

SPATIO-TEMPORAL MODELLING OF VOLUNTEERED GEOGRAPHIC INFORMATION: A PHENOLOGICAL STUDY

EDENGENET MASHILLA DEJENE

February, 2014

SUPERVISORS:

Dr. R.Zurita-Milla

Dr. Ir.Rolf A.de By



SPATIO-TEMPORAL MODELLING OF VOLUNTEERED GEOGRAPHIC INFORMATION: A PHENOLOGICAL STUDY

EDENGENET MASHILLA DEJENE

Enschede, The Netherlands, February, 2014

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.

Specialization: Geo-informatics

SUPERVISORS:

Dr. R. Zurita-Milla

Dr. Ir. Rolf A.de By

THESIS ASSESSMENT BOARD:

Dr. A.A. Voinov (Chair)

Dr. Ir. A.J.H. van Vliet (External Examiner, Wageningen UR, The Netherlands)

DISCLAIMER

This document describes work undertaken as part of a programme of study at the Faculty of Geo-Information Science and Earth Observation of the University of Twente. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the Faculty.

ABSTRACT

Volunteered geographic information (VGI) was used for the acquisition of spatio-temporal information. Several studies show the involvement of volunteers in gathering information about phenological phases of plants. These observations together with climatic variables were used for modelling phenology. The main challenge was the existence of biased phenological observations and the absence of volunteers at important monitoring locations, which could lead to biased parameter values and modelling results.

The objective of this research was assisting with the design of volunteer-based phenological monitoring networks by modelling phenological observations, determining temporal windows of monitoring and detecting spatio-temporal bias. For this purpose, we tested two different types of models, statistical (SW, UNIFORC, UNICHILL) and data-driven (PHASE) for predicting the beginning of the flowering dates of four species and leafing date of one species. The models were calibrated using VGI and raster-based temperature data (for the period 2003 – 2010). The statistical models were optimized by using simulated annealing and the data-driven model used an iterative optimization procedure based on the data. Quality measures (RMSE, efficiency and correlation) were used to choose the best performing model based on VGI data which was split into calibration and validation. Input erroneous phenological observations were identified by minimizing the systematic and unsystematic RMSE of the best performing model. The model was then re-calibrated with the non-erroneous observations and was used to determine areas showing variability (spatial bias) in the date of year (DOY) of the phenological event.

The SW model showed better performance, and the calibration of the model resulted in parameter values with meaningful interpretations. As such, this model accounted for the inter-annual climatic variability and could be used to predict the phenological event of the five plant species. However, phenological responses to annual variations in temperature conditions were not fully tracked, probably due to landscape, topography or coarse gridded input temperature data. Also, the SW model helped in identifying erroneous observations and areas showing low, medium and high climatic variability. Areas showing variability in the prediction DOY of the phenological event differed among the five plant species. For year and plant species combinations, the duration of the phenological event also differed due to climate change. We found that the PHASE model was more robust to outliers than the SW model and could work well in the presence of temporally-biased phenological observations. But, the prediction accuracy of the PHASE model to an independent dataset was low. This indicated the over-fitting of the model to the training (calibration) data. The simulated annealing algorithm determined better optimal parameter values than the optimization process used by the PHASE model. Overall, our findings indicate that the DOY of the phenological event occur later in the northern part and earlier in the southern part of The Netherlands. Besides, the inter-annual variations in the DOY of the phenological event showed the change in climate through the years and revealed differences in the phenology of the five species. The variability in phenological events due to location, time and species were determined in this research. Further similar studies would enable to confirm the validity of the models for precise prediction of the phenological events. In general, we have assessed the value of VGI for monitoring environmental variables and the issues of quality (bias) that are associated with such kinds of data.

Keywords: Volunteered Geographic Information, Phenological modelling, Model comparison, Spatial bias, Temporal bias, Climate change

ACKNOWLEDGEMENTS

First and foremost, I would like to thank the almighty God for His faithfulness, for being patient with me and for holding me tight in times I wanted to give up. Thank You God for being more than enough, more than anything else in the world. With gratefulness, I thank you lord Jesus.

I would like to express my deepest gratitude to my first supervisor, Dr. R. Zurita-Milla, for his guidance, useful comments, remarks and engagement through the entire process of this Master's thesis research. He has been very helpful in introducing me to the topic as well as encouraging and supporting me on the way. Furthermore, I would like to thank my second supervisor, Dr. Ir. Rolf A.de By for the valuable comments starting from the thesis research proposal up to the end of the thesis write-up.

I would like to greatly acknowledge my best friend, Elizabeth Fesseha, for her thoughtfulness, understanding and encouragement throughout the whole thesis work. I would also like to thank my friend and colleague, Kibrewossen Yitbarek, for his encouragement and good friendship.

ITC has been a very enjoyable and luscious place to study. I am very grateful to the members of the faculty as well as fellow students, especially to those in the Department of GFM, who have been cooperating and collaborating with me on various projects, assignments and group works. I would like to extend my special thanks to friends in ITC, Loza Bekalo, Teshale Tadesse, Zelalem Wondimagegn, Mikiyas Mesfin and Qian Liu (Eva).

I would like to acknowledge The Netherlands Fellowship Program (NFP) for giving me financial support. I would also like to extend my deepest gratitude to Hawassa University, the Department of Computer Science at the University and all colleagues for all kinds of support and facilitation towards my personal development and training. I am grateful to my father, Mashilla Dejene, for teaching me to dream big and for all his efforts in making them come true. Thank you father for dedicating your whole life to your family. I really give incredible value for your aspiration and readiness to pave my way for bright future. I would also like to thank my brother, Eskinder Mashilla, for you have been taking a good care of me from childhood to date and my aunt, Seblealem Dessie, for you have been always there for helping me in times of need.

I would like to give special thanks to my husband, Tewodros Awgichew, for your support and kindness, for being there when I am in need indeed. Thank you for doing everything for me. Without you, I could not have survived this life.

Last but not least, I lack words of appreciation to express my special thanks to my mother, Alemtsehay Berihun, for your prayers, love, perseverance, moral support and unconditional care. Mother, thank you for encouraging me and for believing in me in all my day-to-day activities. You are my blessing and a safe keep in my heart.

TABLE OF CONTENTS

| | | |
|------|---|----|
| 1. | Introduction..... | 1 |
| 1.1. | Motivation and Problem Statement..... | 1 |
| 1.2. | Research Identification..... | 2 |
| 1.3. | Method Adopted..... | 4 |
| 1.4. | Overview of the Thesis research..... | 6 |
| 2. | Literature Review..... | 7 |
| 2.1. | Overview..... | 7 |
| 2.2. | Importance of Volunteer-based Geographic Information..... | 7 |
| 2.3. | Quality of Volunteer-based Geographic Information..... | 8 |
| 2.4. | Spatio-temporal Analysis of VGI Data..... | 9 |
| 2.5. | Review of Statistical Phenological Models..... | 10 |
| 2.6. | Review of Data-Driven Phenological Models..... | 12 |
| 3. | Materials and Methods..... | 13 |
| 3.1. | Materials..... | 13 |
| 3.2. | Methods..... | 16 |
| 4. | Results..... | 27 |
| 4.1. | Introduction..... | 27 |
| 4.2. | Model Parameterization..... | 27 |
| 4.3. | Model Performance and Comparison..... | 30 |
| 4.4. | Bias Identification..... | 32 |
| 4.5. | PHASE 2.0 Model..... | 50 |
| 4.6. | Comparison of SW and PHASE 2.0 model..... | 53 |
| 5. | Discussion..... | 55 |
| 5.1. | Model Parameterization..... | 55 |
| 5.2. | Model Performance and Comparison..... | 55 |
| 5.3. | Bias Identification..... | 56 |
| 5.4. | PHASE 2.0 Model..... | 58 |
| 5.5. | Comparison of SW and PHASE 2.0 Model..... | 58 |
| 6. | Conclusions and Recommendations..... | 59 |
| 6.1. | Conclusions..... | 59 |
| 6.2. | Recommendations..... | 60 |
| | List of References..... | 61 |
| | Appendices..... | 67 |
| | Appendix A..... | 67 |
| | Appendix B..... | 73 |

LIST OF FIGURES

| | |
|---|----|
| Figure 1.1 Methods to be adopted for attaining the research objective..... | 5 |
| Figure 3.1. Box plots showing the yearly volunteer observations on the DOY of the phenological stage for the five plant species..... | 15 |
| Figure 3.2. The adopted method for answering the research questions..... | 17 |
| Figure 3.3. Optimization of the PHASE model. $psHU_i$ and $psHU_{max.ph}$ are the location specific heat units accumulated from a starting DOY of i and $max.ph$, respectively. $Start.ph$ is the starting DOY determined by the optimization process..... | 25 |
| Figure 4.1. The linear relationship between daily rate of forcing and average temperature of Cow Parsley..... | 28 |
| Figure 4.2. The non-linear (sigmoidal) dependency between daily rate of forcing and average temperature of Cow Parsley..... | 29 |
| Figure 4.3. Boxplots showing the DOY of the phenological event of the calibration and validation dataset..... | 32 |
| Figure 4.4. Optimal normalized SD threshold values for removal of temporally-biased observations..... | 34 |
| Figure 4.5. Efficiency and correlation values for removal of temporally-biased observations..... | 35 |
| Figure 4.6. Spatial distribution of year based temporally-biased volunteer phenological observations of Cow Parsley (circle symbols). The colour of the map indicates the variability in the DOY of the phenological event between one year and all the other years. For each phenological observation, the sizes of the dot indicates the amount of normalized SD value..... | 36 |
| Figure 4.7 Observed and predicted values of SW model with temporally biased and non-biased volunteer phenological observations..... | 39 |
| Figure 4.8. Observed and Predicted values of SW model with non-temporally biased phenological observations..... | 40 |
| Figure 4.9. Spatial prediction of the DOY of the phenological event of the five plant species in The Netherlands in 2010..... | 41 |
| Figure 4.10. Box plots showing the duration of phenological event between the years 2003 and 2010..... | 42 |
| Figure 4.11. Systematic RMSE, unsystematic RMSE and RMSE (a) and efficiency (b) of the SW model after second round removal of temporally-biased observations of Oak..... | 43 |
| Figure 4.12. The variability in the DOY of the phenological event between Cow parsley and Oak between the years 2003 and 2010..... | 44 |
| Figure 4.13. Prediction results of the SW model for Cow Parsley from the year 2003 to 2010 (a) with different scale depending on DOY of phenological event (b) with common scale ranging between the minimum and maximum DOY of phenological events in all years..... | 46 |
| Figure 4.14. Average annual temperature of The Netherlands between the years 2003 and 2010..... | 47 |
| Figure 4.15. Standard deviation map of DOY of phenological event between the years 2003 and 2010..... | 47 |
| Figure 4.16. Volunteer phenological observations overlaid on areas showing high (green), medium (yellow) and low (pale colour) variability in the prediction of DOY of phenological stage..... | 48 |
| Figure 4.17. Percentages of volunteer phenological observations belonging to high, medium and low variable areas in The Netherlands (a) Cow Parsley (b) Celandine (c) Chestnut (d) Wood Anemone (e) Oak..... | 49 |
| Figure 4.18. Optimization of the PHASE model for (a) Cow Parsley (b) Celandine (c) Chestnut (d) Wood Anemone (e) Oak. Each dot indicates an iteration, which was made for the choice of the starting DOY, from date 1 up to the date of maximum phase. Efficiencies of model 3 (Efficiency3) and model 2 (Efficiency 2) for every iteration are shown on the x and y-axis, respectively. The squared box indicates the iteration that resulted in a maximum $Efficiency_2 * Efficiency_3$ | 52 |

LIST OF TABLES

| | |
|---|----|
| Table 3.1. Name of Species, year and number of observations of the phenological stage of the five plant species in the “Natuurkalender” dataset..... | 13 |
| Table 3.2. A record showing instances of volunteer phenological observation..... | 14 |
| Table 3.3. Computation of psHU from a starting DOY ($T_0 = 1$). Location (x,y), denotes the x and y coordinates of the volunteers. Date, is the date of phenological observation. T is the temperature and the subscripts indicate the DOY..... | 25 |
| Table 3.4. Computation of uHU values from daily average temperatures $T_1, T_2, \dots, T_{\max,ph}$ | 26 |
| Table 4.1. Optimal parameters and statistical analysis of SW model for the calibration dataset. T_b (Base temperature), T_0 (Starting DOY of temperature accumulation) and F^* (Heat requirements) | 28 |
| Table 4.2. Optimal parameters and statistical analysis of UNIFORC model for the calibration dataset. d and e are parameters of sigmoidal function, T_0 (Starting DOY of temperature accumulation) and F^* (Heat requirements)..... | 29 |
| Table 4.3. Optimal parameters and statistical analysis of UNICHILL model for the calibration dataset. a, b, c are parameters of sigmoidal function (R_c), d and e are parameters of sigmoidal function (R_f), C^* (chilling requirements) and F^* (Heat requirements). | 30 |
| Table 4.4. Statistical analysis of the SW, UNIFORC and UNICHILL model for the validation dataset | 31 |
| Table 4.5. Statistical analysis of the ANOVA test between the calibration and validation dataset..... | 31 |
| Table 4.6 Instances of the volunteer phenological observations for Cow Parsley and Celandine. (x, y are the coordinates of the observer). | 33 |
| Table 4.7. Statistical analysis of the SW model after removal of temporally-biased phenological observations in the calibration dataset. | 38 |
| Table 4.8. Statistical analysis of the SW model after removal of temporally-biased phenological observations in the validation dataset..... | 38 |
| Table 4.9. Systematic and unsystematic RMSE of SW model after removal of temporally-biased observations..... | 38 |
| Table 4.10 Percentages of high, medium and low variable areas in the whole area of The Netherlands | 49 |
| Table 4.11 Existing, expected and required number of volunteers in high, medium and low variable areas | 50 |
| Table 4.12. Starting date and efficiencies of oak for model 2 and model 3 for iterations starting from $i=1 \dots$ to $i=\max.ph$ | 51 |
| Table 4.13. Statistical analysis of the PHASE model by using the calibration and validation dataset..... | 53 |
| Table 4.14. Statistical analysis of the PHASE model with a fixed starting DOY for calibration and validation dataset | 54 |
| Table 4.15. Statistical analysis of the PHASE model after removal of temporally-biased phenological observations..... | 54 |

ACRONYMS

| | |
|-----------------|------------------------------------|
| VGI | Volunteered Geographic Information |
| UNIFORC | Unified Forcing Model |
| UNICHILL | Unified Chilling Model |
| SW | Spring Warming |
| RMSE | Root Mean Square Error |
| DOY | Date of the year |
| SD | Standard Deviation |
| OOB | Out of the Bag |

1. INTRODUCTION

1.1. Motivation and Problem Statement

The increased need for collecting spatio-temporal information to detect and forecast ecological and social changes has supported the involvement of volunteer citizens as sensors to monitor natural (Goodchild, 2011; Lovell et al., 2009) and social (Cui, 2013) phenomena. Because of advances in information and communication technology (ICT), scientists can now collaborate with citizens to collect more data about environmental conditions (Tulloch et al., 2013). One can also imagine that citizen science has boomed because of the convergence of two streams. The first one being the stream of citizens collecting large amounts of data in a context of web 2.0 and crowd sourcing applications on the internet (Bordogna et al., 2013). The second stream involves scientists who address problems from a bottom-up perspective, and who require high resolution spatial and temporal observations for their experiments (Jiguet et al., 2012) or models (van Bussel et al., 2011). To address this need, citizen-based observation networks are used for monitoring environmental conditions existing in particular locations.

Phenology, which is the study of seasonal development stages of plants and animals, helps in studying climatic variation (Betancourt et al., 2007). Volunteers can gather spatio-temporal information about the phenological phases (phenophases) of plant species. The collected information can subsequently be used to study the effects of climate change at different locations (Jiguet et al., 2012). The timing of a wide range of phenophases is affected by various environmental variables, like temperature, precipitation, year and location (Polgar et al., 2013). Besides, phenophases display a varying degree of sensitivity to environmental variables, both in space and time. Clark and Thompson (2010) have shown the sensitivity of flowering date, which is a key development stage of a plant, to changes in temperature. Thus, it is important to identify the environmental variables that help explain the phenological phase of the species.

Environmental variables and phenophases exhibit variability with location, time and among species. Also, as discussed by Saracco et al. (2008), environmental variability determines the distribution and abundance of species. Therefore, it is important that volunteers sample the parameters for environmental variables in a non-biased fashion. In other words, missing data from important locations will result in spatially-biased monitoring results that might yield misleading results. Similarly, the absence of volunteers at the appropriate timing of the phenological phase will create temporally-biased monitoring results. In addition, different species display varying levels of response to climatic variation. Some species have phenophases that make a clear response to climate change, while others have phenophases that appear unrelated to climatic variations (Møller et al., 2008). So, the identification of species which are abundant in suitable habitat, with ease of recognition, and phenophases that consistently respond to environmental variables will help decision-makers determine species that are better suited for monitoring purposes (Beaubien and Hamann, 2011). In addition, determination of optimal temporal windows and monitoring locations is useful for decision-makers to give recommendations for volunteers in the design of volunteer-based monitoring networks.

