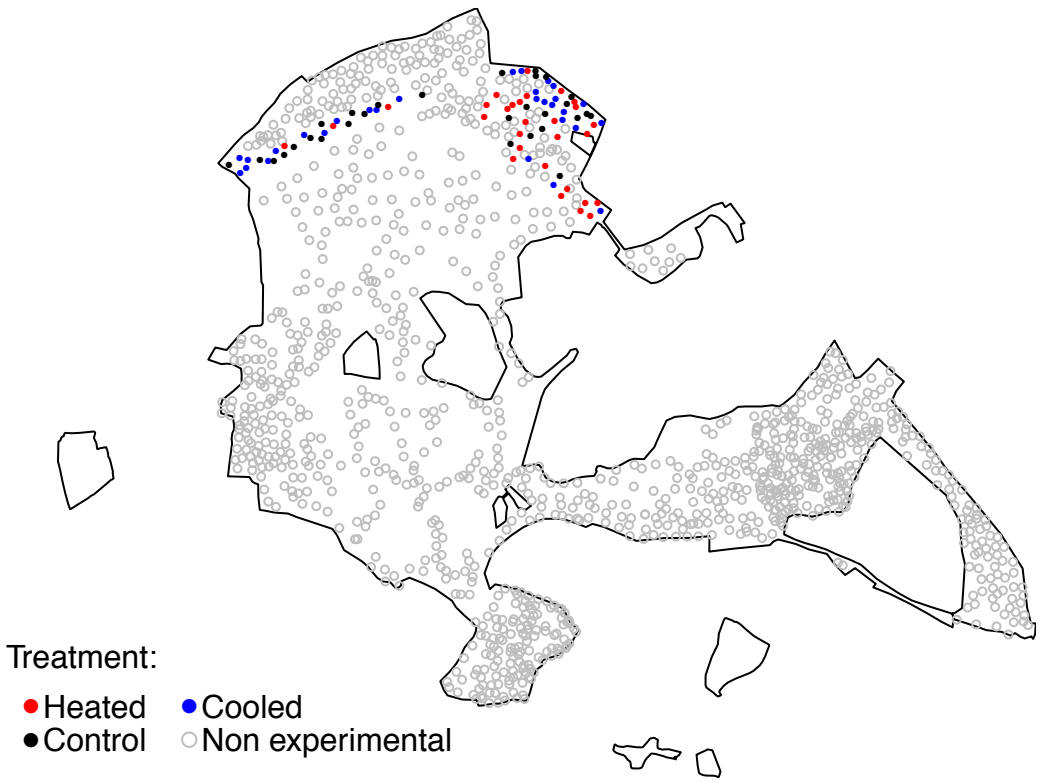

Figure 3.1: Map of position of experimental (filled circles) and non-experimental (open circles) nest boxes in Wytham woods.
Experimental treatments are colour coded.

3.3.3. Experimental Manipulation and Phenological Data Collection

Of the 87 experimental boxes, 29 were assigned to each of the three treatment types: heated, cooled, and control. Boxes were assigned to treatments using a stratified sampling technique, taking account of historical box popularity. Historical popularity was determined by calculating the proportion of years in which each box has been occupied since its placement in the woodland. Boxes were then grouped into bands of popularity in increments of 10 %, within each group treatments were

assigned randomly using R (Team 2008). This assignment method was used to prevent a clustering of treatments by historical popularity as this metric showed a significant negative correlation with mean clutch initiation date of the nest box (EST = -3.18, SE = 1.46, P = 0.03), with the historically most popular boxes having the earliest clutch initiation dates. This sampling approach also created an even likelihood of box occupancy across treatments, therefore allowing us to explicitly test whether our experimental manipulation altered the popularity of individual boxes. We tested whether other spatial components, such as altitude, influenced the mean clutch initiation date of each box, but this was not statistically significant (EST = 0.02, SE = 0.03, P = 0.52). Consequently we did not take account of altitude when assigning treatments. This assignment of treatments does mean individuals can move between different treatments if they do not roost in the same nest box every night, however, it also avoids temporal or spatial clustering of treatments and is essential for the exploration of in-nest temperature influence on occupancy.

Manipulations began on the 27th of March 2015, 17 days before the first egg was laid in the experimental boxes. Manipulations occurred every evening, ending when the first egg was laid in a box or on the 6th May 2015 for unoccupied boxes, 30 days after the first egg was laid in the whole woodland. The cut off of 30 days after the first egg is a common measure used to exclude second clutches and repeat clutches which result from early failures (Van Der Jeugd & McCleery 2002).

The heating device comprised (see Figure 3.2 for details of device construction) of a three-volt filament bulb secured in a mount within a metal topped polystyrene disc (approximately 2.5-3.5 cm thick), which was placed on top of a plastic ring with a 5 cm gap at the front (also 2.5 cm in height). The bulb was powered by two

rechargeable batteries, size C, located beneath the polystyrene disc within the plastic ring (see Figure 3.2). The bulb was covered from above by a 5 cm diameter metal disc, to hide the light produced. Heating was achieved by the bulb heating the metal disc and convectional transfer of the heat up into the nest cavity. These heated boxes also contained four, non-frozen, Thermafreeze ice packs (Thermafreeze EU, black ice packs) in the same configuration as the cooling device (see below) to ensure that all experimental boxes had the same outward appearance.

Figure 3.2: Photograph of experimental set up for a heated box.
Key components of the set up are numbered 1) Ice packs mounted front and back of the box, 2) battery pack, 3) polystyrene insert, 4) plastic ring to hold polystyrene above the battery pack, 5) metal disc topping the bulb. In control and cooled boxes the battery pack would be replaced by a second polystyrene insert.

The cooling device comprised two frozen Thermafreeze ice packs secured at the back of each nest box and two on the nest box door. Ice packs at the back of the box were secured using the same metal topped polystyrene disc as used in the heating device, with another polystyrene disc below (around 2.5 cm thick) to achieve the

same thickness as the heating device but without the plastic ring and batteries. Ice packs on the nest box door were secured using rubber bands. All ice packs had a black outer side to camouflage them from the birds once the nest box was closed.

The control treatments comprised the same construction as the cooled treatment but with non-frozen ice packs. Therefore all boxes had the same appearance, a raised metallic box floor of approximately 5 cm above the true base, with ice packs mounted front and back.

Both the batteries and frozen ice packs had an effective life of around six hours and therefore were replaced each evening around 2.5 to 1.5 hours prior to sunset so that the treatments were running during the night. This period of the day was chosen as female blue tits usually roost overnight within their nest cavity prior to nest building and therefore should be present within the box to experience the temperature manipulation (Kluijver 1950; Perrins 1965a; Kidd et al. 2015). Temperatures were altered on average by ± 0.6 ºC from the control boxes from sunset (around 6 pm) to midnight. Batteries and ice packs were replaced each evening; control boxes were also visited to achieve a uniform disturbance level across boxes. Some damage to equipment did occur (pecks to polystyrene inserts) and equipment was either repaired or replaced to ensure the treatments could continue. Experimental treatments could not be changed on two nights during the experiment (29th March 2015 and 31st March 2015) due to high winds. At the same time that treatments were changed several aspects of breeding phenology were recorded; signs of nest building (moss in box), signs of roosting (feathers and faeces), number of eggs, and whether the eggs were being incubated. From this we could calculate the nest building date (the date when moss first appeared in the nest box), clutch initiation

date (the date on which the first egg was laid in a nest box) and clutch completion date (the date when the last egg was laid in a nest box).

In-nest temperatures were recorded for the entire duration of the experiment using iButton thermometers (DS1921G-F5, accurate to ± 1 ºC, HomeChip Ltd) set to record temperature every 30 minutes. The iButtons were secured to the nest box wall using duct tape approximately 5 cm above the heating or cooling devices and half way between the front ice packs and the back ice packs. The effectiveness of the heating and cooling treatments during the experiment was assessed by calculating the mean temperature difference between treatments during the experimental periods (6 pm to 11:59 pm). The nights of the 29th and 31st March were excluded from these calculations because the experimental treatments did not run on those nights. Occupied nests were also excluded from the calculations after the first egg was laid, as the experimental treatments were halted at the appearance of the first egg. This type of experimental design was based on those previously employed on other nest box breeding passerine populations (Ardia et al. 2009; Pérez et al. 2008; YomTov & Wright 1993; Bryan & Bryant 1999; Vedder 2012; Alvarez & Barba 2014; Nager & van Noordwijk 1992). Previous manipulations have taken the form of heating and cooling usually during incubation (after laying has commenced) with heating devices attached to the outside of the nest box (Ardia et al. 2009; Pérez et al. 2008; YomTov & Wright 1993; Bryan & Bryant 1999) or within the nest box, below the nest itself (Vedder 2012; Alvarez & Barba 2014). Nager and Van Noordwijk (Nager & van Noordwijk 1992) heated and cooled nest boxes prior to laying through ice packs and heating devices located at the back of the nest box.

3.3.4. Statistical Analyses

A power analysis was conducted prior to starting the experiment to determine whether differences were likely to be detected between treatments given the number of nest boxes available. Based on the average historical occupancy of the study boxes (67 %) we would expect an occupancy rate of 19 boxes per treatment. With this number the power analysis showed that we could expect a 75 % chance of detecting a mid to strong signal (an effect size of 0.4 or higher). We would expect a mid to high effect size based on previous work that has found a strong correlation between temperature and clutch initiation date (Perrins 1970a; Perrins 1973). Furthermore, analyses of the relationship between spring temperature (mean of daily minimum and maximum temperatures for March and April) and clutch initiation date from 1960 to 2011 showed close to a 6 day delay in clutch initiation date for every 1 ºC decline in mean temperature (EST = -5.71, SE = 0.08, P = <2e-16), therefore we would expect a strong signal from our experimental manipulation.

We performed statistical analyses to determine the influence of our experimental manipulations on breeding phenology and box occupancy. For the analyses of the influence of experimental treatments on box occupancy, we first tested if there was a difference in the number of boxes occupied between treatments. A chi-squared test of association was performed to determine if the number of occupied boxes differed significantly between treatments. It is also important to consider whether any other environmental factors could be influencing occupancy. Philopatry could have a role here with individuals or pairs returning to the same boxes every year. Across the length of this study 12 % of individuals (8 % of breeding attempts) returned to the same box they had previously bred in, suggesting that for the majority of individuals previous location is not the only driver of nest box choice. Therefore we used one

way ANOVAs and a Welch's T-Test to explore the difference in historical popularity, altitude and ambient temperature between occupied and unoccupied boxes. A Welch's T-Test was used for local ambient temperature due to heteroscedasticity of temperature data. The whole dataset was used for these analyses as there was not enough data to split by treatment.

Analyses of the influence of experimental treatments on breeding phenology were performed using one way ANOVAs to determine if there are differences between phenology measures in each treatment group. The phenology measures considered here were the date of nest building (the date on which moss was first observed in the nest box), the clutch initiation date (date when the first egg is laid), and date of clutch completion (date when the last egg in the clutch is laid). Regression analyses were also performed to assess the influence of environmental variables on clutch initiation date. Environmental variables tested were altitude and local ambient mean temperature in a single linear regression to assess the combined influence of these variables on variance in phenology.

3.3.5. Ethics statement

Work was subject to review by the Department of Zoology ethical committee, University of Oxford. All work adhered to UK standard requirements and was carried out under Natural England licence 20114732. This experimental methodology was discussed with, and approved by the Departmental Animal Welfare Ethical Review Body (AWERB). Field work took place in Wytham Woods (lat. 51°46'N, long. 1°20'W), private land that belongs to the University of Oxford; for permission contact the Conservator, Nigel Fisher. No endangered or protected species were involved in the study. A breakdown of nest building, abandonment and chick mortality by

treatment can be found in supporting information Table S3.8, however these rates did not differ from those in non-experimental boxes.

3.4. Results

3.4.1. The influence of experimental treatment on occupancy

Occupancy in our experimental boxes was 16 %, which was a decline of 73 % from the preceding year (2014). Adjacent non-experimental boxes had a blue tit occupancy of 14 % (Table 3.1), a decrease of 48 % from 2014 occupancy. The number of great tits in these boxes also declined from 2014 to 2015, and to a greater extent, by 54 % (35 occupied in 2014 and 19 in 2015). Occupied boxes had higher historical popularity than unoccupied boxes ($F(1,82) = 6.15$, $P = 0.02$, see Figure 3.3), and this pattern was consistent across treatments. It should be noted that three of the occupied experimental boxes (N = 17) were occupied by marsh tits (*Poecile palustris*). As these are a different species to our study species, this could influence some of our results. Therefore we removed these nests from our analyses, leaving a sample size of N = 14.

Table 3.1: Numbers of occupied and empty boxes by treatment
***** Adjacent non-experimental boxes were defined as those boxes within 100 m of an experimental box. These boxes are those with the larger nest box entrance holes and therefore could be occupied by either great tits or blue tits. However we only include blue tits in this analysis. Great tit occupancy of the adjacent non-experimental boxes was 19/95, 20%

Treatment	Occupied boxes	Empty boxes (not occupied by blue tits)	Total
Cooled	7 (24%)	22	29
Control	5 (17%)	24	29
Heated	2 (7%)	27	29
Adjacent non-experimental *	13 (14%)	82	95

Table 3.2: Numbers of blue tit occupied nest boxes 2014 and 2015
* The whole woodland refers to all boxes in the entirety of Wytham woods (see Fig. 1, N = 1203). This includes all great tit (larger entrance hole, N = 1016) and blue tit only (smaller entrance hole, N = 187) boxes and covers all sections of the woodland. Occupancy numbers are for nests confirmed as blue tits, so including a minimum of 3 eggs. Adjacent non-experimental boxes are defined in using the same criteria as Table 1.
**Excluding marsh tits

	Boxes occupied 2014	Boxes occupied 2015 (year of experiment)	Proportional difference
Whole woodland *	430	258	- 40%
Non-experimental boxes	378	244	- 35%
Experimental boxes	52	14**	- 73%
Adjacent non-experimental boxes	27	13	- 48%

3.4.2. Does in-nest temperature influence nest box choice?

The number of nest boxes occupied during our experimental period varied between treatments (Table 3.1). Cooled boxes had the highest proportion (24 %) of boxes occupied, followed by control boxes (17 %) and heated boxes (7 %) have the lowest levels of occupancy. These differences were not statistically significant (χ^2 = 3.97, DF = 2, P = 0.14). We explored additional factors that may explain variation in occupancy. Neither altitude nor local mean ambient temperature differed significantly between occupied and unoccupied boxes (Altitude results, F(1,82) = 0.98, P = 0.33, local mean ambient temperature results - T = -1.00, P = 0.33, DF = 13.74, see supporting information, 3.7.2).

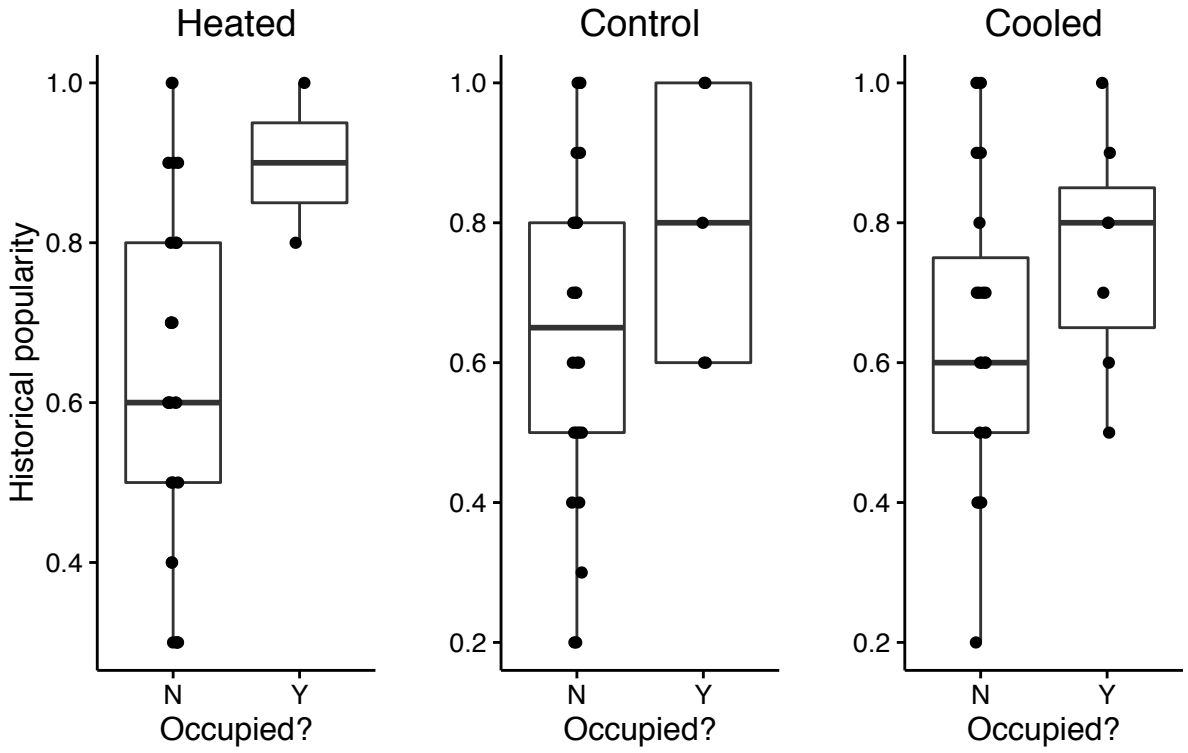

Figure 3.3: Boxplots of the historical popularity of occupied and unoccupied boxes by treatment.
Raw data are plotted as points and the box plots show the median and interquartile and full range for each treatment. Historical popularity represents the proportion of years a box was occupied since its placement in the woodland.

3.4.3. Does in-nest temperature influence breeding phenology?

For these analyses N = 13 (two heated boxes, five control boxes and six cooled boxes), one nest was removed from analyses due to lack of a clutch initiation date as laying occurred after the conclusion of experimental treatments. On average, birds nesting in heated boxes started nest building, laying and completed their clutches earlier than birds nesting in the control or cooled boxes (Figure 3.4), however none of these differences were statistically significant (nest building $F(2,11) = 1.20$, $P = 0.34$, clutch initiation date $F(2,11) = 0.91$, $P = 0.43$, clutch completion date $F(2,11) = 0.91$, $P = 0.43$) (for full ANOVA tables see supporting information, 3.7.3). The influence of environmental variables, altitude (EST = -0.15, SE = 0.08, P = 0.07) and local ambient temperature (EST = -2.52, SE = 4.67, P = 0.60), were also tested but

showed no statistically significant relationship with clutch initiation date, the focal phenological variable here.

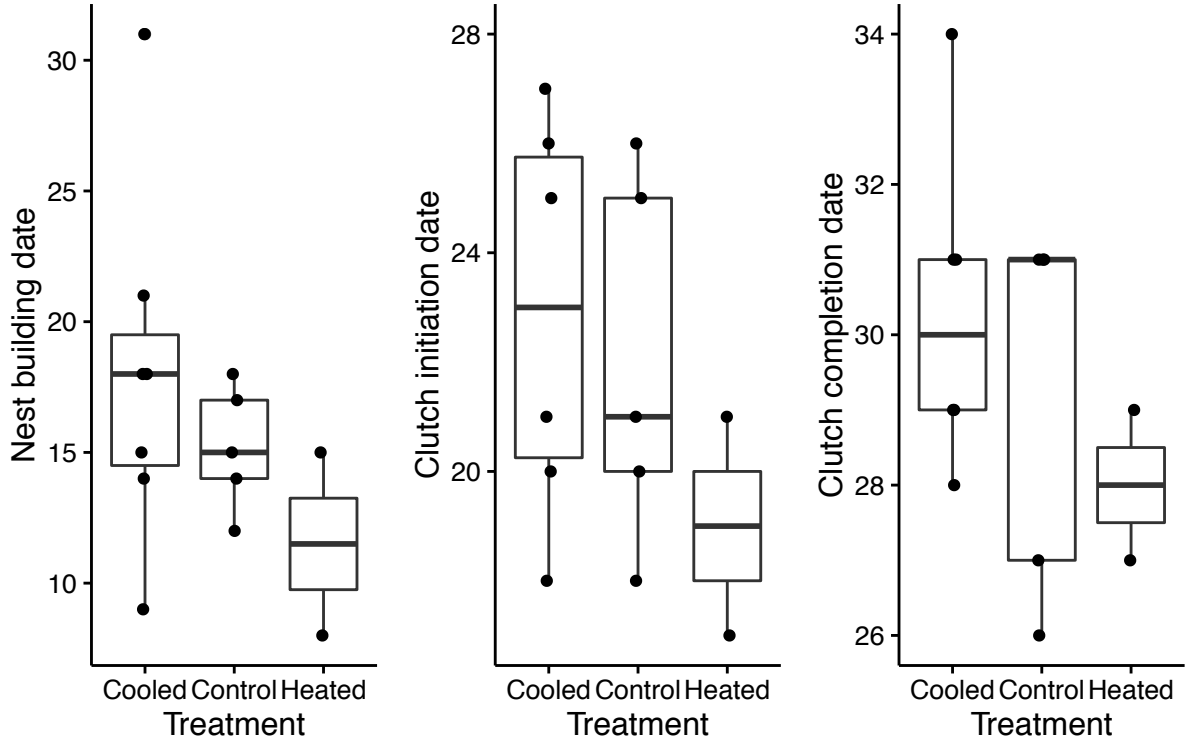

Figure 3.4: Boxplots of the dates of a) nest building, b) clutch initiation and c) clutch completion by treatment.
Raw data are plotted as points and the box plots show the median and interquartile range for each treatment. Dates are all plotted in April days (April 1st = 1)

3.5. Discussion

3.5.1. Experimental manipulation corresponded with low occupancy of nest boxes

This study experimentally tested whether nocturnal in-nest temperatures act as a cue for nest box choice and clutch initiation date in a population of wild blue tits. The experimental manipulation carried out in this study altered the minimum daily temperature experienced by individuals, either elevating it or reducing it dependent on treatment group. One of the defining findings of this study is a substantial and unforeseen decline in the occupancy of our experimental boxes to 16 %. This is considerably below the expected occupancy levels, which since 2008 have averaged 67 %. This 67 % occupancy would have given us approximately 19 boxes per treatment, which according to the preliminary power analysis conducted, would have

been sufficient to detect the mid to high signal we expected (effect size of 0.4 or greater). However, due to circumstances beyond our control, this sample size was not achieved. The potential to increase our sample size was limited by the availability of blue tit only nest boxes. While more blue tit boxes are available in the woodland the remainder are located either in different habitat types or spatially distant from the study boxes. The inclusion of these other boxes would have introduced confounding factors and been logistically difficult to manage. Changes of equipment had to be achieved within a two hour window each evening. However, with the sample size available, we should have been able to detect a signal if one had been present and box occupancy was at the expected levels.

The low occupancy levels achieved in this experiment have resulted from several sources. Firstly, the number of nesting blue tit pairs declined across the whole of Wytham Woods in 2015, reducing by around 40 % (Table 3.2). Secondly, it was also clear that the parent blue tits found the disturbance and/or altering of the box appearance a deterrent from nesting in these boxes. Evidence that the birds were disturbed by the presence of the equipment was the pecking of polystyrene mounts in many of the boxes. This demonstrates the birds were not deterred from entering the box but also appeared to want to remove the heating and cooling devices. Evidence that birds entered the experimental boxes was also found in the form of signs of roost (faeces and feathers) found in several of the experimental boxes. However this use of boxes did not translate into nest building. Smaller nest cavities have been linked to higher nestling mortality (Mertens 1977) and could therefore deter parent birds. Additionally the blue tit occupancy of the experimental nest boxes declined by a larger proportion than the non-experimental boxes (73 % and 48 % respectively, see Table 3.2). However, this cannot definitively be attributed to our

experimental treatment because declines in both blue tit and great tit numbers were seen throughout the whole of Wytham Woods from 2014 to 2015. The lack of increased blue tit occupancy in the surrounding non-experimental boxes, suggests that birds deterred from the experimental boxes did not relocate to nearby unaltered boxes. However, disturbance did not seem to influence marsh tits to the same degree, their numbers remained at a similar level of occupancy to historical records, suggesting they were less deterred by the experimental set up than blue tits. Further refinements of techniques to minimise the amount of equipment required to heat and cool boxes would improve these experimental techniques. Caution should be employed for future blue tit nest box manipulations.

3.5.2. The impact of low sample size on statistical analyses

The sample size achieved in this study was exceptionally low, this heavily shaped our ability to draw statistically supported conclusions from this experimental work. With a sample size of 13 across all three treatments (two heated, five control, and six cooled), very little could be ascertained statistically from the resulting dataset. As analyses were planned prior to running of the experiment, they were duly conducted. However, even if statistically significant results are found, they are unlikely to indicate anything other than chance. The inability to determine statistical results stems from fundamental issues with multiple testing and p-values.

The rise of inappropriate and misleading statistical analyses is a growing concern across many scientific disciplines, in particular psychology, medicine, ecology, and evolution (Mcshane et al. 2017; Horton 2015; Parker et al. 2016; Forstmeier et al. 2016). It is something we must consider when analysing the data generated in this study and interpreting the results. A focus and reliance on p-values and null hypothesis testing, passing of a 0.05 threshold indicating a rejection of the null

hypothesis, can be especially problematic for multiple comparisons on small datasets (Gelman & Loken 2014). When we accept a 5 % change of a Type 1 error (false positive result) over the course of multiple analyses, either on the same dataset or across studies, then 5 % of statistically significant results will actually have arisen if the null hypothesis were true (Forstmeier et al. 2016; Nuzzo 2014). Whether a result is deemed statistically significant or not is particularly changeable for small datasets where errors are high relative to the signal and the influence of individual data points is strong. Given this, there have been several calls to abandon p-values in statistics (Mcshane et al. 2017; Good 1988; Cumming 2014), although this call is not universal (Savalei & Dunn 2015).

Given the small sample size here, and the multiple comparisons conducted, the results should be analysed with caution. Multiple comparisons had to be conducted as there were multiple response variables to a focal explanatory variable (treatment), which were all chosen *a priori*. Our power to distinguish results was diminished with our sample size. As a result, we move away from discussion of statistical significance in this study, as it cannot be relied upon in this instance (Mcshane et al. 2017). Any statistically significant effects are highly likely to be Type 1 error. Instead we discuss any trends found, with the caveat they cannot be distinguished from chance and highlight areas that could be explored in greater detail in replicated studies.

3.5.3. Trends found

Trends shown in this analysis are that historical occupancy levels (popularity) also appear to predict occupancy during the experimental period. It seems logical that the historically most popular boxes are also those occupied most readily during the experimental period. Neither altitude nor local mean temperature appear to explain popularity, either historically or during this experiment. Further analyses of different

box positions would be required, manipulating specific aspects of the box environment, for example, aspect or oak density, to determine the drivers of occupancy/popularity of nesting sites. Although cooled boxes had the highest number of occupied boxes, this was indistinguishable from control or heated. The difference between treatments was a maximum of four boxes. Past studies have found a slight preference for cooler nest boxes (Dhondt & Eyckerman 1979), however, this could also be a chance difference, substantially more analyses would be required to draw definitive conclusions.

No influence was found of in-nest temperature on breeding phenology. There was a strong overlap between the cooled and control groups for all measures of phenology. Additionally, no influence was found of environmental variables on clutch initiation dates. As ambient temperature typically correlates strongly with clutch initiation dates across our long-term data, we would expect that in this case the sample size was too low to distinguish a signal or the influence of ambient temperature was dampened by the experimental manipulation. However, from the current data, we cannot discriminate between the two possibilities. We were unable to find statistical support for the relationships observed by Dhondt and Eyckerman (Dhondt & Eyckerman 1979) and O'Connor (O'Connor 1978), similarly to Nager and Van Noordwijk (Nager & van Noordwijk 1992) we found little effect of in-nest manipulations on clutch initiation date. Due to the difficulty of wild experimental analyses and the low sample size achieved here, this result does not necessarily indicate a statistical null result. We would strongly encourage further replicated analyses of this type with higher sample sizes and across multiple years, before a robust conclusion can be reached.

3.6. Acknowledgements

This study would not have been possible without the hard work of Tony Price and John Hogg, who assisted in the design and manufactured the heating and cooling devices for this experimental work. We would like to extend our gratitude to both Tony and John for all of the time they put in to this project. We are also grateful to all of the Wytham fieldworkers who collected population census data on the Wytham great tits. In particular, we would like to give a special mention to Keith McMahon, Stephen Lang, Koosje Lamers and Nico Alioravainen, who assisted with the changing of batteries and ice packs for this experiment come rain or shine. This work was supported by NERC grant NE/K006274/1 to Ben Sheldon.

3.7. Supporting information

3.7.1. Table of breakdown of nest building, abandonment and chick mortality by treatment

Table S3.8: Table of breakdown of nest building, abandonment and chick mortality by treatment

***** These figures include great tit and marsh tit nesting attempts. Species identification can only occur after 3 eggs have been laid so prior to this point species distinction is not possible. All other figures are blue tit only.

	Nest building	Abandonment pre-laying	Abandonment post-laying	Chick mortality
Cooled	7/27 (26%)	1/7 (14%)	0	2/61 (3%)
Control	5/27 (19%)	0	0	4/37 (11%)
Heated	2/27 (7%)	0	0	0
Non-experimental	43/50* (86%)	9/43* (21%)	0	17/113 (15%)

3.7.2. Supporting tables: Does in-nest temperature influence nest box choice?

Table S3.9: Full ANOVA table from analysis of occupancy as a function of historical popularity

Response = historical popularity					
	DF	Sum of squares	Mean squared	F value	P
Occupied	1	0.28	0.28	6.15	0.02
Residuals	82	3.78	0.05		

Table S3.10: Full ANOVA table from analysis of occupancy as a function of altitude

Response = altitude					
	DF	Sum of squares	Mean squared	F value	P
Occupied	1	125.9	125.92	0.98	0.33
Residuals	82	10533.5	128.46		

Table S3.11: Full table of Welch's T-Test of occupancy as a function of mean temperature

Response = mean temperature	
T	-1
DF	13.74
P	0.33

3.7.3. Supporting tables: Does in-nest temperature influence breeding phenology?

Table S3.12: Full ANOVA table of nest building date as a function of treatment

Response = Nest building date					
	DF	Sum of squares	Mean squared	F value	P
Treatment	2	71.63	35.81	1.19	0.34
Residuals	11	331.3	30.12		

Table S3.13: Full ANOVA table of lay date as a function of treatment

Response = Lay date					
	DF	Sum of squares	Mean squared	F value	P
Treatment	2	22.09	11.05	0.91	0.43
Residuals	11	120.83	12.08		

Table S3.14: Full ANOVA table of clutch completion date as a function of treatment

Response = Clutch completion date					
	DF	Sum of squares	Mean squared	F value	P
Treatment	2	9.097	4.55	0.91	0.43
Residuals	11	50.13	5.01		

Table S3.15: Table of outputs from phenology linear regression

	Estimate	SE	P
Intercept	8.81	44.76	0.85
Local mean temperature	2.52	4.67	0.60
Altitude	-0.15	0.08	0.07

CHAPTER FOUR

The role of phenological mismatch in the population dynamics of wild great tits

Emily G. Simmonds, Ella F. Cole, and Tim Coulson.

4.1. Abstract

Global climate change is altering seasonal timing and affecting biological systems, particularly the timing of life history events. These shifts in phenology, however, are not uniform across species, populations, or even ecosystems. This can lead to disruptions of interspecific relationships with implications for individual fitness. Some species appear to maintain synchrony through phenotypic plasticity, whereas others show little change. Mismatch causes some populations to undergo declines, whereas others appear buffered. Understanding the driving factors of phenological change and the implications of such changes for population dynamics are vital steps to begin to predict the future in a changing environment. In this study we present the first empirical illustration of a new integral projection model (IPM) framework incorporating quantitative genetic inheritance. We parameterise this model for a population of wild great tits, testing the predictive capabilities of different model parameterisations. We find that a fully parameterised model generated from exploratory statistical analyses of the drivers of fundamental demographic rates is needed to capture the combined trait and population dynamics. We use the final model to explore the role of ecological and evolutionary processes in determining the phenology of hatch date. The relative contribution of phenology and phenological synchrony to population dynamics is quantified by perturbation of environmental drivers. Our results showed that while phenotypic plasticity dominates inter-annual changes in phenology, micro-evolution plays a directional role in shifting phenology, visible across five decades. Spring temperature and changes to phenological synchrony were identified as the key drivers of alterations to population size. Synchrony became the largest driver beyond certain thresholds, suggesting that while some populations may remain buffered from negative impacts of mismatch within certain thresholds if these are surpassed rapid population declines could result.

4.2. Introduction

In recent decades, the planet has undergone significant changes in its climate, with change occurring more rapidly than has been experienced in at least the last 1300 years (Mann et al. 2008). One of the most prominent changes in climate is an average increase in global temperatures of 0.85 ± 0.2 ºC from 1880 to 2012 (IPCC 2013). However, seasonal patterns are also shifting, with springs occurring earlier, on average 3.1 days earlier per decade (Parmesan 2007; Parmesan & Yohe 2003) and autumns later (Gallinat et al. 2015). Such changes are already influencing biological populations (Field et al. 2014), and are projected to continue to do so into the next century. One of the most widely observed and well researched responses to the current climatic shifts, has been phenological change; an alteration of the timing of life history events. Phenological change has been observed across almost all taxonomic groups (Thackeray et al. 2010; Thackeray et al. 2016; Menzel et al. 2006; Cleland et al. 2007; Gallinat et al. 2015), and across a multitude of life history events from migration (Lehikoinen & Sparks 2010), reproduction (Both & Visser 2005; Both et al. 2004; Charmantier et al. 2008; Crick et al. 1997; Durant et al. 2007), and the emergence from hibernation (Ozgul et al. 2010), to the senescence of leaves (Gallinat et al. 2015). Shifts in phenology in temperate regions are typically characterised by an advance in spring but a delay in autumn (Gallinat et al. 2015), with spring phenology drawing the majority of research focus. This is a period of the year where there is rapid environmental change, temperatures increase by several degrees in just a few weeks and producing a short windows of optimum conditions, in which species can conclude energetically demanding life history events. Anthropogenic climate change is causing this period of temperature increase to occur earlier and as a result a large number of species are following suit, also advancing their spring phenology (Menzel et al. 2006; Parmesan & Yohe 2003; Cleland et al.

2007; Thackeray et al. 2010; Thackeray et al. 2016). However, the response to such environmental changes is not uniform (Thackeray et al. 2010; Menzel et al. 2006; Parmesan & Yohe 2003). Different species have different degrees of environmental sensitivity and plasticity (Thackeray et al. 2010; Cushing 1969; Durant et al. 2007), consequently rates of advance have been highly variable (Parmesan & Yohe 2003; Menzel et al. 2006; Both et al. 2009). These uneven patterns of response have been hypothesised to lead to mismatches between interacting species, which rely on temporal synchrony between interspecific life history events (Cushing 1969; Reed, Jenouvrier, et al. 2013; Durant et al. 2007; Singer & Parmesan 2010).