Most phenological monitoring networks rely on spatially-distributed volunteers who are involved in gathering the timing of phenophases for a relatively wide range of species. The effectiveness and efficiency of monitoring networks is largely dependent on the survey design, i.e., the distribution of data collection efforts in space and time (Reynolds et al., 2011). As a result, the proper design of the phenological

monitoring network has economic and scientific benefits (Amorim et al., 2012). Properly designed networks consist of optimal spread of volunteers in space and time to sufficiently monitor an environmental phenomenon. This will reduce the economic cost of deploying volunteers for data collection (Bonney et al., 2009; Ferster and Coops, 2013). In addition, the proper design of the monitoring networks is essential for understanding and modelling the relationship between phenological phases and spatio-temporal variation in climatic conditions.

The location of volunteers in phenological monitoring networks influences the quality of collected environmental information (Do et al., 2012). Therefore, an optimized monitoring network with strategically located volunteers is necessary to study better the influence of environmental conditions on the phenology of species. In this way, appropriate temporal windows and spatial variations are identified for a particular phenological phase of a species (Moller et al., 2011). Similarly, optimal design of the monitoring network can be used to predict the distribution of the species in the future and reduce the uncertainty in estimating that distribution (Fink et al., 2010). This information is helpful for decision makers to use their budget wisely to recruit new volunteers by preparing guidelines on which species, when and where citizen sensors should make reports.

1.2. Research Identification

1.2.1. Research Objectives

The main objective of this research is to assist with the design of a citizen-based monitoring network in which volunteers collect spatio-temporal information on the phenological phases of a number of species.

This can be achieved by completing the following sub-objectives:

- Modelling the relationship between environmental information and time-series volunteer phenological observations using three statistical models and a data-driven approach;
- Improving phenological models by removing potentially-erroneous observations and by minimizing the systematic and unsystematic errors (RMSE);
- Identifying the temporal phenological windows of monitoring the phenophase of the species in space using modelling results; and
- Mapping phenological models through time to detect spatially biased areas at large scale.

1.2.2. Research Questions

1. Which statistical phenological modelling approach best fits the species/phenophase combinations of volunteered phenological observations?
 - 1.1. How to fit statistical model parameters?
 - 1.2. Which data-driven approach can be used for modelling phenology of a species?
 - 1.3. What is the added value of data-driven modelling?
 - 1.4. What is the impact of erroneous VGI observations on the output of the phenological models?
2. How can we systematically exclude erroneous observations from phenological models based on volunteered geo-information?
 - 2.1. Which VGI observations are erroneous?
 - 2.2. What is the error threshold that minimizes systematic modelling errors?
3. How can the temporal windows of monitoring a phenophase be determined from the modelling results?
 - 3.1. Which recommendations can be given to volunteers on the timing of monitoring the phenophase in a given location?
4. How can phenological dynamics be used to determine sampling schemes?
 - 4.1. How can phenological dynamics be used to determine temporally biased areas?

- 4.2. Which recommendations can be given to volunteers on the location of monitoring a phenophase of a species?

1.2.3. Innovation aimed at

This research will improve the design of citizen-based phenological observation networks by modeling phenological observations in response to climatic variations and systematically removing erroneous VGI observations. Like this, these networks will be able to identify the species and the phenophases to be monitored as well as when (time) and where (space) to do it.

1.2.4. Related work

The use of volunteered geographic information (VGI) has a long history (Silvertown, 2009). This can be associated with the existence of volunteer-based phenological monitoring networks for a long period. For instance, eBird is a phenological monitoring network, launched in 2002, where birdwatchers collect and store observations on bird populations (Sullivan et al., 2009). Networks of the ebird type (ebird.org in the USA, waarneming.nl in The Netherlands, and many others across Europe) for most part do not deploy a sampling scheme, and simply collect observations that their large volunteer base report. They often do organize, however, specific data collection projects with a fixed runtime. The long-running Christmas Bird Count¹ is one such effort; in The Netherlands, the Great Grey Shrike winter count² is another example. The design of monitoring networks involves different sampling schemes. Delaney et al. (2008) used systematic surveys, using randomly placed quadrats with 1000 volunteers across 52 sites to collect the species, gender, and carapace width of crab. Similarly, Braunisch and Suchant (2010) compared three volunteer-based monitoring schemes using data collected in exact locations; locations aggregated to grid cells and systematic counts conducted within a small subarea.

As the amount of spatio-temporal VGI observation continues to increase, innovative ways of getting the best out of the available big data are required. Phenological models are one of the ways that make use of VGI data to study the effect of climatic variations on the phenophases of plant and animal species and make forecasts of their future distributions. Parker et al. (2011) compared three phenological models to determine the one with the best predictive ability on the flowering and veraison stage of grapevine. Similarly, Bolmgren et al. (2013) modeled a 73 years old data series to predict the first flowering dates of 25 species. Furthermore, Miranda et al. (2013), used two different versions of critical date models and five different versions of critical chilling model to find the model with the best predictive ability. The parameters of the model can either be inferred from literature or fine-tuned. Deligios et al. (2013), obtained parameters of the model, base and optimum temperature, from literature to adapt a new model for predicting growth and yield of rapeseed. Alternatively, parameters can be refined by using a Bayesian approach (Ceglar et al., 2011), simulated annealing (Chmielewski et al., 2013; Parker et al., 2011) or least square parameter estimation methods (Sáez and Rittmann, 1992).

On the other hand, geo-statistical models have been used to provide spatial interpolations and predictions (Júnez-Ferreira and Herrera, 2013). In addition to the statistical approaches, the advancements in data-intensive science have increased the accuracy in estimating the distribution of species (Hochachka et al., 2012b). Thus, data-driven approaches can be used as powerful modeling alternatives (Abdel-Rahman et al., 2012). For instance, Moller et al. (2011) modeled the relationship between climatic variables and ‘Yellow Ripeness’ phase of winter wheat using a machine learning algorithm.

¹ birds.audubon.org/christmas-bird-count

² waarneming.nl/waarnemingen_projecten.php?project=258

Although volunteered geo-information is the source of a vast quantity of data, the quality of the data is questionable. Sullivan et al. (2009) considered lack of identification skills; geographic bias due to absence of standardized location reporting mechanisms and uneven sampling distribution as the cause of bias in volunteer information. Several studies have used different approaches for removing bias in volunteer geographic information. Jun et al. (2012) developed an automated data quality filter which identifies outliers and groups the data into valid and invalid observations to improve data quality. Bonney et al. (2009) proposed that data quality can be achieved by delivering data collection protocols, data forms, and providing help for participants in understanding the protocols.

The proposed research project used three statistical phenological models described by some scholars (Parker et al., 2011) and the data-driven model by Moller et al. (2011). The first two statistical phenological models, Spring Warming and UNIFORC, only consider the effect of forcing temperature on phenological development. Likewise, the third statistical model, UNICHILL uses chilling and forcing temperature to predict the date of the phenophase. On the other hand, the data-driven model (PHASE) made use of a machine learning algorithm, Random Forest, which was based on decision trees to determine the date of the phenophase. The models were used to map phenological dynamics through time to determine spatially and temporally biased volunteer monitoring areas in the design of monitoring networks.

1.3. Method Adopted

The proposed research was carried out by following the workflow shown in Figure 1.1. The research work started with a review of relevant literatures to justify the methods adopted. Statistical and data-driven phenological models, optimization method of model parameters, systematic identification of erroneous VGI observations were based on previous work in the literature. This review of relevant literature was accompanied by data preparation. In this phase, the volunteer phenological observations were checked for incompleteness (missing data) and the environmental variables (e.g., gridded daily temperature files) that were used were selected and prepared. This included extracting the values of the environmental variables at locations of phenological observations and storing them in a manageable way.

Then, the volunteer phenological observations of species/phenophase combinations and environmental variables were fitted on three statistical phenological models taken from literature (Parker et al., 2011) and the data-driven PHASE model (Moller et al., 2011). The performance of each model was determined by validating the models. A separate validation dataset was used for validating the statistical phenological models. On the other hand, the algorithm of the data-driven approach partitioned some percentage of the data for validation and showed the accuracy of the model. Consequently, quantitative quality measures, like RMSE (Root Mean Square Error) and efficiency (percentage of variance explained) were used for comparing the performance of the models (Parker et al., 2011; Zhang and Tao, 2013). Based on these quality measures, the model with the highest efficiency and the smallest RMSE was considered as the best phenological model. The parameters of the best statistical phenological model was then optimized using existing method from literature.

After that, the identification and removal of erroneous VGI observations were done by first partitioning the RMSE components into systematic and unsystematic (Malenovský et al., 2013). Secondly, the error of each observation was determined and an optimal threshold value was selected as the one that minimized the systematic RMSE. Thirdly, the observations above a certain error threshold (model prediction vs. observations) were considered as erroneous. Once the erroneous VGI observations were identified and removed, the performance of the statistical phenological model and of the data-driven model were assessed by calibrating the models with the non-erroneous observations. The data-driven model was also compared with the statistical model taken from literature to see the added value.

Finally, the model was fitted into the time-series data on yearly basis. The standard deviation of the prediction results of each year from all the years were determined and mapped. The variability in prediction result over the years could be associated either with heterogeneity in climate or absence of monitoring volunteers. Hence, areas showing higher variability in prediction results were considered as spatially-biased. In addition, once the erroneous VGI observations were removed, the new sets of observations were used to calibrate the phenological model and to make predictions. Additionally, the modelling results were spatially interpolated to determine the appropriate temporal windows for monitoring the phenophase of a plant species.

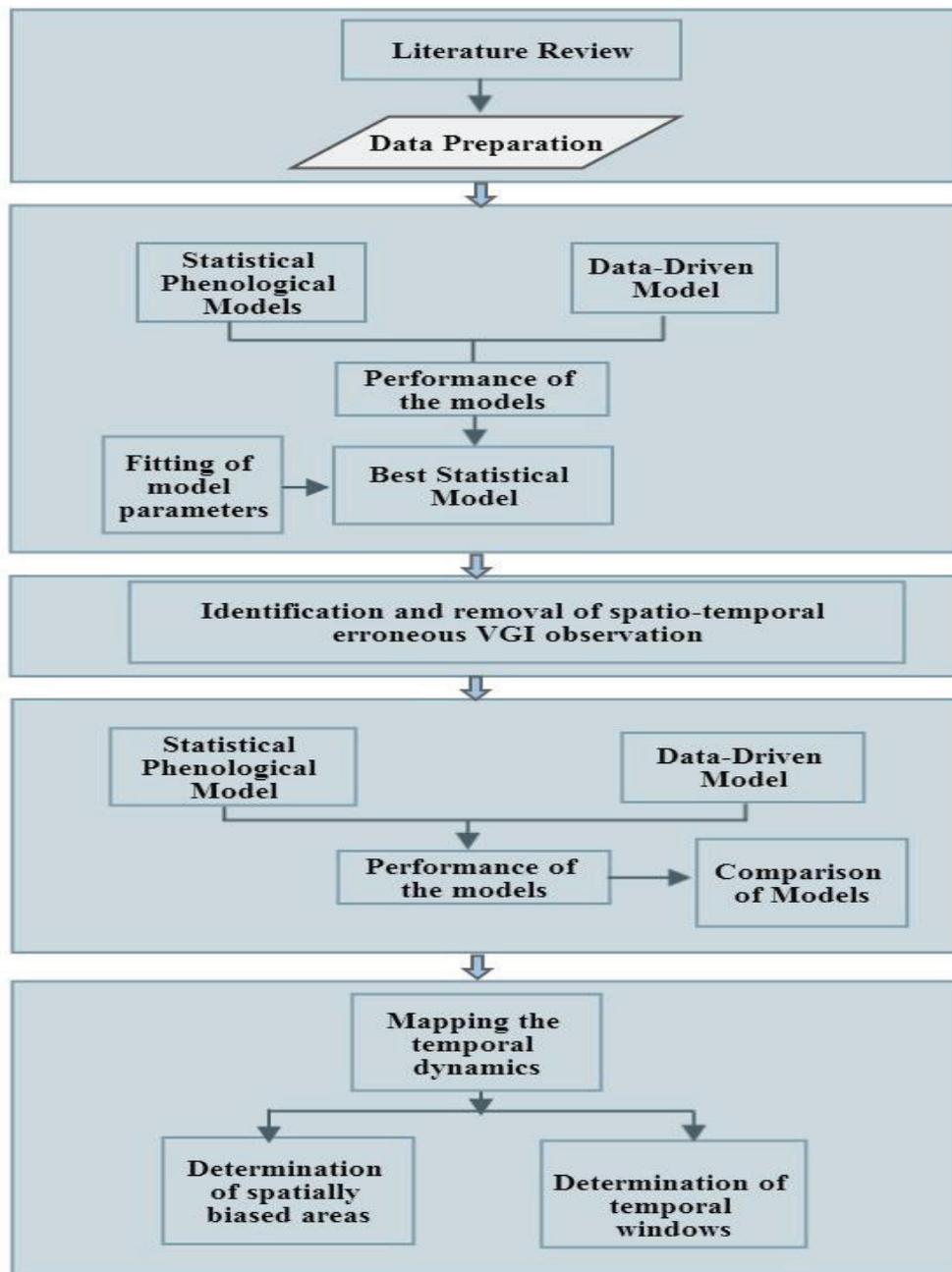


Figure 1.1 Methods to be adopted for attaining the research objective

1.4. Overview of the Thesis research

The thesis is composed of five main chapters. Chapter one of the research explains the motivation and problem statement, research objectives, questions, related work and the method adopted. Chapter two gives a brief explanation of works related to the research and to the methods adopted. Chapter three describes the materials that were used to undertake the research and the methods that were used to answer the research questions. Chapter four presents the results of the major findings of the research. In chapter five, the meanings and importance of the findings are discussed. These findings are also related with other similar findings in literature. Chapter six presents conclusions of the research and future recommendations.

2. LITERATURE REVIEW

2.1. Overview

This chapter gives background information about the proposed research by creating a link with researches that have already been completed. The review of the related works on Volunteered Geographic Information (VGI) justifies the importance of conducting the research and the possible analysis that can be applied on data collected by citizens with no formal expertise in the area. In addition, a sufficient review of existing methods of analysis is discussed to explain the choice on the type of analysis as well as the methods.

2.2. Importance of Volunteer-based Geographic Information

In previous years, geographic information (GI) was used by small group of people for specialized purposes and data were solely accessed by expert users using a particular GI tool. Similarly, data providers, developers and users belonged to the same group of people who were well trained in a specific GI application and who had background knowledge in geographic information (Brown et al., 2013). The use of GI was also limited to paper maps but then shifted to digital maps, spatial databases and web-based applications, respectively (Harding et al., 2009).

The availability of GI through internet and advances in ICT have made a large amount of diverse information and GI applications accessible to a wide range of people for various goals and context of uses (Silvertown, 2009). This has, in turn, contributed to the emergence of VGI, which is a GI shared publicly by a wider community of people (Brown et al., 2013). VGI has enabled users to access and contribute up to date information with wide spatial and temporal coverage to advance scientific knowledge. Global changes in species distribution, climate change and variations of biodiversity, temporal trends for recognizing the decline of species throughout the world are detected using VGI (Silvertown, 2009). Similarly, scientific journals and peer-reviewed papers are published using the data collected by volunteers (Jiguet et al., 2012). In addition, the involvements of volunteers in monitoring activities have economic contributions. Volunteer participation in biodiversity monitoring in France has saved 678,523 and 4,415,251 euros per year which would have been invested in the absence of volunteer participation (Levrel et al., 2010). However, the data collected by volunteers is mostly non-structured and the quality is unknown. Data credibility, non-comparability of the data, data completeness and logistical issues are some of the challenges of VGI (Gouveia et al., 2004).

Citizen science projects engage volunteers for collecting various kinds of information. One of such areas involving volunteers in collection of geo-tagged reports about specific incidents is called VGI (Bordogna et al., 2013). These incidents include reports on natural disasters, crime and environmental phenomena (Yanenko and Schlieder, 2012). For instance, Ushahidi, is an application that enables volunteers to give information on political crisis or natural disaster (Roche et al., 2013). Similarly, NoiseTube project³ involves citizens to monitor noise pollution by collecting noise information about their local environment (Maisonneuve et al., 2009). Designing a citizen science project starts from formulating a scientific question up to measuring the outcome of the project to ensure if the monitoring objective is achieved (Bonney et al., 2009).

³ <http://www.noisetube.net>

Furthermore, volunteer observers participate in phenological monitoring networks as they can only identify organisms reliably to the species level (Hochachka et al., 2007). Phenological monitoring includes tasks that are executed to collect geo-information about phenological phases of species. The phenological monitoring networks are designed for specific monitoring objectives and a particular area ranging from regions to continental or global scale. The observers make reports about phenological phases using web-based maps or a global positioning system (GPS).

Phenological information is important to recognize the potential impact of climate change on the spatial and temporal distribution of the species (Richardson et al., 2013). Additionally, they are useful for decision-makers to organize agricultural activities that require advanced information on the development stages of plants. For example, citizens collect phenological bird data by participating in projects, like hummingbirds.net, Journey North and eBird (Kelling et al., 2012). US National Phenology Network⁴ (Sullivan et al., 2009), Alberta and Canada PlantWatch programs⁵ (Beaubien and Hamann, 2011) engage volunteers for collecting information about the phenological phases of a wide range of plant and animal species. Hence, different kinds of analysis can be applied on volunteer phenological observations to extract valuable information, patterns and trends through time and space. In addition, phenological observations can be associated with other external factors, like climate to determine the spatio-temporal effect of climate change on the phenological stages of organisms.

2.3. Quality of Volunteer-based Geographic Information

Geographic information collected by volunteers have variability or bias as it engages citizens with different level of expertise (Kelling et al., 2012). In addition, the spatial and temporal variability of VGI results in different values of the parameters that are measured. Correspondingly, the credibility of data collected by volunteers is also dependent on the motivations of the contributor (Flanagin and Metzger, 2008). Thus, the credibility of such geo-information should be carefully investigated prior to analysis and presentation of results.

Citizen science projects, which involve a large number of volunteers, have unequal distribution of volunteers in space (spatial bias). Some regions are populated with a high density of volunteers, while others have a small number of volunteer distributions. Kelling et al. (2012) identified under-sampled locations in bird watching by using an active learning approach. The approach considers areas with higher uncertainty and lower density as under-sampled. This was used to give volunteers a context on sampling locations that need monitoring. In addition to the spatial bias, there also exists a temporal bias, which is a deviation from the actual occurrence of events in Volunteered Geographic Information (Bird et al., 2013).

A balance between the quantity and quality of data collected by volunteers is important to get valuable information from citizen science data (Hochachka et al., 2012a). Motivating volunteers by giving rewards and having less complicated data collection protocols can help increase the quantity of data collected by volunteers. On the other hand, many efforts have been made to guarantee the quality of volunteer data collections. In the case of phenological observations, the quality of observation varies with different species, phase, year, location and environmental conditions. Beaubien and Hamann (2011) used a variance partitioning approach to partition the total variance in the phenological datasets into components. The causes of the residual variation in bloom date are partitioned into year, species, phase, location, and

⁴ <http://www.usanpn.org/>

⁵ <http://www.naturewatch.ca/>

observer error considering their years of observing experience. This approach helps in determination of residual variance that is caused by observer error versus by climatic conditions.

Moreover, an automated emergent filter process was developed to improve the quality of data collected by bird watchers in eBird citizen science project (Kelling et al., 2012). In this project, the accuracy of volunteer data on the identification of the bird species is unreliable. The developed emergent filter considers the levels of expertise of the users (experts and novice birders) to identify outlier (unusual) observations that need review.

In addition, the calculation of site level anomaly and trend statistics were suggested as an approach to avoid the spatio-temporal bias existing in volunteer data (Alemu et al., 2013). Statistical analysis between spring, winter, and mean annual temperature and first flowering day in Canada has showed the advances in blooming day (earlier blooming day) due to climatic changes. Inter-annual variability in bloom dates, i.e. earlier or later bloom dates were considered as anomalies. Thus, the identification of these anomalies can help detect the bias in VGI. Moreover, weekend bias, which is a temporal bias caused by having more observations in the weekends than week days, affects the quality of data collected by volunteers. Courter et al. (2013) investigated the temporal bias in volunteered phenological information caused by weekend bias and its implications on phenological studies.

On the other hand, a constraint based approach was used to compensate for the missing and incorrect values existing in volunteer observations (Yanenko and Schlieder, 2012). This approach creates a constraint based on the principle of spatio-temporal proximity. The constraint satisfaction approach has a set of variables and constraints. That is, volunteer observations in same area and a given time delay can confirm each other and are consistent. The constraint graph was drawn for plants of the same species, phenophase and by considering climatic variability in air temperature, elevation, and latitude and plant position. Thus, this approach was used to avoid the spatio-temporal bias existing in VGI. In addition, Bird et al. (2013) used statistical modelling techniques and machine learning tools as the analytical ways of determining the variability and bias existing in data from citizen science projects. They defined random error as the variation in the response variable which is not caused by the explanatory variables. The random error is a mistake that is a result of observation error and which is not included in a model.

In this research, temporal bias is the deviation of the observer report on date of the phenological phase from actual date of the phenological event of the species, whereas spatial bias is the variation of volunteer distributions across space, which is associated with the absence of volunteers at important locations. In addition, for each species, areas showing inter-annual variability in the dates of the phenological event across space are considered as spatially-biased. Hence, it is important to detect the spatio-temporal bias existing in volunteered geographic information to increase the quality and usefulness of the data.

2.4. Spatio-temporal Analysis of VGI Data

The active involvement of volunteers in citizen science projects had increased the availability of spatio-temporal data. In connection to this, there is a dire need for data analysis techniques to extract hidden information from volunteered geographic information. Numerous data-mining tasks can be applied on spatio-temporal data collected by volunteers to achieve the complete potential of citizen science projects. To begin with, identification of spatial and temporal patterns requires exploration of volunteer geographic information. The understanding of the patterns that exist in spatial data is important for determining the cause of the spatial and temporal distributions of events. Siebert and Ewert (2012) used phenological information about oats obtained from volunteer observations to identify spatio-temporal patterns of

phenology in response to temperature and day length. In addition, geo-visualization was used to detect patterns by enabling visual exploration of spatio-temporal datasets (Kim, 2009).

Alternatively, point pattern analysis helps to identify patterns existing in spatio-temporal datasets. The patterns that exist in point data are grouped into complete spatial randomness (CSR), regularity and clustering. The methods used to determine the spatial pattern existing in point datasets are categorized into distance-based and area-based techniques (Haggett et al., 1977). The distance-based techniques consider the spatial dependence (spacing) that exist between neighbouring point patterns. In contrast, area based techniques consider the density of point patterns in a specified region of the study area. Li and Zhang (2007) tested four methods of spatial pattern analysis for analysing the spatial distributions of tree locations. In addition to geographic location, point patterns consist of additional marks (attributes) which characterize them (Stein and Georgiadis, 2006). In the cases of phenological monitoring activities, the additional marks are the times or dates of observations.

Clustering, which involves grouping similar data into one cluster (group), can be applied to spatio-temporal VGI data. Miller and Han (2009) applied cluster analysis algorithms to determine datasets that can be grouped into a cluster. Additionally, classification and regression can be applied to spatio-temporal datasets. Classification involves dividing data into different classes based on attribute values of variables (Rodriguez-Galiano et al., 2012). On the other hand, regression analysis can be used for predicting the values of dependent variable from one or more independent variables in spatio-temporal datasets (Mennis and Guo, 2009). It also helps in determining the type of relation that exists among these variables. Regression analysis can be done by using either the parametric or machine learning (non-parametric) approaches. For instance, phenological models use regression analysis for predicting the date of phenological event based on climatic conditions (Luedeling and Gassner, 2012).

Spatio-temporal analysis methods, which consider the spatial and temporal dimension of data, are applied on volunteered geographic information to extract useful information. These techniques are applied on the phenological information collected by volunteers to determine spatio-temporal patterns existing in phenology through the years. In addition, the analysis techniques help determine the phenological variability existing in different regions that are associated with climatic variability. However, the usefulness of the information extracted from VGI is dependent on the quality of the data. Therefore, it is important to consider the spatio-temporal biases existing in volunteer observations to guarantee the usefulness and quality of data.

2.5. Review of Statistical Phenological Models

Phenological models are used for understanding climate change and its impact on the phenology of a species (Sacks and Kucharik, 2011; Tao et al., 2012; Vitasse et al., 2011). These models are species-specific and are applied to specific species and climatic conditions. In addition, phenological models compensate for the absence of phenological records by predicting the dates of phenological events for areas that don't have information (Miranda et al., 2013). Some phenological models base on the principle that phenological development is solely influenced by temperature (Vitasse et al., 2011). Such models use the chilling or forcing heat requirement of species to determine the day of phenological event. Spring warming is one of the phenological models that consider the effect of forcing temperature (Hunter and Lechowicz, 1992). Spring warming model uses a degree-day approach, which assumes a linear relationship between accumulated temperature and date of the phenological phase. In addition, they assume that a phenological

stage occurs when a certain amount of heat units above a base temperature are accumulated. These models are preferred as they have a good predictive accuracy (Schwartz, 1999).

On the contrary, other phenological models consider the effect of chilling requirement during the dormancy period as a major regulating factor of crop phenological stage. The chilling hours model uses the hourly temperature below 7 °C to calculate the accumulated sum of chilling requirements (Valentini et al., 2001). Another version of the chilling hour's model, Utah model, considers the negative effect of having higher temperature on the accumulation of chilling temperatures. To overcome this, weights were given for temperatures above 15.9 °C (Luedeling, 2012). These models are not much useful for species with medium or low winter chill requirement. Dynamic model assumes sequence of temperature accumulation for chilling requirement (Fishman et al., 1987). The first phase involves accumulation of intermediate chill temperature at low temperature. This phase is changeable as the intermediate temperature can change if high temperature is encountered (second phase) (Luedeling, 2012). On the other hand, sequential models consider both chilling and forcing temperature requirements for predicting phenology of a species (Luedeling, 2012).

Several researches have used phenological models to study phenology of species. Zhang and Tao (2013) compared five phenological models to predict rice phenology in conditions of climatic changes and variability at different regions. The uncertainties of these five phenological models were calculated and the performance of the models varied at different regions. Miranda et al. (2013) used three versions of critical date models and five versions of critical chill models to predict peach bud development. The critical date models only consider the effect of forcing temperature on phenology while the critical chill models consider the effect of chilling temperature. Parker et al. (2011) used spring warming, UNICHILL and UNIFORC model to predict the flowering and verasion stages of grapevine. Spring warming and UNIFORC only consider the effect of forcing temperature while the UNICHILL considers the effect of forcing and chilling temperatures. Unlike the spring warming with linear response, the UNIFORC model has a sigmoid response to temperate (Chuine and Cour, 1999; Vitasse et al., 2011). Lastly, the UNICHILL model considers the effect of chilling temperature until a critical stage of chilling to break dormancy is reached. Following this, the forcing temperature will start to accumulate until critical stage of forcing is reached.

Phenological models take volunteered geographic information (i.e. reports on dates of observation of phenological stages) and climatic variables as input to make predictions on the dates of occurrences of phenological stages. However, these phenological models have systematic defects that reduce the performance of the prediction results. This can be attributed to various factors like the uncertainties in the model parameters leading to non-meaningful values or absence of appropriate predictor variables of phenology. (Blümel and Chmielewski, 2012) improved the performance of spring warming model that predicts apple blossom day by considering day-length in addition to temperature which was solely used in other versions of spring warming models. In addition, they also fitted the model parameters to minimize the RMSE, which is the difference between the observed and predicted values. In addition, the quality of the VGI has a direct influence on the prediction results of phenological models. The existence of spatiotemporal bias in the volunteer observations cause either an over/under predictions in the modelling results. Similarly, the modelling results are dependent on the spatiotemporal analysis method which is used to find the relation between the dependent and independent variables. Therefore, the bias in VGI should be accounted when calibrating phenological models to avoid systematic defects and obtain accurate results.

2.6. Review of Data-Driven Phenological Models

In addition to the statistical modeling techniques, data-driven approaches are used to find the relationship between dependent and one or more independent variables. Data-driven modelling approaches base on computational intelligence and machine learning algorithms to find relationship between dependent and independent variables (Solomatine et al., 2008). They are based on a heuristic algorithm which does not make a prior assumption about the distribution of the data. Hence, data-driven modelling relies on the analysis of the data describing the system under study. For this reason, they can capture more complex relationships than the statistical phenological models. Similarly, they are more robust to the existence of unusual (outliers) observations in the data.

There are different types of data-driven modelling techniques. For instance, artificial neural networks and support vector machines train from empirical datasets to model sophisticated non-linear relationships (Cheng et al., 2012). Similarly, some data-driven modelling techniques like, Random Forest, use decision tree algorithms (DTA) for determining the dependence between variables (Abdel-Rahman et al., 2012). DTA splits the feature space of explaining variables until the resulting partition show the best statistical correlation. Moller et al. (2011) developed a temperature vegetation model (PHASE) by using random forest regression to predict the dates of phenological stages from daily temperature values in Germany. Therefore, data-driven modelling techniques can be applied to model the relationship between volunteer phenological observations and environmental information. They enable the exploration of the spatio-temporal characteristics of VGI data. In conclusion, statistical and data-driven modelling techniques are used to analyse volunteer phenological observations and climatic conditions. In addition, these models are used to detect the spatio-temporal bias existing in the observations.

3. MATERIALS AND METHODS

This chapter explains the data needed for doing the analyses and the corresponding methods that were applied to the dataset.

3.1. Materials

3.1.1. Phenological Data

The main parameters of this study were the flowering dates of four plant species, namely Cow Parsley, Lesser Celandine, White Horse Chestnut, Wood Anemone and one leafing species Common Oak (Table 3.1). Phenological data on the first phenological dates are collected by volunteers from different locations in The Netherlands. A total of data of eight years, from 2003 up to 2010, was obtained from “Natuurkalender”, which is a Dutch phenological monitoring network established in 2001 (Natuurkalendar, 2013).

The Dutch phenological monitoring network engages volunteers for obtaining information on phenological phases to study climatic variations, analyse, forecast and communicate the timing of phenological events. The data obtained from the “Natuurkalender” includes the location and time of observing a given species and phenophase (Table 3.2). In this research, the phenological observations on the first flowering dates of Cow Parsley, Lesser Celandine, White Horse Chestnut and Wood Anemone and the leafing date of Oak were used for calibration and validation of phenological models. In addition, phenological observation were analysed to check the quality of the data and existence of spatio-temporal bias. The phenological observations made by volunteers on the dates of the year (DOY) of the phenological events are shown (Figure 3.1). Also, the phenological data covered the whole area of The Netherlands.