Asynchrony under novel climatic change can arise through a number of processes. Interacting species could have different sensitivities to the environment or capacity for plasticity, leading to their phenology altering to different degrees under the same environmental forcing. This has been shown in terrestrial systems, with lower trophic levels, which often have more direct developmental links to temperature, advancing at a faster rate than higher trophic levels (Thackeray et al. 2010; Walther et al. 2002). It could also be that the exact cue used to predict the optimal phenology in a given year is not shared between interacting species. If the climate then changes heterogeneously in time i.e. some parts of the year increase in temperature while others do not, as is currently the case (on Climate Change 2007; Menzel et al. 2006), then the cue used by one species may change more rapidly than the other, again leading to asynchrony. This has been proposed as a reason for lack of sufficient advance of great tit breeding (*Parus major*) in the Netherlands, despite an advance in their prey species (Visser et al. 2006). Disruption to tight interspecific relationships, such as, predator-prey, plant-pollinators, and parasite-host could result in significant fitness costs (or sometimes benefits) to one or both of the interacting species

(Cushing 1969; Durant et al. 2007; Reed, Jenouvrier, et al. 2013; Plard et al. 2014). However, temporal synchrony can also be maintained through shared sensitivity or homogeneous climatic change, which appears to be the case in great tits in the UK (Charmantier et al. 2008; Perrins 1965a; van Noordwijk et al. 1995).

Whether mismatch will occur under climate change depends on the cues used by interacting species, the capacity for phenotypic plasticity, and the ability to evolve if plasticity reaches a limit and the phenotype no longer changes in response to the abiotic cue. Phenotypic plasticity in phenology is the primary tool used to achieve matching in inter-annually variable environments (Charmantier et al. 2008; Charmantier & Gienapp 2014; Chevin & Lande 2015; Chevin et al. 2010). However, whether plasticity evolved for fluctuating environments will be sufficient under directional change is not known, evolution almost certainly will also be required to maintain synchrony. Several studies have suggested that the contribution of plasticity to phenology is high and heritability low, therefore limiting the capacity to evolve in response to selection pressure (Vedder et al. 2013; Charmantier & Gienapp 2014; Gienapp et al. 2014). Teasing apart the contribution of ecological and evolutionary processes in future responses is a vital step to predicting how phenological synchrony will change.

It has already been shown across a range of species, that asynchrony is occurring (Plard et al. 2014; Singer & Parmesan 2010; Both & Visser 2005; Mayor et al. 2017). However, the impacts of this asynchrony on population dynamics has been varied. Some species show population declines (roe deer, *Capreolus capreolus* - (Plard et al. 2014), migratory birds - (Møller et al. 2008)), whereas others appear to be buffered from the impacts of a failure to retain synchrony (Reed, Jenouvrier, et al.

2013; Reed, Grøtan, et al. 2013; Gienapp et al. 2014; Both 2010). Despite a wealth of research on phenology and its importance for individual fitness (Reed, Jenouvrier, et al. 2013; Plard et al. 2014), our understanding of the role of phenology and phenological synchrony in population dynamics is poorly developed (Miller-Rushing et al. 2010; Bennett et al. 2015; Johansson, Kristensen, et al. 2015). Gaining an in-depth understanding of the demographic consequences of asynchrony can be difficult to achieve. It requires long-term records of individual phenology, synchrony, survival, and reproduction. Data records including all of these components are a rarity. Few studies have analysed the contribution of temporal synchrony in tandem with other demographic processes. Plard et al. (Plard et al. 2014) did combine survival and recruitment rates with projections of trait changes through the construction of an integral projection model (IPM) and explored the population consequences of a failure to advance parturition dates in roe deer. They found a lack of phenological response to warming temperature in the deer, but a strong advance of vegetation they feed on, leading to population declines (Plard et al. 2014). However, such declines are not a universal picture across populations experiencing mismatch (Gienapp et al. 2005; Visser et al. 1998; Reed, Grøtan, et al. 2013; Johansson, Kristensen, et al. 2015; Singer & Parmesan 2010). Further analyses of the causes and consequences of phenological mismatch are needed.

To achieve an understanding of the importance of phenology in population dynamics we need a modelling framework that can pull apart the driving forces of phenology, predict how it will change, and what the impact of this will be for population dynamics. Modelling frameworks now exist, which allow us to access some of these questions. Such models have been introduced in Childs et al (Childs et al. 2016) and Coulson et al (Coulson et al. 2017), which extend the IPM to include quantitative genetics.

However, they have not yet been applied to ask specific questions about a population. These modelling frameworks are relatively easy to parameterise using data collected in field and statistical analyses (Ozgul et al. 2010) and offer the potential to address questions of trait and population dynamics. It has not yet been quantified how useful they are for predictions on actual systems. Predictive models by nature should be as simple as possible, however a balance needs to be struck between simplicity and the accuracy of predictions.

In this study we use a population modelling framework to address questions of the role of ecological and evolutionary processes in phenological change and the role of phenology and synchrony in population dynamics. We present the first empirical use of the extended IPM framework introduced by Coulson (Coulson et al. 2017). This framework allows us to tease apart the contributions of plasticity and evolution to phenology. We parameterise this model using long-term data from a model population of great tits in Wytham Woods. The timing of reproduction in this population plays an important role in reproductive success. Individuals hatching chicks around 13 days prior their prey species peak abundance (winter moth caterpillar, *Operophtera brumata*) fledge the most chicks (Simmonds et al. 2017). Phenology of both the great tits and the winter moth caterpillars are also advancing, about 15 days from 1985 to 2005 (Both 2010). The phenology and demography of this population has been recorded since 1960, providing an exemplary dataset to address three specific questions:

1. How much complexity is needed in a model to capture the trait (phenology) and population dynamics of a real biological population?
2. What is the contribution of micro-evolution versus phenotypic plasticity to the trait dynamics of this population?

3. What is the relative influence of phenological synchrony on population dynamics, compared to other environmental and demographic drivers?

4.3. Methodology

4.3.1. Model description

The model we use here is an extension of the standard Integral Projection Model (IPM) (Easterling et al. 2000). IPMs are constructed from four classes of fundamental demographic functions; survival (S), development (D), recruitment (R), and inheritance (H). These fundamental functions capture the survival of individuals from one year to the next, change in the focal trait for survivors, the number of new individuals that recruit to the population in the next year, and their inherited trait values. The basis of these models has typically been phenomenological, being generated from relationships ascertained from data. Coulson (Coulson et al. 2011) extended the IPM framework to include population genetics. It has been further extended in 2016 by Childs et al. (Childs et al. 2016) to incorporate quantitative genetics into the inheritance function, and by Coulson et al. (Coulson et al. 2017) to explicitly distinguish phenotypic plasticity from micro-evolution. Coulson et al.'s framework (Coulson et al. 2017), the one which we implement here, has several differences to the standard IPM. The primary development, based on Childs et al. (Childs et al. 2016) is the inclusion of a mechanistic inheritance function based on quantitative genetic assumptions. A standard IPM framework tracks a distribution of traits through time, in our model we track a bivariate distribution of breeding values (G) and environmental (E) components with variances defined from the observed additive genetic and residual environmental variance, respectively. At time t = 0 these variances sum to give the phenotypic variance. The two components of the distribution can be combined to create a distribution of phenotypes (Z), we assume Z = G + E. Consequently Z is always a function of G and E, Z(G, E). The E component

of the phenotype changes at each time step based on the environmental conditions at that time (θ), making the trait labile. However, an underlying assumption of this modelling approach is that the breeding value within an individual remains fixed for life, ignoring mutation. The distribution of G can be altered by the survival and recruitment functions that operate on the entire phenotype, Z. As we implement a single sex, clonal model here, with breeding values being passed from mother to daughter without error, both inheritance and development cannot alter breeding values. A single sex model was chosen as males are typically assumed to have little influence on breeding timing in great tits, however, restricting to one sex results in any sex specific differences in offspring recruitment being ignored. If the sex ratio of recruits varies between years based on environmental conditions (as discussed in (Oddie 2000)) then this could lead to over or underestimations of population size in our model. Based on previous analyses (Oddie 2000), which did not find significant sex differences in recruitment in great tits, this should not have a large effect in this population. The development and inheritance functions for E have the effects of redistributing E within each value of G in the bivariate distribution. Through this method of tracking the bivariate distribution of G and E, it is possible to isolate effects of ecological and evolutionary processes involved in quantitative trait and population dynamic change.

The equation for the standard IPM and the extended approach are given below, (Equation 1 and 2 respectively) where Z is the trait at t, Z' is the trait at t+1, N(Z, t) is the distribution of the trait at time t, N(Z', t+1) is the distribution of the trait time t+1. S(Z, θ, t), R(Z, θ, t), D(Z'|Z, θ, t) and H(Z'|Z, θ, t) are the survival, recruitment, development, and inheritance conditional on the phenotype and environment at time t (θ). G and E in Equation 2 are the breeding value and environmental component of

the phenotype at time t, respectively. E' represents the environmental component of the phenotype at time t+1. The survival and recruitment functions, are conditional on both the environment at time t and t-1 (θ') due to lagged effects of spring conditions just prior to the census.

Equation 1:

$$N(Z', t+1) = \int [\, S(Z, \theta, t) D(Z' | Z, \theta, t) + R(Z, \theta, t) H(Z' | Z, \theta, t)] \, N(Z, t) dZ$$

Equation 2:

$$N(G, E', t+1) = = \iint [S(Z(G, E), \theta', t) D(E' | E, G, \theta, t) + R(Z(G, E), \theta', t) H(E' | E, G, \theta, t)] \, N(G, E, t) dG dE$$

4.3.1.1. Functional forms

Survival captures the probability of survival of a particular Z value surviving from time t to t+1. The survival function is assumed to be logistic, following Equation 3. S(Z(G, E), θ', t) is the probability of survival to t+1, given the sum of G and E, and the environmental conditions at t and lagged effects from t-1 (θ'). V(X) is a linear predictor of the form Equation 4, where $ß_0$ is the intercept of a linear regression taking into account effects of any explanatory factors, $ß_1$ is the slope of the relationship between the phenotypic trait of interest and survival, X_1 is the Z value at t, $ß_j$ is the slope of the relationship between explanatory variable j and survival, and X_j is the value of explanatory variable j at this time step.

Equation 3:

$$S(Z(G, E), \theta', t) = \frac{1}{1 + e^{-(V(X)}}$$

Equation 4:

$$V(X) = \beta_0 + \beta_1 X_1 + \beta_j X_j$$

Equation 5:

$$R(Z(G,E), \theta', t) = e^{-(V(X)}$$

Recruitment captures the number of offspring produced for each Z value, which survive to breed in the population at t+1. This can occur in the year following birth as great tits can reproduce from one year of age. The recruitment function takes an exponential form, Equation 5. R(Z(G, E), θ', t) is the number of offspring surviving to enter the population as breeders in t+1, given the Z value of the mother at t, and the environmental conditions at t and lagged effects from t-1. V(X) is a linear predictor of the same form as Equation 4, but where $ß_0$ represents the intercept and $ß_j$ the slopes of the relationship between recruitment and explanatory variable j.

Development captures the change in E from time t to t+1 and inheritance captures the change in E value from the mother's E value at time t to the offspring E value at time t+1. As G values are assigned for life, these functions redistribute E within G values.

Both the development and inheritance functions can be approximated as Gaussian probability functions (Easterling et al. 2000) of the expected values of E at t+1 (E') given E at t. These functions contain two components. Firstly, the expected value of E' given E and θ at time t, this can be approximated using a linear regression following the same linear predictor form as Equation 4. Secondly, the standard deviation around these linear relationships is calculated. These two elements can then be combined into Equation 6 to create the Gaussian probability distribution.

Where V(X) generates the expected value of E' and D(E', θ, t) is the standard deviation around E'. This equation assumes that the distribution of possible E' values at t+1 is normally distributed for any value of E at t. Inheritance follows the same form but with H(E'|E, θ, t) and H(E', θ, t).

Equation 6:

$$D(E'|E,\theta,t) = \frac{1}{D(E',\theta,t)\sqrt{2\pi}} e^{-\frac{(E-V(X)^2}{2V(X)^2}}$$

The timing of the peak abundance of the winter moth caterpillar, half fall date, is also predicted at each time step using a linear predictor (Equation 4). This function is parameterised from a linear regression of half fall date against environmental explanatory variables of spring temperature, spring precipitation, winter temperature, winter precipitation, and beech mast index. Beech mast index is an index of the amount of mast produced by beech trees, observed in winter. Predicting caterpillar timing at each time step allows us to calculate temporal synchrony between great tit hatch timing and caterpillar peak abundance.

4.3.2. Parameterisation

4.3.2.1. Data and the study system

The study system and data collection

Long-term data (from 1960 to 2015) were collected through population censuses of the nest box breeding population of great tits in Wytham Woods (~ 6000 individuals over this period). Censuses have been conducted using a standardised procedure since 1960 (Perrins 1965a; Perrins & McCleery 1989) and provide data on breeding phenology, clutch size, reproductive success, survival, and a detailed social pedigree. In addition to the spring breeding census, mist netting is also conducted throughout the winter within Wytham Woods. Any new birds caught over winter are

fitted with uniquely identifiable metal leg rings and RFID tags and age, sex, and morphometric data are recorded.

The timing of winter moth caterpillar peak abundance is taken to be the median date on which final instar winter moth larvae descend to the ground to pupate. This data have been collected as part of a long-term study in Wytham Woods, supplied by Dr L. Cole (see (Charmantier et al. 2008)).

Beech mast index has been recorded in Wytham Woods annually since 1947 (Chamberlain et al. 2007). The amount of mast produced by beech trees in winter varies annually (Matthews 1955). In Wytham Woods it is recorded on a three point scale from low beech mast years (0) to high beech mast years (2). Data for this study were provided by Dr Andy Gosler.

Precipitation data were provided by the Radcliffe Meteorological Station, Oxford (Radcliffe Meterological Station n.d.). Daily rainfall totals in mm were collected from 1960 to present.

Temperature data were provided by the MET Office, National Climate Information Centre gridded daily data from grid point 447500E 202500N (MET Office 2009). Daily mean temperatures are available from 1960 to 2015. Data are generated from observations from weather stations across the UK and interpolated for missing values, both temporally and spatially.

Data for parameterisation

Cleaned breeding census data (details of data cleaning can be found in section 4.7.1, supporting information) were combined with long term winter ringing records and environmental data to create a working dataset of the format below, with a single breeding attempt being each row. Variables listed below are a complete list of those used in statistical model selection. Not all of these variables are used in the final parameterisation of the IPM.

Response variables:

- *Survival* – whether an individual survived to the next breeding season (1) or not (0).
- *Number recruited* – number of chicks from that breeding attempt, that were seen again as breeders at any later date within Wytham Woods.
- *Hatch date (t+1)* – day on which the first chick hatched in the nest, in days since the 1st April for the year after the focal year. Only available for individuals surviving and breeding beyond the focal year.

Individual explanatory variables:

- *ID* – (factor) uniquely identifiable ring number of breeding female.
- *Nest box* – (factor) identity of the nest box used for breeding attempt.
- *Hatch date (Z)* – (numeric) day on which the first chick hatched in the nest, in days since the 1st April.
- *Synchrony (SN)* – (numeric) hatch date minus caterpillar half fall date.
- *Clutch size (CS)* – (numeric) maximum number of chicks or eggs observed in the nest.
- *Age (A)* – (factor) age of female (mother). Bird ages are from breeding season to breeding season i.e. a bird born in 2000 breeding season would be one in 2001 breeding season and two by the autumn of 2001. Exact age was known

for all birds ringed as nestlings in Wytham Woods. For those born in natural cavities or immigrant birds we did not know exact ages. If an unknown bird was first caught as a juvenile (i.e. in the first year of its life) then exact age can be easily determined by plumage. However, some individuals are first caught and ringed as adults. As great tits rarely disperse after their first year (less than 2 % of adults move between years) and are unlikely to have missed multiple years of breeding, we assume that individuals first caught as adults are aged two on their first observation. However, a few cases exist where there is a discrepancy between breeding data and winter ringing data. When no entry exists for a female until it is an adult, but after it has been recorded breeding in Wytham for more than one year, it is assumed that the female is a first year at the time of first breeding attempt recorded. Exact age is needed in order to calculate birth year, however for the parameterisation of functions a binary coding was used, 1 = first year bird, 2 = older than first year. Across our 55 years of data, only 500 individuals lived over three years (an average number of less than ten per year), therefore we did not distinguish ages beyond three years. From previous analyses it has been shown there is little difference in survival and reproductive success between two and three year old birds in our study system (Bouwhuis et al. 2009; Bouwhuis et al. 2010; Bouwhuis et al. 2012), consequently these two groups were combined to be a single age class (> one year) .

- *Immigrant (I)* – (factor) female was born in a Wytham Woods nest box (0) or not (1).
- *Section* – (factor) section or 'round' of the woodland, determined by management history, habitat, and nest box density.

- *Birth year hatch* – (numeric) mean hatch date of the year in which the female was born.
- *Birth year* – (factor) year in which female was born.
- *Mother ID* – (factor) BTO ring number of the mother of the breeding female.

Environmental explanatory variables:

- *Half fall (HF)* – (numeric) median date on which final instar winter moth larvae descend to ground to pupate (days since 1st April).
- *Year* – (factor) year in which nesting attempt occurred.
- *Population size (N)* – (numeric) number of breeding females, who hatched eggs, in the woodland that year.
- *Spring temperature (ST)* – (numeric) mean temperature for the period 1st March to 9th May. This window of temperature was identified as the optimum critical window in which great tits perceive temperature using an absolute sliding time window method in the Climwin package in R (van de Pol et al. 2016; Bailey & De Pol 2016).
- *Winter temperature (WT)** – (numeric) mean of temperature from the winter following the breeding attempt from (1st December to 28th/29th February).
- *Spring precipitation (SP)* – (numeric) total precipitation from 1st April to 31st May to cover the nesting period and when young chicks are in the nest.
- *Winter precipitation (WP)** – (numeric) total precipitation from 1st December to 28th/29th February.
- *Beech mast index (B)** – (factor) index of winter beech mast production 0 = no mast, 1 = little, 2 = high mast.

* these variables end in the calendar year following the current breeding season e.g. winter temperature would end in February 2001 for the 2000 breeding season.

All numeric variables were scaled across years (mean subtracted and divided by the standard deviation) to ensure comparable effect sizes, including hatch date.

Precipitation, temperature, and beech mast index were chosen as the key environmental variables for this analysis because they have been previously linked to population dynamics, and phenology in passerine birds (Sandvig et al. 2017; Raven et al. 2005) and specifically in tit species (*Paridae*) (Payevsky 2006; Perdeck et al. 2000; Grøtan et al. 2009; Van Balen 1980). While large scale climate indices, such as the North Atlantic Oscillation (Naef-Daenzer et al. 2012) could also play a role in population dynamics, here we are focusing on the influence of local weather manifestations rather than large scale indices which drive them. Our focus is on the environment individuals experience, not the driving force of the environment.

Data from 1961 to 2010 were used to parameterise the functions for the model. 1960 was removed due to a vast amount of missing data on parent IDs, birth years, and previous breeding attempts. Data from 2011 to 2014 were kept aside as a testing dataset for cross validation. Census years run from 1st Junes to 31st May the following calendar year, so from breeding season to breeding season.

4.3.2.2. Statistical analyses and characterisation of fundamental functions

In order to construct an IPM it is first necessary to capture the drivers of four fundamental demographic functions; survival, development, recruitment, and inheritance. We assess the relationships between these demographic rates and explanatory variables of interest through various forms of regression analysis. Model selection is performed for each function. We begin with a full model and use stepwise

reduction until a minimal model is obtained, which best explains variance in the demographic rate.

Each demographic function includes the effects of hatch date (the trait value of interest in this study), synchrony with the caterpillar peak abundance, spring temperature (a known driver of breeding phenology (Husby et al. 2010; Perrins 1965c)), and population size (to ensure density dependence). These variables were included in every function, regardless of significance because they are of biological interest.

Survival

Recapture rate in the study population was 81%, consequently needed to be accounted for when estimating survival. As a result the survival analysis was conducted using the 'marked' package in R (Laake et al. 2013) using a Cormack-Jolly-Seber (CJS) model (Jolly 1965; Seber 1965). This assumes a closed population. While immigration and emigration do occur in the study population, a closed population is still approximated as birds immigrating remain in the population until death and those emigrating do not return. Consequently, emigration does not need to be distinguished from mortality.

The full survival model (CJS model) included a time varying recapture rate and:

- time varying covariates (the same for all individuals); year, spring temperature (t-1), spring precipitation (t-1), winter temperature, winter precipitation, and beech mast index.

- individual covariates (do not vary with time); immigrant status and section of woodland.
- time varying individual covariates (vary with individual and with time); hatch date, hatch date^2, synchrony, synchrony^2, clutch size, and clutch size^2.

Model selection was performed using stepwise model reduction with an Information Theoretic Approach, using the AIC (Akaike Information Criterion) as is commonly employed in the phenology field (Hinks, Cole, Daniels, Wilkin, et al. 2015; Bailey & De Pol 2016). The AIC was used to find the model that best explained the data available. We systematically removed one variable at a time and assessed the impact on the AIC. Variables to be removed were chosen based on their standard errors, those with estimates less than two times the standard error were candidates for removal. Whether a variable was retained in the model or removed permanently was determined by the ΔAIC. Variables were only retained if removal generated a ΔAIC of greater than 2 in comparison to the model containing the variable.

Development

We used a generalized mixed effect model (GLMM) to capture the development of hatch date from time t to t+1. This was run using the 'lme4' package (Bates et al. 2015) in R, with hatch date at time t+1 as response variable.

The full model had a random effect of year and explanatory fixed effects of; hatch date (at t), synchrony (t), spring temperature, spring precipitation, winter temperature (t), winter precipitation (t), beech (t), immigrant status, age, section of woodland, and population size (t+1). Clutch size was not included in this analysis because it is not independent of hatch date and directionality is assumed in the opposite direction.

Due to the need to include hatch dates from t and t+1 for this analysis, we could only include individuals which bred in multiple years (N = 2122).

Model selection was performed in the same way as for survival, systematically removing one variable at a time, based on standard errors. We only retained variables if removal increased the AIC by more than 20.

Recruitment

The recruitment analysis was conducted using a Poisson GLMM with number of recruits as response variable (count data). This was run using the 'lme4' package (Bates et al. 2015) in R. The model was run for 200,000 iterations, with BOBYQA optimisation (the default in the lme4 package). Model selection was performed using stepwise reduction and the ΔAIC, a ΔAIC threshold of 20 was employed for retaining explanatory fixed effects.

The full model had random effects of year and nest box, to take account of multiple observations per year and per nest box and fixed effects of; hatch date, hatch date^2, synchrony, synchrony^2, spring temperature (t-1), spring precipitation (t-1), winter temperature, winter precipitation, clutch size, clutch size^2, beech mast index, immigrant status, age, birth year hatch date, section of the woodland, and population size.

Inheritance

Inheritance was captured mechanistically using a quantitative genetic approach through the animal model (Wilson et al. 2010). This includes the breeding value as a random explanatory variable within the model, allowing estimation of the variance of

breeding values in the population (additive genetic variance). This model is implemented with a Bayesian model optimisation in the 'MCMCglmm' package in R (Hadfield 2010). Models were run for 150000 iterations, burn in 1000 and sampled every 100 iterations to reduce auto-correlation and produce a good posterior distribution. Social pedigree is used to infer relatedness, therefore only individuals that appear in the pedigree could be used here (N = 5855). For females social pedigree should represent accurate genetic relationships, however there are rare instances of females laying eggs in nests other than their own as illustrated by observed mixed species broods (found to be 3 % in blue tits and great tits (Barrientos et al. 2015)). For males on the other hand, a social pedigree will have a higher number of incorrectly assigned relationships. Even in socially monogamous bird species extra pair paternity occurs in on average 11.1 % of offspring (Griffith et al. 2002). For this analysis, with hatch date as the response variable, we first tested different random effect structures; a null model of just individual breeding value, then all combinations of breeding value, maternal effect, birth year of female, and permanent environment effect. Due to the Bayesian nature of the model optimisation, the Deviance Information Criterion (DIC) was used for model selection of the random effects, following Wilson (Wilson et al. 2010). A reduction in the DIC of greater than 20 was taken to indicate a significant improvement in the model fit.

Once the random effect structure had been determined, model selection on the fixed effects was conducted. The full model of fixed effects had explanatory variables of; spring temperature, spring precipitation, winter temperature, immigrant status, age, section of woodland, and population size. Synchrony and clutch size were not included as they are not independent of hatch date. Following the protocol in Wilson (Wilson et al. 2010) model selection was performed by systematically removing the

variable with the largest relative standard error (relative to estimate), until all remaining variables had estimates greater than two times the standard error.

4.3.3. Model testing

4.3.3.1. Cross validation of model parameterisations

The IPM parameterised using a training dataset from 1961 to 2010 was simulated for a test dataset of five years (year beginning and ending 20th May) using observed environmental data from 2011 to 2014. The initial bivariate distribution for this model run, was defined using the observed mean trait value and population size for 2010. The variance of the G and E components of the distribution were determined by the additive genetic and the residual environmental variance calculated from the animal model, respectively. At each simulated time step values of explanatory environmental drivers and caterpillar half fall date were taken from observed data.

Three different model parameterisations and configurations were trialled. Two null parameterisations and one full parameterisation were run, of the form:

- Null 1 (density dependent), only population size and hatch date as explanatory variables, but with spring temperature in the development function.
- Null 2 (synchrony model), synchrony, spring temperature, population size, and hatch date as explanatory variables in all functions.
- Fully parameterised model, using the explanatory variables determined from statistical model selection on each function.

Population size and mean hatch date were calculated at each time step from each model. Simulated outputs were compared to observed values of population size and mean hatch dates from this period, using mean squared error (MSE). MSE is the mean of the squared differences between model estimates and observed values.

4.3.4. Determining the contribution of micro-evolution and phenotypic plasticity to phenological change

The model was simulated for 50 years following the same process as cross validation and using the fully parameterised fundamental functions. The additive genetic variance was then increased ten times and the model simulated again. This was achieved through alteration of the breeding value variance (G) without altering the environmental variance (E) in order to isolate the influence of additive genetic variance. Simulation outputs from the increased, and observed breeding value variance were then compared to assess how sensitive the model is to changes in the evolutionary potential.

For these simulations values of the explanatory drivers were randomly selected, making this a stochastic model run. Values of continuous explanatory variables were selected from a random normal distribution with a mean and standard deviation defined from observed values of that driver or held at their mean (clutch size). Factors were either held at a mean value (section of the woodland), or selected randomly (beech mast index). Incidence of years where beech trees mast are not distributed completely randomly in time. Two full mast years cannot occur consecutively and not more than four years without mast occur successively (Matthews 1955). Therefore, restrictions were placed on this variable. Beech index is chosen from a distribution of no mast (0), little mast (1), and full mast (2) at frequency of the observed data, however, two full beech mast years cannot be selected sequentially and if four years occurred with no full beech mast, then a full mast year is selected. The value of caterpillar half fall, and consequently synchrony, were calculated at each time step from the environmental drivers. The change in G, E, and consequently in the hatch date (Z(G, E)) were recorded at each time step. Therefore

the contribution of G and E to the resulting phenotype can be isolated. As this model is stochastic, pulling values for environmental variables from random distributions, each time it is simulated the exact results differ slightly due to different environmental values being chosen at each time step. Running the model multiple times and then averaging across results can give a more robust demonstration of the simulated results. However, this technique also dampens inter-annual environmental fluctuations across multiple stochastic runs. For the assessment of the contributions of micro-evolution and phenotypic plasticity inter-annual fluctuations in the environment are vitally important. Therefore, we use just a single stochastic run to tease apart the contributions of these components. It should consequently be noted that subsequent runs of the model would produce different results, however the broad trends that we analyse will remain the same.

4.3.5. Perturbation analysis of the drivers of population dynamics

In order to determine the key drivers of population dynamics, we assessed the sensitivity of population size to different explanatory drivers. Values of continuous environmental drivers, population size, and synchrony (through manipulation of caterpillar half fall) were altered by ± 0.05. This perturbation was performed on the scaled variables and therefore represented the same magnitude of change for each driver and did not exceed the observed range for any driver. The focal variable was manipulated in 0.1 increments to explore the sensitivity of population dynamics to perturbations in each driver. Model simulations were run for 50 years for each perturbation, holding all other drivers at their mean. Beech mast index was held at full mast to avoid extinction generated from too frequent masting years. This allowed attribution of the change in population growth rate to the variable of interest. The population size after 50 years was calculated from each perturbed model run.

4.4. Results

4.4.1. Statistical analysis of fundamental functions

Results of the minimal model, determined by model selection for each function are presented below. For full model selection tables please see section 4.7.2.1, supporting information.

4.4.1.1. Survival

The minimal survival model determined from model selection included a time varying recapture rate and:

- time varying covariates (the same for all individuals); spring temperature, spring precipitation, winter temperature, beech mast index, and population size.
- individual covariates (do not vary with time); immigrant status and section of woodland.
- time varying individual covariates(vary with individual and with time); hatch date, synchrony, and synchrony^2.

Table 4.1: Parameter value estimates for survival (Phi component only)
SE missing for average section of the woodland as this estimate is an average of all section estimates, full parameter values can be found in section 4.7.2.3, supporting information

Parameter	Estimate	SE
Intercept	-0.47	0.09
Hatch date	-0.22	0.05
Synchrony	0.03	0.04
Synchrony^2	0.04	0.02
Spring temperature	-0.10	0.05
Spring precipitation	0.10	0.03
Winter temperature	0.12	0.03
Beech mast index = 0	-0.06	0.08
Beech mast index = 1	0.38	0.09
Resident	0.25	0.07
Average section of woodland	-0.21	
Population size	-0.09	0.04

4.4.1.2. Development

The minimal model for development included a random effect of year and fixed effects of hatch date, synchrony, synchrony^2, spring temperature, age, section of the woodland, and population size.

Table 4.2: Parameter value estimates for development
SE missing for average section of the woodland as this estimate is an average of all section estimates, full parameter values can be found in section 4.7.3.3, supporting information

Parameter	Estimate	SE
Intercept	-0.11	0.07
Hatch date	0.13	0.07
Synchrony	0.21	0.06
Synchrony^2	-0.04	0.01
Spring temperature	-0.73	0.03
Age = 2	0.11	0.02
Average section of the woodland	0.012	
Population size	0.10	0.04

4.4.1.3. Recruitment

The minimal recruitment model included random effects of year and nest box and fixed effects of; hatch date, synchrony, synchrony^2, spring temperature, clutch size, clutch size^2, beech mast index, section of the woodland, and population size.

Table 4.3: Parameter value estimates for recruitment
SE missing for average section of the woodland as this estimate is an average of all section estimates, full parameter values can be found in section 4.7.4.3, supporting information

Parameter	Estimate	SE
Intercept	-1.35	0.09
Hatch date	-0.16	0.06
Synchrony	-0.17	0.05
Synchrony^2	-0.12	0.02
Spring precipitation	0.13	0.05
Clutch size	0.13	0.02
Clutch size^2	-0.04	0.01
Beech = 1	0.07	0.14
Beech = 2	0.50	0.12
Average section of the woodland	0.03	
Population size	-0.17	0.07

4.4.1.4. Inheritance

The minimal inheritance model included fixed effects of spring temperature, spring precipitation, winter temperature, winter precipitation, section of the woodland, and population size.

Table 4.4: Parameter value estimates for inheritance
Table showing mean estimate from posterior distribution and the upper and lower confidence intervals (CI) for each explanatory variable. SE missing for average section of the woodland as this estimate is an average of all section estimates, full parameter values can be found in section 4.7.5.3, supporting information

	Post mean estimate	Lower CI	Upper CI
Parameter	Estimate	Std. Error	Std. Error
Intercept	0.10	0.06	0.14
Spring temperature	-0.64	-0.66	-0.62
Spring precipitation	0.15	0.13	0.16
Winter temperature	0.02	0.00	0.04
Winter precipitation	-0.07	-0.08	-0.05
Average section of the woodland	-0.05		
Population size	-0.06	-0.08	-0.04

For model simulations the effect of section of the woodland is held at its mean, only the dynamics of resident birds are simulated, and all environmental variables are assigned values at each time step. Age structure was included in the model because development rates of the trait value differed for first year and adult birds. The population distribution vector was split into an adult and a juvenile population, while they were parsed through the same survival, recruitment, and inheritance functions, the development function differed for each population. These two vectors were combined at the end of each time step to calculate the population size and trait values across the whole population. Population size is calculated from the previous time step. Synchrony is calculated based on the half fall timing, which is generated at each time step based on the environmental conditions.

4.4.2. Cross validation of model parameterisations

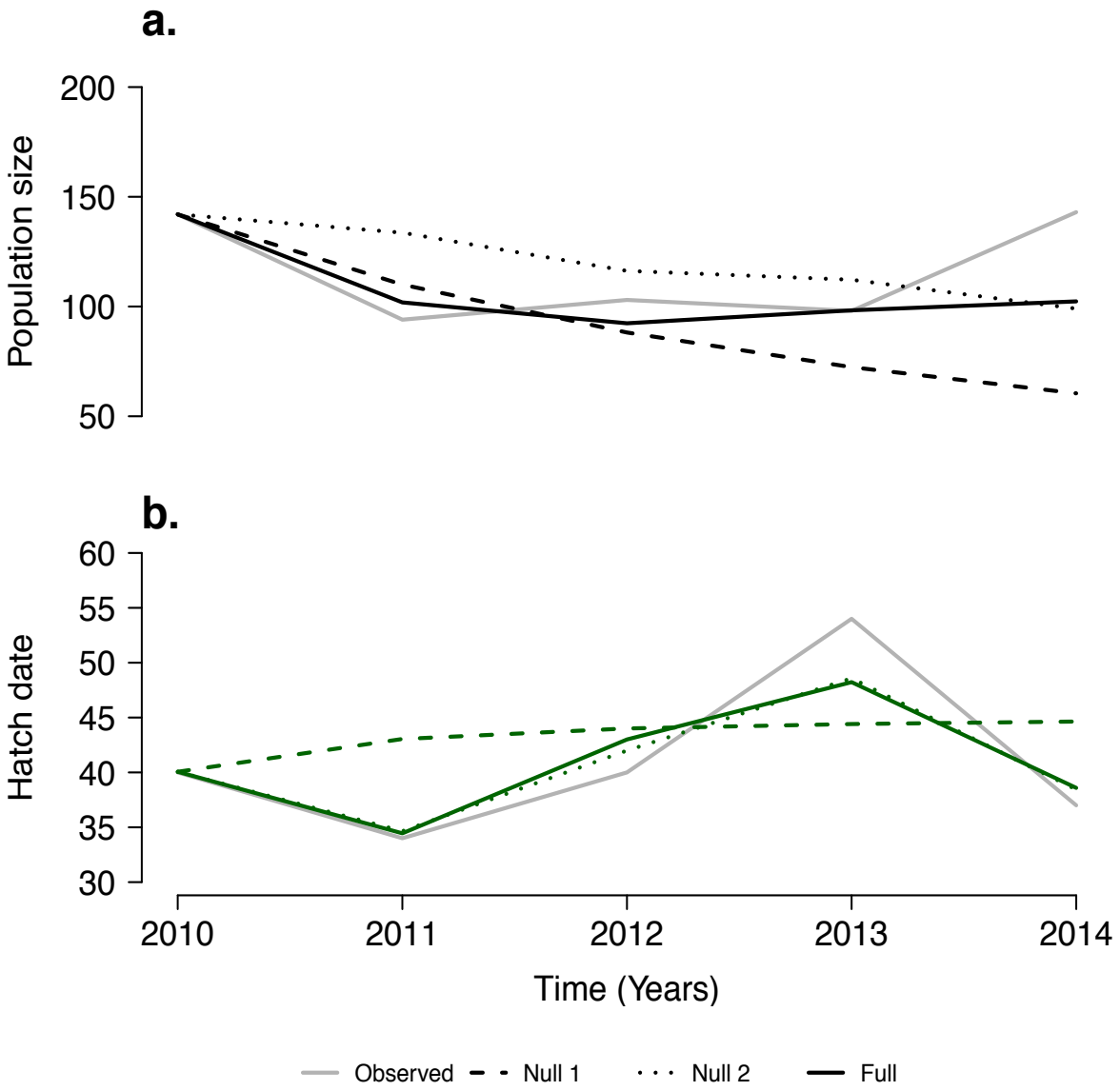

Figure 4.1: Cross validation of projected population size and mean trait values.
a) Projected resident population size and b) projected mean resident trait value against observed values from 2010 to 2014. Model simulations were run for a density dependent (Null 1), density and synchrony (Null 2), and fully parameterised model (Full).

All models trialled capture the observed resident population dynamics of the 2010 to 2014 period to varying degrees but with all models performing poorly for 2014. Null model one, density dependence only, declines at an almost constant rate across the five simulated years. While this parameterisation captures the population decline from 2010 to 2011, it fails to capture the increases or more gradual declines from 2011 to 2014.

Null model two, including synchrony and density dependence, projects a higher population size than observed from 2011 to 2013. This model parameterisation shows population dynamics following an opposite pattern to observed. The model projects declines in population size in 2012 and 2014 while observed population size increases in these years.

The fully parameterised model captures the observed population dynamics more closely than the two null parameterisations, remaining within 11 individuals of the observed values from 2010 to 2013. In 2014 both the observed population size and model projections showed an increase in individuals, however the observed population increased to a greater extent that model projections. The better fit of the projections from the fully parameterised model to the observed data is demonstrated in Table 4.5, both null parameterisations have higher mean squared errors (MSE) than the full model (approximately twice and five times as high for null parameterisations two and one, respectively).

Table 4.5: Discrepancy between simulated results and observed values. Mean squared error (MSE) for population size and trait values (phenology) for both null models and the fully parameterised model simulation.

Model	MSE	
	Population size	Phenology
Null 1 (density dependent)	1589.66	53.43
Null 2 (synchrony)	776.47	47.82
Full	365.74	48.06

For trait dynamics there is also a varying picture of model parameterisation performance. Null parameterisation one does not capture the observed phenology, projected mean hatch dates increase (delaying to later in the year) steadily from 2011 to 2014, while observed values fluctuate over this period. Null model two and

the fully parameterised model both capture the trait dynamics well across the five year test dataset. Both capture the mean hatch date exactly for 2011 and follow the same pattern from 2012 to 2014. Model projections overestimate the mean trait value in 2012 and 2014 and underestimate in 2013. The fit of each parameterisation to the observed test dataset is shown in Table 4.5, null model one has MSE of more than four times the other model parameterisations. Null model two has the lowest MSE, with the fully parameterised model MSE only 0.24 greater.

Across both population and trait dynamics, the fully parameterised model performs the best, and null model one the least well.

4.4.3. Determining the contribution of micro-evolution and phenotypic plasticity to phenological change

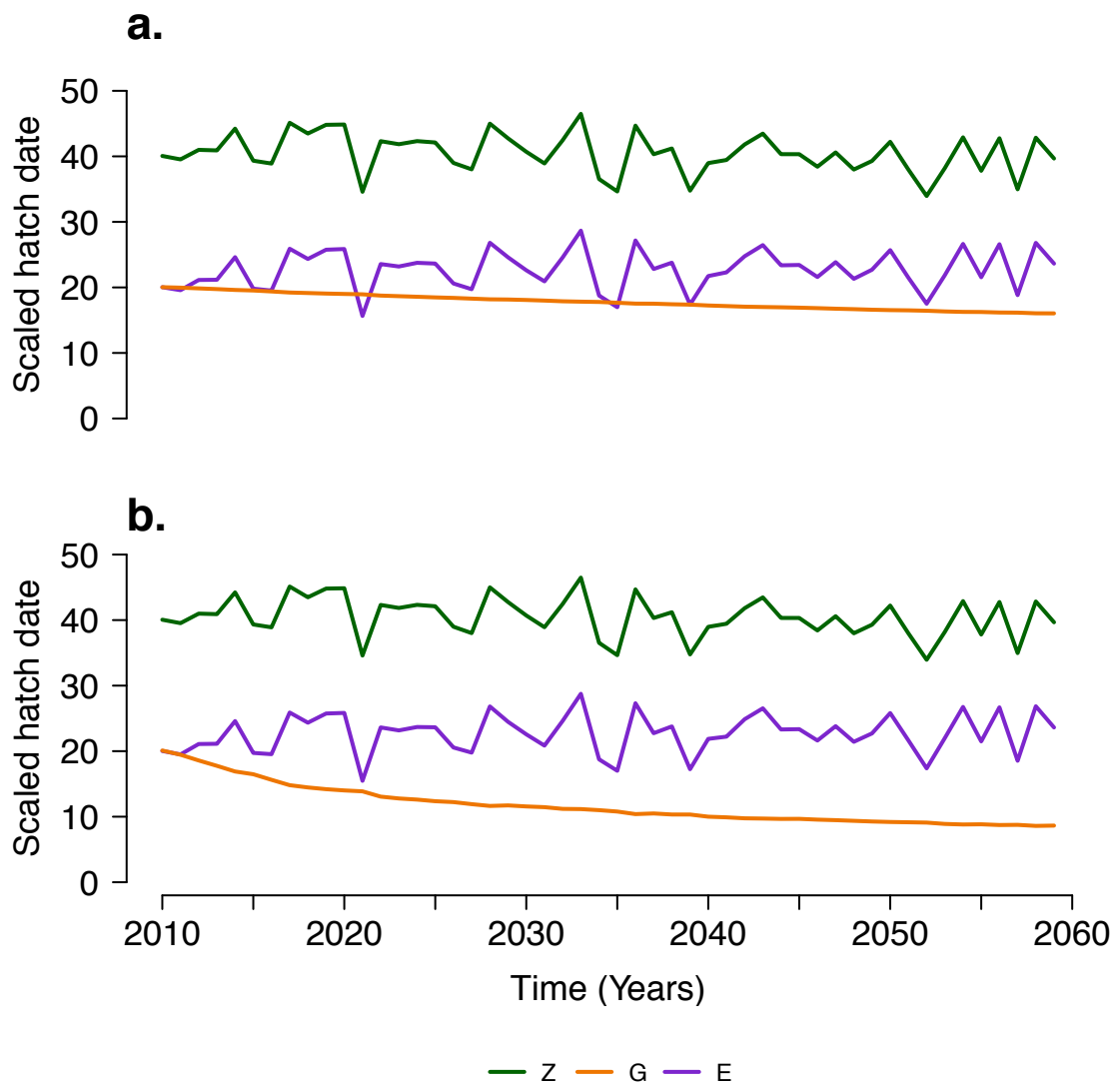

Figure 4.2: Simulated mean values of G, E, and Z for a single stochastic run at different levels of additive genetic variance
a) with observed G variance b) with G variance ten times greater than observed

Under the observed additive genetic variance, over a 50 year time period, only a small amount of genetic change is shown. The mean value of the G distribution advanced by four calendar days over the course of the stochastic model simulation. In contrast the mean E component of the phenotype showed considerable interannual fluctuations, see Figure 4.2a and b. These interannual variations were an order of magnitude greater than the change in G component, with a range of 13 calendar days.

When the additive genetic variance is increased ten fold, while maintaining environmental variance, a more substantial directional change in the genetic component can be seen, Figure 4.2b. With increased evolutionary potential directional change over the 50 year simulation reaches 11.5 calendar days advance through genetic change, comparable to the influence of phenotypic plasticity.

4.4.4. Perturbation analysis of the drivers of population dynamics

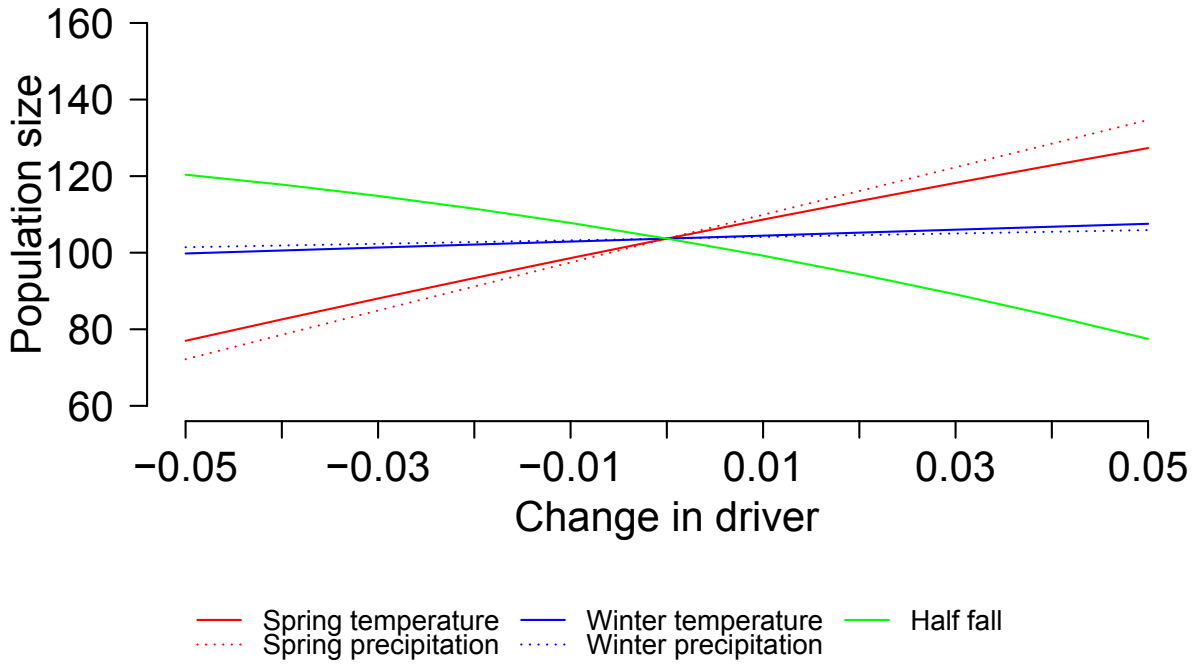

Figure 4.3: Population size after 50 years for changes in explanatory drivers

The population size after 50 years of perturbed simulation plotted against the change in explanatory driver, with each plotted line representing a different driver

Population size was least sensitive to changes in winter conditions. The population size after 50 years of simulation was approximately 100 individuals for all perturbations of both winter temperature and winter precipitation. Winter temperature shows a weak positive relationship with population size, with higher winter temperatures producing slightly higher population sizes. However the relationship between winter precipitation and population size is flat.

The drivers with the greatest influence on population size are spring temperature, spring precipitation and caterpillar half fall (and consequently synchrony). Spring temperature and spring precipitation show strong positive relationships with population size. Higher values of spring conditions result in higher population sizes, ranging from a 20 female decrease for a 0.05 reduction in either spring temperature or spring precipitation to the opposite pattern for increases in temperature and rainfall. Half fall/synchrony shows a slightly curved relationship with population size. Positive manipulations of half fall equate to delays in half fall relative to hatching dates, with negative perturbations indicating earlier caterpillar timing. All increases in half fall, result in a rapid decline in the population size. Declines in half fall, however result in an increase in population size at the same magnitude as spring condition effects. The sensitivity of population size to perturbations, as represented by the slope of the lines in Figure 4.3, is greater for increases in caterpillar timing.

4.5. Discussion

4.5.1. The final model

The final model parameterisation, determined by statistical analyses and model selection on the fundamental functions, included the influence of both winter and spring environmental conditions. Winter conditions, including food supply (beech mast index) played a significant role in both survival and recruitment as well as inheritance of hatch date. The role of winter conditions confirms previous work suggesting a key role of winter conditions on overwinter survival for both adults and first year birds (Van Balen 1980). Development of hatch date, however, only had temperature in spring as a significant environmental driver demonstrating that breeding phenology is primarily driven by spring conditions, as we would expect for this system (Husby et al. 2010; Perrins 1965a). Section of the woodland was a highly important factor for all fundamental functions, suggesting demographic rates vary

significantly across different areas of the woodland. This could be a result of philopatry and territorial nature of the study species, with less than 2 % of females moving between sections of the woodland during their mature life. Synchrony between great tit hatch dates and winter moth caterpillar peak abundance, and its quadratic, were also significant across survival, recruitment, and development. This demonstrates the importance of temporal matching for demographic rates and as we expected, indicates that mismatches occur in either direction (too early and too late) as shown by the significance of a quadratic effect. Age was only found to be important in the development of the phenological trait, suggesting little age based variation in survival or reproductive success. This could indicate a role of experience in breeding timing of great tits. Older (beyond first year) females are known to breed earlier than first year females (Perrins 1970b). However, these could be indirectly influenced through different plasticity of first years and older birds. Clutch size only influenced reproductive success and with a quadratic relationship indicating both small and very large clutch sizes produced fewer recruits.

4.5.2. Model parameterised by model selection captures the population and trait dynamics better than null parameterisation

The fully parameterised model captured the combined trait and population dynamics of the test dataset better than either null model. No model fully captured the observed population dynamics, however, the mean squared error (MSE) of the full model was considerably lower than either of the null models, Table 4.5. Model projections were particularly poor for 2014, despite close tracking from 2011 to 2013 for the full model. The reason for the universally poor fit in 2014 could be because the transition from 2013 to 2014 was unusual for this population. The spring of 2013 was the second coldest spring on record with a spring temperature in the critical window identified for hatch date of 5.97 ºC. The only spring with a lower temperature was 1962 with a

temperature of 5.87 ºC. In contrast 2014 was one of the warmest springs on record (9.13 ºC), only five other years since 1960 have exceeded 9 ºC. This jump from a cold to a warm spring is the second largest difference between consecutive spring temperatures over the past 55 years, and is the largest increase in temperature across our whole long-term record (3.17 ºC). The rarity of this type of event in our training dataset could have contributed to the poor model fit to the test dataset year. If such years were to become more frequent into the future, our model could fail to predict the full extent of the dynamics that would be generated. However, the general patterns observed do seem to be captured, even in 2014. The model predictions tend in the same direction as the observed population size, even if not to the same magnitude. Some caution should still be exercised when interpreting results from highly variable years, the rate of which could increase through increased variability under climate change (IPCC 2013).

For projections of trait dynamics from 2011 to 2014 both null model two and the full model performed well, capturing the pattern of observed phenological change for the whole of the test dataset. However, null model one failed to capture any of the trait dynamics from the test dataset. The projections generated from null model two, including synchrony and spring temperature but no other environmental drivers, and the full model are almost indistinguishably similar, both visually and in terms of MSE (Figure 4.1 and Table 4.5). This suggests that it is spring temperature and previous synchrony which predominantly determine phenology, with winter conditions playing only a minor or even negligible role. This is supported by previous work (Charmantier et al. 2008; van Noordwijk et al. 1995; Lack 1958a; Visser et al. 1998; Husby et al. 2010; Perrins 1965a; Kluijver 1950) and by the results of model selection on the

development function (Section 4.4.1.2), which is one of the key drivers of interannual variation in the trait values expressed.

Despite the superior fit of null model two to the observed trait dynamics, and a reasonable fit of null model one to some of the observed population dynamics neither of these models managed to capture both the trait and population dynamics. This suggests that winter conditions, which are not present in null model two, play a significant role in the population dynamics of the study population. Winter conditions (temperature, precipitation, and beech mast index), in addition to habitat and individual differences (section of the woodland, age, and clutch size) are needed in order to capture the combined trait and population dynamics. In order to create meaningful predictions of the future of real biological populations detailed exploratory statistical analyses of the demographic rates for these populations is required.

4.5.3. Phenotypic plasticity is the primary driver of interannual population dynamics

Across a 50 year simulation, changes in the environmental component of the phenotype drive the majority of the resulting hatch date fluctuations (Figure 4.2a). This demonstrates that when looking at inter-annual fluctuations it is this environmentally driven component of the phenotype, phenotypic plasticity, which is the predominant factor determining the hatch date expressed (Vedder et al. 2013; Charmantier et al. 2008). This is partly caused by only two functions influencing G but four functions influencing E, resulting in more rapid change in E relative to G. Phenotypic plasticity has been proposed in several previous studies to be the key driver of adaptive phenotypic changes, typically lay date, in great tits (Charmantier et al. 2008; Teplitsky et al. 2009; Hoffmann & Sgrò 2011; Anderson et al. 2012). Our findings suggest that the same is largely true for hatch dates however, we also find

an influence of genetic change, evolution. Even across a 50 year simulation with observed additive genetic variance we see an advance in hatch dates driven by genetic change. This change is emphasised when the additive genetic variance is increased ten fold (Figure 4.2b). While there is selection for earlier hatch dates in great tits (Brinkhof et al. 1993; Verhulst et al. 1995; Reed et al. 2016; Charmantier et al. 2008), heritability of phenological traits is moderate to low (Vedder et al. 2013; Van Der Jeugd & McCleery 2002; Sheldon et al. 2003) resulting in an assumed slow evolutionary change (Gienapp et al. 2006). In this study the observed additive genetic variance produced a change equating to an advance in hatch date of four calendar days in 50 years. This suggests a micro-evolutionary response to warmer springs that could be seen on the timescale of several decades, although this is across many tens of generations. Consequently, micro-evolutionary change of the scale shown here could still play an important role in maintaining synchrony with peak caterpillar abundance under directional climate change, i.e. progressively warmer springs (Gienapp et al. 2008; Ghalambor et al. 2007; Visser 2008). However, it is likely that such an effect is often masked by phenotypic plasticity which alters hatch dates by a maximum of 13 calendar days, an order of magnitude higher change than micro-evolution. Nevertheless, smaller alterations to the mean phenotype via micro-evolutionary change could play an important role in shifting the limits of plasticity and retaining synchrony under changing conditions.

4.5.4. Population size is most sensitive to changes in phenological synchrony and spring temperature

Each of the continuous environmental drivers included in our model showed a different effect on population size (Figure 4.3). Population size was least sensitive to changes in winter conditions, changes in both precipitation and temperature had little influence on the population size after 50 years. This suggests that the influence of

each winter conditions in isolation does not translate into alterations of population size. Winter conditions have been shown on multiple occasions to influence adult survival in great tits (Perrins 1965a; Van Balen 1980). We would therefore expect more favourable winters, warmer winters, to result in higher survival and consequently increase population size (Grotan et al. 2009; Van Balen 1980). However, winter temperature in our model also showed a positive relationship with inheritance, resulting in later hatch dates for new recruits in warmer winters. Later hatch dates correspond, based on mean environmental conditions, to reduced reproductive success (Charmantier et al. 2008; Verhulst et al. 1995), consequently counteracting the increased survival. Winter precipitation was only significant in the inheritance function, therefore its influence on population dynamics is likely limited. These opposing effects likely combine to explain why winter conditions do not appear to have strong influences on population dynamics in isolation, combined changes could have a greater influence.

The environmental drivers with the greatest impact on population size were spring temperature, spring precipitation, and synchrony with caterpillar peak abundance, manipulated through alterations to half fall timing. Despite a negative relationship between spring temperature and survival, spring temperature shows a positive relationship with population size after 50 years. This could be driven by spring temperature's influence on development and inheritance, therefore having an indirect effect on survival and reproduction through phenotype expression. Although warmer springs result in lower survival, they also advance hatch dates. Earlier hatch dates tend to have higher reproductive success and higher survival, as do earlier lay dates (Charmantier et al. 2008; Verhulst et al. 1995). As a result, a complex pattern of influence of spring temperature emerges. It does not only have its direct effect on

survival but also has indirect effects on both survival and reproductive success via both absolute hatch dates and synchrony with the caterpillar peak abundance. These combined effects result in a positive relationship between spring temperature and population size.

Spring precipitation has a positive relationship with both survival and recruitment, therefore unsurprisingly has a positive influence on population size. Higher levels of precipitation during spring lead to larger population sizes after 50 years.

Synchrony with caterpillar peak abundance showed a curved with population size. All delays to half fall relative to great tit hatch dates (increases) caused sharp population declines of more than 20 individuals for a + 0.05 change in half fall. In contrast decreases/advances in half fall closer to great tit hatch dates initially caused population increases but to a lesser extent. Population increases are driven by a tightening of synchrony as half fall advances. As the synchrony has a quadratic relationship to both survival and reproduction a continued advance in half fall, beyond the small perturbation applied here, would gradually cause a reduction in population size as caterpillars advance to be earlier than great tits. This curved relationship between changes in synchrony and population size is of the same magnitude as the relationships between spring conditions and population size. This suggests that synchrony does have an important effect on population dynamics and to the same level as key environmental drivers. It should be noted that the perturbations explored here have all been run in isolation. In reality all of the environmental variables will change at the same time and often in differing directions. The combined impact of changes in multiple environmental variables could act to magnify or dampen the impacts of the others, potentially leading to even greater or reduced population

impacts. All of these signal that alterations to phenological synchrony have the potential to play a non-negligible role in future population dynamics.

The patterns of relationship between population size and perturbations to both spring temperature and synchrony shown here, could explain why population declines do not always accompany phenological mismatch resulting from rising spring temperatures (Reed, Grøtan, et al. 2013; Reed, Jenouvrier, et al. 2013). Despite our statistical models showing significant relationships between individual fitness (survival and recruitment) and phenological synchrony (Table 4.1 and Table 4.3), supporting previous work on this species (Visser et al. 2006; Reed, Jenouvrier, et al. 2013; Reed, Grøtan, et al. 2013), this did not translate into all perturbations of synchrony driving population declines (Figure 4.3). This is partly driven by the shape of the relationship between synchrony and population size but could also be driven by interactions between the temperature change and other demographic processes, such as earlier hatch dates improving fitness.

4.5.5. Conclusions

In this study we have explored the causes and potential consequences of changes in phenology and phenological synchrony in a population of wild great tits. We applied a new modelling framework to this population, demonstrating how such frameworks can be applied to real biological populations and combined with detailed statistical analyses to produce predictions of population and trait dynamics. A fully parameterised model, based on statistical analyses to identify the key drivers in each demographic function was needed in order to capture both the population and trait dynamics of the population. Using this fully parameterised model we showed that both phenotypic plasticity and micro-evolution are likely to play a role in the response of this population to climatic change. Phenotypic plasticity plays key role in

determining hatch date inter-annually and maintaining synchrony between years. Micro-evolution plays more of a long-term role, altering phenotypes in a consistent direction across several decades and potentially shifting the limits of plasticity (Gienapp et al. 2008). Both have the potential to act well within the next century and drive significant phenological change.

This has also been one of the first studies to explore the potential role of changes in phenological synchrony in driving population dynamics, in conjunction with other key environmental drivers. It appears that both spring temperature and phenological synchrony are key drivers of changes in population dynamics and that the relationship between these drivers and the population size are complex. The lack of response to mismatch seen in many populations (Reed, Jenouvrier, et al. 2013; Reed, Grøtan, et al. 2013; Gienapp et al. 2014; Both 2010) could result from thresholds have not yet been passed or influences of environmental change on other components of demography that compensate for any fitness reduction caused by mismatch (Johansson, Nilsson, et al. 2015). In order to know how these different processes will play out into the future and how populations will respond, directed model simulations are required to test realistic scenarios of future climate.

4.6. Acknowledgements

We are grateful to all of the Wytham fieldworkers who collected population census data on the Wytham great tits. This work was supported by NERC grant NE/K006274/1 to Ben Sheldon.

4.7. Supporting information

4.7.1. Data cleaning

Long term breeding census data were cleaned prior to use. Breeding attempts were restricted to first attempts. All known second attempts in any single box and all nests with a lay date more than 30 days after the first egg date laid in the woodland (Van Der Jeugd & McCleery 2002) were removed. Several checks were performed to reduce data entry error; we ensured all hatch dates occurred after lay dates, all fledge dates after hatch dates, all clutch sizes were larger than or equal to chick numbers, and all chick numbers greater than or equal to the number fledging. Any breeding attempts not meeting these criteria were removed as it is not possible to know which of the conflicting numbers is the correct entry. All duplicate entries were also removed. Six birds appear in historical ringing data as both great tits and other species. This is likely a result of data entry error, however, as it is not possible to definitively know which entry is correct, the individuals were removed.

4.7.2. Parameterisation of the survival function

4.7.2.1. Model selection - survival

Table S4.1: AIC and ΔAIC values for survival model selection

AIC	ΔAIC *in comparison to model on the preceding row	Variable removed
14223.19	NA	Null model one
14203.72	-19.47	Null model two
14119.92	-83.8	Full model
14118.58	-1.34	removed time varying survival
14130.13	-11.55	removed winter precipitation
14129.28	-0.85	removed clutch size
14128.90	-0.38	removed quadratic hatch

4.7.2.2. Null model parameters

Table S4.2: Full model output for null survival model one (density dependent)

		Estimate	Standard Error	Lower confidence interval	Upper Confidence interval
Phi	Intercept	-0.28	0.02	-0.33	-0.24
	Population size	-0.06	0.02	-0.10	-0.01
p	Intercept	3.35	4.34	-5.15	11.85
	1962	-0.69	4.33	-9.18	7.80
	1963	-1.68	4.36	-10.22	6.86
	1964	-2.57	4.35	-11.10	5.95
	1965	-3.02	4.35	-11.54	5.50
	1966	-2.44	4.35	-10.98	6.09
	1967	-2.72	4.35	-11.26	5.81
	1968	-3.16	4.35	-11.69	5.36
	1969	-3.24	4.36	-11.78	5.29
	1970	-3.34	4.35	-11.87	5.19
	1971	-2.43	4.36	-10.98	6.13
	1972	-2.24	4.36	-10.79	6.31
	1973	-2.80	4.35	-11.32	5.73
	1974	-1.93	4.36	-10.48	6.61
	1975	-3.28	4.35	-11.81	5.24
	1976	-2.66	4.35	-11.19	5.86
	1977	-2.00	4.36	-10.54	6.54
	1978	-2.71	4.35	-11.24	5.81
	1979	-2.42	4.35	-10.95	6.11
	1980	-2.11	4.35	-10.64	6.42
	1981	-2.04	4.35	-10.57	6.48
	1982	-1.51	4.35	-10.04	7.02
	1983	-2.87	4.35	-11.38	5.65
	1984	-2.25	4.35	-10.79	6.28
	1985	-3.00	4.35	-11.52	5.52
	1986	-0.79	4.37	-9.35	7.76
	1987	-1.60	4.36	-10.15	6.95
	1988	-2.21	4.35	-10.73	6.31
	1989	-2.74	4.34	-11.26	5.77
	1990	-3.29	4.35	-11.81	5.22
	1991	-2.34	4.36	-10.88	6.20
	1992	-1.27	4.36	-9.82	7.27
	1993	-1.65	4.35	-10.17	6.87
	1994	-2.29	4.34	-10.80	6.23
	1995	-1.63	4.35	-10.16	6.89
	1996	-2.13	4.35	-10.65	6.40
	1997	-2.21	4.35	-10.73	6.31
	1998	-2.40	4.34	-10.92	6.11
	1999	-1.24	4.36	-9.78	7.31
	2000	-2.71	4.34	-11.22	5.80

	2001	-1.55	4.35	-10.07	6.97
	2002	-2.99	4.34	-11.50	5.51
	2003	-2.44	4.34	-10.96	6.07
	2004	-1.58	4.35	-10.11	6.96
	2005	-1.96	4.35	-10.48	6.56
	2006	-1.77	4.35	-10.29	6.74
	2007	-1.74	4.34	-10.25	6.78
	2008	-1.36	4.36	-9.89	7.18
	2009	-1.26	4.39	-9.87	7.35

Table S4.3: Full model output for null survival model two (density dependent, spring temperature, and synchrony)

		Estimate	Standard Error	Lower confidence interval	Upper Confidence interval
Phi	Intercept	-0.29	0.03	-0.35	-0.23
	Hatch date	-0.14	0.04	-0.22	-0.07
	Synchrony	-0.02	0.03	-0.09	0.05
	Synchrony^2	0.02	0.02	-0.02	0.06
	Population size	-0.11	0.03	-0.17	-0.05
p	Intercept	2.10	1.45	-0.74	4.95
	1962	0.52	1.56	-2.54	3.58
	1963	-0.48	1.52	-3.46	2.50
	1964	-1.25	1.49	-4.17	1.67
	1965	-1.75	1.48	-4.65	1.16
	1966	-1.05	1.51	-4.01	1.91
	1967	-1.50	1.50	-4.45	1.45
	1968	-1.94	1.49	-4.86	0.98
	1969	-1.92	1.52	-4.91	1.07
	1970	-2.05	1.50	-5.00	0.89
	1971	-0.95	1.56	-4.01	2.11
	1972	-1.02	1.52	-3.99	1.95
	1973	-1.53	1.49	-4.45	1.39
	1974	-0.68	1.53	-3.67	2.31
	1975	-2.02	1.49	-4.95	0.91
	1976	-1.45	1.49	-4.37	1.47
	1977	-0.68	1.51	-3.65	2.29
	1978	-1.32	1.52	-4.29	1.65
	1979	-1.04	1.50	-3.98	1.90
	1980	-0.91	1.49	-3.84	2.02
	1981	-0.72	1.51	-3.67	2.23
	1982	-0.23	1.49	-3.16	2.69
	1983	-1.58	1.52	-4.55	1.40
	1984	-0.57	1.57	-3.64	2.50
	1985	-1.79	1.48	-4.69	1.11
	1986	0.38	1.53	-2.63	3.39
	1987	-0.08	1.58	-3.17	3.01
	1988	-0.82	1.49	-3.74	2.10
	1989	-1.51	1.47	-4.40	1.37

	1990	-2.24	1.46	-5.10	0.62
	1991	-1.28	1.49	-4.19	1.64
	1992	-0.07	1.50	-3.02	2.88
	1993	-0.51	1.49	-3.42	2.41
	1994	-0.98	1.48	-3.88	1.93
	1995	-0.46	1.48	-3.36	2.44
	1996	-1.15	1.47	-4.02	1.73
	1997	-0.94	1.51	-3.90	2.02
	1998	-1.14	1.48	-4.04	1.76
	1999	-0.09	1.51	-3.05	2.88
	2000	-1.50	1.46	-4.36	1.37
	2001	-0.16	1.49	-3.08	2.76
	2002	-1.87	1.47	-4.75	1.00
	2003	-1.20	1.47	-4.08	1.68
	2004	-0.51	1.49	-3.42	2.41
	2005	-0.70	1.49	-3.62	2.22
	2006	-0.46	1.49	-3.37	2.46
	2007	-0.55	1.47	-3.43	2.32
	2008	-0.03	1.51	-2.99	2.93
	2009	-0.55	1.49	-3.48	2.38

4.7.2.3. The final survival model

Table S4.4: Full model output for the final survival model

		Estimate	Standard Error	Lower confidence interval	Upper Confidence interval
Phi	Intercept	-0.47	0.09	-0.64	-0.30
	Hatch date	-0.22	0.05	-0.31	-0.13
	Synchrony	0.03	0.04	-0.05	0.11
	Sychrony^2	0.04	0.02	-0.01	0.08
	Beech = 0	-0.06	0.08	-0.21	0.09
	Beech = 1	0.38	0.09	0.21	0.55
	Spring temperature	-0.10	0.05	-0.19	0.00
	Spring precipitation	0.10	0.03	0.03	0.17
	Winter temperature	0.12	0.03	0.06	0.18
	Immigrant	0.25	0.07	0.12	0.39
	Section of woodland C	-0.02	0.09	-0.19	0.15
	Section of woodland CP	-0.36	0.10	-0.56	-0.16
	Section of woodland EX	-0.24	0.07	-0.38	-0.09
	Section of woodland MP	-0.09	0.09	-0.26	0.09
	Section of woodland O	-0.27	0.08	-0.43	-0.11
	Section of woodland P	-0.53	0.18	-0.88	-0.18
	Section of woodland SW	-0.19	0.09	-0.36	-0.02
	Section of woodland W	0.04	0.08	-0.12	0.19
	Population size	-0.09	0.04	-0.16	-0.01
p	Intercept	5.04	25.20	-44.34	54.43

1962	-2.36	25.16	-51.67	46.94
1963	-3.25	25.20	-52.64	46.13
1964	-4.08	25.20	-53.47	45.30
1965	-4.68	25.19	-54.07	44.70
1966	-4.07	25.19	-53.43	45.30
1967	-4.27	25.18	-53.63	45.09
1968	-4.97	25.20	-54.36	44.42
1969	-4.77	25.21	-54.17	44.64
1970	-5.25	25.21	-54.67	44.17
1971	-3.97	25.22	-53.41	45.46
1972	-3.86	25.18	-53.22	45.50
1973	-4.46	25.19	-53.84	44.92
1974	-3.96	25.21	-53.37	45.45
1975	-4.95	25.20	-54.34	44.44
1976	-4.48	25.20	-53.88	44.91
1977	-3.51	25.20	-52.90	45.88
1978	-3.74	25.20	-53.14	45.66
1979	-3.95	25.20	-53.35	45.45
1980	-4.08	25.23	-53.52	45.36
1981	-3.29	25.19	-52.67	46.09
1982	-3.29	25.20	-52.68	46.10
1983	-4.69	25.22	-54.13	44.74
1984	-3.74	25.26	-53.24	45.77
1985	-4.65	25.20	-54.05	44.74
1986	-2.52	25.20	-51.92	46.88
1987	-3.42	25.23	-52.88	46.04
1988	-3.75	25.20	-53.14	45.64
1989	-4.57	25.19	-53.94	44.80
1990	-4.67	25.17	-54.01	44.67
1991	-4.15	25.19	-53.53	45.23
1992	-2.95	25.18	-52.31	46.41
1993	-3.45	25.19	-52.82	45.92
1994	-3.92	25.20	-53.30	45.47
1995	-3.50	25.20	-52.89	45.89
1996	-4.24	25.22	-53.68	45.19
1997	-3.59	25.21	-53.01	45.82
1998	-4.06	25.20	-53.46	45.34
1999	-2.72	25.17	-52.06	46.62
2000	-4.72	25.19	-54.09	44.65
2001	-3.24	25.21	-52.65	46.16
2002	-4.76	25.20	-54.15	44.64
2003	-3.89	25.19	-53.27	45.49
2004	-3.94	25.19	-53.32	45.44
2005	-3.35	25.19	-52.72	46.03
2006	-3.49	25.21	-52.90	45.92
2007	-3.54	25.19	-52.92	45.84
2008	-2.98	25.20	-52.37	46.40
2009	13.88	3416.07	-6681.62	6709.39

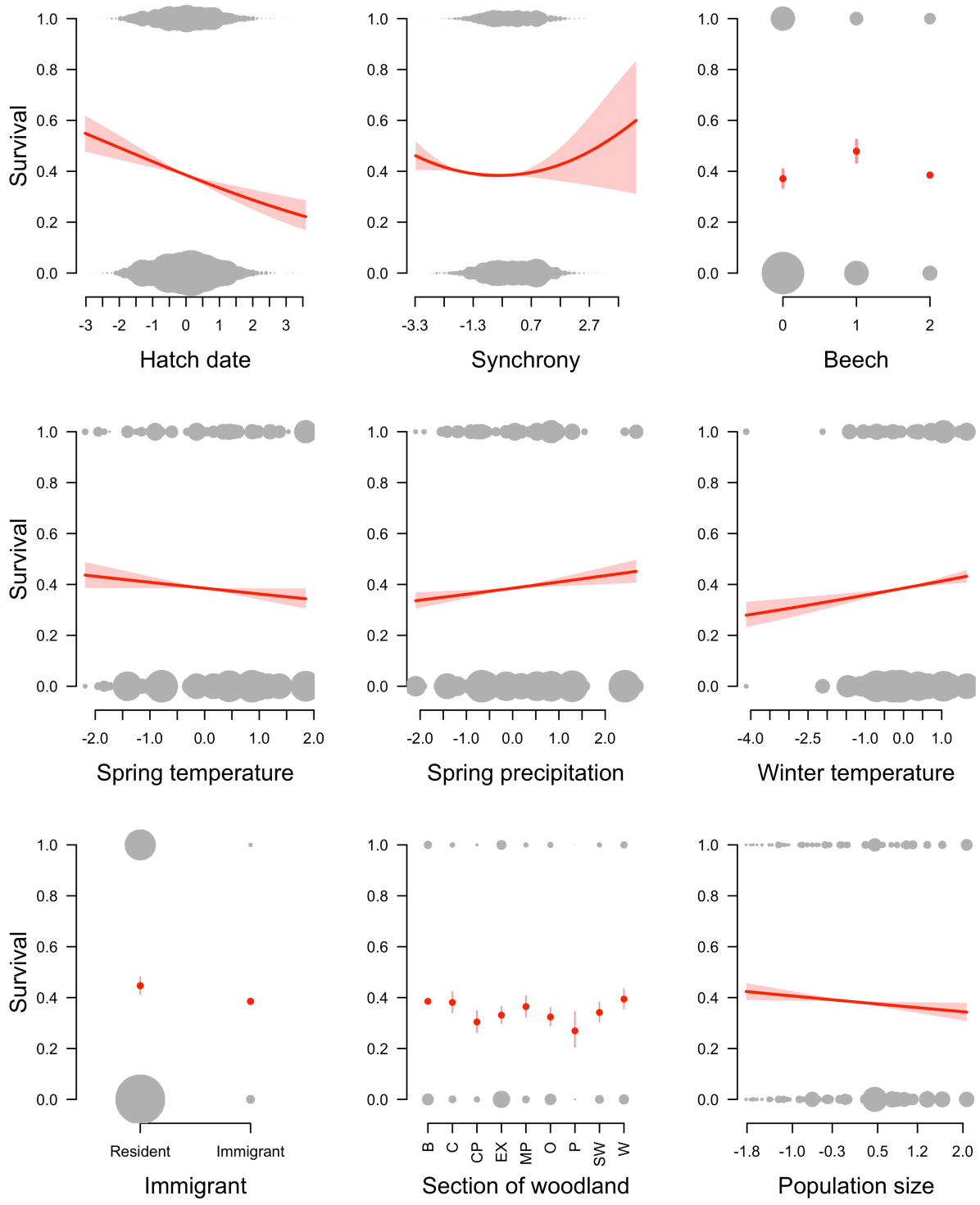

Figure S4.1: Relationship between probability of survival and significant explanatory variables identified by model selection
Plotted point size is scaled by the number of data points that have that value, larger indicating a greater number of points. Shaded areas around lines and lines around point estimates represent the 95% confidence interval of each relationship.