Table 3.1. Name of Species, year and number of observations of the phenological stage of the five plant species in the “Natuurkalender” dataset

| Species Name (English) | Species Name (Dutch) | Species Name (Scientific) | Years | No of Observations | Stage |
|---------------------------|---------------------------|-----------------------------------|-----------|-----------------------|--------------|
| Cow Parsley | Fluitenkruid | <i>Anthriscus sylvestris</i> | 2003-2009 | 1185 | First flower |
| Lesser Celandine | Speenkruid | <i>Ranunculus ficaria</i> | 2003-2009 | 2080 | First flower |
| White Horse Chestnut | Paardekastanje (witte) | <i>Aesculus hippocastanum</i> | 2003-2009 | 601 | First flower |
| Wood Anemone | Bosanemoon | <i>Anemone nemorosa</i> | 2003-2009 | 827 | First flower |
| Common Oak | Eik (zomer) | <i>Quercus robur</i> | 2003-2009 | 384 | First leaf |

Table 3.2. A record showing instances of volunteer phenological observation.

| Species Name | X (Coordinate) | Y (Coordinate) | Phenophase | Date | Place | Year |
|---------------------|-------------------|-------------------|--------------|------|-----------------------------------|------|
| Cow Parsley | 109000 | 402000 | First Flower | 86 | Breda | 2004 |
| Lesser Celandine | 168000 | 391000 | First Flower | 69 | gemeente LAARBEEK, de schop | 2007 |

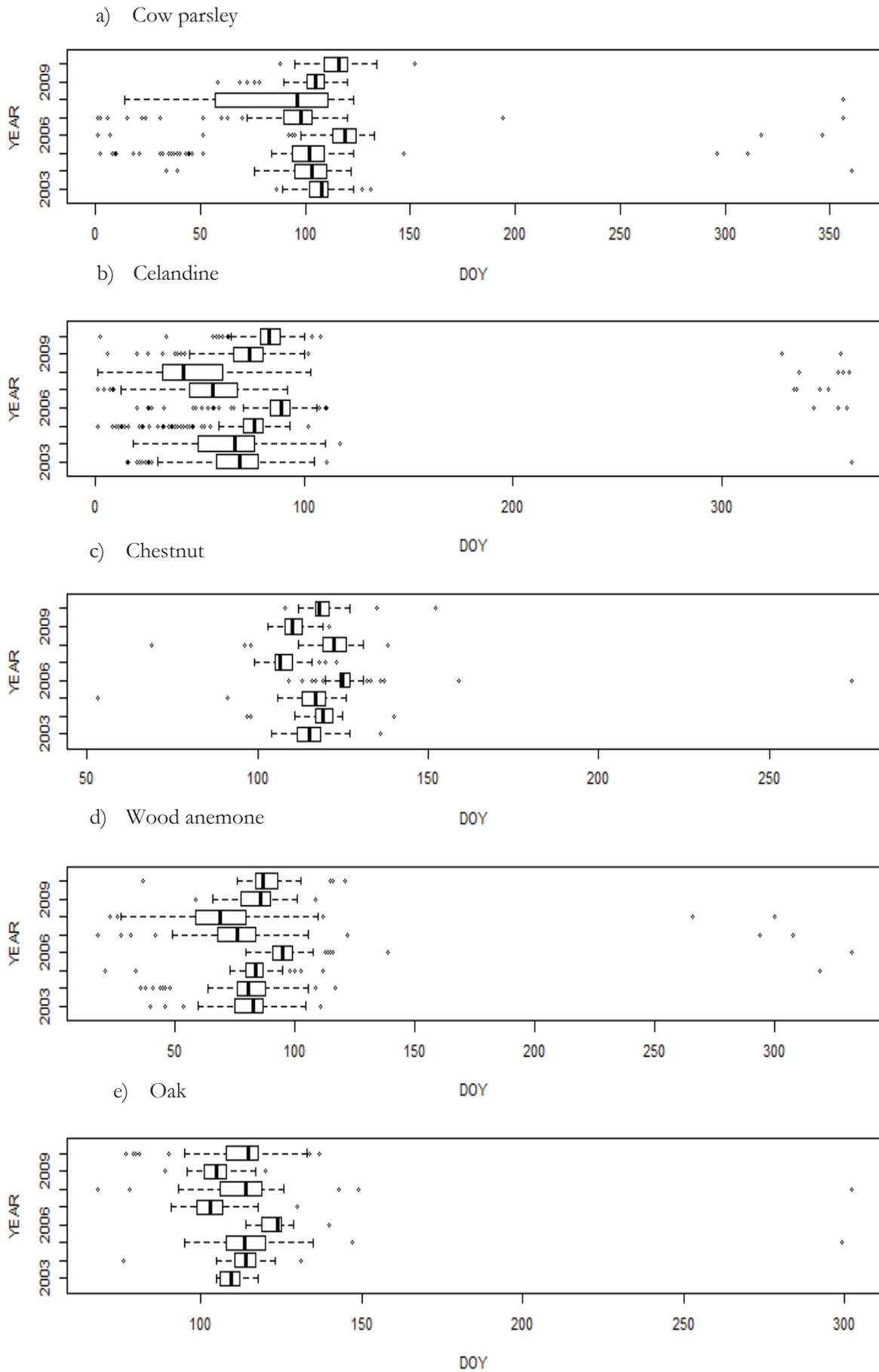


Figure 3.1. Box plots showing the yearly volunteer observations on the DOY of the phenological stage for the five plant species

3.1.2. Observed meteorological data

To model the date of flowering or leafing and to calibrate the models, we used air temperatures observed by the Royal Netherlands Meteorological Institute (KNMI) from 2003 up to 2010. KNMI is a national institute that forecasts weather information to the public (KNMI, 2013). It collects daily maximum, minimum and average values of different climatic variables, like temperature, precipitation and evapo-transpiration.

This research used the daily average temperatures that were interpolated in a raster file of size 1km for the whole area of The Netherlands. The interpolation is based on the measurements made by 500 meteorological stations. This interpolated temperature is used to calibrate the phenological models and later on to predict the DOY of phenological event across the entire study area.

3.2. Methods

This section describes the methods that were applied to do data analysis and complete the objective of the research (Figure 3.2). The data analysis is done by using the statistical computing and graphics software R (Grunsky, 2002). In addition, the codes for reproducing all analyses are provided as supplementary materials of the research.

Firstly, observations that were incomplete (i.e. '0' values for x and y coordinates) were removed from the dataset. Then, the dataset was split per species for an independent analysis within a species. For each species, the dataset was further divided into model building and validation. Secondly, phenological models were calibrated to model volunteer phenological observations in response to daily temperature values. Location information provided by volunteers was used to extract daily temperature values at the year of phenological observation. Three statistical and one data-driven phenological models were calibrated with daily temperature values and phenological observations to predict the flowering dates of four plant species and the leafing date of one plant species. In addition, the parameters of the statistical models were fitted by using an optimization algorithm, i.e. simulated annealing. After that, quality measures (RMSE, efficiency and correlation) were used to compare the three statistical models and to choose the best performing model.

Thirdly, temporally-biased volunteer phenological observations were identified by using the phenological models. Specifically, the volunteer phenological observation dates were compared with dates predicted by the models and the error of each observation was analysed to detect temporally-biased observations. Then, biased observation sets were discarded and the improvement in the systematic and unsystematic root mean square error (RMSE) of the model was checked. Fourthly, the phenological models were used to predict the DOY of flowering of the four plant species and leafing of one species between the years 2003 and 2010. As a result, the phenological dynamics was detected to determine the spatial bias across different regions of The Netherlands. Generally, phenological models were used to detect temporal bias in the volunteer phenological observations and spatial bias in the whole area of The Netherlands.

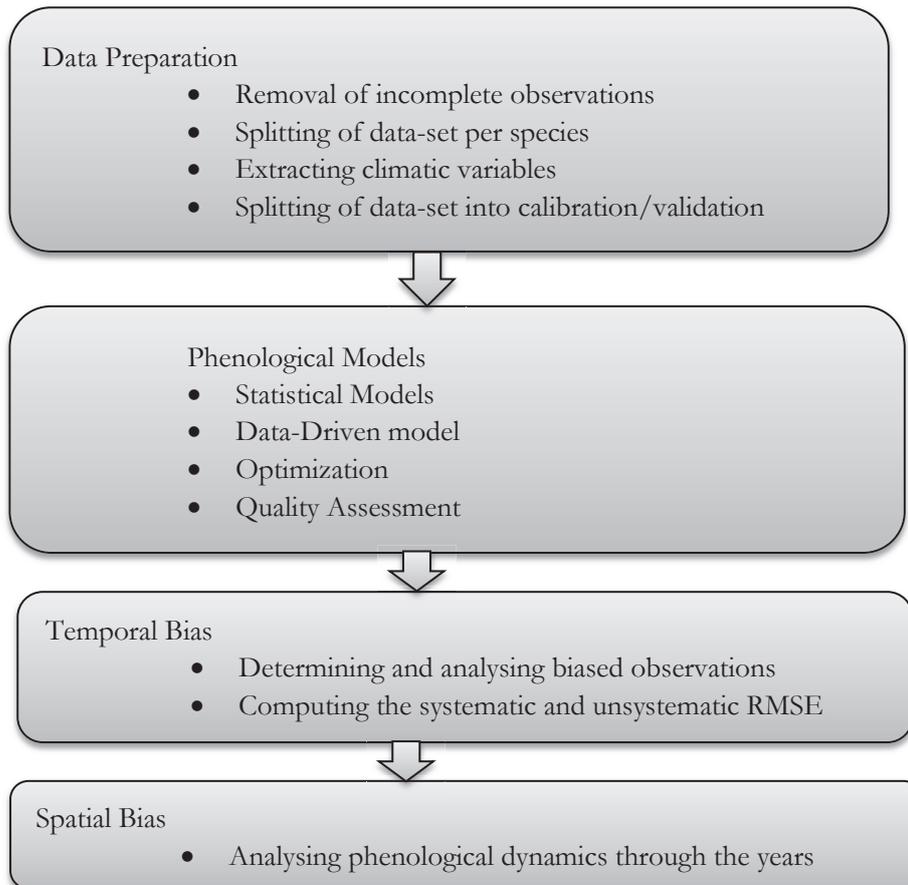


Figure 3.2. The adopted method for answering the research questions

3.2.1. Data preparation

In this step, incomplete phenological observations with ‘0’ values for the x and y coordinates were discarded from the dataset. Then, the dataset was split into two independent datasets, calibration and validation, per species. Splitting the data into calibration and validation helped in assessing the robustness of the models. The calibration data was solely used for building the model and the validation data were independent set of data to check the predictive accuracy of the model. About 70% of the data from all years were used for model calibration. The remaining 30% of the data from even years were used for validation purpose. Finally, climatic variables, in particular daily temperature values for the locations and years of the phenological observations, were extracted.

3.2.2. Statistical phenological models

Phenological models provide an assessment on the adaptation of species to climatic variations and the impact of climate change on their future performance. Two different phenological modelling approaches, statistical (parametric) and data-driven (non-parametric), are tested in the study. The phenological models are used to predict the first flowering dates of four plant species and leafing date of one plant species based on daily average temperature values. These models are calibrated and tested using independent sets of volunteer phenological observations.

3.2.2.1. Forcing model

Temperature is the main factor affecting the development stages of plants (Chuine et al., 1999). In connection to this, some phenological models consider the effect of warm (forcing) temperature required to reach a phenological stage. Two forcing models that consider the effect of warm temperature on phenological development of plant species are used in this study. The models are described in section 3.2.2.1.1 and 3.2.2.1.2.

3.2.2.2. Thermal time (Spring Warming) model

Spring warming model assumes that a phenological stage occurs when a plant accumulates a sufficient amount of heat sums (growing degree days) above a base temperature (Chuine et al., 2003). Hence, it uses a linear sum of daily temperature above the threshold temperature to calculate the heat sum requirement of a phenological stage (i.e. flowering or leafing). If daily temperature of the plant is below the base temperature, the plant stops growing and has a rate of forcing of zero (Eq. 1). Otherwise, the non-zero daily rate of forcing would be accumulated from a starting day until the date of phenological stage (Eq. 2). This model implicitly assumed that cold temperature was fully released before the starting date of forcing units' accumulation. The rate of forcing of the spring warming model is calculated as follows:

$$R_f(T_d) = \begin{cases} 0 & \text{if } T_d < T_b \\ T_d - T_b & \text{if } T_d \geq T_b \end{cases} \quad (1)$$

where T_b is the base temperature (minimum temperature required for development), T_d is daily average temperature and $R_f(T_d)$ is the daily rate of forcing, which is a function of T_d . The state of forcing, which is the summation of daily rate of forcing, from T_0 until T_s is calculated as in equation 2.

$$S_f(t_s) = \sum_{T_0}^{T_s} R_f(T_d) \Delta t \geq F^* \quad (2)$$

where Δt denotes the time span of calculating the rate of forcing, i.e. 1 day; T_0 indicates the starting day of accumulating a temperature; T_s represents the date where a critical phenological stage (flowering or leafing) reached and F^* shows the amount of heat that needs to be accumulated by the plant to reach the state of forcing ($S_f(t_s)$). Hence, spring warming model had three parameters, T_0 , T_b and F^* , which were fitted and determined by an optimization process.

3.2.2.3. Thermal Time (UNIFORC) Model

The UNIFORC model considers the effect of forcing temperature on phenological stage (flowering or leafing) of plants. However, the rate of forcing in the UNIFORC model was calculated by assuming a non-linear (sigmoidal dependency) of rate of forcing on temperature (Chuine and Cour, 1999; Parker et al., 2011) (Eq. 3). The sigmoidal function maps the entire range of daily temperature values to a domain that ranges between 0 and 1.

$$R_{f(T(d))} = \begin{cases} 0 & \text{if } T_d < 0 \\ \frac{1}{1 + e^{d(T_d - e)}} & \text{if } T_d \geq 0 \end{cases} \quad (3)$$

Therefore, the UNIFORC model had four parameters (d , e , T_0 , F^*) that were fitted by an optimization process. The parameters d and e correspond to the slope at the inflection point (width) and the temperature of mid-response (centre) of the sigmoidal function, respectively (Vitasse et al., 2011).

3.2.2.4. Sequential (UNICHILL) Model

Many plants are assumed to require cold temperatures during the dormancy period to initiate flowering (Luedeling et al., 2009). In connection to that, sequential models consider the effect of cold and warm temperature required by a plant to reach a critical phenological stage (Cannell and Smith, 1983; Luedeling, 2012). A plant first accumulates sufficient cold temperature starting from a fixed date (September 1) until it reaches a state of chilling ($Sc(tc)$). The state of chilling is computed between two dates, T_0 (start) and T_c (end), by assuming a nonlinear dependency of rate of chilling on temperature. The state and rate of chilling were calculated as in equation 4 and 5, respectively. Once the state of chilling is reached, the plant starts accumulating warm temperature to reach a state of forcing. The rate of forcing of the UNICHILL model was calculated by using the sigmoidal function of the UNIFORC model (Eq. 3).

$$S_c(T_c) = \sum_{T_0}^{T_c} R_c(T_d) \geq C^* \quad (4)$$

with T_d the daily temperature, $R_c(T_d)$ the daily rate of chilling calculated as follows:

$$R_c(T_d) = \frac{1}{1 + e^{a(T_d - c)^2 + b(T_d - c)}} \quad (5)$$

In sequential models, chilling and forcing temperatures have independent and successive effects on phenological stages of the plant. Once the chilling requirements are met, forcing temperature start to accumulate from the date where the critical stage of chilling (T_c) has reached. In this case, the forcing requirement (F^*) is not fixed rather dependent on the amount of chilling temperature (C^*) that has been accumulated.

The UNICHILL model has a, b, c and C^* as fitted parameters for determining the chilling requirement and d, e and F^* were fitted for computing the forcing requirement of the plant. Therefore, this model enabled us to determine the amount of chilling units that were accumulated starting from September 1 to the date of dormancy and the amount of forcing units accumulated from the date of dormancy to the date of the phenological event.

3.2.3. Simulated annealing

An optimization algorithm was needed to optimize the parameters of the three statistical models. For this reason, the global optimization algorithm, simulated annealing was chosen. Classical optimization algorithms have problems in convergence as they get trapped in local minima or maxima in the parameter space (Kramer, 1994). However, simulated annealing is known to find global optimal solutions of an objective function and it does not get trapped into local minima or maxima.

The name simulated annealing is derived from the annealing (cooling) process of metals or from the liquid freezing and crystallization processes (Bohachevsky et al., 1986). In metallurgy, molten metal is annealed to reach at a state of equilibrium by lowering the temperature slowly. Similarly, the objective function in simulated annealing resembles the energy function to be minimized in the cooling process of a metal. Hence, simulated annealing has a temperature that is first set to be high and cooled slowly to find optimal results of the objective function (Kirkpatrick et al., 1983). In this research, the optimization algorithm minimizes the RMSE between the observed and predicted values and returns parameter values that gave the minimum results (Eq. 6).

$$\text{RMSE}_{(\text{opt})} = \frac{1}{N} \sum [(t_s(\text{predicted}) - t_s(\text{observed}))^2]^{1/2} \quad (6)$$

Where t_s is the date of phenological event (i.e. flowering or leafing) and N is total number of observations

The configuration setting of the algorithm affects its performance and choice of parameters (Park and Kim, 1998). Firstly, the annealing or cooling schedule of the algorithm has to be set. The schedule involves setting the initial temperature, stopping condition, lower and upper bounds of the parameters of the function to be minimized. Once the initial temperature is set, the algorithm computes the solution with the specified condition. The solution will then be compared with the objective function that needs to be minimized. This process continues iterating until some defined stopping condition is met.

In this research, we used the package GenSA (Generalized Simulated Annealing) for determining the optimal parameter values of SW, UNIFORC and UNICHILL model. The GenSA package finds the global optimum of the objective function among all the possible local optimum values (Yang Xiang, 2012). Once the optimal parameters are determined, the three statistical models were used to predict the first flowering dates of the four plant species and leafing date of one plant species. Hence, the optimization algorithm help determine the optimal parameter sets which are different for each model and species.