4.7.3. Parameterisation of the development function

4.7.3.1. Model selection - development

Table S4.5: AIC and ΔAIC values for development model selection

AIC	ΔAIC *in comparison to model on the preceding row	Variable removed
6390.62	NA	Null model one
5848.69	-541.93	Null model two
5808.92	-39.77	Full model
5805.09	-3.82	removed beech mast index
5799.65	-5.44	removed winter precipitation
5794.30	-5.36	removed winter temperature
5789.43	-4.87	removed spring precipitation
5783.33	-6.10	removed immigrant

4.7.3.2. Null model parameterisations

Table S4.6: Full model output for null development model one (density dependent)

	Estimate	Standard error	T value
Intercept	-0.02	0.10	-0.18
Hatch date	0.40	0.02	22.39
Population size (t+1)	0.00	0.05	0.00

Table S4.7: Full model output for null development model two (density dependent, spring temperature, and synchrony)

	Estimate	Standard error	T value
Intercept	-0.06	0.07	-0.85
Hatch date	0.14	0.07	1.93
Synchrony	0.22	0.06	3.78
Synchrony^2	-0.04	0.01	-4.43
Spring temperature	-0.73	0.03	-24.67
Population size (t+1)	0.10	0.04	2.38

4.7.3.3. The final development model

Table S4.8: Full model output for the final development model

	Estimate	Standard error	T value
Intercept	-0.11	0.07	-1.51
Hatch date	0.13	0.07	1.83
Synchrony	0.21	0.06	3.73
Synchrony^2	-0.04	0.01	-4.20
Spring temperature	-0.73	0.03	-25.03
Section of woodland C	-0.15	0.04	-3.85
Section of woodland CP	0.03	0.05	0.65
Section of woodland EX	0.01	0.03	0.27
Section of woodland MP	-0.20	0.04	-4.97
Section of woodland O	0.08	0.04	2.04
Section of woodland P	0.18	0.09	2.02
Section of woodland SW	0.15	0.04	3.86
Section of woodland W	0.01	0.04	0.28
Age = 2	0.11	0.02	5.66
Population size (t+1)	0.10	0.04	2.46

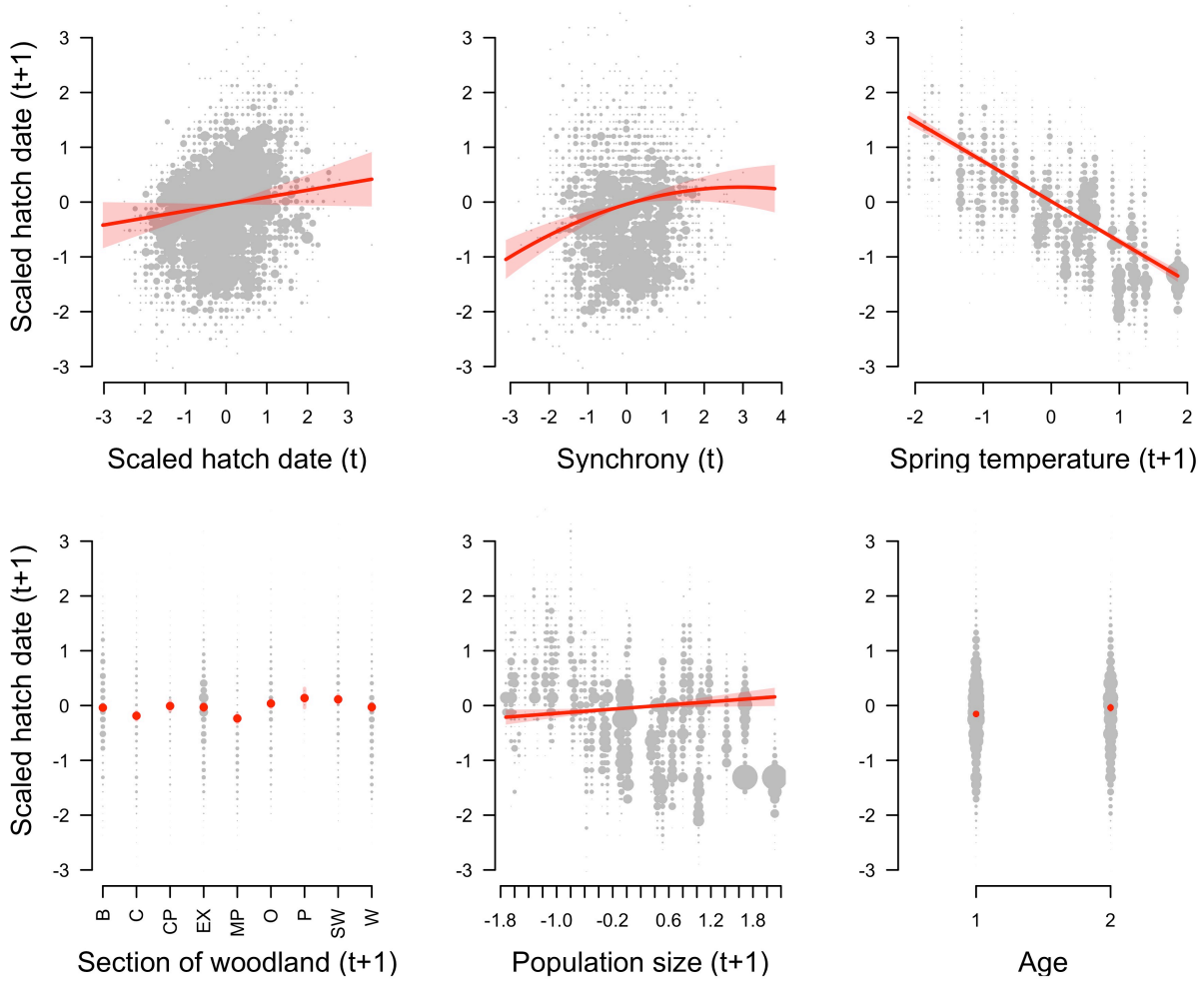

Figure S4.2: Relationship between hatch date at t+1 and significant explanatory variables identified by model selection
Plotted point size is scaled by the number of data points that have that value, larger indicating a greater number of points. Shaded areas around lines and lines around point estimates represent the 95% confidence interval of each relationship.

4.7.4. Parameterisation of the recruitment function

4.7.4.1. Model selection - recruitment

Table S4.9: AIC and ΔAIC values for recruitment model selection

AIC	ΔAIC *in comparison to model on the preceding row	Variable removed
13213.63	NA	Null model one
13077.22	-136.41	Null model two
12983.66	-93.56	Full model
12981.66	-2.00	removed birth year hatch
12979.72	-1.94	removed hatch date^2
12977.83	-1.89	removed spring temperature
12976.09	-1.74	removed immigrant
12974.30	-1.79	removed age
12973.75	-0.55	removed winter temperature
12973.66	-0.09	removed winter precipitation

4.7.4.2. Null model parameterisations

Table S4.10: Full model output for null recruitment model one (density dependent)

	Estimate	Standard error	P value
Intercept	-1.33	0.07	0.00
Population size	-0.16	0.07	0.03

Table S4.11: Full model output for null recruitment model two (density dependent, spring temperature, and synchrony)

	Estimate	Standard error	P value
Intercept	-1.26	0.07	0.00
Hatch date	-0.13	0.08	0.09
Synchrony	-0.19	0.06	0.00
Synchrony^2	-0.10	0.02	0.00
Population size	-0.19	0.07	0.01

4.7.4.3. The final recruitment model

Table S4.12: Full model output for the final recruitment model

	Estimate	Standard error	P value
Intercept	-1.35	0.09	0.00
Hatch date	-0.16	0.06	0.01
Synchrony	-0.17	0.05	0.00
Synchrony^2	-0.12	0.02	0.00
Beech = 1	0.07	0.14	0.64
Beech = 2	0.50	0.12	0.00
Spring precipitation	0.13	0.05	0.02
Clutch size	0.13	0.02	0.00
Clutch size^2	-0.04	0.01	0.00
Section of woodland C	-0.30	0.09	0.00
Section of woodland CP	0.08	0.09	0.37
Section of woodland EX	0.22	0.07	0.00
Section of woodland MP	-0.14	0.09	0.10
Section of woodland O	-0.10	0.08	0.23
Section of woodland P	0.13	0.15	0.40
Section of woodland SW	0.17	0.08	0.03
Section of woodland W	0.16	0.08	0.03
Population size	-0.17	0.07	0.01

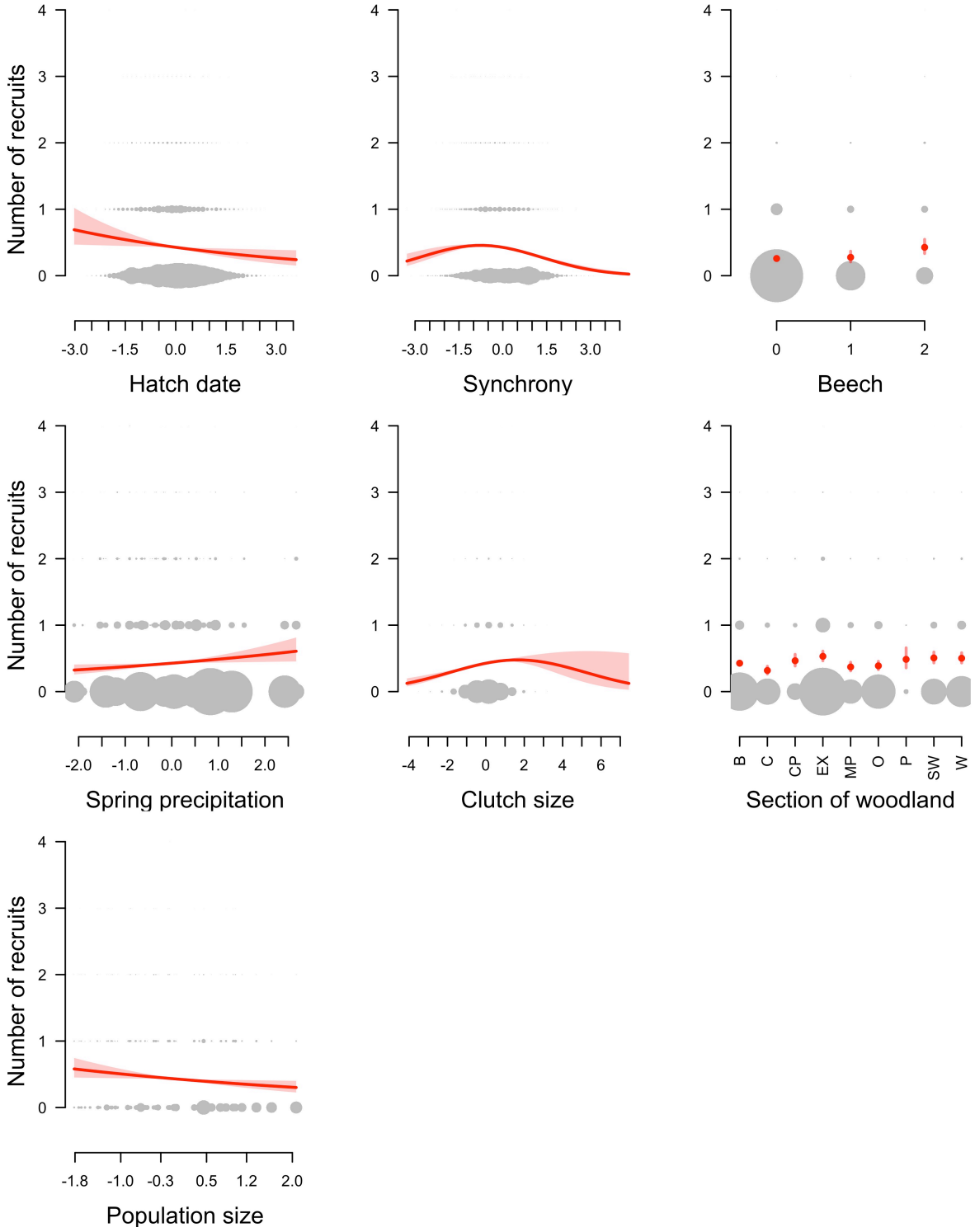

Figure S4.3: Relationship between number of recruits and significant explanatory variables identified by model selection
Plotted point size is scaled by the number of data points that have that value, larger indicating a greater number of points. Shaded areas around lines and lines around point estimates represent the 95% confidence interval of each relationship.

4.7.5. Parameterisation of inheritance function

4.7.5.1. Model selection - inheritance

Table S4.13: DIC and change in DIC values for selection of random effects in the inheritance model selection

DIC	Change in DIC *in comparison to null	Random effect
24028.96	NA	Null model
24038.16	9.2	Permanent environment (PE)
21930.55	-2098.41	Birth year (BY)
24033.03	4.07	Maternal effect (ME)
21937.14	-2091.82	BY + PE
21929.82	-2099.14	BY + ME
24040.49	11.53	ME + PE
21934.12	-2094.84	BY + PE + ME

While the addition of birth year to the random effect structure did generate a large reduction in the DIC, the posterior distribution generated by the addition of this effect was very poor. High autocorrelation was present in the estimation of the birth year parameter, demonstrating that this variable was not being well estimated. Increases in the number of iterations and reduction in sampling frequency did not rectify this issue. Consequently, despite the reduction in DIC birth year was not included as a random effect, we do not believe the reduction represents explanatory power of this variable but rather an inability to properly estimate it.

The only fixed effect removed from the full inheritance model, based on the posterior mode confidence intervals was ‘Immigrant status’.

4.7.5.2. Null model parameterisations

Table S4.14: Full model output for null inheritance model one (density dependent)

	Posterior mean	Lower confidence interval	Upper confidence interval
Intercept	0.05	0.03	0.08
Population size	-0.35	-0.37	-0.33

Table S4.15: Full model output for null inheritance model two (density dependent, spring temperature, and synchrony)

	Posterior mean	Lower confidence interval	Upper confidence interval
Intercept	0.06	0.04	0.08
Spring temperature	-0.65	-0.67	-0.64
Population size	-0.05	-0.07	-0.03

4.7.5.3. The final inheritance model

Table S4.16: Full model output for the final inheritance model

	Posterior mean	Lower confidence interval	Upper confidence interval
Intercept	0.1	0.06	0.14
Spring temperature	-0.64	-0.66	-0.62
Spring precipitation	0.15	0.13	0.16
Winter temperature	0.02	0	0.04
Winter precipitation	-0.07	-0.08	-0.05
Section of woodland C	-0.17	-0.24	-0.11
Section of woodland CP	-0.15	-0.22	-0.08
Section of woodland EX	-0.07	-0.12	-0.01
Section of woodland MP	-0.34	-0.4	-0.28
Section of woodland O	0.08	0.02	0.14
Section of woodland P	0.11	0	0.23
Section of woodland SW	0.18	0.12	0.24
Section of woodland W	-0.02	-0.08	0.04
Population size	-0.06	-0.08	-0.04

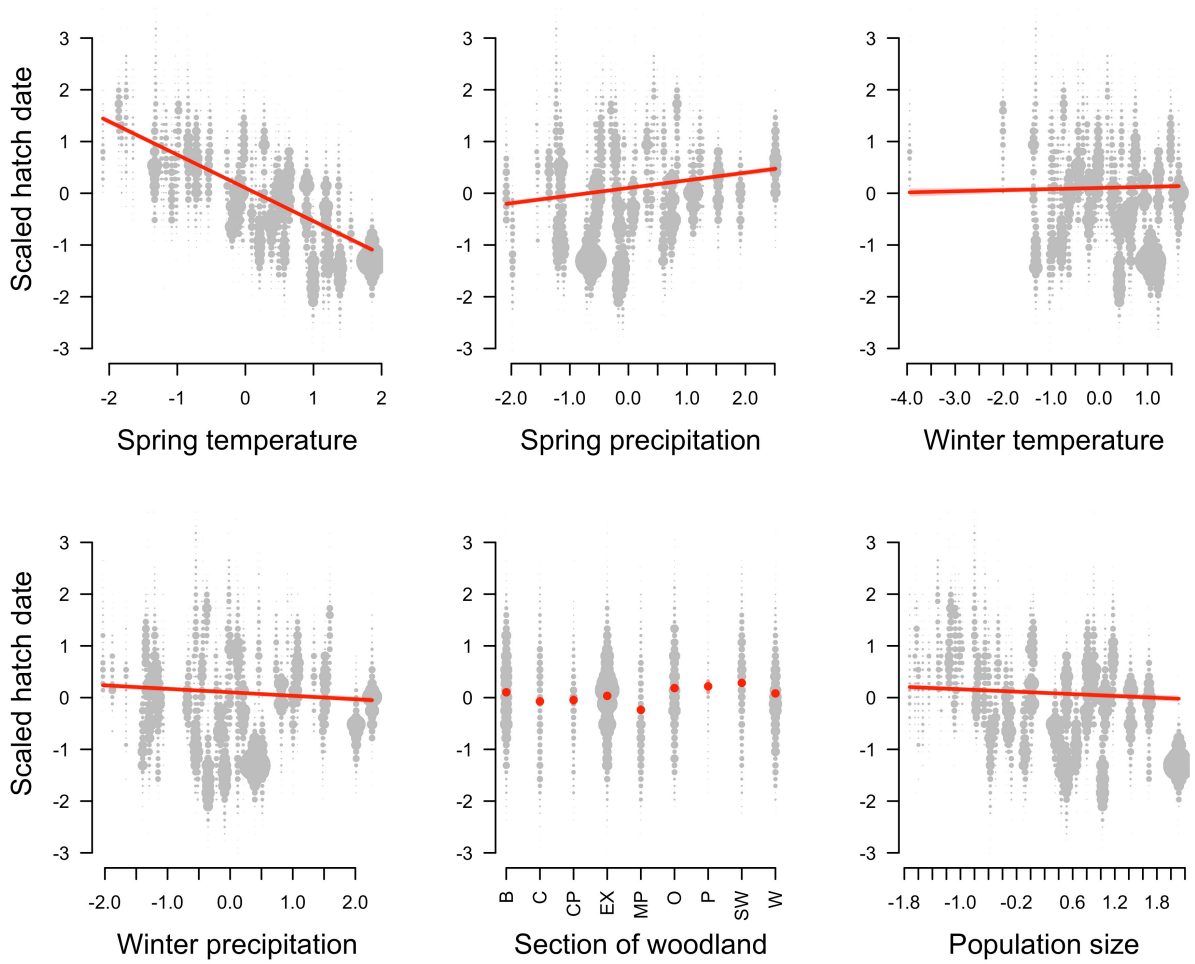

Figure S4.4: Relationship between hatch date and significant explanatory variables identified by model selection
Plotted point size is scaled by the number of data points that have that value, larger indicating a greater number of points. Shaded areas around lines and lines around point estimates represent the 95% confidence interval of each relationship.

CHAPTER FIVE

Shared cues, phenotypic plasticity, and micro-evolution are key to population prosperity under climate change

Emily G. Simmonds, Ella F. Cole, and Tim Coulson

5.1. Abstract

Rapid climate change is altering biological systems across the globe. With these changes projected to continue, predicting how populations will respond is a key challenge. Alterations to the timing of life history events are some of the most prominent biological responses to a changing climate. However, species and populations are not shifting their timing uniformly, resulting in disruptions to interspecific interactions. Predicting the occurrence of temporal mismatches and their impacts on population dynamics is a vital step for estimating population persistence. Previous work has largely isolated phenological influence on population dynamics. To generate accurate predictions, we need a consideration of how alterations to environmental drivers, in tandem with phenology, might impact key demographic rates. In this study we use an integral projection model (IPM) parameterised for a model population of wild great tits (*Parus major*) and their winter moth caterpillar prey (*Operophtera brumata*). We generate directed predictions of future great tit population dynamics to the end of the 21st Century by combining the population model with climate projections from the UKCP09 estimates under low, medium, and high greenhouse gas emissions scenarios. We also explore how different cue usage by interacting species can influence the likelihood of mismatch under climate change. Our results show that if phenological cues are shared the great tit population will be able to maintain synchrony with their prey species, in all but the highest emissions scenario, even showing population increases. However, if phenological cues are not shared, mismatch occurred accompanied by population plateaus or declines. Overall, this study highlights the benefits of using a statistically based predictive model and considering the influence of multiple environmental drivers can generate a picture of future population persistence. Via this approach we managed to show that Wytham Woods great tits are likely to persist into the next century.

5.2. Introduction

Rapid climate change is altering biological populations across the globe (on Climate Change 2007). Predicting which biological populations may be at risk of extinction is a key challenge for conservation, management, and policy makers. Changes in phenology are some of the most well documented species responses to climate change (Parmesan 2007; Parmesan & Yohe 2003; Cleland et al. 2007; Menzel et al. 2006; Thackeray et al. 2016; Thackeray et al. 2010). These shifts have been observed across systems, taxa, and life history events; from flowering (Menzel et al. 2006; Cleland et al. 2007), to migration (Lehikoinen & Sparks 2010; Møller et al. 2008), and breeding (Plard et al. 2014; Visser et al. 1998; Charmantier et al. 2008; Visser & Both 2005; Sparks & Crick 1999). However, the observed phenological changes are not uniform, resulting in mismatches between temporally interacting species. Such mismatches are hypothesised to cause fitness reductions in one or both interacting species (Cushing 1969), leading to population declines. Yet neither mismatches nor population declines are ubiquitous (Singer & Parmesan 2010; Johansson, Kristensen, et al. 2015; Menzel et al. 2006; Reed, Jenouvrier, et al. 2013; Reed, Grøtan, et al. 2013).

Temporal mismatches occur when interacting species respond differently to the same environmental change. Responses can diverge due to different use of environmental cues, such as annually varying temperature and stronger reliance on invariant photoperiod (Plard et al. 2014), or differing amounts of environmental sensitivity (Visser & Both 2005; Gienapp et al. 2006; Gienapp et al. 2014). Mismatches, are generally assumed to link closely with individual fitness (Visser et al. 2006; Reed, Jenouvrier, et al. 2013; Charmantier & Gienapp 2014). In spite of these links, mismatches do not always translate into reductions in population size (Visser &

Both 2005). Buffering can occur through density-dependent feedbacks (Grøtan et al. 2009; Both 2010; Johansson, Kristensen, et al. 2015) or positive effects of other environmental drivers, such as winter conditions (Perrins 1965b; Van Balen 1980). Reduced competition or increased winter food availability can decrease over winter mortality and counteract any mismatch induced fitness costs (Grøtan et al. 2009; Both 2010; Perrins 1965b; Van Balen 1980).

To predict the future shape of populations, identifying those which may be negatively impacted by climatic change, we need to ascertain the conditions which cause mismatches and the implications of those on population dynamics. Despite a wealth of research into phenology and phenological change, an empirical understanding of exactly why mismatches occur and their influence on population persistence is lacking (Miller-Rushing et al. 2010; Johansson, Kristensen, et al. 2015; Bennett et al. 2015). Some previous work has considered the role that phenology and phenological mismatch plays in population dynamics (Vedder et al. 2013; Chevin et al. 2010; Chevin & Lande 2015; Childs et al. 2016; Gienapp et al. 2014; Johansson, Kristensen, et al. 2015; Plard et al. 2014). These have either been theoretical (Gienapp et al. 2014; Johansson, Kristensen, et al. 2015) or consider the influence of phenology on fitness in isolation of other explanatory drivers (Vedder et al. 2013; Chevin et al. 2010). Theoretical work suggests there are several mechanisms through which populations can experience mismatch without accompanied declines in population size; density-dependence, adaptive mismatch, alternative interacting species, and low fitness implications of mismatch (Johansson, Kristensen, et al. 2015). However, direct assessments of fitness impacts suggest strong individual costs for mismatch (Reed, Jenouvrier, et al. 2013; Reed, Grøtan, et al. 2013). In order to disentangle the causes and consequences of phenological mismatch we

need to combine detailed observational data with appropriate, and predictive models (Miller-Rushing et al. 2010), which simultaneously consider the influence of key drivers of fitness and trait change. To date there have been few examples of analyses which achieve this (Childs et al. 2016; Plard et al. 2014). Plard et al (Plard et al. 2014) provided a consideration of the likelihood and consequences of mismatch in a population of roe deer (*Capreolus capreolus*). They showed that a lack of advance of parturition dates, relative to vegetation abundance, would lead to declines in population size (Plard et al. 2014). More studies, making use of available long-term datasets and considering the role of phenology in conjunction with other drivers of population dynamics, are needed to build up a more complete picture of the future of biological systems. Furthermore, such studies should be directed by actual climate change predictions to generate projections of population dynamics based on our current best estimates of the future climate.

In this study we employ a general framework of population model to explore the causes and consequences of mismatch in a model system of wild great tits and their prey species the winter moth caterpillar (*Operptera brumata*) from Wytham Woods. This system is currently experiencing matching between the interacting species' spring phenology. However, great tit fitness has been shown to be reduced by mismatch (Reed, Jenouvrier, et al. 2013). Consequently, population declines could occur if climate change induced mismatch impacts these populations. We use an integral projection model (IPM) [34,35, Chapter 4] combined with detailed long-term census data to project future trait and population dynamics for the resident Wytham Woods great tit population. This model was previously parameterised and cross validated for the study population in Chapter 4. Projections span to the end of this century. We direct the model using climate predictions generated by the UKCP09

project, following different greenhouse gas emission scenarios; low, medium, and high. We test three hypotheses:

1. Great tit hatch dates will advance due to ecological and evolutionary over this century.
2. When cues are shared between interacting species synchrony will remain under climate change.
3. Population size will be regulated by both spring and winter conditions and phenological synchrony.

To be able to ask questions of what conditions could facilitate the onset of mismatches, we explored two biological scenarios. One representing shared cue usage and one representing a discrepancy in environmental sensitivity. It is highly unlikely that any two species use the exact same environmental cue to determine timing, particularly if one of the interacting species has a direct relationship between temperature and development (Buse et al. 1998). This is the case for the great tit and winter moth caterpillar system. As a result, we would expect that winter moth caterpillars respond to temperature right up until their peak abundance (van Noordwijk et al. 1995; Visser et al. 2004; Buse et al. 1998), whereas great tits use temperature indirectly, as more of a predictive cue (van Noordwijk et al. 1995; Visser et al. 2004; Charmantier et al. 2008). Therefore, we set up one scenario of a shared environmental cue to represent environmental cues changing at the same rate. We also set up a scenario of differently changing cues to capture what could happen if the environmental cues used by species change at different rates or if the species themselves have different environmental sensitivities.

By addressing these questions and scenarios we can begin to gain insights into the possible future form of populations under novel climatic change. We can assess not only how the population dynamics will change but also explore the drivers of such changes. In doing so, we can isolate the influence of phenology and phenological synchrony, as well as exploring under which circumstances we would expect mismatches to arise.

Methodology

5.2.1. The study system

The great tit is a small passerine bird, resident in UK all year. The study population is a nest box breeding population from Wytham Woods, Oxfordshire, which has been monitored continuously since 1960 (Perrins 1965b; Perrins & McCleery 1989). Data have been collected on the timing of reproduction (lay date of first egg, lay date of last egg, hatch date), individual identity, and reproductive success. The timing of chick hatching relative to the peak abundance of the primary breeding season food source (winter moth caterpillars), has significant fitness implications in this population (Reed, Jenouvrier, et al. 2013; Van Balen 1980; Perrins & McCleery 1989). A failure to time optimally can reduce reproductive success (van Noordwijk et al. 1995; Charmantier et al. 2008; Simmonds et al. 2017). Climate change has already driven phenological advance in this population, on average 0.25 days per year since 1960. Caterpillar peak abundance has also advanced by 0.28 days per year over the same period. Despite the differing rates of advance synchrony between breeding timing and the prey abundance appears to have been maintained through phenotypic plasticity (Charmantier & Gienapp 2014; Gienapp et al. 2014; Vedder et al. 2013). However, it is not known whether this will continue to be the case under persistent directional climate change.

Environmental data was taken from the MET Office UKCP09 interpolated dataset (MET Office 2009). For further detail of population and environmental data used for the model, please see Chapter 4.

5.2.2. Climate projections

Projections from the UK Climate Projection 2009 estimates (UKCP09) (MET Office 2009), grid square 4500210, were used as the basis for our climate projections. We used the MET office Weather Generator tool (MET Office 2009) to generate 1000 equally plausible daily mean temperature and precipitation projections from 2015 to 2100. These daily environmental records were generated in three sets using the UKCP09 projections for 2010 to 2039 as the basis for our 2015 to 2040 daily predictions, UKCP09 projections for 2040 to 2069 as the basis for our 2041 to 2070 daily predictions, and UKCP09 projections for 2070 to 2099 as the basis for our 2071 to 2100 daily predictions. The 90 year period was split into three sections to match the time periods projections exist for rather than assuming the environmental conditions for the first three decades will persist for the remainder of the 21st century.

The daily predictions for each period were combined to generate 1000 equally plausible daily environmental predictions stretching continuously from 2015 to 2100. As the 1000 different predictions for each period were generated from random sampling of the climate model parameters, each set of daily predictions did not have an exact equivalent in the other time periods. However, as each set of daily temperatures was equally plausible, it is equally likely that any of the 2015 to 2040 predictions are followed by any of the 2041 to 2070 predictions. Therefore, the exact set of predictions combined did not matter when combining the different decades of prediction to create continuous daily predictions.

From these daily predictions we calculated annual values of the environmental drivers used in model parameterisation:

- **Spring temperature** - mean temperature for the period 1st March to 9th May. This window of temperature was identified as the optimum critical window in which great tits perceive temperature using an absolute sliding time window method in the 'Climwin' package in R (Bailey & De Pol 2016; van de Pol et al. 2016).
- **Spring precipitation** - total precipitation from 1st April to 31st May to cover the nesting period and when young chicks are in the nest. This period spans from just over a week prior to the earliest mean lay date to three days prior to the latest mean hatch date on record.
- **Winter temperature** - mean temperature from the winter following the breeding attempt from (1st December to 28th/29th February).
- **Winter precipitation** - total precipitation from 1st December to 28th/29th February.

Observed values of the environmental drivers from 2010 to 2014 were added to the beginning of all climate predictions in order to begin each model simulation on known values.

5.2.2.1. Emissions scenarios

The process of climate projection assembly was repeated for climate models run using different levels of anthropogenic greenhouse gas emissions scenarios; low, medium, and high. Projections of future climate are made using our best knowledge of the climate system in conjunction with anthropogenic forcing. As human decisions are a large source of uncertainty, multiple levels of anthropogenic greenhouse gas

emissions are usually generated to take account of some of the uncertainty and variability that could result.

Table 5.1: Summary statistics of climate predictions.
The 5th and 95th percentiles of climate predictions for each environmental driver.

Climate variable	Observed	Predicted		
		Low	Medium	High
Spring temperature (ºC)	7.9	8.5 - 9.8	8.7 - 10.2	9.0 - 10.7
Spring precipitation (mm)	106	140 - 166	140 - 165	139 - 167
Winter temperature (ºC)	4.3	5.1 - 6.7	5.1 - 6.7	5.5 - 7.5
Winter precipitation (mm)	164	169 - 202	170 - 209	172 - 217

5.2.2.2. Environmental drivers

To isolate the effects of different environmental drivers on the trait and population dynamics we ran the coupled, high emissions scenario four additional times. In each of the additional simulations, one of the environmental drivers (spring temperature, spring precipitation, winter temperature, and winter precipitation) was held fluctuating within the observed range.

5.2.3. The model

5.2.3.1. Model description and parameterisation

The model we employ here was parameterised and cross validated in Chapter 4. It takes the form of an IPM, which incorporates a mechanistic inheritance function based on quantitative genetic principles. These models are based on four fundamental functions, survival (S), recruitment (R), development (D), and inheritance (H). Exploratory statistical analyses and model selection were used to identify the significant explanatory drivers of each of these demographic processes (see Chapter 4, Methodology for full details on model parameterisation). The

equations for the final functional forms are shown in Equations 7 to 11. β_1 is the intercept. Z is hatch date, SN is synchrony, ST is spring temperature, SP is spring precipitation, WT is winter temperature, WP is winter precipitation, CS is clutch size, and N is population size and their respective slopes (β_i). B is the effect of beech mast index. $V^D(E', \theta, t)$ and $V^H(E', \theta, t)$ are the standard deviation around E' for the development and inheritance functions.

Equation 7:

$$S(Z(G,E),t) = \frac{1}{1+e^{-(\beta_1+Z.\beta_z+SN.\beta_{sn}+SN^2.\beta_{sn2}+ST\prime.\beta_{st}+SP\prime.\beta_{sp}+WT.\beta_{wt}+N.\beta_n+B)}}$$

Equation 8:

$$R(Z(G,E),t) = e^{-(\beta_1+Z.\beta_z+SN.\beta_{sn}+SN^2.\beta_{sn2}+SP\prime.\beta_{sp}+CS.\beta_{cs}+CS^2.\beta_{cs2}+N.\beta_n+B)}$$

Equation 9:

$$D(E'|E,\theta,t) = \frac{1}{V^D(E\prime,\theta,t)\sqrt{2\pi}} e^{-\frac{(E-(\beta_1+Z.\beta_z+SN.\beta_{sn}+SN^2.\beta_{sn2}+ST.\beta_{st\prime}+N.\beta_n))^2}{2V^D(E\prime,\theta,t)^2}}$$

Equation 10:

$$H(E'|E,\theta,t) = \frac{1}{V^H(E\prime,\theta,t)\sqrt{2\pi}} e^{-\frac{(E-(\beta_1+Z.\beta_z+ST\prime.\beta_{st\prime}+SP\prime.\beta_{sp\prime}+WT.\beta_{wt}+WP.\beta_{wp}+N.\beta_n))^2}{2V^H(E\prime,\theta,t)^2}}$$

The extended IPM tracks a bivariate distribution of breeding values (G) and environmental components (E), which combine to a phenotype (Z(G, E)) we assume Z = G + E. This model formulation allows us to track the influence of ecological and evolutionary processes on the expressed phenotype, in addition to tracking the overall trait and population dynamics. The full model equation is shown in **Equation 11**, where G and E are the breeding value and environmental components of the phenotype at time t, respectively. E' represents the environmental component of the

phenotype at time t+1. dG and dE are the change in G and the change in E from t to t+1. S(Z(G, E), θ', t), R(Z(G, E), θ', t), D(E'|E, G, θ, t, a) and H(E'|E, G, θ, t) are the survival, recruitment, development, and inheritance conditional on the phenotype (Z) and environment at time t (θ). Development is also conditional on age (a), making the model age structured. Therefore the population size at t+1 is calculated by summing the age 1 (N_{a1}) and age 2 (N_{a2}) population sizes at each time step. The survival and recruitment functions, are conditional on both the environment at time t and t-1 (θ') due to lagged effects of spring conditions just prior to the census.

Equation 11:

$$N(t+1) = \sum N_{a1} + N_{a2}$$

$$N_a(G, E', t+1)$$

$$= \iint [S(Z(G,E),\theta',t)D(E'|E,G,\theta,t,a) + R(Z(G,E),\theta',t)H(E'|E,G,\theta,t)]\, N(G,E,t)dGdE$$

The emergent parameters used the final model are as follows:

- Survival - hatch date, synchrony, synchrony^2, beech mast index, spring temperature, spring precipitation, winter temperature, immigrant, section of woodland, and population size.
- Development - hatch date, synchrony, synchrony^2, spring temperature, section of woodland, age, and population size.
- Recruitment - hatch date, synchrony, synchrony^2, beech mast index, spring precipitation, clutch size, clutch size^2, section of woodland, and population size.
- Inheritance - spring temperature, spring precipitation, winter temperature, winter precipitation, section of woodland, and population size.

5.2.3.2. Model simulation

The model was simulated for 91 years (from 2010 to 2100). At each time step values for continuous environmental drivers (spring temperature, spring precipitation, winter temperature, winter precipitation) were chosen from either observed values (2011 to 2014) or climate projections (2015 to 2100). Beech mast index could not be predicted as the exact drivers of mast years have not yet been identified and therefore was selected stochastically at each time step based on observed frequency from 1961 to 2010 (following the method of Chapter 4). Beech mast values were selected prior to model simulation and consequently were the same for each model simulation run within each scenario. Values of zero (no mast), one (low mast), or two (high mast) were chosen at a frequency of the observed data, however two high mast years could not be allocated concurrently. Additionally if four consecutive non-mast years occur then the next year is a high mast year (Matthews 1955).

Caterpillar peak abundance timing is determined as the date on which half of the observed late instar larvae have descended to the group to pupate, also termed half fall date. This was estimated at each time step using a linear predictor (Equation 12) parameterised from a linear model. The linear model was first defined in Chapter 4 and was created to interpolate historical half fall timings, in addition to predicting future values. The difference between the timing of caterpillar peak abundance and great tit hatch dates is used to calculate synchrony. This generates an index where 0 indicates perfect matching, positive numbers represent hatch dates being later than caterpillar timing, and negative numbers represent hatch dates prior to caterpillar timing.

Equation 12:

$$HF = \beta_1 + ST'.\beta_{st'} + SP'.\beta_{sp'} + WT.\beta_{wt} + WP.\beta_{wp} + B$$

The model was simulated for six different climate and biological scenarios:

Table 5.2: Details of the composition of the six climate-biological scenarios.

Details of each test scenario including greenhouse gas emissions scenario and whether spring temperature cues are shared between great tits and caterpillars.

Scenario	Emissions level	Shared cue?
1	Low	Yes
2	Medium	Yes
3	High	Yes
4	Low	No
5	Medium	No
6	High	No

The model was simulated in this way 1000 times for each scenario, each of the 1000 runs using an equally likely projection of future climate for the period of 2010 to 2100. At each time step of the simulation we calculate and save the population size (resident breeding females) and the G and E distribution. The coupled scenario involves the same value of spring temperature being chosen to predict both caterpillar peak abundance timing and great tit hatch dates. In the decoupled scenario, the spring temperature value used to predict caterpillar half fall timing is taken from the climate change projections. In contrast, the spring temperature value used for great tit hatch dates is chosen stochastically from a distribution defined by observed values i.e. not experiencing climate change.

5.3. Results

5.3.1. Coupled climate cues for great tits and winter moth caterpillars

5.3.1.1. Predicted population and trait dynamics 2010 to 2100

In scenarios one to three (Table 5.2), with coupled environmental cues, a general pattern of an increase in the resident breeding female population size after an initial decline can be seen (Figure 5.1). The initial decline is generated from the observed environmental conditions from 2010 to 2014, which did see population declines over this period. Later increases in population size are accompanied by maintenance of synchrony, right up to 2100. Synchrony between caterpillar peak abundance and great tit hatch dates fluctuates around zero, perfect synchrony. Synchrony remains tight for the entirety of the low emissions scenario. However, the medium and high emissions scenarios show an increase in inter-annual variability towards the end of the century, with hatch dates beginning to lag behind caterpillar peak abundance (increasingly positive values).

Population growth rates were similar across all emissions scenarios (Figure 5.1), with 90 % of simulation runs for each scenario showing average population growth rates of between 2 and 3 % (Table 5.3). All emissions scenarios produced standard deviations, which span the observed standard deviation (69.9 individuals from 1960 to 2010). The standard deviation of population size within each individual model simulation was greatest for the medium emissions scenario (standard deviation of 60.8 to 104.2). The simulated standard deviation range stretched further above observed than below for all emissions scenarios, suggesting a general trend to an increased variability in population size through the 21st century.

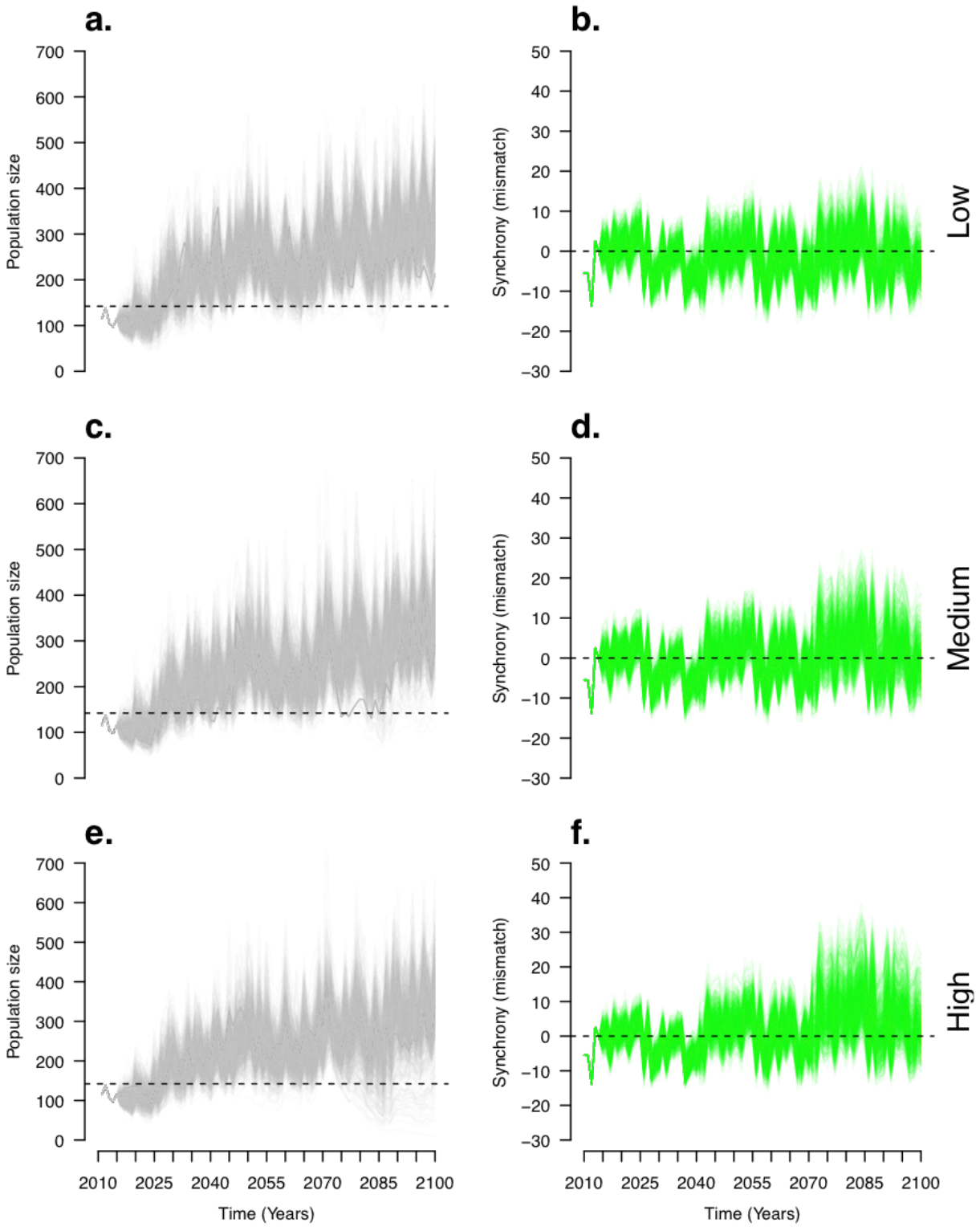

Figure 5.1: Simulated population size and synchrony from 2010 to 2100. Each plot shows the simulated results for population size and synchrony from a. and b.1000 low emissions runs, c. and d. 1000 medium emissions runs, and e. and f. 1000 high emissions runs. Horizontal dotted line indicates start value (a, c, e) or 0 (b, d, f).

Table 5.3: Summary of simulated population dynamics.
Summary statistics from simulated population size under different emissions scenarios for coupled winter moth caterpillar and great tit cues. All statistics stated are the 5th and 95th percentile of the 1000 equally probable climate predictions.

	Emissions scenario		
	Low	Medium	High
Mean population growth rate	1.02 - 1.03	1.02 - 1.03	1.02 - 1.03
Standard deviation of population size (breeding females)	52.2 - 92	60.8 - 104.2	54.6 - 104.1

In the coupled scenario hatch dates (phenotype) advanced across all simulations and emissions scenarios (Figure 5.2). Phenotypic advance was greatest for the high emissions scenario, with the advance of average hatch date equivalent to an advance of 0.07 to 0.24 days per year (Table 5.4). The standard deviation increased with increasing emissions levels from a range of 3.2 to 6.3 up to 4.1 to 9.3. The observed hatch date standard deviation (7.5) sits within, but at the upper end of the 5th and 95th percentile of predictions for high and medium emissions scenarios and above 90 % of the predictions for the low emissions scenario. The phenotypic trends shown were driven by a combination of genetic change and phenotypic plasticity (Figure 5.2). Both the genetic and environmental components of the phenotype advance over the course of the model simulation. While the environmental component fluctuates heavily, the genetic component advances gradually, driving a directional shift in hatch date and leading to a comparable level of advance to that generated by the environmental component of the phenotype.

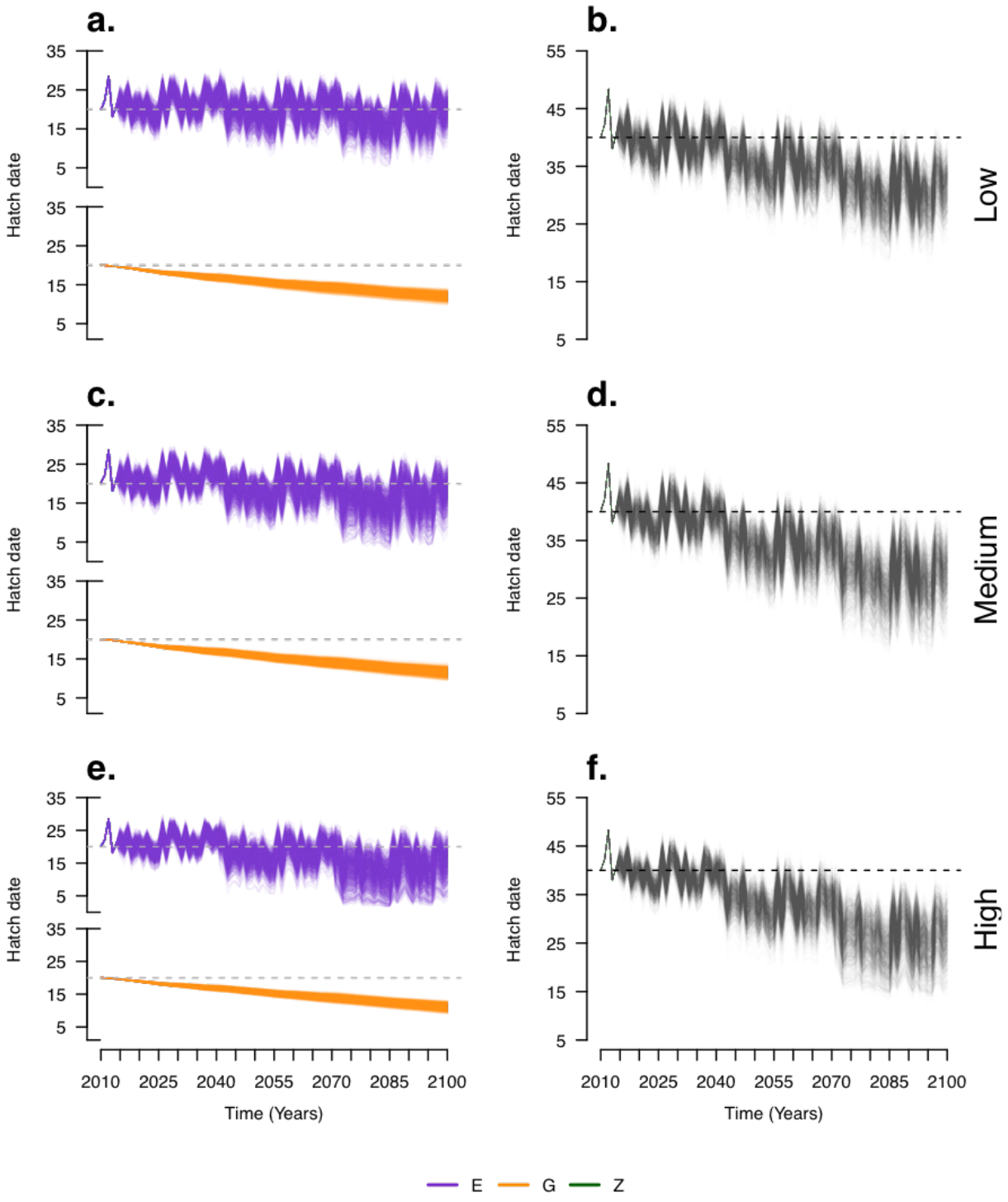

Figure 5.2: Predicted dynamics of the genetic (G) and environmental (E) component of the phenotype (Z) from 2010 to 2100.
a. and b. for low emissions scenarios, c. and d. for medium emissions scenarios, and e. and f. for high emissions scenarios. Horizontal dotted line indicates start value.

Table 5.4: Summary of simulated phenotype dynamics.
Summary statistics from simulated hatch date under different emissions scenarios for coupled winter moth caterpillar and great tit cues. All statistics stated are the 5th and 95th percentile of the 1000 equally probable climate predictions.

	Emissions scenario		
	Low	Medium	High
Mean slope of phenotype change (days per year)	-0.15 - -0.04	-0.19 - -0.05	-0.24 - -0.07
Standard deviation of hatch date (days)	3.2 - 6.3	3.5 - 7.6	4.1 - 9.3

5.3.1.2. The influence of changes in different environmental drivers on population and phenotype dynamics

For the coupled cue scenario, we also tested the influence of environmental drivers on predicted population sizes and hatch date change. No simulation managed to reproduce the level of population increase shown under full climate change (Figure 5.1). Removing climate change for any one of the key environmental drivers (spring temperature, spring precipitation, winter temperature, and winter precipitation) led to substantially reduced population growth. A slight increase in population size is seen in all trialled combinations of environmental drivers, with the greatest increase occurring when spring temperature is held at observed levels. In contrast, advances in hatch date are shown to be driven primarily by spring temperature. When spring temperatures are held at observed levels then hatch dates remain fluctuating around 2010 values. However, when spring temperature is allowed to experience climate change, hatch dates always advance.

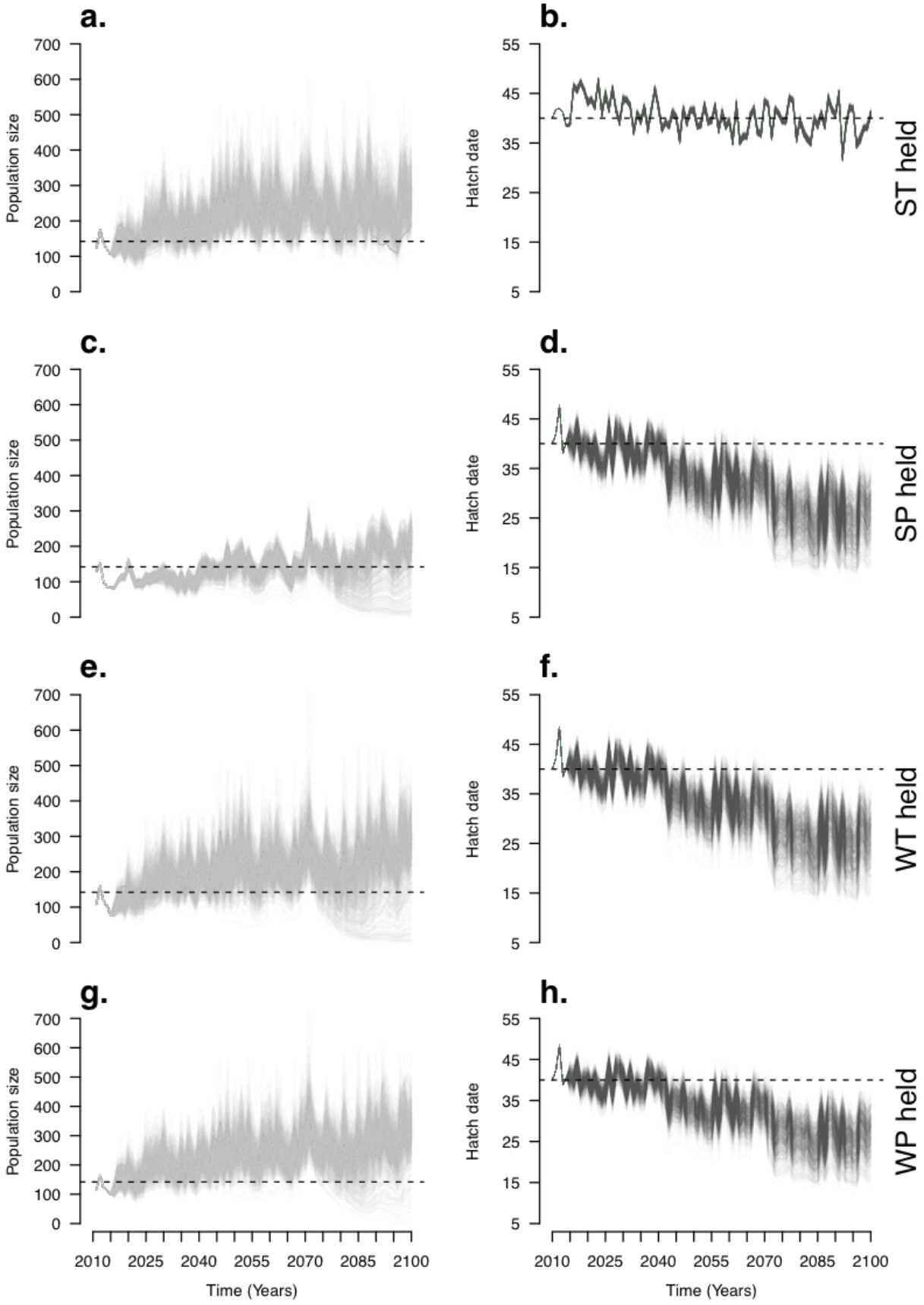

Figure 5.3: Predicted population and hatch date dynamics from 2010 to 2100.

Simulated population and trait dynamics for models with one environmental driver held at 1960 to 2010 levels. a. and b. spring temperature (ST), c. and d. spring precipitation (SP), e. and f. winter temperature (WT), g. and h. winter precipitation (WP).

5.3.2. Decoupled

5.3.2.1. Changes in population and phenotype dynamics 2010 to 2100

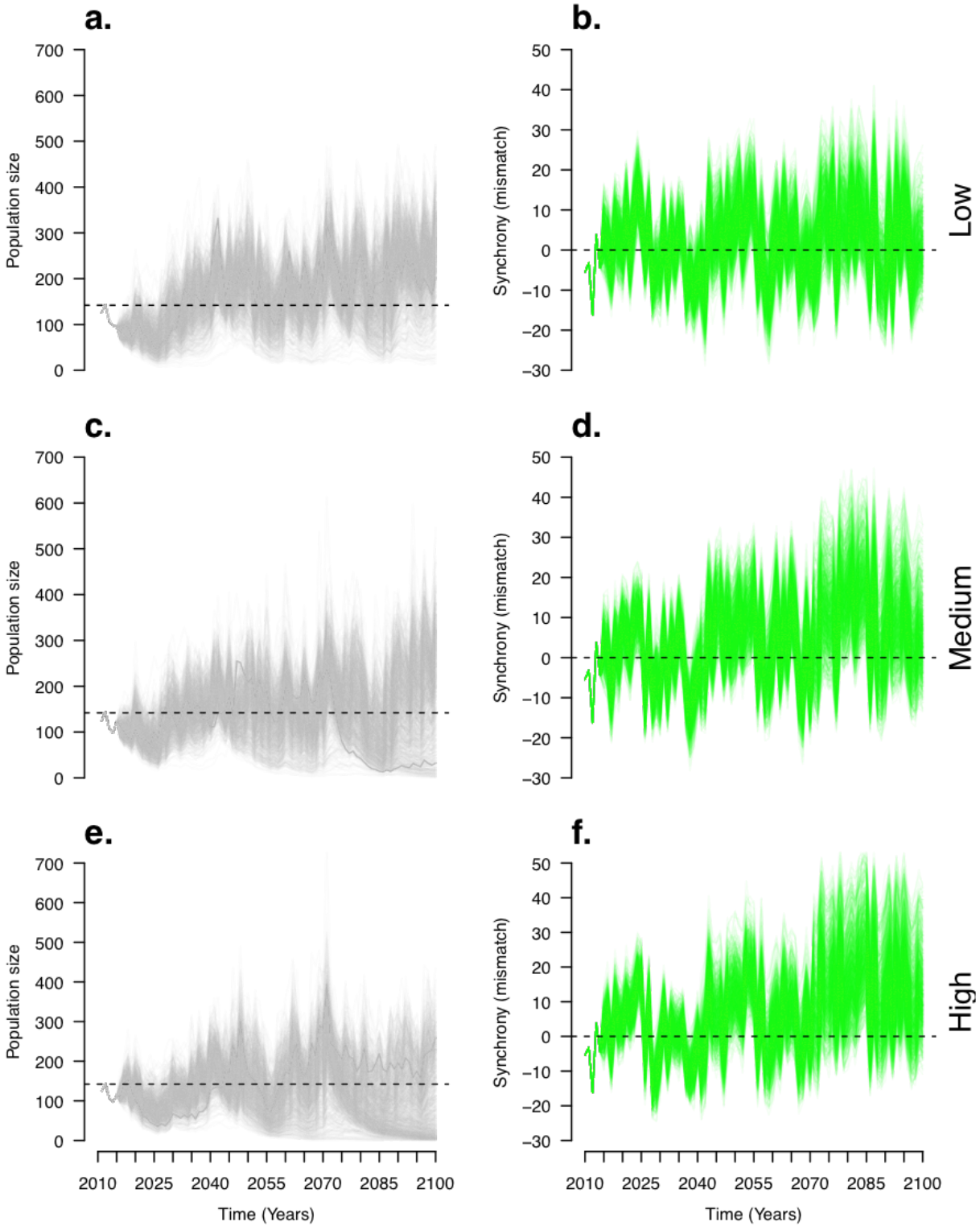

Figure 5.4: Simulated population size and synchrony from 2010 to 2100, decoupled cues.

Each plot shows the simulated results for population size and synchrony from a. and b.1000 low emissions runs, c. and d. 1000 medium emissions runs, and e. and f. 1000 high emissions runs. Horizontal dotted line indicates 2010 value (a, c, e) or 0 (b, d, f).

In scenarios four to six (Table 5.2), with decoupled cues, a general pattern can be seen across emissions scenarios, of a plateau in population size. The population fluctuates around 2010 levels after an initial decline. Again, the initial decline is driver by the observed environmental conditions from 2010 to 2014. Some individual simulations in each scenario saw a slight increase in population size and others a decrease, some to extinction. This is further supported by Table 5.5, which shows population growth rates spanning 1 (no population growth) for the medium and high emissions scenarios. While variability in population size appears to increase over this century in Figure 5.4 between individual model simulations, the standard deviation of population size within model simulations largely decreased, with the observed standard deviation (69.9 breeding females) falling close to the 95th percentile for all emissions scenarios.

Synchrony between caterpillar peak abundance and great tit hatch dates showed double the variability of the coupled scenario, with hatch dates fluctuating between 20 days prior to and 20 days after the caterpillar peak. This increased variability was accompanied by a directional shift in synchrony for hatch dates to increasingly lag behind caterpillar timing the further into the future we simulate. This trend was most pronounced in the high emissions scenario (Figure 5.4.e. and f.), which also showed a number of declining populations (218/1000 simulations).

Phenotypic change in the decoupled simulations was considerably slower than for the coupled simulations, with less variability between model simulations. However, a general trend of an advance of hatch dates can still be seen (Figure 5.4). The rate of advance shown for all emissions scenarios was almost identical and amounted to 0.01 days per year, an order of magnitude lower than rates reached in the coupled

scenario. Phenotypic change here was driven almost entirely by genetic change. The environmental component of the phenotype fluctuates around the 2010 value (Figure 5.5). The genetic component advances a similar amount to the coupled simulations and almost identically across emissions scenarios.

Table 5.5: Summary of simulated population dynamics.
Summary statistics from simulated population size under different emissions scenarios for coupled winter moth caterpillar and great tit cues. All statistics stated are the 5th and 95th percentile of the 1000 equally probable climate predictions.

	Emissions scenario		
	Low	Medium	High
Mean population growth rate	1.01 - 1.03	0.99 - 1.03	0.98 - 1.03
Standard deviation of population size (breeding females)	35.7 - 84.3	32.7 - 76.6	38.2 - 82.2

Table 5.6: Summary of simulated phenotype dynamics.
Summary statistics from simulated hatch date under different emissions scenarios for coupled winter moth caterpillar and great tit cues. All statistics stated are the 5th and 95th percentile of the 1000 equally probable climate predictions.

	Emissions scenario		
	Low	Medium	High
Mean slope of phenotype change (days per year)	-0.01 - -0.01	-0.01 - -0.01	-0.01 - 0.00
Standard deviation of hatch date (days)	0.4 - 0.5	0.4 - 0.5	0.5 - 0.5

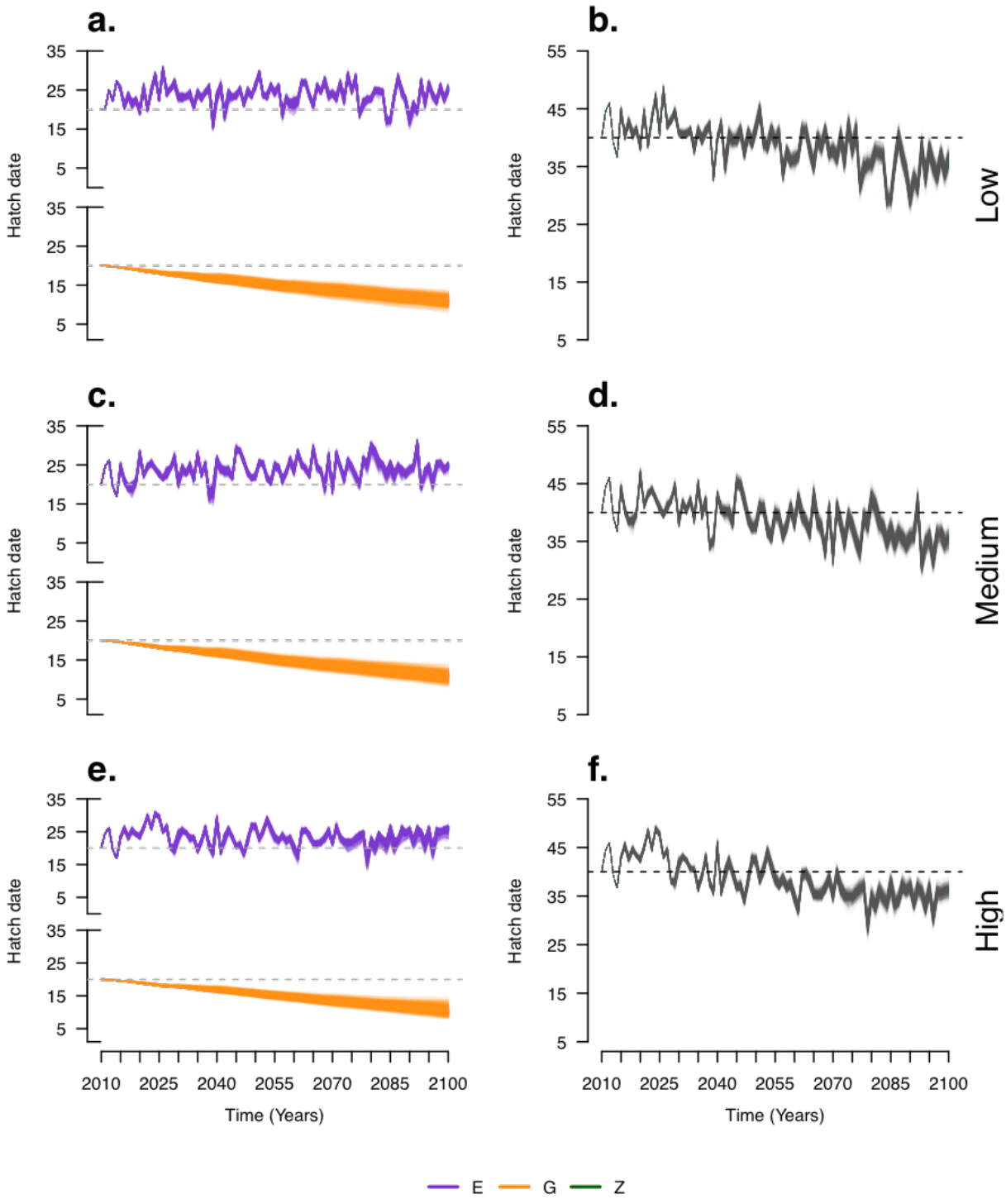

Figure 5.5: Predicted dynamics of the genetic (G) and environmental (E) component of the phenotype (Z) from 2010 to 2100.
a. and b. for low emissions scenarios, c. and d. for medium emissions scenarios, and e. and f. for high emissions scenarios. Horizontal dotted line indicates 2010 value.

5.4. Discussion

5.4.1. Great tit hatch dates advance under climate change, driven by both phenotypic plasticity and evolution

All scenarios tested in this study (low, medium, and high greenhouse gas emissions scenarios and both coupled and decoupled environmental cues) resulted in an advance of great tit hatch dates. Rates of advance varied from on average 0.01 to 0.24 days per year, with the fastest rates achieved for the coupled, high emissions scenario (scenario three, Table 5.2). Advances in hatch dates were shown to be predominantly driven by changes in spring temperature. A halting of hatch date advance could be driven by holding spring temperatures within the observed range (Figure 5.3). This suggests that spring temperatures are a key cue or good proxy cue for the timing of hatching in great tits, supporting previous analyses on the drivers of reproductive phenology in this population (Husby et al. 2010; Perrins 1965b; Perrins & McCleery 1989; Charmantier et al. 2008).

The advances shown in our results were driven by a mixture of phenotypic plasticity and micro-evolutionary change. Previous work on great tit responses to climate change has suggested that despite strong selection for earlier breeding, low heritability of phenological traits leads to slow evolutionary change (Gienapp et al. 2008; Teplitsky et al. 2009; Hoffmann & Sgrò 2011; Anderson et al. 2012; Vedder et al. 2013). In contrast, our results demonstrate that while phenotypic plasticity plays an important role in driving inter-annual fluctuations in hatch dates, the majority of directional phenotypic change is driven by evolutionary change. This evolutionary change can generate advances visible even within this century - approximately 90 great tit generations (Figure 5.2, Figure 5.5). The importance of evolutionary change in hatch date advance is highlighted by the decoupled cue scenarios (Figure 5.4). In

these scenarios, phenotypic plasticity cannot act to advance hatch dates as the cue used by great tits does not experience climate change. However, in these scenarios we still found an advance of hatch date. This was driven entirely by evolutionary change likely resulting from selection for maintenance of synchrony as caterpillar peak abundance advances. The evolutionarily driven hatch date advance in the decoupled scenario was considerably slower than the rate of advance seen in the coupled cue scenarios. This indicates that it is the combination of phenotypic plasticity and micro-evolution that produces hatch date advances sufficient to maintain synchrony for the majority of this century. It could also indicate that non-adaptive plasticity occurs in the decoupled scenario. If mismatch were to increase as existing plasticity drives phenotypes further from the new optimum, this could accelerate evolutionary change (Robinson & Dukas 1999; Badyaev 2005).

5.4.2. Synchrony between great tit hatch dates and caterpillar peak abundance is only retained if environmental cues are shared

Results of our model simulations showed that optimal synchrony was reduced even when following a coupled cue scenario (Figure 5.1). While synchrony values in our simulations do fluctuate around zero, an exact match with caterpillar peak abundance is not the optimal timing for great tit hatch dates. Maximum reproductive success is reached for hatch dates around 13 days prior to caterpillar peak abundance (Simmonds et al. 2017). Indeed, mean synchrony across the observed dataset shows hatch dates 12.5 days prior to the caterpillar peak. The shift away from optimal synchrony timing indicates some mismatch arising between great tits and their spring food source. This suggests the similar level of environmental sensitivity in both winter moth caterpillars and great tits is not sufficient to fully keep pace under changing climatic conditions. Increasing mismatch is particularly clear in the highest emissions scenario and increases towards the end of the 21st century.

Failure to retain synchrony under prolonged climatic change could be driven from three sources; limits to phenotypic plasticity (Visser 2008; DeWitt et al. 1998; Auld et al. 2009), an erosion of additive genetic variance, or different environmental responses from interacting species (Gienapp et al. 2014). Existing phenotypic plasticity, which evolved to cope with inter-annually fluctuating conditions may become non-adaptive when pushed into novel environmental conditions (Ghalambor et al. 2007). Similarly, persistent directional selection can erode genetic variance and slow evolutionary change (Van Straalen & Timmermans 2002) but can also generate new mutations (Van Straalen & Timmermans 2002). Either or both of these processes could lead to a slowing of hatch date advancement towards the end of the century. It has also been proposed that inherent differences in the responsiveness of predator and prey species to temperature changes can drive mismatch, even when correlated cues are used by interacting species (Gienapp et al. 2014). Our results appear to support this proposal and indicate even when identical cues are shared, mismatch can still arise.