3.2.4. Quality Assessment

3.2.4.1. Statistical Models

Once the model parameters were fitted, the best statistical model was selected based on two criteria (Efficiency and RMSE), which are widely used in plant models (Bellocchi et al., 2010). Efficiency (EF) is a measure of the proportion of variance that the model has explained (Eq. 7). This value ranges from $-\infty$ to 1. If the efficiency is 1, then there is a perfect match between the observed and predicted values. If it is 0, the predictive ability of the model is as accurate as the mean value. Otherwise, a value less than 0 indicate that the observed mean values are better than values predicted by the model. Hence, an efficiency value closer to 1 is more accurate than 0 or negative values.

$$\text{EF} = 1 - \left(\frac{\sum_{i=1}^n (S_i - O_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \right) \quad (7)$$

with S_i , simulated values and O_i , observed values and n , number of observations

On the other hand, the RMSE indicates the average error that is produced by the model. RMSE is divided into its systematic and unsystematic components to determine the sources of error produced by the model (Willmott, 1981) (Eq. 8, 9 and 10). The systematic component indicates the bias between predicted and observed values of the model. Conversely, the unsystematic error is related with the quality of input observations. In this research, the quality of input observations is related with the biases (i.e. spatio-temporal biases) that exist in volunteer phenological observations.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}} \quad (8)$$

$$\text{RMSE}_s = \sqrt{\frac{\sum_{i=1}^n (\hat{p}_i - O_i)^2}{n}} \quad (9)$$

$$\text{RMSE}_u = \sqrt{\frac{\sum_{i=1}^n (P_i - \hat{P}_i)^2}{n}} \quad (10)$$

where O_i : observed value; P_i : predicted value of the model; $\hat{p}_i = a + bO_i$ [a (intercept) and b (slope) of linear regression between P_i and O_i]; n : total number of observations.

Hence, systematic and unsystematic components of RMSE were used to detect and remove temporally-biased phenological observations as explained in section 3.2.4.3. In addition to efficiency and RMSE, statistical correlation was used in comparing the performance of the three statistical models used in the study. Specifically, Pearson's correlation coefficient was used to calculate the degree of linear dependence that exist between observed and predicted values (Plata, 2006). Pearson's correlation coefficient ranges between -1 (negative correlation) and 1 (positive correlation). The Pearson correlation's coefficient is calculated as follows:

$$\rho_{xy} = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y} \quad (11)$$

Where ρ_{xy} (rho) is Pearson's correlation coefficient; $\text{cov}(x, y)$ is the covariance between x and y ; σ_x , σ_y is the standard deviation of x and y , respectively.

3.2.4.2. One factor analysis of variance (ANOVA) test

One factor ANOVA test was used for determining the variability among the mean of groups of data sets. ANOVA uses F-statistic, which is the ratio between group variance to within group variance, to check if the groups that are being tested have the same mean (Bluman, 2012). In the test, the null hypothesis assumes that the means of the two groups are the same and the alternative hypothesis assumes that the two groups have different means. In this research, ANOVA test was done between the calibration and validation dataset to evaluate if there is a difference in their mean values. Hence, if there is no difference in the mean values of the two groups, the null hypothesis would not be rejected.

3.2.4.3. Temporal bias

Liu et al. (2004) pointed out that the noise in the data should be reduced to avoid biased parameter estimation and incorrect modelling results. Additionally, the quality of input data should be assessed to give an interpretation of the phenomena under study. For this reason, the best statistical model was used to identify phenological observations with temporal bias. The phenological model was assumed to account for spatio-temporal variability in the DOY of the phenological event of the five plant species. Therefore, temporally-biased observations were not outliers that deviated from the data rather they were determined by a model that accounted for the environmental variability.

To determine the temporal bias, the error of each volunteer phenological observation was first computed as follows:

$$E = O_i - P_i \quad (12)$$

Where E represents the error of each phenological observation, O_i denotes the observed value and P_i shows the predicted value.

Secondly, the SD of the error of each observation from the mean of the errors of all observation sets was calculated by using equation 13. Thirdly, the errors of each observation were normalized by the standard deviation as shown in equation 14. The normalized SD had a negative value if the error is below the mean (zero) and positive if the error is above the mean.

$$SD = \sqrt{\frac{1}{N} \sum_{i=1}^N (E - \mu)^2} \quad (13)$$

Where N is the total number of observations and μ is the mean of observed values.

$$\text{Normalized_SD} = \frac{E - \mu}{SD} \quad (14)$$

The exclusion of outliers from phenological observation sets could be done in different ways. For instance, it is a common practice to remove observations above 3 (very conservative), 2.5 (moderately conservative) or 2 (poorly conservative) SD (Miller, 1991) or above 1.5 SD (Scheifinger et al., 2002) depending on the justification given by the researcher. We selected an optimal threshold value for normalized SD to detect temporally-biased phenological observations. However, we varied the SD threshold for each plant species by considering two conditions.

The first condition was the point where the systematic and unsystematic RMSE of the model were improved and there was a smaller systematic RMSE than the unsystematic RMSE. Systematic errors of the model showed the bias between the observed and predicted values. They were computed by assuming a linear relationship between the observed and predicted values, while non-linear errors were unsystematic. If the systematic error of the model was smaller than the unsystematic error, then the model could be assumed to have good predictive accuracy (Malenovsky et al., 2013; Willmott, 1981). The second condition was the point where the smallest numbers of observations were removed given that the first condition was fulfilled. This was because we want to keep phenological observations with useful spatio-temporal information. Hence, the normalized SD with systematic RMSE less than the unsystematic RMSE and the smallest number of observations removed was chosen as an optimal threshold value.

Once the threshold for the normalized SD was selected, observations above the threshold were discarded from the observation sets and were considered as erroneous (temporally-biased). Then, the non-biased observations were used to recalibrate the best statistical phenological model and obtain new parameter values.

3.2.4.4. Spatial bias

Inter-annual variability (spatial-bias) was measured in terms of inter-annual standard deviation in the DOY of the phenological events between the years 2003 and 2010 for the five plant species. The best performing model, together with its optimal parameter values, was used to predict the DOY of the phenological event (i.e. flowering or leafing) of each year in the whole area of the Netherlands. For each

species, the inter-annual variation in the DOY of the phenological event with respect to location was determined and used to determine areas with climatic variability (Vilhar et al., 2013). In our research, this was done by computing the standard deviation of the prediction results of each year at a given location from the other years in the corresponding location. This assisted in the determination of areas that had variability in the DOY of the phenological event, which were also considered as areas with climatic variability (spatially-biased). Once the variability (SD) in the DOY of the phenological event at each location was computed, we grouped the areas into high, medium and low variable based on the SD values. Thus, areas showing high, medium and low variability had a higher, medium and lower SD values, respectively.

The spatial distributions of volunteer phenological observations affect the quality of collected environmental information (Do et al., 2012; Reynolds et al., 2011). Hence, understanding the relationship between phenological phases and spatio-temporal variations in climate conditions require optimal spread of volunteers in space. We analysed the spatial spread of volunteers to identify areas that lack phenological observations. For this, the numbers of volunteers in the high, medium and low variable areas were first determined by calculating the average number of volunteers which were in the high, medium and low variable areas between the years 2003 and 2010. In addition, for each species, the areal coverage of high, medium and low variable areas in the whole of The Netherlands were calculated. As the medium variable areas were the average, it could be assumed that they had sufficient number of volunteers required for phenological monitoring. Thus, we used the numbers of volunteers in the medium variable areas to determine the expected number of volunteers in the low and high variable areas (Eq. 15 and Eq.16, respectively). Once the expected numbers of volunteers were determined, the required numbers of volunteers for low variable (Eq. 17) and high variable (Eq. 18) areas were calculated.

$$E_{low} = \frac{N_{Medium} * A_{low}}{A_{Medium}} \quad (15)$$

$$E_{high} = \frac{N_{Medium} * A_{high}}{A_{Medium}} \quad (16)$$

$$R_{low} = E_{low} - N_{low} \quad (17)$$

$$R_{high} = E_{high} - N_{high} \quad (18)$$

where E_{low} and E_{high} are the expected numbers of volunteers in the low and high variable areas, respectively. N_{low} , N_{medium} and N_{high} are the existing number of volunteers in the low, medium and high variable areas, respectively. A_{low} , A_{medium} and A_{high} are the areal coverage (in percentages) of the low, medium and high variable areas, respectively. R_{low} and R_{high} are the required number of volunteers in the low and high variable areas, respectively.

3.2.5. Data-Driven Model

The data-driven phenological model that was used for predicting the DOY of the phenological events is known as PHASE 2.0 (Section 3.2.5.2.). This model is an integration of three models which were developed by using Random Forest regression and ensemble-based decision tree algorithm (Section 3.2.5.1.).

3.2.5.1. Random Forest Regression

Random forest regression is an ensemble-based regression which uses decision trees to determine the relationship between predictor and response variables (Breiman, 2001). The ensemble-based regression obtains prediction results from multiple decision tree models. This gives it a better predictive accuracy than having a single decision tree model. Random forest regression uses 2/3 (70%) of the data for model calibration and the remaining 1/3 (30%) of the data is an Out of Bag data (OOB) to be used for model validation (Mendez and Lohr, 2011). Decision rules, which are randomly selected subsets of the predictor variables, are used to split the nodes of the tree. After the nodes of the tree are split, the terminal nodes are the ones that will be used for prediction purposes. The number of predictor variables (m_{try}) and trees (n_{tree}) are specified by the user.

3.2.5.2. PHASE 2.0

PHASE 2.0 model integrates three models that are built by using random forest regression (Moller et al., 2011). Model 1 interpolates the temperature values at unknown locations based on known temperature values at some meteorological stations. On the other hand, model 2 predicts the phase specific heat units (poHU) that need to be accumulated by a plant to reach a phenological stage (flowering or leafing). This model is developed by assuming relations between location-specific heat units (psHU) and unspecific heat units (uHU). For each location in the volunteer phenological observation set, location-specific heat unit was the summation of daily average temperature from a starting DOY (T_0) until the date of phenological event (Table 3.3). Similarly, uHU values were computed for each location as summations of daily average temperatures from day 1 up to the date of the maximum phase (Table 3.4). For each psHU, random forest regression was used to find the best set of explanatory variables among all the uHU values to predict the poHU.

Finally, model 3 is developed by having a random forest regression between psHU and volunteer phenological observations to predict the DOY of the phenological event. In this research, model 2 and 3 were used to determine the phase specific heat units and phenological dates (flowering date of four plant species and leafing date of one plant species), respectively. However, we did not use model 1 as interpolated temperature values for the entire area of The Netherlands were available.

3.2.5.3. Optimization of PHASE 2.0

The first version of PHASE model, PHASE 1.0, used expert knowledge to determine the optimal starting DOY of accumulating phase-specific heat units. Contrarily, PHASE 2.0, used an optimization procedure to determine the optimal starting DOY (Figure 3.3). The optimization process involved an iterative step for determining the optimal starting DOY of accumulating a temperature. Firstly, for each location of the phenological observation, location-specific heat units (psHU) were calculated by summing daily temperature values from day 1 (T_1) up to the date of phenological observation (T_{Date}) (Table 3.3). Then, model 2 found the relationship between psHU and uHU to predict the phase specific heat units (poHU) based on uHU values. Similarly, model 3 found the relationship between psHU and date of the phenological observation to predict the DOY of the phenological event based on poHU values. For each model and tree in the forest, the efficiency of the model was calculated for all the OOB samples (Eq. 7). If the forest had 500 trees, then the efficiency would have 500 values. The i^{th} efficiency of the forest corresponds to the i^{th} tree and the last value is the efficiency value determined by the whole forest.

Secondly, psHU values were calculated by summing temperature values from day 2 (T_2) up to T_{Date} . Then, model2 was built basing on the random forest regression between psHU and uHU. In the same way, model3 used the new psHU to find the random forest regression between psHU and T_{Date} . Lastly, the efficiency values of model 2 and 3 were computed. In general, psHU values were calculated from $T_i \dots T_{Date}$, where i was between 1 and maximum date of phenological event (max.ph). In addition, the

efficiency values of model 2 and 3 were calculated for every iteration in i . The values of Efficiency2 and Efficiency3 which were above zero were only considered. In this case, the optimal starting day was the i^{th} date where there was a maximum Efficiency2 * Efficiency3 value in all iterations of i . After obtaining the optimal starting DOY, location-specific heat units were accumulated by summing daily temperature values from the starting DOY until T_{Date} . Then, model 2 and model 3 were built by using the new psHU values.

Finally, the quality of random forest regression for model 2 and model 3 was assessed. Random forest regression returns the mean square error and the total amount of variance of the observed variable (efficiency) that was explained by the model for the OOB dataset. Additionally, we assessed the performance of PHASE model by calculating RMSE and efficiency of the model on the calibration and validation (test) dataset.

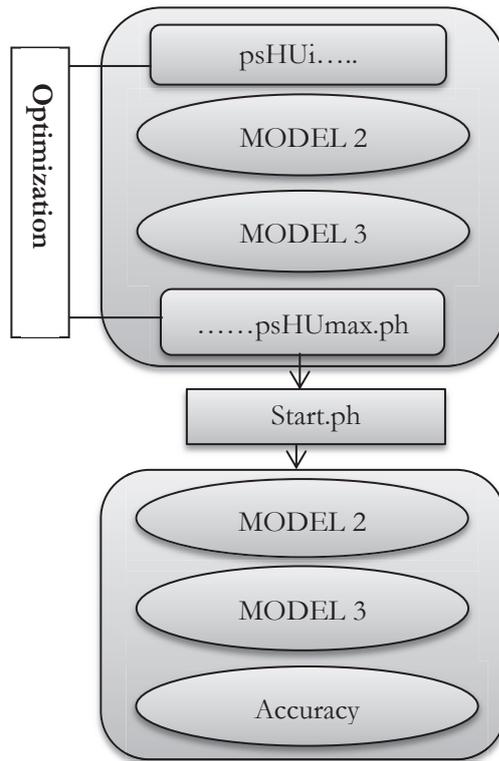


Figure 3.3. Optimization of the PHASE model. $psHU_i$ and $psHU_{\text{max.ph}}$ are the location specific heat units accumulated from a starting DOY of i and max.ph , respectively. Start.ph is the starting DOY determined by the optimization process.

Table 3.3. Computation of psHU from a starting DOY ($T_0 = 1$). Location (x,y) , denotes the x and y coordinates of the volunteers. Date, is the date of phenological observation. T is the temperature and the subscripts indicate the DOY.

| Location(x,y) | Date | T_1 | T_2 | T_3 | T_4 | T_5 | T_6 | T_{Date} | psHU |
|---------------|------|-------|-------|-------|-------|-------|-------|-------------------|--|
| x_1, y_1 | 80 | 4 | 5 | 7 | 9 | 2 | 5 | 8 | $T_1 + T_2 + T_3 + T_4 + T_5 + T_6 \dots + T_{80}$ |
| x_2, y_2 | 70 | 3 | 2 | 1 | 4 | 9 | 3 | 6 | $T_1 + T_2 + T_3 + T_4 + T_5 + T_6 \dots + T_{70}$ |

Table 3.4. Computation of uHU values from daily average temperatures $T_1, T_2, \dots, T_{\max,ph}$

| Location (x, y) | T_1 | T_2 | T_3 | | $T_{\max,ph}$ | $uHU_{1..2}$ | $uHU_{1..3}$ | $uHU_{1.....}uHU_{\max,ph}$ |
|------------------------|-------|-------|-------|-------|---------------|--------------|-------------------|---------------------------------|
| x_1, y_1 | 4 | 2 | 3 | | 4 | $T_1 + T_2$ | $T_1 + T_2 + T_3$ | $T_1 + T_2 + T_3 + T_{\max,ph}$ |
| x_2, y_2 | 3 | 4 | 2 | | 5 | $T_1 + T_2$ | $T_1 + T_2 + T_3$ | $T_1 + T_2 + T_3 + T_{\max,ph}$ |

4. RESULTS

4.1. Introduction

This section presents the results obtained after applying the research methods. The data analysis involved data cleaning of incomplete observation sets and splitting of data per species into two independent sets for model calibration and validation. Calibration data were used to build the model and the results of the three statistical models for these data are discussed in section 4.2. The results included optimal parameter values determined by simulated annealing, RMSE, efficiency and correlation between the observed and predicted values. The validation data were independent dataset that were used to assess the performance of the model. The results of the statistical models with the validation dataset and the comparison of the three statistical models are presented in section 4.3.

The results of the temporal and spatial bias are explained (Section 4.4.). The optimal parameter values and spatial prediction results of the best statistical model are documented after removal of temporally biased observations (Section 4.4.2.). On the other hand, the results of the data-driven (PHASE) model are discussed thoroughly and depicted (Section 4.5). In this case, the optimization and prediction results (Section 4.5.1) as well as the overall performance of the model are discussed accordingly (Section 4.5.2). Finally, the best statistical model is compared with data-driven (PHASE) model (Section 4.6).

4.2. Model Parameterization

4.2.1. Spring warming

In the case of spring warming (SW) model, a fixed sum of forcing units (F^*) was accumulated above a base temperature (T_b). As such, the phenological event would be reached when F^* amount of heat was accumulated from a starting day (T_0). Daily rate of forcing units were calculated by assuming linear relationship between base and mean daily temperature values (Figure 4.1). In addition, the optimal parameters were determined by fitting the models with simulated annealing. Simulated annealing was properly fine-tuned to have successful search results in the parameterization space. For this, we configured the algorithm to start at random initial point in the lower/upper bounds of the parameter space. It then continued with another random point until the stopping criterion was reached. The lower and upper bound threshold for the base temperature was 0 and 20, respectively. Similarly, optimal T_0 values were searched in a range between the year of phenological observation and the year before the observation was made. Hence, the lower bound was a date in the previous year of the phenological observation and the upper bound was in the year of observation. F^* was also set in a bound between 20 and 2000. Moreover, the initial temperature and stopping condition of simulated annealing were configured.