In contrast to the coupled scenarios, the decoupled cue scenarios showed early and more dramatic onset of mismatch in all emissions scenarios. In the decoupled scenarios, hatch dates advance at a slower rate as the cue is not advancing, the change we do see occurs from genetic change. However the rate of micro-evolution alone is not enough to keep pace with a plastic prey species. Although our findings show a level of mismatch arising in all scenarios tested, persistent mismatch does not occur until after 2070, showing genetic change does succeed in retaining some synchrony for several decades. In particular the combination of phenotypic plasticity

and micro-evolutionary change in the coupled scenarios holds synchrony at a plateau for the majority of the 21st century.

5.4.3. Synchrony, winter conditions, and spring precipitation are key drivers of population change

For scenarios one to three, with coupled cues (Table 5.2) the resident breeding female population size increased across the 21st century. These increases in population size are driven by combinations of increases in winter conditions, and spring precipitation. If any one of these variables is held at observed levels then we do not see the same level of population growth (Figure 5.3). Spring temperatures appear to play little role in the population dynamics of these simulations, despite its strong influence on hatch dates. This results from low effect sizes in the survival and recruitment functions (see supporting information, Chapter 4). Population increases are predicted for the study population under future climate change. However, milder winter conditions lowering energetic costs and potentially increasing food supplies could act to reduce overwinter mortality (Perrins 1965b; Van Balen 1980). This combined with increased spring precipitation elevating spring food supplies, could quickly generate substantial population growth. These advances occur despite low level mismatch occurring in the population, suggesting thresholds of mismatch may not yet have been passed (Chapter 4).

Population sizes of the number predicted by the coupled scenarios are greater than any resident breeding female population size recorded in Wytham Woods (maximum recorded, 306 individuals). Typically the resident population of Wytham Woods comprises around 88 % of the total population, with the additional 12 % (mean proportion) being immigrants. The proportion of immigrants in the population varies

from less than 5 % to around 50 %. If the same proportion of immigrants was retained up until 2100, where predictions frequently just exceed 306 breeding females, population sizes could reach up to around 460 total breeding females. These levels are 244 % greater than the current mean population size of 188 breeding females. Whether these levels would actually be reached is uncertain and extrapolating current proportions of the population can be problematic. It is highly likely that rapid increases in the resident population size of Wytham Woods, would have knock on impacts to the proportion of immigrants to the population. However, the exact form of such alterations cannot be predicted from the current modelling format. Further uncertainties arise the further into the future we simulate. Aside from a consideration of temporal change in the winter moth caterpillar lifecycle, changes to other species are not incorporated into this modelling framework. Climate change will undoubtedly also influence other species within Wytham Woods, be these competitors of the great tit (blue tits (*Cyanistes caeruleus*)), predators (Sparrow hawks (*Accipiter nisus*)), or food resources (beech mast, and invertebrates). Alterations to any of these components of the system could act to either increase or decrease the carrying capacity of the great tit population. Any of these processes could moderate the population growth predicted by our model and should be considered when interpreting model results. However, to determine if they could actually occur and the exact impact they may have, further models including other elements of the interacting system are required.

In strong contrast to the population increases predicted from the coupled cue scenarios, the decoupled scenarios showed population plateaus, or even declines. The decoupled scenarios still experienced amelioration of winter conditions and increased in spring precipitation. However, in this case, population increases did not

accompany these environmental changes. Here fitness impacts of increased mismatch between great tit hatch dates and winter moth caterpillar peak abundance cancelled out mortality reductions from positive climate changes. This resulted in a stable, rather than an increasing, population size that in some cases goes extinct. Population declines resulting from mismatch have been observed in some systems (Plard et al. 2014; Both 2010), however, they are not always the case. Some populations appear to be buffered from severe population declines through processes such as density-dependence or compensatory responses to environmental conditions (Gienapp et al. 2014; Grøtan et al. 2009; Perrins 1965b; Van Balen 1980). Our study population of great tits does appear to be partially buffered from mismatch induced population declines, with only 218 of 1000 high emissions scenario simulations seeing population extinction. This buffering would be expected based on the results found in Chapter 4. Synchrony became the key driver of population dynamics beyond certain thresholds of mismatch. These potentially only being reached in some of the highest emission scenario simulations.

5.4.4. Conclusions

Here we have considered not only the direct impact of phenology and phenological synchrony on the population dynamics of a predator species, but also incorporated the influence of key environmental drivers on demographic rates. This holistic approach has yielded insights into the interplay between different environmental drivers in conjunction with phenological change. Such approaches are required in order to generate accurate predictions of the future population dynamics. Only under the highest emissions scenarios and decoupled cues do we predict the study population of great tits might experience population declines and extinction. Generally the population appears resilient to climate change over this century.

5.5. Acknowledgements

We are grateful to all of the Wytham fieldworkers who collected population census data on the Wytham great tits. This work was supported by NERC grant NE/K006274/1 to Ben Sheldon.

CHAPTER SIX

Cue identification in phenology: a comparison and contrast of our current statistical toolkit

Emily G. Simmonds, Ben C. Sheldon, and Ella F. Cole

6.1. Abstract

Changes in the timing of life history events are a widespread consequence of climate change. However, the degree of alteration varies between species due to different environmental sensitivities, leading to temporal mismatch between trophic levels. To predict population resilience under climate change, we need to predict how timing will change. This requires identification of the environmental cues (in temperate regions, primarily temperature) used to determine phenology. The most common cue identification approach is to test candidate cues using regression-based or mechanistic models. However, the usefulness of these approaches for identifying precise biological cues, or generating meaningful predictions of timing, has not been assessed. Here, we compare and contrast the predictive ability of five commonly applied statistical methods for phenological cue identification. We explore how the timing and aggregate statistic of the identified cue changes based on the method used, the number of years of data included, and the timing of this data relative to present. We do so by applying all methods to the same model dataset of lay date timing of wild great tits (*Parus major*) from Wytham Woods. The predictive capacity of each method was assessed for a five-year test dataset and the robustness of predictive capacity to sample size was tested. We show that the critical window identified differed considerably depending on the method, as well as the timespan and length of the data period used. The five methods inconsistently and inaccurately predicted lay dates, with predictions generally lagging one to fifteen days behind observed timings. These findings suggest that the current toolkit does not very accurately capture the biological cue used by great tits to time their breeding. We discuss how such methods could be improved in order to identify precise biological cues and generate accurate predictions. While the tools available can be useful for explaining patterns of phenological change, using them predictively requires caution.

6.2. Introduction

Rapid climate change drives changes in populations and species by shifting the timing of annual peak resource availability. Species can respond to these changes, either by plastically altering the timing of energetically demanding life history events, or through evolutionary change. Phenotypic plasticity has often been indicated to be the primary driver of interannual phenological change (Charmantier & Gienapp 2014; Gienapp et al. 2014; Charmantier et al. 2008; Vedder et al. 2013; Brommer et al. 2008). In temperate regions, changes in temperature are a key driver of plastic changes in phenology. Species either respond directly to temperature changes, or use closely related cues as a proxy to predict optimal timing. As a result, temperature sensitivity differs between species, leading to the potential disruption of interspecific interactions, such as temporal mismatch between different trophic levels. Such mismatches have close links to individual fitness (Visser et al. 2006; Reed, Jenouvrier, et al. 2013), with poor matching reducing survival and reproductive success, and potentially having implications for population resilience.

Predicting exactly how phenology will change is an important component in assessing population resilience into the future. These predictions can indicate where trophic mismatches might occur and therefore highlight populations that may be at risk in the future. Before predictions can be generated, however, it is necessary to build an understanding of how species time their life history events. Variation in the timing of many life history events has been linked to temperature variables (for example; (Visser et al. 2009; Visser et al. 2004; Husby et al. 2010; Charmantier et al. 2008; Thackeray et al. 2016; Lawson et al. 2015; Ardia et al. 2006; Menzel et al. 2006)). However, identifying the exact temperature cues used to time events, is challenging. Many species – particularly endotherms - largely do not have direct

physiological relationships with the temperature (McNab 2012; Khaliq et al. 2014), instead being thought to use combinations of photoperiod and/or environmental conditions as indirect cues to begin a particular life history event (Thackeray et al. 2016; Schaper et al. 2012; Phillimore et al. 2013; Visser et al. 2006). The only way to determine causal links between a proposed environmental cue and phenology is to conduct experiments, preferably in the wild. However, experiments in the majority of wild animal systems are often logistically, and ethically, impractical. Wild experiments to alter animal phenology have been attempted for cavity breeding birds ((Nager & van Noordwijk 1992), Chapter 3), but with little success. Captive experiments, such as in temperature controlled aviaries, have offered an alternative for hard-to-manipulate natural systems (Schaper et al. 2011; Schaper et al. 2012; Lambrechts et al. 1999; Visser et al. 2009). These experiments have succeeded in advancing laying through the manipulation of temperatures, indicating a potentially causal relationship between certain temperature patterns and laying (for example, temperature (Visser et al. 2009) and more specifically temperature gradient (Schaper et al. 2012)). However, the cues and patterns identified in these studies, can be contrary to those observed in natural systems (Lambrechts et al. 1999) and consequently also require testing in natural environments.

For the majority of long-term phenological studies, the only viable tool for the identification of environmental cues is statistical analysis. There are many different methodological approaches that have been employed to address these questions, broadly fitting into two overarching categories: phenomenological and mechanistic (Roberts et al. 2015). Mechanistic approaches are based on biological processes. Phenomenological approaches are based on statistical associations between observed data (Roberts et al. 2015). All of these statistical assessments of

phenological cues identify correlative rather than causative relationships. This can make it difficult to tease apart actual cues from confounding variables or good proxies. This is particularly challenging as many weather variables are highly auto-correlated temporally and spatially. Within each of these overarching categories, there are a multitude of different methods employed. Some of the most common are the phenomenological sliding time window and smooth function based methods, and the mechanistic growing degree day (GDD) method. Other, Bayesian, approaches also exist (Hudson 2010); however these are not so widely used and are therefore not discussed here.

Sliding time window analyses are widely applied (van de Pol et al. 2016; van de Pol & Cockburn 2011; Phillimore et al. 2013; Perrins & McCleery 1989; Husby et al. 2010) regression based approaches (Hudson 2010). They statistically identify the critical time window in which temperature, or any other environmental variable, can explain the most variance in phenology. They can take the form of either simple linear, or more complex multivariate and stepwise, regressions. Typically, the explanatory variable in the regression is some measure of temperature for instance mean, minimum or maximum across a temporal window. The response variable is the phenological variable of interest. The duration and temporal position of the window can be altered and explanatory power of each is compared. The temporal position of the window is either tied to a reference Julian day (for example, days prior to 20th May) or to the phenological event timing itself (days prior to event), these approaches are absolute sliding time window analysis (SWA) and relative sliding time window analysis (SWR), respectively (van de Pol et al. 2016). Absolute time windows vary in the annual lag between the window and the event, whereas relative time

windows occur at different Julian days each year but remain fixed relative to the phenological event.

Smooth function methods have been more recent introductions into the phenological toolkit (Roberts 2008; Roberts et al. 2015; Thackeray et al. 2016). They are also regression based statistical analyses but employ smoothing functions and penalties to generate sensitivity profiles (Roberts 2010). These analyses involve the consideration of the influence of all days of temperature during a year rather than bounded windows. Mean temperature from a single Julian day is iteratively regressed against the phenological variable of interest, beginning at a reference Julian day and moving backwards in time. This generates 365 coefficients which are then smoothed, using penalties for difference, to create an annual profile of temperature influence. These two processes can either be done concurrently, as in the p-spline signal regression method (PSR) (Roberts 2008) or in two steps using a generalised additive model (GAM) on the coefficient values to create climate sensitivity profiles (CSP) (Thackeray et al. 2016). Once coefficients are smoothed, it is possible to tease apart the most influential days of the year, based on the effect size and amount of variance explained.

GDD models are based on the assumption of linear relations between temperature and development due to enzyme activity (Bonhomme 2000). Any temperatures exceeding a particular threshold contribute to development, the influence of each degree over the threshold is cumulative until a second threshold is reached and the phenological event occurs. GDD models can be developed to be more complex, also including chilling requirements or being initiated by photoperiod. This model is used extensively for plant species (Kramer 1995; Chuine 2000; Rötzer et al. 2004), which

have development which is directly linked to temperature. They can also be applied to animal systems, however, endotherms tend to have fewer direct developmental links to temperature (Khaliq et al. 2014; McNab 2012).

The commonly applied methods discussed above are not an exhaustive list of the existing cue identification toolkit, however, they do provide an insight into the diversity of statistical approaches which currently exist. A few previous studies have compared the performance of different methods in capturing environmental sensitivity (Phillimore et al. 2013; Roberts et al. 2015), finding largely congruent results across different methods. Similar cues were identified using GDD, SWA, and PSR (Phillimore et al. 2013; Roberts et al. 2015). While predictive potential was inferred in these studies through R^2 values, the predictive capacity and accuracy of different methods has not yet been compared.

A fundamental assumption consistent across the phenological statistical toolkit is that the environmental cues which drive phenology remain consistent across time. Many studies of phenology in long-term systems continue to use the same cue identified years previously to inform later analyses (Charmantier et al. 2008; Visser et al. 1998; Visser et al. 2006). Alternatively, cues are defined using all data available or without consideration of the temporal position of the data (Perrins & McCleery 1989; Husby et al. 2010; Phillimore et al. 2013; van de Pol et al. 2016; Thackeray et al. 2016). Both approaches assume that the cue, or the relationship between a proxy and timing, does not change over the time period considered. This fails to account not only for the potential of changing cues over time but also for the influence of sample size on the cues identified.

In this study, we test five of the most commonly used statistical methods (SWA, SWR, CSP, PSR, and GDD) for cue identification using a single dataset of great tit (*Parus major*) laying dates spanning 55 years. Our data are taken from a long-term breeding population study from Wytham Woods. Clutch initiation timing, hereafter lay date, has been extensively studied in this population and strongly linked to spring temperature (Charmantier et al. 2008; Perrins & McCleery 1989; Husby et al. 2010; van Noordwijk et al. 1995). Two previous attempts at cue identification have occurred for this population, though neither has been exhaustive. Perrins and McCleery (Perrins & McCleery 1989) conducted an analysis on 29 years of data. They identified a critical window of the 1st March to the 25th April, from a trial of two temporal windows (Perrins & McCleery 1989). Both warmth-sum (sum of mean of daily minimum and daily maximum temperatures) and sum of daily maximum temperatures were tested, with the latter providing the highest explanatory power (Perrins & McCleery 1989). Husby et al. (Husby et al. 2010) conducted a more extensive sliding time window assessment on 33 years of data. They tested windows from 1st January to 30th April spanning 10 to 120 days in length, all for mean temperature, and found 15th February to 25th April was the best predictor. Despite the similar methodologies and data span of these two studies on the same system, the identified cues differ by greater than two weeks and in the temperature aggregate statistic. Further exploration of the influence of methodological approach on the precise temperature cue identified and in predictive ability is required.

In this study we address the following specific objectives, all with the response variable of annual mean lay date for the population:

1. To compare and contrast cues identified by five commonly used methods (SWA, SWR, CSP, PSR, and GDD).

2. To assess how these identified cues vary depending on the length of the dataset, and the precise time period used.
3. To compare the variance in timing explained by the five methods.
4. To use the cues identified by the different methods and lengths of dataset to predict phenology for a five year test dataset of observed data.

6.3. Methodology

6.3.1. The study system

The Wytham Woods great tit population study has been conducted in a standardised way since 1960 (Perrins 1979). Annual breeding season censuses of the next box breeding population are conducted by a team of fieldworkers. Weekly nest box checks from the beginning of April provide data on nest stage, number of eggs, and the onset of incubation. Lay date is determined by assuming a lay rate of one egg per day, therefore it is possible to count backwards from the number of eggs counted on the weekly check to obtain an initial lay date. Species is confirmed by weighing eggs, when at least three are present, and in the vast majority of cases by subsequently seeing the parents.

6.3.2. Environmental data collection

Temperature data were collected by the MET Office, National Climate Information Centre gridded daily data (grid point 447500E 202500N) (MET Office 2009). Daily mean temperatures were recorded in weather stations across the UK, with spatial and temporal interpolation for missing data. Environmental data are available from 1960 to 2015.

Daily rainfall totals (mm) were collected by the Radcliffe Meteorological Station, Oxford (Radcliffe Meterological Station n.d.).

6.3.3. Statistical analyses

6.3.3.1. The cue identification methods

The response variable used in all the below analyses is the mean annual lay date for the Wytham Woods great tit population.

Absolute sliding time window (SWA)

The SWA analysis is run using the R package 'climwin' (van de Pol et al. 2016; Bailey & De Pol 2016). This package was designed to standardise the process of fitting sliding time windows to phenological data. It allows for exhaustive exploratory analyses testing windows of durations from 1 to 365 days and spanning the whole year preceding an event and across a variety of aggregate statistics, reducing bias introduced by *a priori* choices. For this analysis we use a reference day of the 20th May, windows are allowed to vary in length from a single day up to 365 days, aggregate temperature statistics tested are the mean, minimum, maximum, and slope gradient of temperature across each window. The 'best' temporal window and aggregate statistics are chosen based on their AICc compared to the baseline model, an intercept only linear regression of annual lay date means.

Relative sliding time window (SWR)

SWR analyses also use the 'climwin' package in R. For this analysis no reference day is used as the temporal windows are referenced to the phenological event itself. In the same way as for SWA analyses the time windows are allowed to vary in length from one to 365 days, we tested aggregate temperature statistics of the mean, minimum, maximum, and slope gradient of temperature across each window. The 'best' temporal window and aggregate statistics are chosen based on their AICc compared to the baseline model, an intercept only linear regression of annual lay date means.

Climate sensitivity profile (CSP)

The CSP technique was first introduced by Thackeray (2016). Here we follow their general methodology. The reference calendar day for this analysis remains consistent with SWA, being the 20th May. We also use annual mean lay date to remain consistent with other methods trialled here, whereas this method is proposed to use the date on which 95% of the individuals have initiated laying. We begin with mean temperature for the 20th May and regress this against the annual mean lay dates (using the lm function in R), we repeat this iteratively going backwards in time up to 365 days prior to 20th May. For each regression we save the coefficient value and the R^2. These coefficient and R^2 values are then each passed through a GAM using the R package 'mgcv' (Wood 2001; Wood & Wood 2015) to smooth the values across time. These smoothed functions are used to define the calendar days of greatest influence on phenology. This period is identified as a period of consecutive days on which the coefficient and the R^2 values exceed the lower and upper quantiles, defined as greater or equal to the lowest 2.5 % and highest 97.5 % following the method used in Thackeray (Thackeray et al. 2016).

P-spline signal regression (PSR)

PSR was introduced by Roberts (Roberts 2008). This method works on a similar principle to the CSP methodology, however, in this analysis the GAM is run on the raw annual mean lay date data combining the coefficient and smoothing elements, therefore not producing daily R^2 values. To run the PSR we run a GAM 'mgcv' R package, with degree of difference and B-splines set as in Roberts (Roberts 2008). The GAM is run on the raw phenology data and climate data indexed to the reference date of 20th May. The resulting smoothed coefficients are used to define calendar

days of greatest temperature influence, in order to use these for prediction. These were defined as the coefficient values which exceeded the mean coefficient value ± two times the standard deviation of coefficients, representing the extreme 5 % tails of the coefficient distribution.

Growing degree day (GDD)

The GDD model we use here is a simple three-parameter thermal-time model, used in Phillimore (Phillimore et al. 2013). It does not include any chilling requirement, which is unlikely to play a developmental role for great tits. The parameters used in this model are; start date, minimum threshold temperature, and the cumulative GDD requirement. Mean daily temperatures are recorded from the start date onwards, every degree above the minimum threshold temperature is cumulatively summed until the cumulative GDD requirement is reached, at which point the phenological event is predicted to occur. These parameters were optimised to minimise the sum of squared differences between the predicted annual mean lay date and the observed mean lay dates using a simulated annulating optimiser in the 'GenSA' R package (Xiang et al. 2013). A wide area of parameter space is searched with bounds for each parameter of; start dates from 90 to 365 (year day on which temperature starts being counted), minimum temperature from 4 ºC to 8 ºC, and cumulative GDD requirement of 50 ºC to 120 ºC.

6.3.3.2. Identification of critical time windows and aggregate statistics

Each of the study methods detailed above were run for various subsections of the 55-year dataset, retaining the most recent five years (2011-2015) as a test dataset for predictive analyses. The remaining 50-year training dataset was split into three groups; the whole 50 year training dataset (1961-2010), the first 25 years (1961-

1985), the second 25 years (1986-2010). This subdivision of data allows exploration of the influence of the time period of data relative to present on cue identification, distinguishing if cues might change over time. For each method and data subsection combination the 'best' cue was identified following the appropriate optimisation procedure for each method. The length, position, and aggregate statistic of the cue recorded.

6.3.3.3. Estimation of cue-lay date relationships

Linear models (LM) were run with annual mean lay date as a response variable and the cues identified in 6.3.3.2 as the explanatory variables, for the phenomenological methods (SWA, SWR, PSR, and CSP) the cue identified is a temperature variable, for the GDD model the cue identified is the date on which the cumulative GDD requirement is reached. For each of the LMs the adjusted R^2 value was calculated as an indication of the amount of variance in the annual mean lay date that has been explained by the focal cue.

6.3.3.4. Prediction of lay dates using identified cues

Annual mean lay dates were predicted for a test dataset of observed lay dates from 2011 to 2015. Only four methods were used predictively in this study because the SWR cue is defined based on the timing of the phenological event, to identify the cue, the event must have already occurred. As a result, prediction using this method was not possible. For the remaining phenomenological methods (SWA, PSR, and CSP) predictions of annual mean lay date timing were generated from LMs of the identified cues against phenology, based on the observed temperature values from 2011 to 2015. For the GDD model the temperature data from 2011 to 2015 were passed through the GDD equation and to identify the date in each year on which the cumulative GDD requirement was reached.

To tease apart the influence of dataset length and time period of data on the accuracy of predictions, different subsets of the training dataset were used. For each predictive method we used nine different datasets covering different lengths and time periods beginning with the whole training dataset (1961-2010) and decreasing in 10 year increments in both directions (i.e. 1971-2010 and 1961-2000), down to two 10 year datasets of the earliest and latest decade. This bi-directional variation in data length allowed the influence of years of study to be distinguished from the time period that the dataset covers. The temperature cues for each of these periods were identified following the same methodology as 6.3.3.2.

To distinguish the accuracy of the phenological predictions and compare across different methods and dataset combinations, the mean squared error (MSE) was calculated for each set of five predicted years. The MSE is the squared discrepancy between the predicted lay date and the observed lay date summed across the five test years. The larger the MSE the greater the discrepancy between predicted and observed phenology.

6.4. Results

6.4.1. Critical time windows of sensitivity

The critical time windows of environmental sensitivity identified differ by the method used and the time period of the dataset. When using data from the earliest half of the long-term study (1961 to 1985) the identified critical windows were typically shorter than those identified from the latter half of the data (1986 to 2010) (Figure 6.1). The exception to this is the GDD model, which produced a longer window for the earlier data. The timing of the critical windows was broadly similar between the SWA, PSR,

and GDD methods, for all time periods and for CSP method with the exception of the earliest data. However, the SWR method identified windows many days earlier than all other methods. Both SWR and CSP showed large differences in window timing depending on the dataset used, with older data producing earlier windows. Time windows identified using the whole long-term training dataset (1961 to 2010) were typically mid-way between the early data and late data, with the exception of SWR and CSP where the window was different from both early and late. No consistent pattern can be seen of cue change based on the timing of the data used to identify the cue.

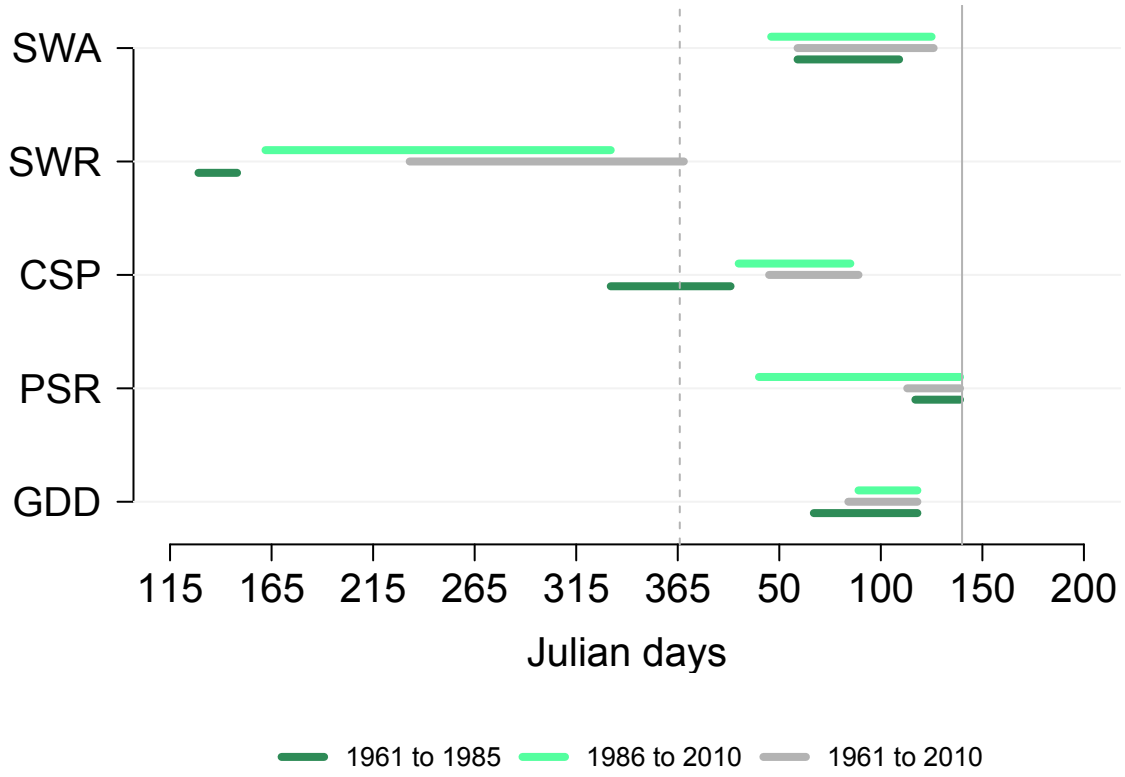

Figure 6.1: Temporal critical windows identified by different statistical methods and amounts of data.
SWA is the absolute sliding window, SWR is the relative sliding window, CSP is the Thackeray method, Spline is the Roberts method and GDD is growing degree days. Vertical dotted line shows the 1st January and solid vertical line indicates 20th May, the reference day for absolute methods (SWA, CSP, PSR).

Table 6.1: Summary statistics for linear models of identified cues and lay date for all methods.

Shows time period of data used, intercept of the regression, slope of the relationship, standard error (SE), adjusted R^2 (R^2), window open, and window close (in days prior to reference day). The minimum threshold temperature and cumulative GDD requirement are also presented.

Method	Time period	Aggregate statistic	Intercept	Slope	SE	R^2	Window open	Window close	
SWA	Whole dataset	Mean	164	-6.1	0.5	0.7	81	14	
SWA	Early	Mean	141	-2.8	0.5	0.5	81	31	
SWA	Late	Mean	151	-5.0	0.7	0.7	94	15	
SWR	Whole dataset	Mean	167	-5.2	0.5	0.7	250	115	
SWR	Early	Slope	122	-14.6	2.9	0.5	354	335	
SWR	Late	Slope	86	458	70.2	0.6	321	151	
CSP	Whole dataset	Mean	136	-3.2	0.5	0.5	95	51	
CSP	Early	Mean	125	-0.4	0.5	0.0	173	114	
CSP	Late	Mean	50	-2.8	0.7	0.4	110	55	
PSR	Whole dataset	Mean	146	-2.6	0.8	0.2	27	1	
PSR	Early	Mean	122	0.1	1.0	0.0	23	1	
PSR	Late	Mean	149	-4.4	0.8	0.5	1	100	
								ºC min	**ºC tot**
GDD	Whole dataset	Sum	5	1.0	0.1	0.5	85	4.1	134.9
GDD	Early	Sum	53	0.6	0.2	0.4	84	4.1	149.8
GDD	Late	Sum	23	0.8	0.2	0.4	89	4.2	105.3

6.4.2. Ability of the identified cue to explain variance in annual mean lay date

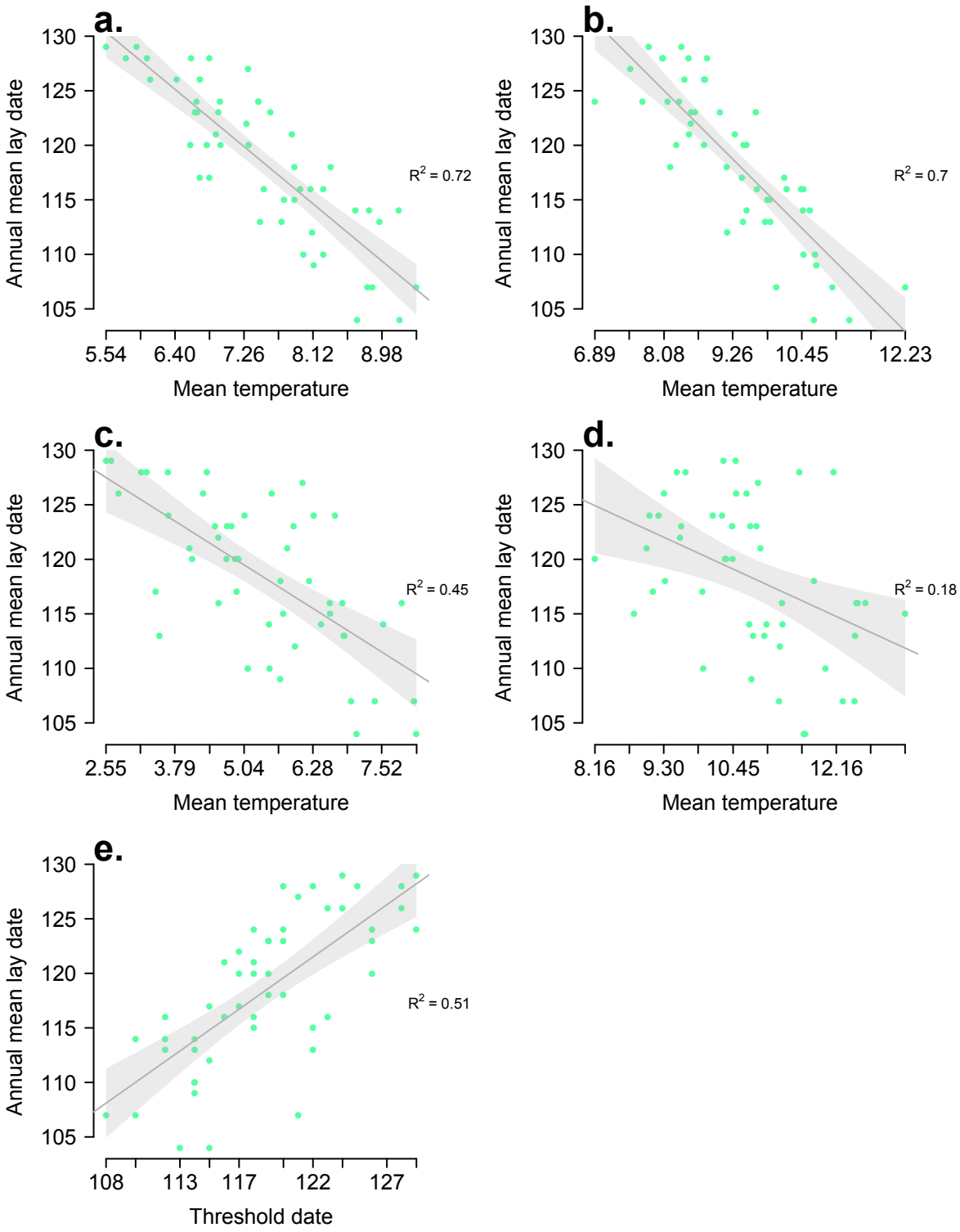

Figure 6.2: The relationships between identified cues and mean annual lay date.

Plotted lines generated from the model fit and shaded areas representing the 95% confidence intervals. Using the whole long-term dataset. a) SWA, b) SWR, c) CSP, d) PSR, e) GDD.

Relationships between the identified cue and the mean annual lay date vary from explaining 18% (PSR) to 72% (SWA) of the variance in the response variable. Despite the different lengths and aggregate statistics of the time windows identified for the different methods, the SWA, SWR, and GDD methods all explained a similar amount of lay date variance. SWA and SWR also have similar slope values. PSR explains almost no variance in mean annual lay date. The amount of variance explained by the identified cue does not vary substantially with the time period covered by the data or the method used (see Figure 6.2 and Table 6.1). All cues identified for SWA, SWR, and GDD had R^2 values of greater than 0.5. CSP and PSR had lower R^2 values, between 0.18 and 0.45.

6.4.3. The ability of different methods, dataset lengths, and time periods of data to predict future phenology

No method accurately and precisely predicted the annual mean lay dates from 2011 to 2015 (Figure 6.3). The influence of the dataset length on the mean squared error (SE) of predictions differed greatly based on the time period of data used. For datasets beginning at the most recent decade, the MSE is roughly half of that for the earliest decade. This pattern is consistent across all methods, except GDD, which retains the same MSE regardless of data length or time period. Across both time periods, the error associated with SWA predictions declined as dataset length increased over 20 years, this trend was more pronounced for the data beginning earlier. CSP and PSR methods showed a similar response in MSE to increasing data and different time periods. Both methods saw a decline in MSE for increasing amounts of data, when beginning at the earliest decade. However, this pattern was reversed when beginning at the most recent decade, with MSE increasing for 40 and 50 years of data. Largely all methods performed similarly for the most recent data but predictive differences amplified further into the past.

Figure 6.4 demonstrates that all methods tend to over predict annual mean lay dates i.e. predicting lay dates later than observed. The observed values of the test dataset (2011 to 2015) usually fall within the prediction intervals for all methods, with the exception of 2011. However, the prediction intervals for all methods are wide, spanning around 15 to 30 days, which is substantial in terms of phenology. The range of annual mean lay dates across the 50 year training dataset is 26 days, meaning the prediction intervals for half of the methods (CSP and PSR) exceed this variability, and SWA prediction intervals cover almost the entire range.

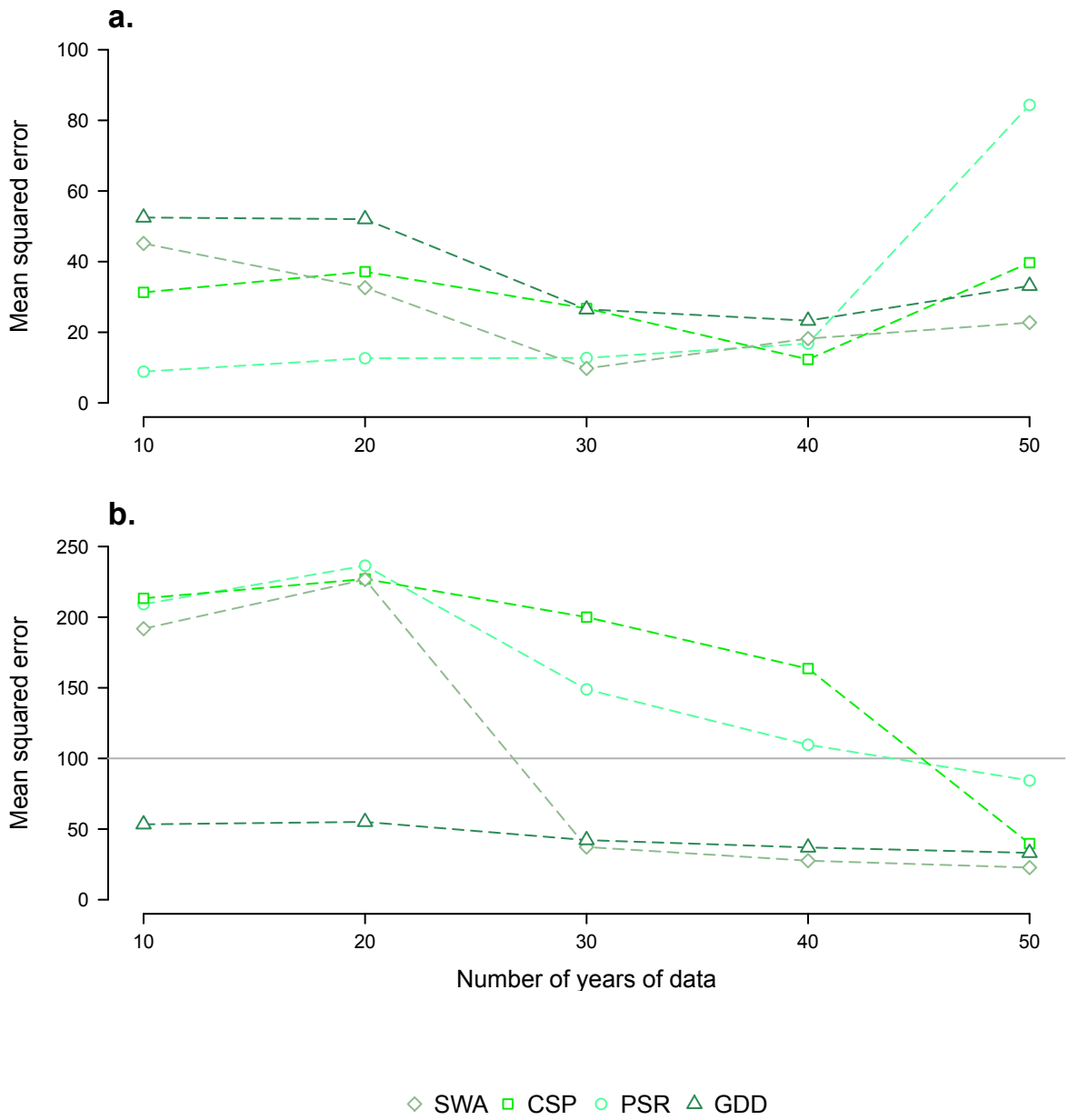

Figure 6.3: Mean squared error (MSE) of model predictions using different amounts of data.

Four predictive methods were used to generate predictions of mean annual lay dates for the period 2011 to 2015. The amount of data used to generate predictions varied from 10 to 50 years, in 10 year increments beginning at 1991 to 2010 and extending backwards in time, **a**, and in 10 year increments beginning at 1961 to 1970 and extending forwards in time, **b**. Due to difference in scales, the highest y-value from **a** is highlighted on **b** as a grey horizontal line.

Figure 6.4:Predicted and observed mean annual lay dates. Predictions generated from different methods, using the whole long-term dataset, are plotted against the observed lay dates. Vertical lines represent 95% prediction intervals, these were not available for the GDD model.

6.5. Discussion

6.5.1. Temperature cues identified differ markedly between methods and over time

We have shown here using multiple analyses on the same dataset that temperature cues identified differ by the statistical method and the time period of data used. The temporal windows of environmental sensitivity varied in their temporal position and duration (Figure 6.1). While there is some temporal overlap between the windows identified by the SWA, CSP, PSR, and GDD methods, the SWR method identified critical windows of temperature sensitivity that were 50 to 250 days earlier than other methods (except CSP using the earliest 25 years of data). The degree of temporal overlap in cues identified by different methods was highly influenced by the time period of data used to identify the cue. If only the most recent 25 years had been used, there would be largely congruent results shown between the SWA, CSP, PSR, and GDD methods, as found in previous comparisons of statistical methods for cue identification (Phillimore et al. 2013; Roberts et al. 2015). However, when different

time periods are used, such as the whole 50 year training dataset, or the earliest 25 years of data, marked between-method differences can be seen. This appears to be driven by different susceptibility of the methods to the input data used. While cues identified by the SWA and GDD methods do not experience substantial temporal shifts based on the time period used, the PSR method sees a widening of the window identified as input data moves closer to the present, and the CSP and SWR methods see a forward temporal shift of the cue. Variation based on the data used is greater for CSP and PSR than between method variation. SWR has a large discrepancy with other methods and a high susceptibility to the time period of data used. For the most commonly applied methods, SWA and GDD, the cue identified does not shift substantially with the time period of data used.

The shift in cue position with different time periods of data could stem from several sources. Change over time could indicate that the temperature cue itself is shifting temporally, potentially evolving as the climate changes. However, we found no consistent pattern across time periods of a directional shift in the cue, suggesting that evolution of the cue used is probably not the driving force behind the observed pattern. An alternative possibility is that the statistical methods we use could be identifying proxies for the actual cue used by great tits and that under climatic change the relationship between these proxies and the actual cue are shifting, leading to a change in the proxy identified as the best predictor. It could also be that none of the statistical methods we use identify either the precise cue or a consistent proxy and that the cue identified is simply the best predictor of variance in lay dates for that particular time period and that this is altered depending on the exact years included. The lack of shift in the cue for the SWA and GDD methods lends support to the idea that the cue and or proxies for the Wytham great tits are not evolving substantially.

It is striking in Figure 6.1 how different the cue identified by SWR method is compared to all other techniques, which identified very similar windows here and in previous comparison studies (Phillimore et al. 2013; Roberts et al. 2015; Hudson 2010). The SWR has been proposed (van de Pol & Cockburn 2011; van de Pol et al. 2016) as an alternative to absolute methods to allow for individual or inter-annual variation in weather conditions experienced. Instead of being fixed to a Julian day SWR methods are fixed to the phenological event, covering different calendar days each year (or for each individual) but having a fixed lag time between the cue and event. In nature, it is highly unlikely that the exact same period of calendar days e.g. May temperature, influences laying in all individuals or in all years. In a particularly warm year the mean lay date may fall prior to or within May. Consequently, the laying decision could not be influenced by this cue, even if it were to correlate strongly with it. Absolute methods can identify good proxies for the actual cue, which can even be useful for prediction, if the relationship between the cue and proxy is stable. However, they cannot identify a true biological cue. Relative windows provide some alternative to this in an attempt to access a more biologically meaningful cue by having temporally variable windows e.g. temperature in the month prior to laying would influence lay date each year.

The SWR approach has been used successfully to look at climate predictors of egg size in fairy wrens (Langmore et al. 2016). A variant on a relative approach was also successfully used to look at the timing of incubation onset relative to clutch completion (Simmonds et al. 2017); here the temperature windows were tied to clutch completion rather than a calendar date but the lag between the cue and incubation onset remained variable. For studies of phenology, however, relative

approaches should be used with caution. Due to the linear regression basis of these methods the strongest effect size, R^2, and lowest AIC values will be produced when there are is variability in the explanatory variable and this variability correlates with the response variable, i.e. high values of temperature correspond with either early or late lay dates and low temperatures with the opposite. When using relative windows, it is likely that the time period preceding the phenological event will have similar temperature values for all years, if laying commences soon after particular temperatures are reached regardless of their exact yearly timing. In this case the explanatory power of windows identified close to the lay date will be low. Only when a difference between temperatures for early and late years occurs will the explanatory power increase. This is likely to occur at periods of seasonal transition for example, the onset of spring or winter. At this point, years with early lay dates will have their relative window cross these transitions prior to later years generating strong temperature differences, a linear relationship, and potentially temperatures which can explain variance in lay date. However, what has been identified is not a cue for laying, it is simply a statistical artefact of the method being used. The location of the SWR windows identified here, around the autumn and winter onsets suggests that this might be the cause of the erroneous cues identified in this study.

Being aware of the statistical limitations of any methods used is vital for conducting such analyses. If used in a threshold rather than regression based format, relative windows might provide a suitable alternative to absolute methods, which will never identify the precise cues being used. Equally, a failure to identify a precise cue, if the proxy is good, may not be an impairment to many analyses.

6.5.2. The amount of variance in annual mean lay date explained differs by the cue used

Despite the variation in the cues identified across the different methods and time periods of data, the amount of variance explained by each method is different, some are similar despite highly divergent cues whereas other more similar cues explain different amounts of variance (Figure 6.2). GDD, SWA, and SWR all have adjusted R^2 values of greater than 0.5, with both SWA and SWR explaining around 70 % of the variation in lay dates despite their very different cues. PSR method had a similar temporal cue window to the SWA, GDD, and CSP methods but had substantially lower explanatory power. The similar, and generally very high, explanatory power across the time periods of data and methods highlights the difficulty in accurately identifying phenological cues. Correlation between different environmental variables can result in many different cues, which all explain a similar amount of variance and are themselves highly correlated. Distinguishing the actual cue used from a closely related proxy is impossible without experimental manipulation. Using metrics of model power such as AIC or R^2 may not be sufficient to distinguish between correlated cues. It is perhaps better to assess how well the identified cues perform for prediction to determine if the identified cues hold for future years or novel conditions.

6.5.3. The predictive capacity of methods improves slightly with increasing dataset length and proximity to present

The length of dataset and its proximity to present both influenced predictive accuracy of all methods Figure 6.3. The mean squared error (MSE) of all predictions is greatest for shorter datasets further from the present. As the time period of data used reaches closer to the present the MSE decreases, however, this interacts with dataset length producing the lowest MSE for 30 years of data. Further data only

improves predictions if it closer to the present, gaining more data from further into the past (Figure 6.3a) does not improve. Between method predictive ability is quite similar, GDD is the most consistent method having almost identical MSEs across all data lengths and time periods. PSR appears to be the most sensitive method to the time period and length of data, showing the greatest variability in MSE. SWA has the lowest MSE, so the greatest predictive ability when all data is used and the most recent 30 to 40 years. The cue that is identified by the SWA method, although it must be proxy rather than the precise cue, appears to capture variability in lay dates well, even into the future.

6.5.4. All methods tend to over predict lay date timing

Figure 6.4 shows that all methods used here have a tendency to over predict lay date, i.e. predict lay dates which are later in the year than observed. In addition the prediction intervals for the majority of methods (SWA, PSR, and CSP) are also as large as the entire range of lay date values over the 50 years of observed data. If any value within the range of observed data is equally as likely to be true as the predicted value this suggests the models are not fully capturing the cue-lay date relationship.

These results have implications for the use of such predictions. If predictions fail to capture the advance of phenology by over predicting lay dates, it could appear that great tits will lag behind their caterpillar prey when in reality they will keep pace. Misleading predictions with large amounts of uncertainty (due to wide prediction intervals) could paint a false picture of the future of populations. However, with the right consideration of how the underling statistical methods operate, the influence of data length and time period, and cross validation of predictions, usable predictive outputs could still be generated.

6.5.5. Conclusions

This study has demonstrated differences in the cue identification and predictive ability of five different statistical methods. None of the methods trialled in this study successfully identified the precise biological cue used by great tits to time their laying. In order to achieve this, experimental manipulations such as those by Schaper and Lambrechts (Schaper et al. 2011; Schaper et al. 2012; Lambrechts et al. 1999) need to be conducted and where possible translated into natural systems. However, statistical methods can still produce useful insights if they are used with caution and take into account the limits and drawbacks. Overall the SWA method performed best in terms of stability to dataset length and time period, explanatory power, and predictive ability. However, all methods discussed here and the majority of studies (Visser et al. 1998; Perrins & McCleery 1989; Husby et al. 2010; Charmantier et al. 2008) focus solely on the role of abiotic cues. It seems highly unlikely that biotic cues play no role in the phenology of other species. Further study into the joint role of biotic and abiotic cues may bring us closer to the actual driving factors behind phenological change.

When employing such methods caution should be employed. Careful consideration of the aim of such methods and what exactly they tell us, is needed. The methodologies discussed here can be useful for prediction but are unable to determine exact cues of phenological change. We recommend several key steps when implementing cue identification methods:

- Previously identified cues from populations should be reassessed when new data become available, removing the assumption that cues are static across time and the cues identified previously remain true (Charmantier et al. 2008; Visser et al. 2006).

- To account for the influence of data length, time period, and method used, different combinations of data and different methods should ideally be implemented in all analyses. The discrepancies between these can then be considered and errors in our estimations quantified.
- Cross validation of prediction from each trialled method and dataset should also be conducted, when enough data exists.
- More flexible approaches allowing windows that vary in length and timing across years should be explored.

Cue identification models are increasingly being used predictively ((Roberts et al. 2015; Thackeray et al. 2016; van de Pol et al. 2016; Morin et al. 2009) Chapter 4, Chapter 5, and others) consequently, consideration of the accuracy of such predictions is timely. This study has shown that phenological cue identification methods have a tendency to over predict lay dates of the Wytham Woods great tits. This predictive ability is sensitive to both the length and time period of data used. Careful consideration of the input data, as well as the statistical method used and the variability around the predicted outcomes is needed if usable predictions are to be generated.

All of the results here are generated from one study system and may not hold for other species or systems. However, our results do highlight some fundamental issues with our current toolkit. For future assessments of phenological cues more nuanced and flexible approaches may be required to take account of cues which potentially change over time.

6.6. Acknowledgements

We are grateful to all of the Wytham fieldworkers who collected population census data on the Wytham great tits. This work was supported by NERC grant NE/K006274/1 to Ben Sheldon. We are also grateful to Martijn van de Pol and Liam Bailey for their support in using their new climwin package

CHAPTER SEVEN

General Discussion: The future of phenological research

Emily G. Simmonds

7.1. How has this thesis added to the body of phenology research

The five data chapters in this thesis have shown:

- That synchrony between hatch dates in great tits and winter moth caterpillar peak are a better predictor of reproductive success than clutch initiation synchrony. This is because plasticity in incubation behaviour allows fine-tuning of hatch dates based on temperatures right up until hatching, to produce better matching with the resource species. **Chapter 2**.
- That experimental manipulations in the wild are difficult and that despite perceived resilience to disturbance, blue tits do appear to be deterred from breeding with experimental apparatus in their nest boxes. **Chapter 3**.
- That while phenotypic plasticity dominates inter-annual changes in phenology, micro-evolution plays a directional role in shifting phenology, that will be needed to maintain synchrony with winter moth caterpillars under systematic climate change. **Chapter 4**.
- Over certain thresholds of mismatch, synchrony with winter moth caterpillars becomes the key driver of population dynamics in the Wytham Woods great tits. Within these thresholds, winter conditions, beech mast index, and spring temperature can play larger roles than synchrony. **Chapter 4**.
- Under climate change scenarios forecast over the 21st Century, if cues are shared between great tits and their winter moth prey, synchrony can be maintained. In this case population size will increase due to improved winter conditions. However, if cues are not shared, mismatch will result with associated population plateaus and, under some conditions, declines. **Chapter 5**.
- That our current statistical toolkit for identifying phenological cues is failing to capture precise biological cues. This work highlights the variation between

methods and data subsets, which can lead to differing results. It has also demonstrated how these methods could be used predictively and some of the cautions that must be associated with this. **Chapter 6**.

This work has investigated the causes and consequences of phenological change, building up from understanding the levels of plasticity present in great tit breeding, to predicting under what circumstances temporal mismatch might occur and what the consequences of this will be, and to an assessment of the methods we are currently employing as a field. To achieve this aim novel data collection, experimental, statistical, and modelling techniques have all been used, furnishing me with a diverse research toolkit. My work has touched on several key issues in phenology and has made several steps to further our pursuit of answers to them. However, outstanding questions still exist.

7.2. Benefits and limitations of long-term data in phenology

Long-term datasets, those covering several decades, undoubtedly have a vital role to play in ecological research. There are copious accounts of their benefits (Clutton-Brock & Sheldon 2010; Callahan 1984; Likens 1989; Krebs 2008). Long-term datasets allow us to access important questions in ecology and evolution, which can only be answered with data spanning many years, for example, studying response to disturbances such as climate change (Lindenmayer & Likens 2009), slow processes such as the population dynamics of long-lived species (Franklin 1989), and predictive analyses which require cross validation (Lindenmayer & Likens 2009; Franklin 1989). In the field of phenology a lot of our understanding has been generated from long-term studies. Several National and International Phenology Networks and databases exist (summarised in (Koch 2010)), in addition to a growing number of long-term

studies of individual systems, covering plants (for example (Gordo & Sanz 2009)), invertebrates (for example the bay checkerspot butterfly (Singer & Parmesan 2010) (*Euphydryas editha bayensis*)), birds (for example great tits in the UK and the Netherlands (Lack 1947; Perrins 1970b; Visser et al. 1998)), and mammals (for example red deer (*Cervus elaphus*) (Clements et al. 2010) from the Isle of Rum). Such resources provide an opportunity to examine change over time, giving a baseline against which to determine if phenology is shifting. However, these datasets, such as the one used here, are not without limitations. Consideration of these limitations is perhaps too rarely undertaken but could influence findings derived from such data.

Long-term studies have an inherent bias towards populations that persist across multiple decades. By their very definition, populations that fluctuate heavily or go locally extinct will not generate sufficient data to become a long-term study. Coupled with this could be a researcher bias to choose systems which are more likely to yield stable data and from high quality habitats, excluding marginal populations. As a result, in terms of phenology, we may find that mismatch and severe population decline is less likely to occur in the systems we study in comparison to those that we do not. This is likely to be emphasised by a geographic bias that exists across all ecological studies - overrepresentation of protected areas, temperate deciduous woodlands, and wealthy countries (Martin et al. 2012). These areas do not correspond with those that are predicted to experience the most dramatic changes in climate over the next century (IPCC 2013), potentially leading to an underestimation of the influence of climate change on phenological shifts.

In addition to the general biases that occur across all long-term ecological studies, there are also more specific assumptions that can generate uncertainty and biases in data generated from these systems. I Illustrate some of these with the Wytham Woods great tit population study as an example. Temporal and spatial assumptions underpin research using long-term monitoring systems. Spatially it is known that individuals immigrate and emigrate into and out of the population. While immigrants to the population enter the dataset if they successfully hatch young in a nest box and are subsequently included in analyses, little is known about those individuals that emigrate. Individuals never recorded again are often considered as mortalities. This can bias estimates of reproductive success if newly hatched individuals disperse outside of the study area. It is very difficult to recapture these individuals, though it has been shown in other populations that fledglings from late broods are more likely to emigrate (van Noordwijk et al. 1995; Dhondt & Hublé 1968; Van Balen & Hage 1989). As a result, failure to track emigrating birds could lead us to underestimate reproductive success in late broods, consequently overestimating the costs of late breeding. However, recapture rates of greater than 0.8 for this population [19, Chapter 4] suggest permanent emigration is rare.

Further data limitations and biases can enter long-term population monitoring studies from incomplete capturing of individuals within the study area. As well as losing individuals to emigration, not all individuals that enter the study area or remain there are captured every year. For studies of nest box breeding birds, only breeding attempts within these nest boxes are recorded. However, there are some birds that will breed each year in natural cavities rather than nest boxes and these individuals will be missed from the population census. There have been several studies that have inferred from recapture rates (Lachish et al. 2011; Kidd et al. 2015; Bouwhuis et

al. 2010) that the majority of blue tits and great tits use nest boxes to breed in Wytham Woods. If it is a particular subset of the population, which breed outside of nest boxes this could add a bias to population estimates, for example, assuming a lower proportion of immigrants than is actually present (Kidd et al. 2015). As the proportion of nests where a female is successfully identified is very close to the recapture rate, it appears as if those individuals not using nest boxes is a fixed and low proportion of the population (Kidd et al. 2015; Perrins 1979; Bouwhuis et al. 2012). Whether this group of individuals follow the same phenological patterns as the nest box breeding population, is not known.

One source of bias in the Wytham Woods study that has already been quantified, is the sampling bias generated by individuals that fail early in breeding attempts (Kidd et al. 2015). Following the standardised protocol (Perrins 1965c), breeding adults are not identified until nestlings are six (for passive radio frequency identification) to eleven (for catching of adults using spring loaded traps) days old. Therefore, parents from nesting attempts that fail prior to this stage will not be identified. It has been shown that females failing early are more likely to be immigrants to the population (Kidd et al. 2015), and that failure to account for this leads to overestimation of the reproductive success of immigrants. All of the sources of potential bias mentioned above are specific to the study system used in this thesis, though they also might apply to other long-term studies of cavity breeding birds. Sampling biases similar or different to those discussed here will be present across all long-term monitoring studies and consideration of their influence on conclusions drawn needs to be more widely considered (Hadfield 2008). If a particular subset of your population more readily avoids capture (Kidd et al. 2015) or is more often missing data (Nakagawa & Freckleton 2008), conducting analyses without these individuals will increase

estimation bias (Nakagawa & Freckleton 2008). However, dealing with these issue is a difficult challenge. For the Wytham Woods population at least, studies such as Kidd et al (Kidd et al. 2015) make a good start to quantifying groups of missing individuals. With advances in RFID (Radio-frequency identification) readers it may be possible to identify natural cavity breeders if they are caught at other times in the year. This combined with new ways to handle missing data, such as multiple imputation (Nakagawa & Freckleton 2008) may allow us to begin to address some of these biases.

Further to the spatial assumptions, which lead to sampling biases, temporal assumptions can also produce limitations for long-term data. A primary assumption in phenological studies is a view that studies across several decades are capturing the same system during this time. The potential for a system and the relationships and interactions within it to change over time is not usually considered. However, abiotic conditions are not temporally static, and it does not make sense to assume that biological components of a system (especially considering the influence of evolutionary processes) will be either. The potential for change over time in the cues used to determine phenology has been demonstrated in **Chapter 6**. We showed that a failure to consider these shifts can impact the accuracy of our predictions and identification of cues (**Chapter 6**). It follows that other aspects of a system, for instance interspecific interactions, could also show temporal shifts. For future analyses it is important to consider long-term systems not as static entities but as developing and evolving systems and tailor analyses accordingly.

Despite the biases and limitations detailed above it should still be emphasised that long-term studies do play a vital role in phenological research. Without them, we

would have no baseline against which to measure phenological change. Nor would we be able to explore evolutionary as well as ecological responses to climate change. However, when using such studies researchers should bear in mind their limitations and the particular biases and assumptions that are present in their systems.

7.3. Uncertainty in predictive modelling

Predictions were generated in **Chapters 4** and **5** for the future phenology and population dynamics of the Wytham Woods great tits. When creating models of natural systems, the models are always simplifications of reality. As a result, some variance will always remain unexplained, and therefore uncertainty and error will enter the model output. For predictive population models, this uncertainty accumulates at several different levels; in data and in parameter estimates (within-model uncertainty), in the choice of model used (between-model uncertainty), and from climate predictions used to direct predictions (Koo et al. 2017). Two levels of uncertainty have been accounted for in **Chapters 4** and **5**: first, within-model uncertainty through the statement of standard errors around parameter estimates, and second, some uncertainty in climate predictions via the consideration of 1000 equally likely climate realisations for three levels of emissions scenario. However, these are not the only sources of uncertainty in such models. A more detailed consideration could have been conducted, for example considering the influence of the uncertainty generated by different cue identification methods (as discussed in **Chapter 6**) and other sources of parameter uncertainty. All of these considerations quickly add to create highly complex models with wide prediction ranges. Even the limited consideration of uncertainty shown here generates a wide band of equally plausible outcomes.

The component of uncertainty that has not been considered here is the between-model uncertainty. Predictive modelling in this thesis was conducted using only one type of model; altering the type of predictive model used and the composition of the model, such as trialling a matrix population model or a two sex construction, could influence the predicted outcomes. A very clear area that has been neglected is the influence of other species, both from other trophic levels and competitive interactions. Incorporating the population dynamics as well as the timing of winter moth caterpillars into the model would give a consideration of how their dynamics will be influenced by climate change. If winter moth caterpillars fail to keep pace with changes in oak bud burst then population declines could be seen in great tits, even if they remain matched to the winter moth caterpillars. Competitive interactions could also significantly alter predictions, as discussed in **Chapter 5** the unprecedentedly high population sizes predicted for the great tit population could be modulated by competitor species such as blue tits (Stenseth et al. 2015). Incorporating the influence of other species would produce a fuller picture of the future population dynamics, although the population dynamics of each other species will come with its own uncertainty which would need to be quantified and considered. In order to test the influence of the particular variables included as well as the type of model chosen and fully disentangle the influence of between-model uncertainty, predictions from different model constructions would need to be compared.

There is currently no standardised procedure to quantify and communicate the multiple levels of uncertainty in population predictions, despite their ability to substantially alter predictive conclusions. A consideration of the major sources of uncertainty will generate a wide range of equally possible outcomes, giving a fuller

picture of possible future realisations. However, even if this is achieved there is always the possibility that the actual outcome will fall outside of the predicted range. It is not possible to capture all uncertainty in model predictions and therefore we can never be completely confident that our predictions will be realised. Given this, is there any benefit to predictive modelling? Despite their drawbacks and limits predictive models are often our only insights into the potential future of populations. For many wild systems, particularly large vertebrates, experimental manipulations of environmental conditions are not ethically or logistically possible. It is important to take the best guesses we have at the time based on our best knowledge of a system but work to constantly improve and update them as data and techniques change and improve. However, with any predictive modelling it is important to consider results in the light of their limits.

7.4. Future directions

This thesis is a step in a chain of analyses exploring the causes and consequences of phenological change, it adds to a large historical body of such work with novel approaches and it will be followed with further new advances. There are several areas of further work that could follow on directly and complement the work presented here. The first is to conduct experimental manipulations to test the limits of phenotypic plasticity across the great tit breeding cycle. In **Chapter 2** I began to explore flexibility in several components of the breeding cycle through an observational study. Experimental manipulations, particularly under novel environmental conditions, would demonstrate whether flexibility can continue indefinitely or if there are limits to current plasticity. A consideration of this during the whole breeding cycle rather than focusing on particular components is essential to determining the actual impact of novel climate changes.

Alongside these manipulations there is always a need for further experimental study into the cues used by individuals to time life history events. In particular a focus on more flexible and complex cues should be explored, including biotic as well as abiotic drivers and considering temporal changes in their relative roles. I have demonstrated in **Chapter 3** how difficult experimental manipulations in the wild can be, however, they will be necessary if we are to establish causal links between phenology and its drivers. A combination of captive experiments, building on from those conducted by Schaper (Schaper et al. 2011; Schaper et al. 2012), and wild experiments will likely be required (Nager & van Noordwijk 1992; Vedder 2012). This should also be supported by statistical analyses, which should also be formed in a flexible manner, as discussed **Chapter 6**. However, to explore the role of biotic cues we require more detailed data on bud burst of trees and winter moth phenology. In Wytham Woods, we do have detailed bud burst data from six species of tree (*Quercus robur*, *Fraxinus excelsior*, *Fagus sylvatica*, *Betula pendula*, *Corylus avellana*, and *Acer pseudoplatanus*) for two years, however, these are focused around 200 sample sites within the woodland (Cole & Sheldon 2017). Previous work has suggested that biotic cues could operate at very local scales (Hinks, Cole, Daniels & others 2015), consequently to accurately explore the role of other species we require detailed spatial data over a wider time span. Winter moth phenology is currently only recorded at a single site, four oak trees with different phenology, within Wytham Woods. As tree phenology has been shown to vary spatially in the woodland (Cole & Sheldon 2017) winter moth caterpillar phenology is highly likely to do the same. Expanding our assessment of the primary food resource phenology would be a key step to understanding the role of biotic cues, and developing detailed individual specific measures of phenological mismatch.

The final area of further development I would suggest leading on from this thesis would be an exploration and consideration of how limitations to our data and statistical techniques influence our results. Building on the analysis by Kidd et al (Kidd et al. 2015) to further quantify biases generated by missing individuals, either through observational work or statistical simulations, would allow quantification of biases and their impacts. Following this through to every step of analyses, quantifying uncertainty our statistical analyses, and in our predicted outcomes is a necessary step to achieving openness in scientific reporting and illustrating the true variability in our results.

By combining all of the areas mentioned above we may one day be able to achieve predictions of phenological change for the Wytham Woods population that include quantified relationships between phenology and the precise cues that drive it, the influence of changes in other trophic levels and interacting species, individual level estimates of phenological synchrony, all with quantified multiple layers of uncertainty. This is an ambitious aim, and may not be completely feasible but if we do not aim high, we will not get progress. However, if we can set out an ideal, we can aim to make what progress we can towards it.

Bibliography

Alvarez, E. & Barba, E., 2014. Behavioural responses of great tits to experimental manipulation of nest temperature during incubation. *Ornis Fennica*, 91(4), pp.220–230.

Álvarez, E. & Barba, E., 2014. Incubation and hatching periods in a Mediterranean Great Tit Parus major population. *Bird Study*, 61(2), pp.152–161.

Anderson, J.T. et al., 2012. Phenotypic plasticity and adaptive evolution contribute to advancing flowering phenology in response to climate change. *Proceedings of the Royal Society B: Biological Sciences*, 279(1743), pp.3843–3852.

Ardia, D.R. et al., 2009. Temperature and life history: Experimental heating leads female tree swallows to modulate egg temperature and incubation behaviour. *Journal of Animal Ecology*, 78(1), pp.4–13.

Ardia, D.R., Cooper, C.B. & Dhondt, A.A., 2006. Warm temperatures lead to early onset of incubation, shorter incubation periods and greater hatching asychrony in tree swallows at the extremes of their range. *Journal of Avian Biology*, 37, pp.137–142.

Ardia, D.R., Pérez, J.H. & Clotfelter, E.D., 2010. Experimental cooling during incubation leads to reduced innate immunity and body condition in nestling tree swallows. *Proceedings. Biological sciences / The Royal Society*, 277, pp.1881–1888.

van Asch, M. et al., 2007. Predicting adaptation of phenology in response to climate change, an insect herbivore example. *Global Change Biology*, 13(8), pp.1596–1604.

van Asch, M., 2007. *Seasonal synchronization between trophic levels under climate change*. The University of Groningen.

van Asch, M. & Visser, M.E., 2007. Phenology of forest caterpillars and their host trees: the importance of synchrony. *Annual review of entomology*, 52, pp.37–55.

Asher, G.W., 2007. Gestation length in red deer: genetically determined or environmentally controlled? Reproduction in Domestic Ruminants. *Society of Reproduction and Fertility Supplement*, 64, pp.255–260.

Asher, G.W. et al., 2005. Influence of level of nutrition during late pregnancy on reproductive productivity of red deer. Adult primiparous hinds gestating red deer calves. *Animal Reproductive Science*, 86(3–4), pp.261–283.

Auld, J.R., Agrawal, A.A. & Relyea, R.A., 2009. Re-evaluating the costs and limits of adaptive phenotyic plasticity. *Proceedings. Biological sciences / The Royal Society*, 277, pp.503–511.

Badyaev, A. V., 2005. Role of stress in evolution: From individual adaptability to evolutionary adaptation. In *Variation*. pp. 277–302.

Bailey, L.D. & De Pol, M. Van, 2016. Climwin: An R Toolbox for Climate Window Analysis. *PLoS ONE*, 11(12).

Balen, J.H. & Balen van, H., 1973. A comparative study of the breeding ecology of the great tit Parus major in different habitats. *Ardea*, 61, pp.1–93.

Van Balen, J.H., 1980. Population fluctuations of the Great Tit and feeding conditions in winter. *Ardea*, 68(1–4), pp.143–164.

Van Balen, J.H. & Hage, F.J., 1989. The effect of environmental factors on tit movements. *ORNIS SCAND.*, 20(2), pp.99–104.

Barrientos, R. et al., 2015. Facultative interspecific brood parasitism in tits: a last resort to coping with nest-hole shortage. *Behavioral Ecology and Sociobiology*, 69(10), pp.1603–1615.