After configuring simulated annealing, we obtained species-specific optimal parameter values (Table 4.1). That is, the optimization results were species-dependent and differed among species. T_b generally varied between 0.34 to 7.81 °C, and T_0 ranged between late January (for Cow Parsley and Celandine) and start of April (for Oak) depending on the species. The amount of F^* accumulated by the plant species varied from 52 to 236. For instance, wood anemone started accumulating warm temperature on date 53 (February 22) and needed to accumulate 165 forcing units to flower. The number of days required to accumulate the forcing units was dependent on daily temperature values.

Moreover, the performance of the SW model was evaluated by using the calibration dataset (Table 4.1). RMSE, which showed the average error produced by the model, ranged between 9.5 and 31.7 days.

Similarly, the efficiency of the model varied between 0.08 and 0.41, and the correlation coefficient values varied between 0.54 and 0.88.

Table 4.1. Optimal parameters and statistical analysis of SW model for the calibration dataset. T_b (Base temperature), T_0 (Starting DOY of temperature accumulation) and F^* (Heat requirements)

| Species Name | T_b | T_0 | F^* | RMSE | Efficiency | Correlation |
|--------------|-------|-------|-------|-------|------------|-------------|
| Cow Parsley | 3.52 | 22 | 236 | 28.18 | 0.10 | 0.65 |
| Celandine | 0.34 | 27 | 177 | 31.78 | 0.12 | 0.68 |
| Chestnut | 2.31 | 58 | 341 | 9.53 | 0.41 | 0.88 |
| Wood Anemone | 1.15 | 53 | 165 | 25.6 | 0.08 | 0.54 |
| Oak | 7.81 | 96 | 52 | 21.30 | 0.15 | 0.71 |

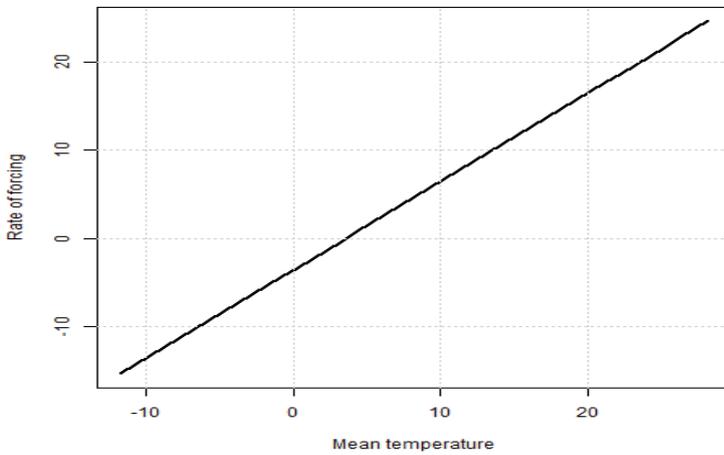


Figure 4.1. The linear relationship between daily rate of forcing and average temperature of Cow Parsley

4.2.2. UNIFORC

The daily rate of heat accumulation (forcing) was computed by assuming a sigmoidal (non-linear) dependency between rates of forcing and mean daily temperature values (Figure 4.2). Optimal parameters of the sigmoidal function (d and e), T_0 and F^* , were determined by using simulated annealing algorithm. The parameters d and e were restricted to ranges, such that $d < 0$ and $e > 0$, respectively. Additionally, we used the same lower/upper bound combinations of SW model for selecting the optimal T_0 and F^* values in UNIFORC model.

In the results, the starting date of temperature accumulation ranged between November of the previous year (for celandine) and February (for chestnut and oak) of the current year. For instance, Celandine started accumulating warm temperature in November of the previous year and flowered when it had accumulated a total of 93 °C temperature. The optimal parameter values, RMSE, efficiency and correlation of the model after being applied to the five plant species are shown in Table 4.2. It was observed that the optimal parameter values were different for all species.

Table 4.2. Optimal parameters and statistical analysis of UNIFORC model for the calibration dataset. d and e are parameters of sigmoidal function, T_0 (Starting DOY of temperature accumulation) and F^* (Heat requirements).

| Species Name | d | e | T_0 | F^* | RMSE | Efficiency | Correlation |
|--------------|------|------|-------|-------|-------|------------|-------------|
| Cow Parsley | -1.0 | 1.3 | 1 | 85 | 29.1 | 0.04 | 0.50 |
| Celandine | -1.0 | 1.0 | 317 | 93 | 32.1 | 0.09 | 0.68 |
| Chestnut | -1.0 | 2.37 | 40 | 64 | 11.08 | 0.20 | 0.65 |
| Wood Anemone | -1.8 | 1.03 | 20 | 53 | 26.5 | 0.01 | 0.48 |
| Oak | -5.4 | 3.16 | 44 | 57 | 21.7 | 0.12 | 0.57 |

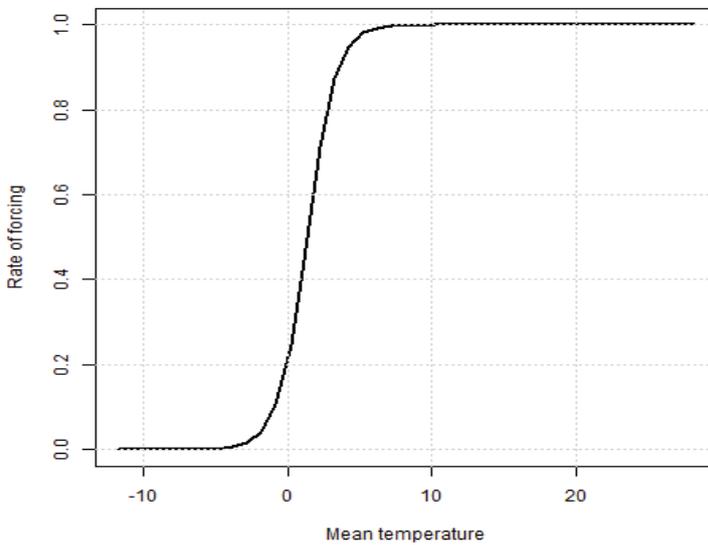


Figure 4.2. The non-linear (sigmoidal) dependency between daily rate of forcing and average temperature of Cow Parsley

4.2.3. UNICHILL

The parameters a , b and d were bounded to have values less than zero and c and e were greater than zero. In addition, C^* and F^* were fitted for determining the accumulated chilling and forcing temperature requirement, respectively (Table 4.3). The lower and upper bounds were 1 and 100 for C^* , 10 and 1000 for F^* , respectively. After setting the lower and upper bounds of simulated annealing, optimal values of the seven fitted parameters were determined. With these optimal parameter values, the RMSE of the model ranged between 27 and 43 days and the efficiency was between -3.74 and -0.45. The correlation between the observed and predicted values of the model ranged between 0.31 and 0.60, respectively.

Moreover, the starting DOY of temperature accumulation was forced to September 1. For instance, 74 °C of temperature was first accumulated by Oak to break dormancy. Then, the plant started accumulating warm temperature from T_d until it reached 115 °C of temperature to start leafing.

Table 4.3. Optimal parameters and statistical analysis of UNICHILL model for the calibration dataset. a, b, c are parameters of sigmoidal function (R_c), d and e are parameters of sigmoidal function (R_f), C*(chilling requirements) and F*(Heat requirements).

| Species Name | a | b | c | d | e | C* | F* | RMSE | Efficiency | Correlation |
|--------------|-------|-------|------|-------|------|-----|-----|------|------------|-------------|
| Cow Parsley | -4.23 | -2.84 | 6.05 | -1.01 | 3.23 | 94 | 65 | 38.4 | -0.66 | 0.55 |
| Celandine | -2.99 | -2.23 | 9.39 | -1.00 | 1.85 | 67 | 70 | 43.3 | -0.64 | 0.60 |
| Chestnut | -2.10 | -1.02 | 3.48 | -1.00 | 2.07 | 108 | 77 | 27.1 | -3.74 | 0.36 |
| Wood Anemone | -3.99 | -1.18 | 8.95 | -1.00 | 1.28 | 113 | 53 | 32.2 | -0.45 | 0.48 |
| Oak | -1.13 | -4.53 | 3.88 | -1.06 | 1.00 | 74 | 115 | 30.3 | -0.73 | 0.31 |

4.3. Model Performance and Comparison

The three statistical models, namely SW, UNIFORC and UNICHILL, had differences in design and in the type and initial set up of the parameters used. Thus, the aim of model comparison was to determine a biologically meaningful model that accounted the variability in the data and that had optimal parameters with physiological interpretation. The best model was chosen by comparing the RMSE, efficiency and correlation of the three statistical models using the model building and validation dataset. The results of the three statistical models for the model calibration dataset are presented for comparison (Table 4.1, 4.2 and 4.3). These results indicated that the SW model had a better performance than the other two statistical models in terms of RMSE, efficiency and correlation coefficient values.

Moreover, the results of the three statistical models were compared for the validation dataset (Table 4.4). The validation dataset was independent of the calibration data and could be used for comparing the performance of the models. In the result, SW model showed a better prediction for three of the plant species, namely Oak, Celandine and Chestnut. However, the UNICHILL model showed an improvement in the prediction by 2.1 days for Wood Anemone and 2.52 for Cow Parsley than SW model.

In conclusion, SW model showed better performance for all the five plant species in the model calibration dataset than UNIFORC and UNICHILL models. On the other hand, the UNICHILL model had a slight improvement over SW model in prediction results of Cow Parsley (2.52 days) and Wood Anemone (2.16 days) for the model validation dataset. However, the improvements of the UNICHILL model were considered insignificant as compared to the improvements of the SW model for Oak, Celandine and Chestnut by 11.5, 11.9 and 16 days, respectively. Finally, we selected the SW model as the best statistical model for predicting the phenological stages of the five plant species in the ‘Natuurkalender’ dataset.

Table 4.4. Statistical analysis of the SW, UNIFORC and UNICHILL model for the validation dataset

| Name of Plant | RMSE | Efficiency | RMSE | Efficiency | RMSE | Efficiency |
|---------------|-----------------------|------------|----------------|------------|-----------------|------------|
| | Spring Warming | | UNIFORC | | UNICHILL | |
| Cow Parsley | 25.94 | 0.14 | 26.48 | 0.10 | 23.42 | 0.29 |
| Celandine | 20.31 | 0.24 | 25.49 | -0.19 | 32.19 | -0.91 |
| Chestnut | 8.19 | -0.69 | 8.70 | -0.92 | 24.21 | -13.83 |
| Wood | 15.26 | 0.04 | 16.06 | -0.06 | 13.10 | 0.29 |
| Anemone | | | | | | |
| Oak | 8.92 | 0.34 | 9.52 | 0.25 | 20.44 | -2.46 |

Furthermore, we compared the means of the calibration and validation dataset by using one factor ANOVA test. The result of the ANOVA test for the five plant species are shown in Table 4.5. The test showed insignificant results for all species and the null hypothesis was rejected. This indicates that the calibration and validation dataset of the species have differing means. Besides, the F-value, which is the ratio between group variability and within group variability, showed higher values. If the F-value is greater than 1, then the between group variability is greater than the within group variability and the datasets have different mean. Moreover, these results are visualized by boxplots of the calibration and validation dataset (Figure 4.3).

Table 4.5. Statistical analysis of the ANOVA test between the calibration and validation dataset

| Species Name | F-value | P-Value | Significance level |
|--------------|---------|----------|--------------------|
| Cow parsley | 57.69 | 6.29e-14 | 0 |
| Celandine | 61.45 | 7.28e-15 | 0 |
| Chestnut | 1.256 | 0.263 | 0.1 |
| Wood anemone | 38.81 | 7.53e-10 | 0 |
| Oak | 3.165 | 0.0761 | 0.05 |

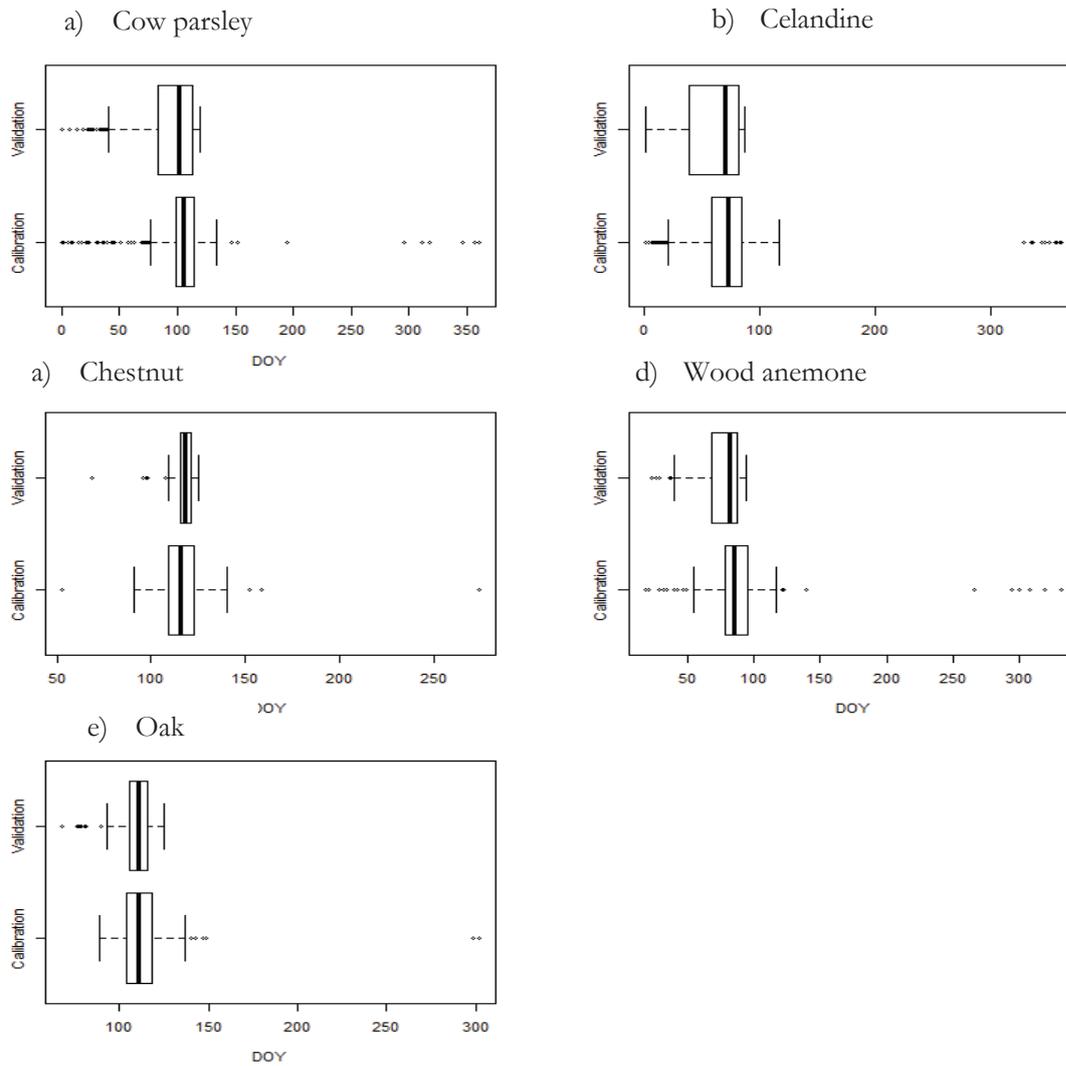


Figure 4.3. Boxplots showing the DOY of the phenological event of the calibration and validation dataset

4.4. Bias Identification

4.4.1. Temporal Bias

The presence of temporal bias in the volunteer phenological observations prohibited the model from explaining the variance existing in the data. For this reason, the RMSE was high and the efficiency of the model was lower than what had been expected. Hence, the best phenological model was used to identify temporal bias in the volunteer phenological observations.

To detect biased observations, the SW model was used to predict the DOY of the phenological event at locations of phenological observations. For each phenological observation, the error (Predicted - Observed) and the normalized SD were computed. Two instances of volunteer phenological observations that were made for Cow parsley and Celandine are presented for illustration (Table 4.6). The error of the observation for Cow parsley indicated a latency of 4 days. Similarly, the normalized SD was positive, which means that the error of the observation was above the mean. In addition, Celandine showed an advance of 6 days in the prediction results than the observation date. Also, the normalized SD was negative indicating that the error of the observation was below the mean. After calculating the normalized SD of each observation, observations between 0.5 and 4 SD were removed and the systematic and unsystematic RMSE values of the model were compared.

Table 4.6 Instances of the volunteer phenological observations for Cow Parsley and Celandine. (x, y are the coordinates of the observer).

| Species Name | X | Y | Observed | Predicted | Error | Normalized_SD |
|--------------|--------|--------|----------|-----------|-------|---------------|
| Cow Parsley | 191000 | 440000 | 104 | 108 | 4 | 0.13 |
| Celandine | 160000 | 426000 | 80 | 74 | -6 | -0.20 |

The systematic RMSE, unsystematic RMSE and RMSE (Figure 4.4), correlation and efficiency (Figure 4.5) of the SW model were assessed by removing observations within a range of 0.5 and 4 normalized SD. The values were species-dependent. As a result, the RMSE, systematic and unsystematic RMSE of each species that gave an optimal normalized SD values were surrounded by a square box (Figure 4.4). Cow Parsley had a systematic RMSE smaller than the unsystematic RMSE for normalized SD ranging between 0.1 and 2. However, we chose 2 SD with 5% of observations removed as optimal threshold rather than 0.1, 0.5 or 1 SD where 71%, 15%, 6% of observations, respectively, were removed from the calibration dataset. Similarly, normalized SD of 0.7, 4, 0.55 and 1 were chosen as optimal threshold values for Celandine, Chestnut, Wood Anemone and Oak, respectively.

As the percentage of removed observations increases, more outliers were removed and there would be a better fit to the model. As a result, the efficiency and correlation values of the model got better and better (Figure 4.5). For all species, the efficiency of the model got closer to 1 as observations above 0.1 standard deviations were removed. Hence, as the percentage of removed observations increase, the systematic and unsystematic RMSE of the model got lower and the efficiency and correlation of the model get improved.

Finally, phenological observations above the chosen species-specific normalized SD threshold were considered as temporally-biased observations. The spatial distributions of species-dependent temporally-biased observations are shown on yearly basis (Figure 4.6). The observations were overlaid on a standard deviation map of the corresponding year of observation. This map showed the standard deviation of the phenological date of one year from the other years and the colours indicated the amount of deviation which was different for year/species combination. Positive values of the normalized standard deviation indicated an over-prediction, whereas the negative values indicated under-prediction. The temporally-biased observations were visualized by three different sizes. Those observations that were above 0, 1 and 2 standard deviation were shown with small, medium and biggest size circles, respectively (Appendix A). In the case of cow parsley, observations above 2 standard deviation were removed. Hence, the biased observations were shown by the biggest sized circles (Figure 4.6).

Furthermore, the number and spatial distribution of biased observations varied from year to year. Most of the observations that were regarded as temporally-biased for cow parsley were in the year 2005, whilst there were no temporally-biased observations in 2009 and 2010. Lastly, temporally-biased phenological observations were removed from the calibration dataset and the remaining data were fitted to the SW model to determine new optimal parameter values and make predictions (Section 4.4.2).

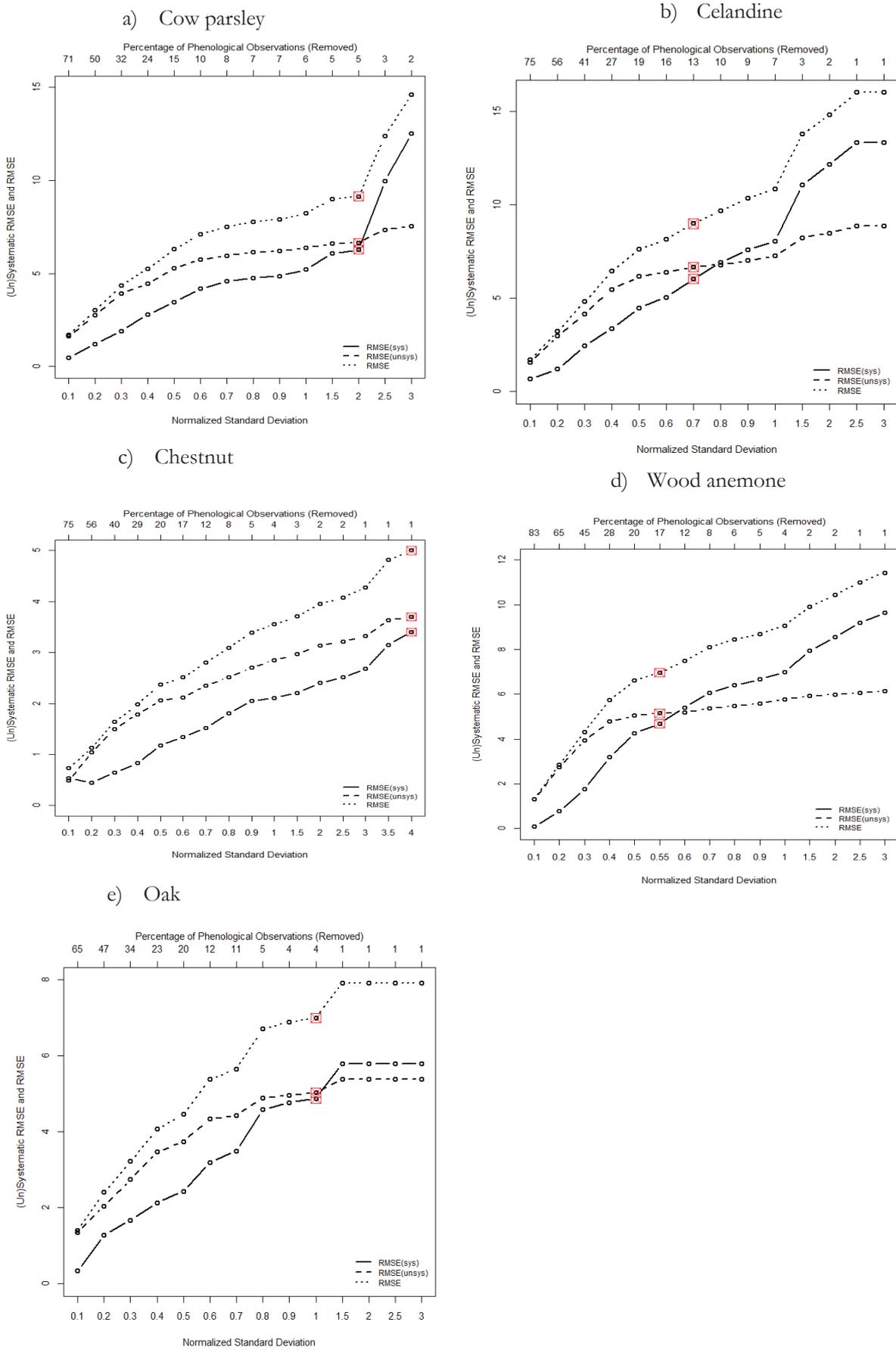


Figure 4.4. Optimal normalized SD threshold values for removal of temporally-biased observations

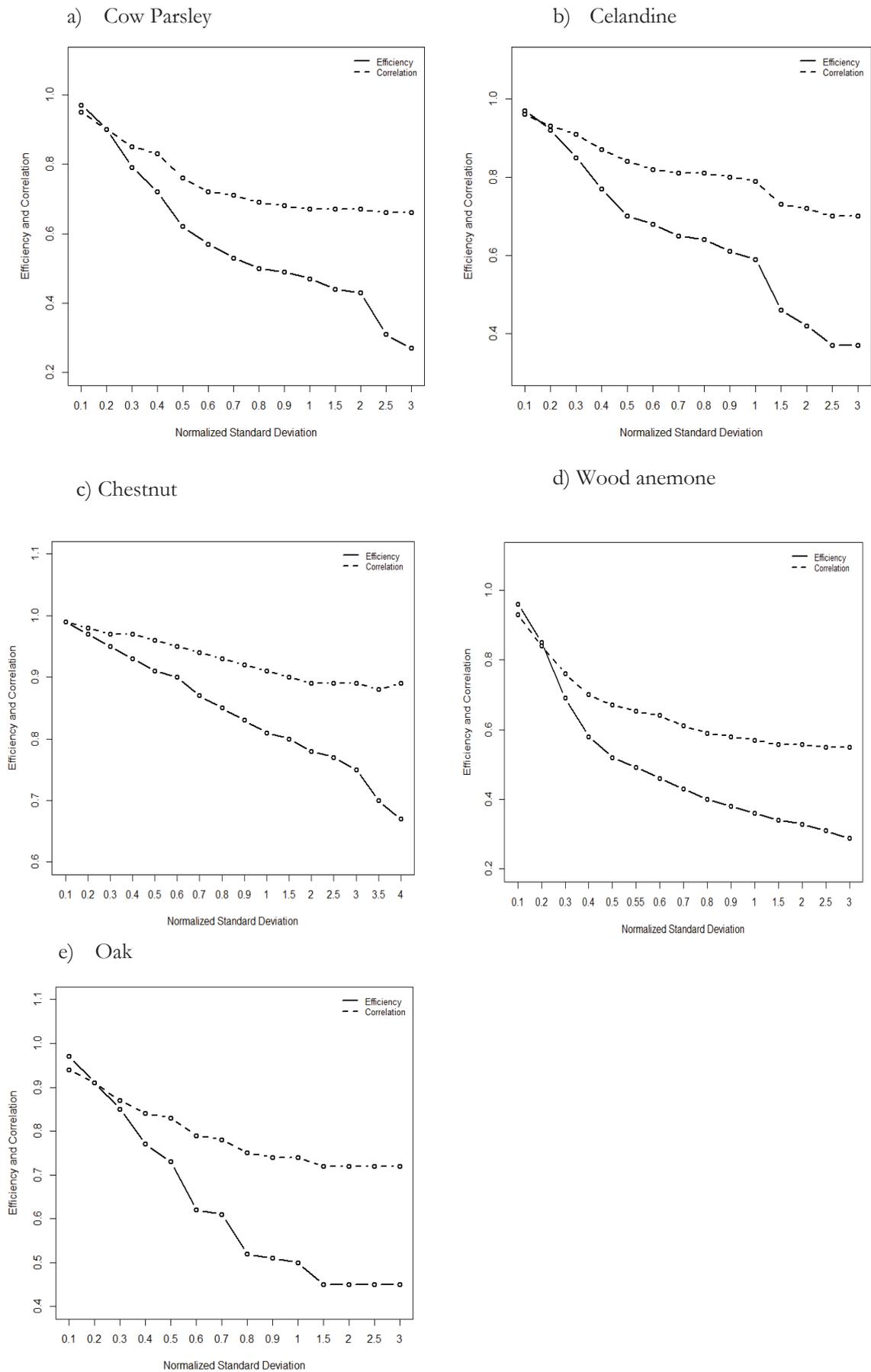


Figure 4.5. Efficiency and correlation values for removal of temporally-biased observations

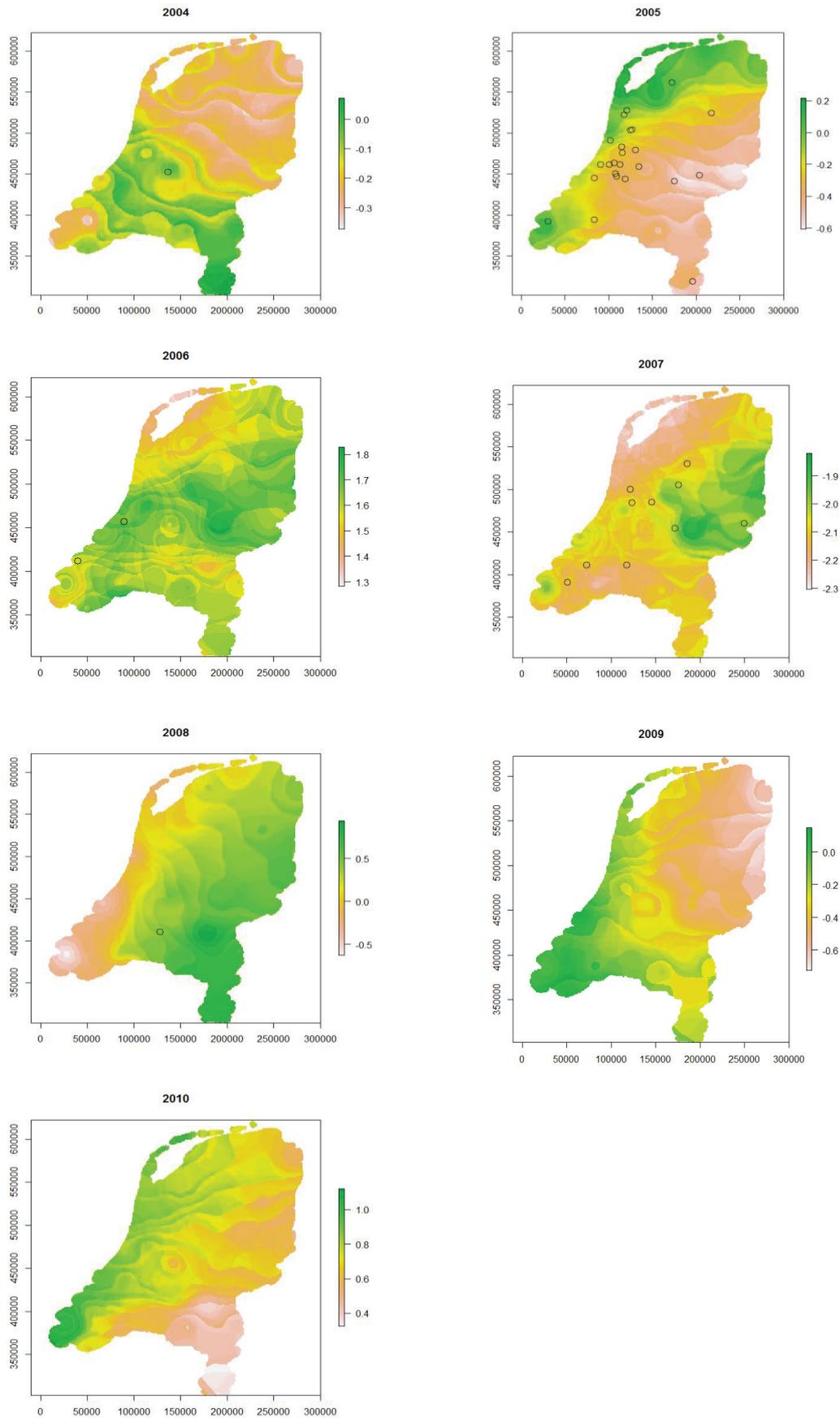


Figure 4.6. Spatial distribution of year based temporally-biased volunteer phenological observations of Cow Parsley (circle symbols). The colour of the map indicates the variability in the DOY of the phenological event between one year and all the other years. For each phenological observation, the sizes of the dot indicates the amount of normalized SD value.

4.4.2. Parameterization of the best model after removal of temporally-biased observations

The temporally-biased observations (red triangles) were spread away from the regression line where as the non-biased (black circles) were well aligned to the regression line (Figure 4.7). As shown in the figure, celandine had the highest and chestnut had the smallest number of temporally-biased observations. Once the temporally-biased observations were removed, we fitted the SW model by using non-biased phenological observations. Then, new optimal parameter values for T_b , T_0 and F^* were determined and presented (Table 4.7). The new optimal parameter values were different from the optimal parameter values determined by simulated annealing in the presence of temporally-biased observations. There was a smaller difference in the T_b , T_0 and F^* values of Wood anemone, Celandine, Cow parsley and Chestnut. However, the optimal parameter values of oak showed a larger difference in T_0 from date 96 to 54, T_b from 7.81 to 1.98, and F^* from 52 to 326.

Moreover, the removal of temporally-biased observations improved the RMSE, efficiency and correlation of the SW model both for the calibration and validation dataset (Table 4.7 and 4.8, respectively). As expected, the RMSE and efficiency of the calibration dataset were better than the validation dataset except for oak where there was a better RMSE and efficiency for the validation dataset than the calibration. The RMSE of the calibration dataset ranged between 5.1 and 9.0 days and the efficiency was between 0.45 and 0.68. Similarly, the correlations between the observed and predicted values ranged between 0.65 and 0.88. On the other hand, the RMSE of the SW model for the validation dataset ranged between 4.1 to 18.9 days. The correlation was in a range between 0.20 and 0.87. Also, the efficiency of the model ranged between -1.31 and 0.64.