Bates, D. et al., 2015. Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software*, 67(1), pp.1–48.
Bennett, N.L. et al., 2015. Geographic mosaics of phenology, host preference, adult size and microhabitat choice predict butterfly resilience to climate warming. *Oikos*, 124(1), pp.41–53.
Bonhomme, R., 2000. Bases and limits to using “degree.day” units. *European Journal of Agronomy*, 13(1), pp.1–10.
Both, C. et al., 2009. Climate change and unequal phenological changes across four trophic levels: constraints or adaptations? *The Journal of animal ecology*, 78(1), pp.73–83.
Both, C., 2010. Food availability, mistiming , and climatic change. In A. P. Møller, W. Fiedler, & P. Berthold, eds. *Effects of Climate Change on Birds*. Oxford University Press, pp. 129–147.
Both, C. et al., 2004. Large-scale geographical variation confirms that climate change causes birds to lay earlier. *Proceedings. Biological sciences / The Royal Society*, 271(1549), pp.1657–62.
Both, C. & Visser, M.E., 2005. The effect of climate change on the correlation between avian life-history traits. *Global Change Biology*, 11(10), pp.1606–1613.
Bourgault, P. et al., 2010. Spring vegetation phenology is a robust predictor of breeding date across broad landscapes: a multi-site approach using the Corsican blue tit (Cyanistes caeruleus). *Oecologia*, 162(4), pp.885–892.
Bouwhuis, S. et al., 2009. Great tits growing old: selective disappearance and the partitioning of senescence to stages within the breeding cycle. *Proceedings of the Royal Society B: Biological Sciences*, 276(1668), pp.2769–2777.
Bouwhuis, S. et al., 2010. Similar patterns of age-specific reproduction in an island and mainland population of great tits Parus major. *Journal of Avian Biology*, 41(6), pp.615–620.
Bouwhuis, S. et al., 2012. The forms and fitness cost of senescence: age-specific recapture, survival, reproduction, and reproductive value in a wild bird population. *The American Naturalist*, 179(1), pp.E15–E27.
Brinkhof, M.W.G. et al., 1993. Timing of reproduction and fledging succes in the coot Fulica atra: evidens for a causal relationship. *Journal of Animal Ecology*, 62(3), pp.577–587.
Brommer, J.E., Rattiste, K. & Wilson, A.J., 2008. Exploring plasticity in the wild : laying date – temperature reaction norms in the common gull Larus canus. *Proceedings of the Royal Society B: Biological Sciences*, (January), pp.687–693.
Bryan, S.M. & Bryant, D.M., 1999. Heating Nest-Boxes Reveals an Energetic Constraint on Incubation Behaviour in Great Tits, Parus major. , 266(1415), pp.157–162.
Buse, A. et al., 1998. Effects of Elevated Temperature and Carbon Dioxide on the Nutritional Quality of Leaves of Oak (Quercus robur L.) as Food for the Winter Moth (Operophtera brumata L.). *Functional Ecology*, 12(5), pp.742–749.
Buse, A. et al., 1999. Effects of elevated temperature on multi-species interactions: The case of Pedunculate Oak, Winter Moth and Tits. *Functional Ecology*, 13(SUPPL. 1), pp.74–82.
Buse, A. & Good, J.E.G., 1996. Synchronization of larval emergence in winter moth (Operophtera brumata L) and budburst in pedunculate oak (Quercus robur L) under simulated climate change. *Ecological Entomology*, 21, pp.335–343.
Callahan, J.T., 1984. Long-Term Ecological research. *Bioscience*, 34(6), pp.363–367.
Cao, Y. et al., 2016. Simulated warming shifts the flowering phenology and sexual reproduction of Cardamine hirsuta under different planting densities. *Scientific*

Reports, 6(27835), pp.1–9.
Chamberlain, D.E., Gosler, A.G. & Glue, D.E., 2007. Effects of the winter beechmast crop on bird occurrence in British gardens. *Bird Study*, 54(1), pp.120–126.
Charmantier, A. et al., 2008. Adaptive Phenotypic Plasticity in Response to Climate Change in a Wild Bird Population. *Science*, 320(5877), pp.800–803.
Charmantier, A. & Gienapp, P., 2014. Climate change and timing of avian breeding and migration: Evolutionary versus plastic changes. *Evolutionary Applications*, 7(1), pp.15–28.
Chevin, L.-M. & Lande, R., 2015. Evolution of environmental cues for phenotypic plasticity. *Evolution*, 69(10), pp.2767–2775.
Chevin, L.-M., Lande, R. & Mace, G.M., 2010. Adaptation, plasticity, and extinction in a changing environment: Towards a predictive theory. *PLoS Biology*, 8(4).
Childs, D.Z., Sheldon, B.C. & Rees, M., 2016. The evolution of labile traits in sex- and age-structured populations. *Journal of Animal Ecology*, 85(2), pp.329–342.
Cholewa, M. & Wesołowski, T., 2011. Nestling Food of European Hole-Nesting Passerines: Do We Know Enough to Test the Adaptive Hypotheses on Breeding Seasons? *Acta Ornithologica*, 46(2), pp.105–116.
Chuine, I., 2000. A Unified Model for Budburst of Trees. *Journal of Theoretical Biology*, 207(3), pp.337–347.
Cleland, E.E. et al., 2007. Shifting plant phenology in response to global change. *Trends in ecology & evolution*, 22(7), pp.357–65.
Cleland, E.E., Esch, E. & Mckinney, J., 2015. Priority effects vary with species identity and origin in an experiment varying the timing of seed arrival. *Oikos*, 124(1), pp.33–40.
Clements, M.N. et al., 2010. Getting the timing right: Antler growth phenology and sexual selection in a wild red deer population. *Oecologia*, 164(2), pp.357–368.
Clutton-Brock, T.H. & Sheldon, B.C., 2010. Individuals and populations: The role of long-term, individual-based studies of animals in ecology and evolutionary biology. *Trends in Ecology and Evolution*, 25(10), pp.562–573.
Cole, E.F. & Sheldon, B.C., 2017. The shifting phenological landscape: Within- and between-species variation in leaf emergence in a mixed-deciduous woodland. *Ecology and Evolution*, 7(4), pp.1135–1147.
Conway, C.J. & Martin, T.E., 2000. Effects of ambient termperature on avian incubation behavior. *Behav Ecol*, 11(2), pp.178–188.
Coulson, T. et al., 2017. Modeling Adaptive and Non-adaptive Responses of Populations to Environmental Change. *bioRxiv*.
Coulson, T. et al., 2011. Modeling effects of environmental change of wolf population dynamics, trait evolution, and life history. *Science*, 334, pp.1275–1279.
Cresswell, W. & McCleery, R.H., 2003. How great tits maintain synchronization of their hatch date with food supply in response to long-term variability in temperature. *Journal of Animal Ecology*, 72, pp.356–366.
Crick, H., 2004. The impact of climate change on birds. *Ibis*, 146(s1), pp.48–56.
Crick, H. et al., 1997. UK birds are laying eggs earlier. *Nature*, 388, p.526.
Cumming, G., 2014. The new statistics: Why and how. *Psychological Science*, 25(1), pp.7–29.
Cushing, D.H., 1969. The Regularity of the Spawning Season of some Fishes. *Journal du Conseil international pour l'Exploration de la Mer*, 33(1), pp.81–97.
DeWitt, T.J., Sih, A. & Wilson, D.S., 1998. Cost and limits of phenotypic plasticity. *Trends in Ecology & Evolution*, 13(97), pp.77–81.
Dhondt, A.A. & Eyckerman, R., 1979. Temperature and Date of laying by tits parus spp. *Ibis*, 121(3), pp.329–331.
Dhondt, A.A. & Hublé, J., 1968. Fledging-date and sex in relation to dispersal in

young great tits. *Bird Study*, 15(3), pp.127–134.
Dunn, P.O. & Møller, A.P., 2014. Changes in breeding phenology and population size of birds. *Journal of Animal Ecology*, 83(3), pp.729–739.
Dunn, P.O. & Winkler, D.W., 1999. swallows America North. *Proceedings: Biological Sciences*, 266(1437), pp.2487–2490.
Dunne, J. a, Harte, J. & Taylor, K.J., 2003. Subalpine Meadow Flowering Phenology Responses to Climate Change : Integrating Experimental and Gradient Methods. *Ecological Monographs*, 73(1), pp.69–86.
Durant, J.M. et al., 2007. Climate and the match or mismatch between predator requirements and resource availability. *Climate research*, 33, pp.271–283.
Easterling, M.R., Ellner, S.P. & Dixon, P.M., 2000. Size- Specific Sensitivity: Applying a New Structured Population Model. *Ecology*, 81(3), pp.694–708.
ESRI, 2010. ArcGIS desktop.
Falconer, D.S., 1990. Selection in different environments: effects on environmental sensitivity (reaction norm) and on mean performance. *Genet. Res., Camb*, 56, pp.57–70.
Feeny, P., 1970. No TitSeasonal Changes in Oak Leaf Tannins and Nutrients as a Cause of Spring Feeding by Wintr Moth Caterpillars. *Ecological society of America*, 51(4), pp.565–581.
Field, C.B. et al., 2014. *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects*, Cambridge University Press.
Forstmeier, W., Wagenmakers, E.J. & Parker, T.H., 2016. Detecting and avoiding likely false-positive findings - a practical guide. *Biological Reviews*, pp.000–000.
Franklin, J.F., 1989. Importance and justification of long-term studies in ecology. *Long-term studies in Ecology: approaches and alternatives*, pp.136–157.
Gaillard, J.-M. et al., 1993. Timing and Synchrony of Births in Roe Deer. *Journal of Mammalogy*, 74(3), pp.738–744.
Gallinat, A.S., Primack, R.B. & Wagner, D.L., 2015. Autumn, the neglected season in climate change research. *Trends in Ecology and Evolution*, 30(3), pp.169–176.
García-Navas, V. & Sanz, J.J., 2011. Short-Term Alterations in Songbird Breeding Schedule Lead to Better Synchronization in Food Availability. *The Auk*, 128(1), pp.146–155.
Gelman, A. & Loken, E., 2014. Data-dependent analysis—a “garden of forking paths”— explains why many statistically significant comparisons don’t hold up. *American Scientist*, 102(40). Available at: https://www.americanscientist.org/article/the-statistical-crisis-in-science.
Ghalambor, C.K. et al., 2007. Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Functional Ecology*, 21(3), pp.394–407.
Gibb, J., 1950. The breeding biology of the Great and Blue Titmice. *Ibis*, 92, pp.507–539.
Gienapp, P. et al., 2008. Climate change and evolution: Disentangling environmental and genetic responses. *Molecular Ecology*, 17(1), pp.167–178.
Gienapp, P., Hemerik, L. & Visser, M.E., 2005. A new statistical tool to predict phenology under climate change scenarios. *Global Change Biology*, 11(4), pp.600–606.
Gienapp, P., Postma, E. & Visser, M.E., 2006. Why breeding time has not responded to selection for earlier breeding in a songbird population. *Evolution*, 60(11), pp.2381–2388.
Gienapp, P., Reed, T. & Visser, M.E., 2014. Why climate change will invariably alter selection pressures on phenology. *Proceedings of the Royal Society B: Biological Sciences*, 281(20141611), pp.1–8.

Good, I.J., 1988. The Interface Between Statistics and the Philosophy of Science. *Statistical Science*, 3(4), pp.386–412.
Gordo, O. & Sanz, J.J., 2009. Long-term temporal changes of plant phenology in the Western Mediterranean. *Global Change Biology*, 15(8), pp.1930–1948.
Griffith, S.C., Owens, I.P.F. & Thuman, K.A., 2002. Extra pair paternity in birds: A review of interspecific variation and adaptive function. *Molecular Ecology*, 11(11), pp.2195–2212.
Grotan, V. et al., 2009. Spatial and temporal variation in the relative contribution of density dependence, climate variation and migration to fluctuations in the size of great tit populations. *Journal of Animal Ecology*, 78, pp.447–459.
Grøtan, V. et al., 2009. Spatial and temporal variation in the relative contribution of density dependence, climate variation and migration to fluctuations in the size of great tit populations. *Journal of Animal Ecology*, 78, pp.447–459.
von Haartman, L., 1969. *The nesting habits of Finnish birds. I. Passeriformes.*, Societas scientiarum Fennica.
Hadfield, J.D., 2008. Estimating evolutionary parameters when viability selection is operating. *Proceedings of the Royal Society B: Biological Sciences*, 275(1635), pp.723–734.
Hadfield, J.D., 2010. MCMC methods for multi-respoinse generalized linear mixed models: The MCMCglmm R package. *Journal of Statistical Software*, 33(2), pp.1–22.
Haftorn, S., 1981. Incubation during the Egg-Laying Period in Relation to Clutch-Size and Other Aspects of Reproduction in the Great Tit Parus major. *Ornis Scandinavica*, 12(3), pp.169–185.
Hänninen, H., Slaney, M. & Linder, S., 2007. Dormancy release of Norway spruce under climatic warming: testing ecophysiological models of bud burst with a whole-tree chamber experiment. *Tree Physiology*, 27(2), pp.291–300.
Henry, G.H.R. & Molau, U., 1997. Tundra plants and climate change: the International Tundra Experiment (ITEX). *Global Change Biology*, 3(Suppl. 1), pp.1–9.
Hepp, G.R., 2004. Early onset incubation by wood ducks. *Condor*, 106(1), pp.182–186.
Hepp, G.R., Kennamer, R.A. & Johnson, M.H., 2006. Maternal effects in Wood Ducks: incubation temperature influences incubation period and neonate phenotype. *Functional Ecology*, 20, pp.307–314.
Hinde, R.A., 1952. The behaviour of the Great Tit (Parus major) and some other related species. *Behaviour supplement*, II, pp.1–201.
Hinks, A.E., Cole, E.F., Daniels, K.J. & others, 2015. Phenotypic plasticity in a spatially heterogenous environment: scale-dependent phenological synchrony between songbirds and their caterpillar food source. *American Naturalist*, 186(1), pp.84–97.
Hinks, A.E., Cole, E.F., Daniels, K.J., Wilkin, T.A., et al., 2015. Scale-Dependent Phenological Synchrony between Songbirds and Their Caterpillar Food Source. *The American Naturalist*, 186(1), pp.84–97.
Hoffmann, A.A. & Sgrò, C.M., 2011. Climate change and evolutionary adaptation. *Nature*, 470(7335), pp.479–485.
Horton, R., 2015. Offline: What is medicine's 5 sigma? *The Lancet*, 385(9976), p.1380.
Hudson, I.L., 2010. Interdisciplinary approaches: Towards new statistical methods for phenological studies. *Climatic Change*, 100(1), pp.143–171.
Husby, A. et al., 2010. Contrasting patterns of phenotypic plasticity in reproductive traits in two great tit (Parus major) populations. *Evolution*, 64(8), pp.2221–2237.

Inouye, D.W. et al., 2000. Climate change is affecting altitudinal migrants and hibernating species. *Proceedings of the National Academy of Sciences*, 97(4), pp.1630–1633.
IPCC, 2013. IPCC Fifth Assessment Report. *IPCC*.
Van Der Jeugd, H.P. & McCleery, R.H., 2002. Effects of spatial autocorrelation, natal philopatry and phenotypic plasticity on the heritability of laying date. *Journal of Evolutionary Biology*, 15(3), pp.380–387.
Johansson, J., Kristensen, N.P., et al., 2015. The eco-evolutionary consequences of interspecific phenological asynchrony - a theoretical perspective. *Oikos*, 124(1), pp.102–112.
Johansson, J., Nilsson, J.-Å. & Jonzén, N., 2015. Phenological change and ecological interactions: An introduction. *Oikos*, 124(1), pp.1–3.
Johnson, L.S. et al., 2013. Variation in incubation effort during egg laying in the Mountain Bluebird and its association with hatching asynchrony. *Journal of Field Ornithology*, 84(3), pp.242–254.
Jolly, G.M., 1965. Explicit estimates from capture-recapture data with both death and immigration-stochastic model. *Biometrika*, 52(1–2), pp.225–248.
Khaliq, I. et al., 2014. Global variation in thermal tolerances and vulnerability of endotherms to climate change. *Proceedings of the Royal Society B: Biological Sciences*, 281(1789), pp.20141097–20141097.
Kidd, L.R. et al., 2015. Who escapres detection? Quantifying the causes and consequences of sampling biases in a long-term field study. *Journal of Animal Ecology*, 84(6), pp.1520–1529.
Kluijver, H.N., 1950. Daily routines of the Great Tit, Parus m. major L. *Ardea*, 38(December), pp.99–135.
Koch, E., 2010. Global Framework for Data Collection - Data Bases, Data Availability, Future Networks, Online Databases. In M. R. Keatley & I. L. Hudson, eds. *Phenological Research: Methods for Environmental and Climate Change Analysis*. Springer, pp. 23–62.
Kokko, H., López-Sepulcre, A. & Morrell, L.J., 2006. From Hawks and Doves to Self-Consistent Games of Territorial Behavior. *The American Naturalist*, 167(6), pp.901–912.
Koo, K.A. et al., 2017. Potential climate change effects on tree distributions in the Korean Peninsula: Understanding model & climate uncertainties. *Ecological Modelling*, 353, pp.17–27.
Kramer, K., 1995. Phenotypic plasticity of the phenology of seven European tree species in relation to climatic warming. *Plant, Cell & Environment*, 18(2), pp.93–104.
Krebs, C., 2008. *An Ecological World View*, CSIRO Publishing.
Laake, J.L., Johnson, D.S. & Conn, P.B., 2013. marked: An R package for maximum likelihood and Markov Chain Monte Carlo analysis of capture-recapture data. *Methods in Ecology and Evolution*, 4(9), pp.885–890.
Lachish, S. et al., 2011. Fitness effects of endemic malaria infections in a wild bird population: The importance of ecological structure. *Journal of Animal Ecology*, 80(6), pp.1196–1206.
Lack, D., 1958a. A quantitative breeding study of British tits. *Ardea*, pp.91–124.
Lack, D., 1958b. A quantitative breeding study of British tits. *Ardea*, 46, pp.91–124.
Lack, D., 1955. British tits (Parus spp.) in nesting boxes. *Ardea*, 43, pp.50–84.
Lack, D., 1968. *Ecological adaptations for breeding in birds.*, London: Methuen.
Lack, D., 1947. The Significance of Clutch-size. *Ibis*, pp.302–352.
Lambrechts, M.M. et al., 1999. Do experiments with captive non-domesticated animals make sense without population ® eld studies ? A case study with blue

tits ' breeding time. *Proceedings of the Royal Society B: Biological Sciences*, 266, pp.1311–1315.
Lane, J.E. et al., 2012. Delayed phenology and reduced 306 tness associated with climate change in a wild hibernator. *Nature*, 489, pp.554–557.
Langmore, N.E. et al., 2016. Egg size investment in superb fairy-wrens: helper effects are modulated by climate. *Proceedings of the Royal Society B: Biological Sciences*, 283(20161875), pp.1–8.
Lawson, C.R. et al., 2015. Environmental variation and population responses to global change. *Ecology Letters*, 18(7), pp.724–736.
Lehikoinen, E. & Sparks, T.H., 2010. Changes in migration. In A. P. Møller, W. Fiedler, & P. Berthold, eds. *Effects of Climate Change on Birds*. Oxford University Press, pp. 89–112.
Likens, G.E. ed., 1989. *Long-Term Studies in Ecology: Approaches and Alternatives*, Springer-Verlag.
Lindenmayer, D.B. & Likens, G.E., 2009. Adaptive monitoring: a new paradigm for long-term research and monitoring. *Trends in Ecology and Evolution*, 24(9), pp.482–486.
Loos, E.R. & Rohwer, F.C., 2004. Laying-stage nest attendance and onset of incubation in prairie nesting ducks. *Auk*, 121(2), pp.587–599.
Lord, A.M., McCleery, R.H. & Cresswell, W., 2011. Incubation prior to clutch completion accelerates embryonic development and so hatch date for eggs laid earlier in a clutch in the great tit, Parus major. *Journal of Avian Biology*, 42, pp.187–191.
Mann, M.E. et al., 2008. Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia. *Proceedings of the National Academy of Sciences*, 105(36), pp.13252–13257.
Martin, L.J., Blossey, B. & Ellis, E., 2012. Mapping where ecologists work: Biases in the global distribution of terrestrial ecological observations. *Frontiers in Ecology and the Environment*, 10(4), pp.195–201.
Matthews, J.D., 1955. The influence of weather on the frequency of beech mast years in England. *Forestry*, 28(2), pp.107–116.
Matthysen, E., Adriaensen, F. & Dhondt, A.A., 2010. Multiple responses to increasing spring temperatures in the breeding cycle of blue and great tits (Cyanistes caeruleus, Parus major). *Global Change Biology*, 17(1), pp.1–16.
Mayor, S.J. et al., 2017. Increasing phenological asynchrony between spring green-up and arrival of migratory birds. *Scientific Reports*, 7(1), p.1902.
McCleery, R.H. & Perrins, C.M., 1985. Behavioural Ecology. Ecological Consequences of Adaptive Behaviour. In Oxford: Blackwell Scientific Publications, pp. 353–373.
McClintock, M.E., Hepp, G.R. & Kennamer, R.A., 2014. Plasticity of incubation behaviors helps Wood Ducks (*Aix sponsa*) maintain an optimal thermal environment for developing embryos. *The Auk*, 131(4), pp.672–680.
McNab, B.K., 2012. *Extreme measures: the ecological energetics of birds and mammals.*, The University of Chicago Press.
Mcshane, B.B. et al., 2017. Abandon Statistical Significance. *eprint arXiv:1709.07588*, pp.1–12.
Menzel, A. et al., 2006. European phenological response to climate change matches the warming pattern. *Global Change Biology*, 12(10), pp.1969–1976.
Merilä, J. & Hendry, A.P., 2014. Climate change, adaptation, and phenotypic plasticity: The problem and the evidence. *Evolutionary Applications*, 7(1), pp.1–14.
Mertens, J.A.L., 1977. Thermal conditions for successful breeding in Great Tits

(Parus major L.) - I. Relation of growth and development of temperature regulation in nestling great tits. *Oecologia*, 28(1), pp.1–29.
MET Office, 2009. UKCP09 gridded observation datasets. Available at: https://www.metoffice.gov.uk/climatechange/science/monitoring/ukcp09/.
Miller-Rushing, A.J. et al., 2010. The effects of phenological mismatches on demography. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 365(1555), pp.3177–3186.
Møller, A.P., Rubolini, D. & Lehikoinen, E., 2008. Populations of migratory bird species that did not show a phenological response to climate change are declining. *Proceedings of the National Academy of Sciences of the United States of America*, 105(42), pp.16195–16200.
Morin, X. et al., 2009. Leaf phenology in 22 North American tree species during the 21st century. *Global Change Biology*, 15(4), pp.961–975.
Moyes, K. et al., 2011. Advancing breeding phenology in response to environmental change in a wild red deer population. *Global Change Biology*, 17, pp.2455–2469.
Murren, C.J. et al., 2015. Constraints on the evolution of phenotypic plasticity: limits and costs of phenotype and plasticity. *Heredity*, 115(4), pp.293–301.
Naef-Daenzer, B. et al., 2012. Cascading climate effects and related ecological consequences during past centuries. *Climate of the Past*, 8(5), pp.1527–1540.
Nager, R.G. & van Noordwijk, A.J., 1992. Energetic Limitation in the Egg-laying Period of Great Tits. *Proceedings. Biological sciences / The Royal Society*, 249, pp.259–264.
Nakagawa, S. & Freckleton, R.P., 2008. Missing inaction: the dangers of ignoring missing data. *Trends in Ecology and Evolution*, 23(11), pp.592–596.
van Noordwijk, A.J., McCleery, R.H. & Perrins, C.M., 1995. Selection for the Timing of Great Tit Breeding in Relation to Caterpillar Growth and Temperature. *Journal of Animal Ecology*, 64(4), pp.451–458.
Nussey, D.H. et al., 2005. Selection on heritable phenotypic plasticity in a wild bird population. *Science (New York, N.Y.)*, 310(5746), pp.304–306.
Nuzzo, R., 2014. Scientific method: Statistical errors. *Nature*, 506(7487), pp.150–152.
O'Connor, R.J., 1978. Nest-Box Insulation and the Timing of Laying in the Wytham Woods Population of Great Tits Parus Major. *Ibis*, (120), pp.534–537.
Oddie, K.R., 2000. Size matters : competition between male and female great tit o spring. *Journal of Animal Ecology*, 69, pp.903–912.
on Climate Change, I.P., 2007. *The Physical Science Basis: Contribution of Working Group 1 to the Fourth Assessment Report of the IPCC.*, Cambridge University Press.
Ozgul, A. et al., 2010. Coupled dynamics of body mass and population growth in response to environmental change. *Nature*, 466(7305), pp.482–5.
Parker, T.H. et al., 2016. Transparency in Ecology and Evolution: Real Problems, Real Solutions. *Trends in Ecology and Evolution*, 31(9), pp.711–719.
Parmesan, C., 2006. Ecological and Evolutionary Responses to Recent Climate Change. *Annual Review of Ecology, Evolution and Systematics*, 37, pp.637–669.
Parmesan, C., 2007. Influences of species, latitudes and methodologies on estimates of phenological response to global warming. *Global Change Biology*, 13(9), pp.1860–1872.
Parmesan, C. & Yohe, G., 2003. A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, 399, pp.579–583.
Payevsky, V. a., 2006. Mortality rate and population density regulation in the great tit, Parus major L.: A review. *Russian Journal of Ecology*, 37(3), pp.180–187.
Perdeck, A.C., Visser, M.E. & Van Balen, J.H., 2000. Great tit Parus major survival

and the beech crop cycle. *Ardea*, 88, pp.99–108.
Pérez, J.H. et al., 2008. Experimental heating reveals nest temperature affects nestling condition in tree swallows (Tachycineta bicolor). *Biology letters*, 4(5), pp.468–471.
Perrins, C.M., 1979. *British Tits*, London: Collins.
Perrins, C.M., 1965a. Fluctuations and Clutch-Size in the Great Tit, Parus major L. *Journal of animal ecology*, 34(3), pp.601–647.
Perrins, C.M., 1965b. Population Fluctuations and Clutch-Size in the Great Tit. *Journal of Animal Ecology*, 34(3), pp.601–647.
Perrins, C.M., 1965c. Population Fluctuations and Clutch-Size in the Great Tit, Parus major L. *Journal of Animal Ecology*, 34(3), pp.601–647.
Perrins, C.M., 1973. Some effects of temperature on breeding in the great tit and manx shearwater. *Journal of Reproduction and Fertility*, 19, pp.163–173.
Perrins, C.M., 1970a. The timing of birds' breeding season. *Ibis*, 112, pp.242–255.
Perrins, C.M., 1970b. The Timing of Birds' Breeding Seasons. *Ibis*, 112(2), pp.242–255.
Perrins, C.M., 1991. Tits and Their Caterpillar Food-Supply. *Ibis*, 133(Perrins 1976), pp.49–54.
Perrins, C.M. & McCleery, R.H., 1989. Laying Dates and Clutch Size in the Great Tit. *The Wilson Journal of Ornithology*, 101(2), pp.236–253.
Phillimore, A.B. et al., 2010. Differences in spawning date between populations of common frog reveal local adaptation. *Proceedings of the National Academy of Sciences*, 108(18), pp.8292–8297.
Phillimore, A.B. et al., 2012. Dissecting the Contributions of Plasticity and Local Adaptation to the Phenology of a Butterfly and Its Host Plants. *The American Naturalist*, 180(5), pp.655–670.
Phillimore, A.B. et al., 2013. Inferring local processes from macro-scale phenological pattern: A comparison of two methods. *Journal of Ecology*, 101(3), pp.774–783.
Pigliucci, M., 2001. *Phenotypic Plasticity: Beyond Nature and Nurture*, Baltimore, MD: Johns Hopkins University Press.
Plard, F. et al., 2014. Mismatch between Birth Dates and Vegetation Phenology Slows the Demography of Roe Deer. *PLoS Biology*, 12(4), pp.1–8.
van de Pol, M. et al., 2016. Identifying the best climatic predictors in ecology and evolution. *Methods in Ecology and Evolution*, 7(10), pp.1246–1257.
van de Pol, M. & Cockburn, A., 2011. Identifying the critical climatic time window that affects trait expression. *The American naturalist*, 177(5), pp.698–707.
Powell, L.E., 1987. Plant Hormones and Their Role in Plant Growth and Development. In Dordrecht: Nijhoff, pp. 539–552.
Price, T.D., Qvarnström, A. & Irwin, D.E., 2003. The role of phenotypic plasticity in driving genetic evolution. *Proceedings of the Royal Society B: Biological Sciences*, 270(1523), pp.1433–1440.
Racey, P.A. & Swift, S.M., 1981. Variations in gestation length in a colony of pipistrelle bats (Pipistrellus pipistrellus) from year to year. *Journal of Reproduction and Fertility*, 61(1), pp.123–129.
Radcliffe Meterological Station, Meterological data.
Raven, M.J., Noble, D.G. & Baillie, S.R., 2005. *The Breeding Bird Survey 2004*, Thetford.
Reed, T., Grøtan, V., et al., 2013. Population growth in a wild bird is buffered against phenological mismatch. *Science*, 340, pp.488–491.
Reed, T., Gienapp, P. & Visser, M.E., 2016. Testing for biases in selection on avian reproductive traits and partitioning direct and indirect selection using quantitative genetic models. *Evolution*, 70(10), pp.2211–2225.

Reed, T., Jenouvrier, S. & Visser, M.E., 2013. Phenological mismatch strongly affects individual fitness but not population demography in a woodland passerine. *Journal of Animal Ecology*, 82(1), pp.131–144.
Roberts, A.M.I., 2008. Exploring relationships between phenological and weather data using smoothing. *International Journal of Biometeorology*, 52(6), pp.463–470.
Roberts, A.M.I. et al., 2015. Predicting a change in the order of spring phenology in temperate forests. *Global Change Biology*, 21(7), pp.2603–2611.
Roberts, A.M.I., 2010. Smoothing Methods. In M. R. Keatley & I. L. Hudson, eds. *Phenological Research: Methods for Environmental and Climate Change Analysis*. Springer, pp. 255–270.
Robinson, B.W. & Dukas, R., 1999. The influence of phenotypic modifications on evolution: the Baldwin effect and modern perspectives. *Oikos*, 85(3), pp.582–589.
Rötzer, T., Grote, R. & Pretzsch, H., 2004. The timing of bud burst and its effect on tree growth. *International Journal of Biometeorology*, 48(3), pp.109–118.
Rowe, L., Ludwig, D. & Schluter, D., 1994. Time, Condition, and the Seasonal Decline of Avian Clutch Size. *American Naturalist*, 143(4), pp.698–722.
Saether, B.-E. et al., 2002. Density dependence and stochastic variation in a newly established population of a small songbird. *Oikos*, 99, pp.331–337.
Sandvig, E.M. et al., 2017. The influence of climatic variation and density on the survival of an insular passerine Zosterops lateralis. *PLoS ONE*, 12(4).
Savalei, V. & Dunn, E., 2015. Is the call to abandon p-values the red herring of the replicability crisis? *Frontiers in Psychology*, 6(MAR).
Schaper, S. V. et al., 2012. Increasing temperature, not mean temperature, is a cue for avian timing of reproduction. *The American naturalist*, 179(2), pp.E55-69.
Schaper, S. V. et al., 2011. Spring phenology does not affect timing of reproduction in the great tit (Parus major). *The Journal of experimental biology*, 214(Pt 21), pp.3664–71.
Scott, I.C. et al., 2008. The effect of conception date on gestation length of red deer (Cervus elaphus). *Animal Reproduction Science*, 109(1–4), pp.206–217.
Seber, G., 1965. A note on multiple recapture census. *Biometrika*, 52(1–2), pp.249–260.
Sheldon, B.C., Ruuk, L.E.B.K. & Merilä, J., 2003. Natural selection and inheritance of breeding time and clutch size in the collared flycatcher. *Evolution*, 57(2), pp.406–420.
Simmonds, E.G. et al., 2017. Incubation behaviour adjustments, driven by ambient temperature variation, improve synchrony between hatch dates and caterpillar peak in a wild bird population. *Ecology and Evolution*, pp.1–11.
Singer, M.C. & Parmesan, C., 2010. Phenological asynchrony between herbivorous insects and their hosts: signal of climate change or pre-existing adaptive strategy? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1555), pp.3161–3176.
Sparks, T.H. & Carey, P.D., 1995. The responses of species to climate over two centuries : an analysis of the Marsham phenological record , 1736-1947. *Journal of Ecology*, 83(2), pp.321–329.
Sparks, T.H. & Crick, H., 1999. Opinion: The times they are a-changing? *Bird Conservation International*, 9(May 2010), pp.1–7.
Sparks, T.H. & Yates, T.J., 1997. The effect of spring temperature on the appearance dates of British butterflies 1883-1993. *Ecography*, 20(4), pp.368–374.
Stenning, M.J., 2008. Hatching Asynchrony and Brood Reduction in Blue Tits, Cyanistes caeruleus, May be a Plastic Response to Local Oak, Quercus robur,

Bud Burst and Caterpillar Emergence. *Acta Ornithologica*, 43(1), pp.97–106.
Stenseth, N.C. et al., 2015. Testing for effects of climate change on competitive relationships and coexistence between two bird species. *Proceedings of the Royal Society B*, 282(1807), p.20141958.
Van Straalen, N.M. & Timmermans, M.J.T.N., 2002. Genetic Variation in Toxicant-Stressed Populations: An Evaluation of the "Genetic Erosion" Hypothesis. *Human and Ecological Risk Assessment: An International Journal*, 8(5), pp.983–1002.
Strathdee, A.T. et al., 1995. Climatic severity and the response to temperature elevation of Arctic aphids. *Global Change Biology*, 1, pp.23–28.
Tansey, C.J., Hadfield, J.D. & Phillimore, A.B., 2017. Estimating the ability of plants to plastically track temperature-mediated shifts in the spring phenological optimum. *Global Change Biology*, 23(8), pp.3321–3334.
Team, Q.D., 2016. QGIS Geographic Information System.
Team, R.D.C., 2008. R: A language and environment for statistical computing.
Teplitsky, C. et al., 2009. Heritability of fitness components in a wild bird population. *Evolution*, 63(3), pp.716–726.
Thackeray, S.J. et al., 2016. Phenological sensitivity to climate across taxa and trophic levels. *Nature*, 535(7611), pp.241–245.
Thackeray, S.J., Sparks, T.H. & Frederiksen, M. et al., 2010. Trophic level asynchrony in rates of phenological change for marine, freshwater and terrestrial environments. *Global Change Biology*, 16(12), pp.3304–3313.
Tomás, G., 2015. Hatching date vs laying date: What should we look at to study avian optimal timing of reproduction? *Journal of Avian Biology*, 46(1), pp.107–112.
Torti, V.M. & Dunn, P.O., 2005. Variable effects of climate change on six species of North American birds. *Oecologia*, 145, pp.486–495.
Vedder, O., 2012. Individual birds advance offspring hatching in response to increased temperature after the start of laying. *Oecologia*, 170(3), pp.619–628.
Vedder, O., Bouwhuis, S. & Sheldon, B.C., 2013. Quantitative assessment of the importance of phenotypic plasticity in adaptation to climate change in wild bird populations. *PLoS biology*, 11(7), p.e1001605.
Verhulst, S., Van Balen, J.H. & Tinbergen, J.M., 1995. Seasonal decline in reproductive success of the great tit: Variation in time or quality? *Ecology*, 76(8), pp.2392–2403.
Verhulst, S. & Tinbergen, J.M., 1991. Experimental evidence for a causal relationship between timing and success of reproduction in the great tit Parus m. major. *Journal of Animal Ecology*, 60, pp.269–281.
Visser, M.E. et al., 2011. Genetic variation in cue sensitivity involved in avian timing of reproduction. *Functional Ecology*, 25(4), pp.868–877.
Visser, M.E., 2008. Keeping up with a warming world; assessing the rate of adaptation to climate change. *Proceedings of the Royal Society B: Biological Sciences*, 275(1635), pp.649–659.
Visser, M.E. et al., 2010. Phenology, seasonal timing and circannual rhythms: towards a unified framework. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 365(1555), pp.3113–27.
Visser, M.E. et al., 2003. Variable responses to large-scale climate change in European Parus populations. *Proceedings. Biological sciences / The Royal Society*, 270(1513), pp.367–72.
Visser, M.E. et al., 1998. Warmer springs lead to mistimed reproduction in great tits (Parus major). *Proceedings. Biological sciences / The Royal Society*, 265(May), pp.1867–1870.

Visser, M.E. & Both, C., 2005. Shifts in phenology due to global climate change: the need for a yardstick. *Proceedings. Biological sciences / The Royal Society*, 272, pp.2561–2569.
Visser, M.E., Both, C. & Lambrechts, M., 2004. Global climate change leads to mistimed avian reproduction. *Advances in Ecological Research*, 2504(4), pp.89–110.
Visser, M.E. & Holleman, L.J.M., 2001. Warmer springs disrupt the synchrony of oak and winter moth phenology. *Proceedings. Biological sciences / The Royal Society*, 268(1464), pp.289–94.
Visser, M.E., Holleman, L.J.M. & Caro, S.P., 2009. Temperature has a causal effect on avian timing of reproduction. *Proceedings. Biological sciences / The Royal Society*, 276, pp.2323–2331.
Visser, M.E., Holleman, L.J.M. & Gienapp, P., 2006. Shifts in caterpillar biomass phenology due to climate change and its impact on the breeding biology of an insectivorous bird. *Oecologia*, 147, pp.164–172.
Walther, G.R. et al., 2002. Ecological responses to recent climate change. *Nature*, 416(6879), pp.389–395.
Wilkin, T.A. et al., 2006. Density effects on life-history traits in a wild population of the great tit Parus major: analyses of long-term data with GIS techniques. *Journal of Animal Ecology*, 75(2), pp.604–615.
Wilson, A.J. et al., 2010. An ecologist's guide to the animal model. *Journal of Animal Ecology*, 79(1), pp.13–26.
Wolkovich, E.M. et al., 2012. Warming experiments underpredict plant phenological responses to climate change. *Nature*, 485(7399), pp.494–497.
Woltereck, R., 1909. Weitere experimentelle Untersuchungen über Artveränderung, speziel über das Wesen quantitativer Artunterschiede bei Daphniden. *Verhandlungen der deutschen zoologischen Gesellschaft*, 19, pp.110–173.
Wood, S.N., 2001. mgcv: GAMs and generalized ridge regression for R. *R News*.
Wood, S.N. & Wood, M.S., 2015. Package "mgcv." *R package version*, pp.1–7.
Xiang, Y. et al., 2013. Generalized simulated annealing for global optimization: the GenSA Package. *R Journal*, 5(June), pp.13–28.
YomTov, Y. & Wright, J., 1993. Effect of heating nest boxes on egg laying in the blue tit (Parus caeruleus). *Auk*, 110(1), pp.95–99.