The results of the accuracy measures on the five plant species were depicted on the scatter diagrams (Figure 4.8). The diagram showed a positive relationship between the observed and predicted values along the regression function. However, the annual variations existing in one year were not well explained by the model. Hence, the outliers in the scatter diagram were caused by annual climatic variations of phenological dates which could not be captured by the model. On the other hand, the model had explained inter-annual variation that existed between the years. For instance, the box plots of all species between the years 2003 and 2010 showed three periods for a phenological event, namely, early, medium and late periods (Figure 4.10). The year 2007 showed an early occurrence of the phenological event for four of the species except for celandine where there was an early phenological stage in 2007 and 2008. On the contrary, the year 2006 showed a later period for the phenological event and the remaining years had a medium period. These periods could be visualized from the scatter diagrams where there were three clusters for the early, medium and late phenological periods (Figure 4.8). This indicated that the model had captured the inter-annual variability in the phenological event.

At this juncture, temporally-biased observations were removed from the calibration dataset. So, the systematic RMSE of the model was expected to be smaller than the unsystematic RMSE. For this reason, we computed the systematic and unsystematic RMSE of the model after removal of temporally-biased observations (Table 4.9). In that case, the model showed a systematic RMSE smaller than the unsystematic RMSE for the four plant species. But Oak showed a systematic RMSE larger than the unsystematic RMSE. This could be related to the incorrect parameter values of SW model, which was used in determination of temporally-biased observations. Specifically, the SW model had shown a large difference in the optimal parameter values for oak after removal of temporally-biased observations. Therefore, we did a second round analysis on the components of RMSE for oak by fitting SW model with the non-biased observations. Phenological observations between 0.1 and 3 SD were removed and the quality of the model was assessed (Figure 4.11). Twelve percent of the observations above a normalized SD of 1.5 (i.e. optimal threshold) were removed. As a result, the systematic RMSE (2.9) was smaller than the

unsystematic RMSE (3.36). Consequently, the SW model was fitted with the non-biased observations and new parameter values of 2.33, 52, 304.81 were obtained for T_b , T_0 and F^* , respectively. Similarly, the RMSE, efficiency and correlation of the SW model for the calibration dataset were improved to 4.35, 0.73, and 0.83, respectively. On the other hand, the accuracy of the model on the validation dataset did not change as the observations had normalized SD values less than 1.5.

In conclusion, the new SW model could be used to accurately predict the dates of the phenological event of the five plant species. For this reason, we predicted the DOY of flowering of the four plant species and leafing of one species for the whole area of The Netherlands for the year 2010 (Figure 4.9). The phenological dates of all species in the year 2010 were in spring, which ranged between mid-March and mid-May. Early phenological events were encountered in the southern part of The Netherlands and late phenological events in the northern part. Moreover, the prediction results of each species varied from one year to another. Hence, the prediction results of all years were compared to determine inter-annual variations in climate and spatial bias (Section 4.4.3).

Table 4.7. Statistical analysis of the SW model after removal of temporally-biased phenological observations in the calibration dataset.

| Name of species | T_b | T_0 | F^* | RMSE (days) | Efficiency | Correlation |
|-----------------|-------|-------|-------|-------------|------------|-------------|
| Cow Parsley | 3.03 | 20 | 272 | 9.04 | 0.45 | 0.69 |
| Celandine | 0.73 | 20 | 183 | 8.67 | 0.68 | 0.83 |
| Chestnut | 3.27 | 57 | 290 | 5.14 | 0.67 | 0.88 |
| Wood Anemone | 2.25 | 55 | 116 | 6.77 | 0.52 | 0.65 |
| Oak | 1.98 | 54 | 326 | 6.43 | 0.58 | 0.78 |

Table 4.8. Statistical analysis of the SW model after removal of temporally-biased phenological observations in the validation dataset.

| Name of species | RMSE | Efficiency | Correlation |
|-----------------|-------|------------|-------------|
| Cow parsley | 18.99 | 0.26 | 0.33 |
| Celandine | 14.81 | -0.05 | 0.87 |
| Chestnut | 7.67 | -1.31 | 0.2 |
| Wood Anemone | 7.66 | -0.41 | 0.86 |
| Oak | 4.12 | 0.64 | 0.77 |

Table 4.9. Systematic and unsystematic RMSE of SW model after removal of temporally-biased observations

| Name of species | Systematic RMSE | Unsystematic RMSE |
|-----------------|-----------------|-------------------|
| Cow parsley | 6.06 | 6.7 |
| Celandine | 4.76 | 7.25 |
| Chestnut | 3.36 | 3.89 |
| Wood anemone | 4.56 | 5.01 |
| Oak | 4.98 | 4.06 |

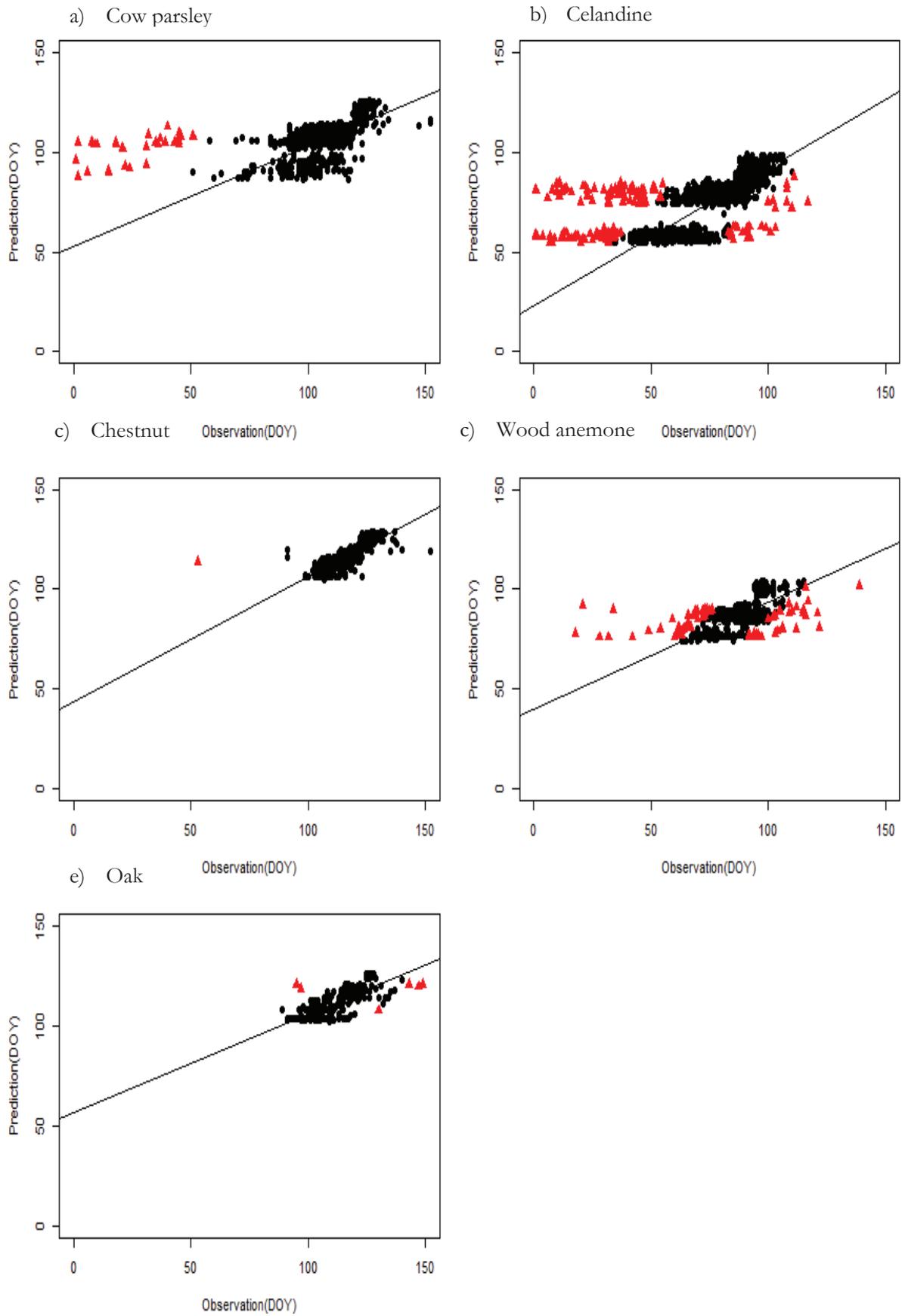


Figure 4.7 Observed and predicted values of SW model with temporally biased (▲) and non-biased (●) volunteer phenological observations

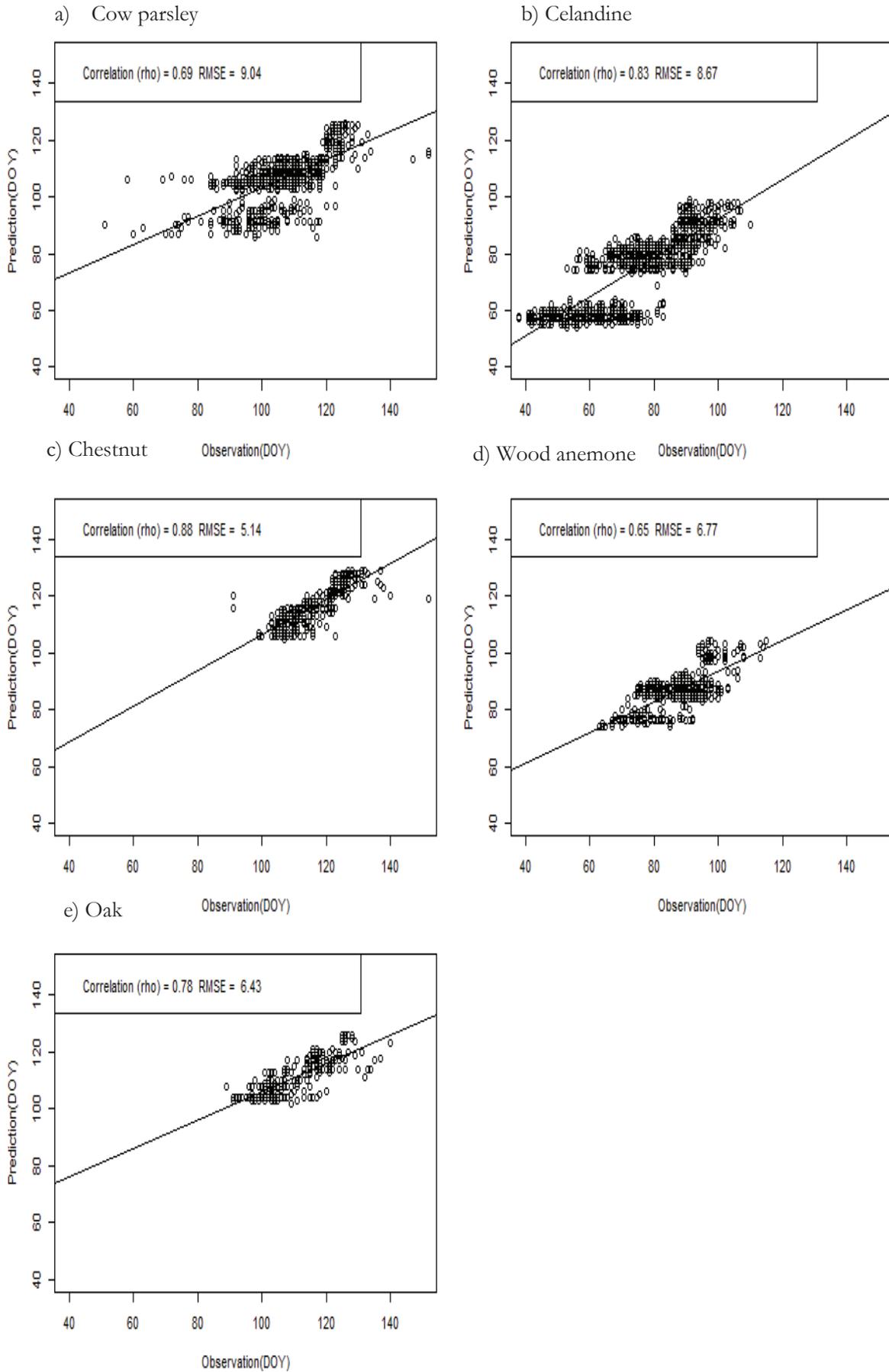


Figure 4.8. Observed and Predicted values of SW model with non-temporally biased phenological observations

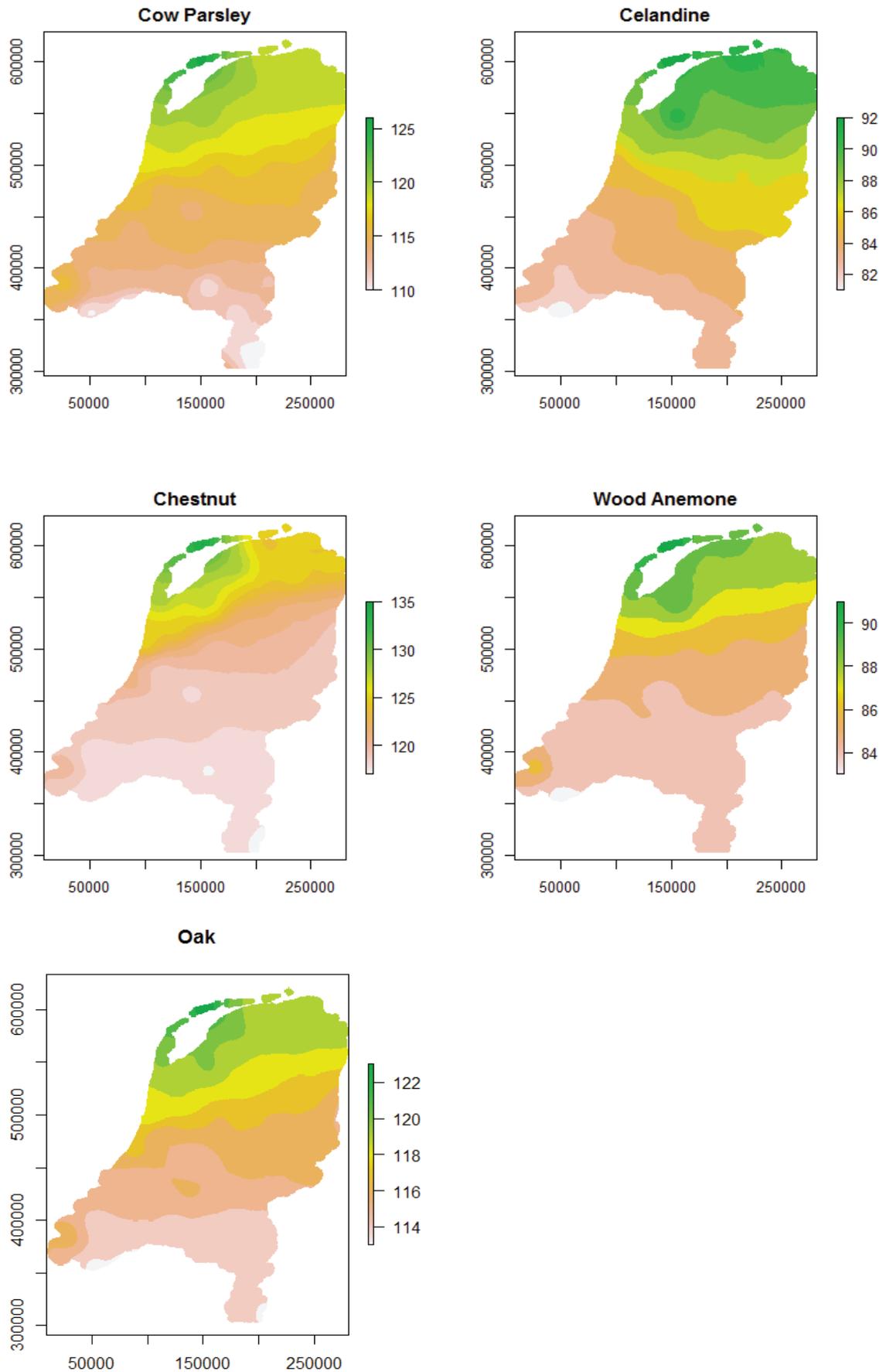


Figure 4.9. Spatial prediction of the DOY of the phenological event of the five plant species in The Netherlands in 2010.

In addition to the DOY of phenological events, the SW model helped in determining the duration (temporal windows) and annual variability of phenological events for a period of 8 years (i.e. between 2003 and 2010). The duration of phenological event, which was species and year dependent, is shown by the box plots (Figure 4.10). The difference between the minimum and maximum value of the box plots could be considered as the duration of the phenological period. In addition, the median showed the DOY where there was 50% of the phenological event. Similarly, the lower and upper quartile showed the DOY where there was 75% and 25%, respectively, of phenological event above that DOY. Some species, like Cow parsley (2008, 2009, 2010), Celandine (2004, 2005), Chestnut (2008, 2009 and 2010), and Oak (2005, 2007, 2010) had outliers in the DOY. These outliers were due to annual variations in DOY of the phenological event at different locations of The Netherlands, caused by climatic variations.

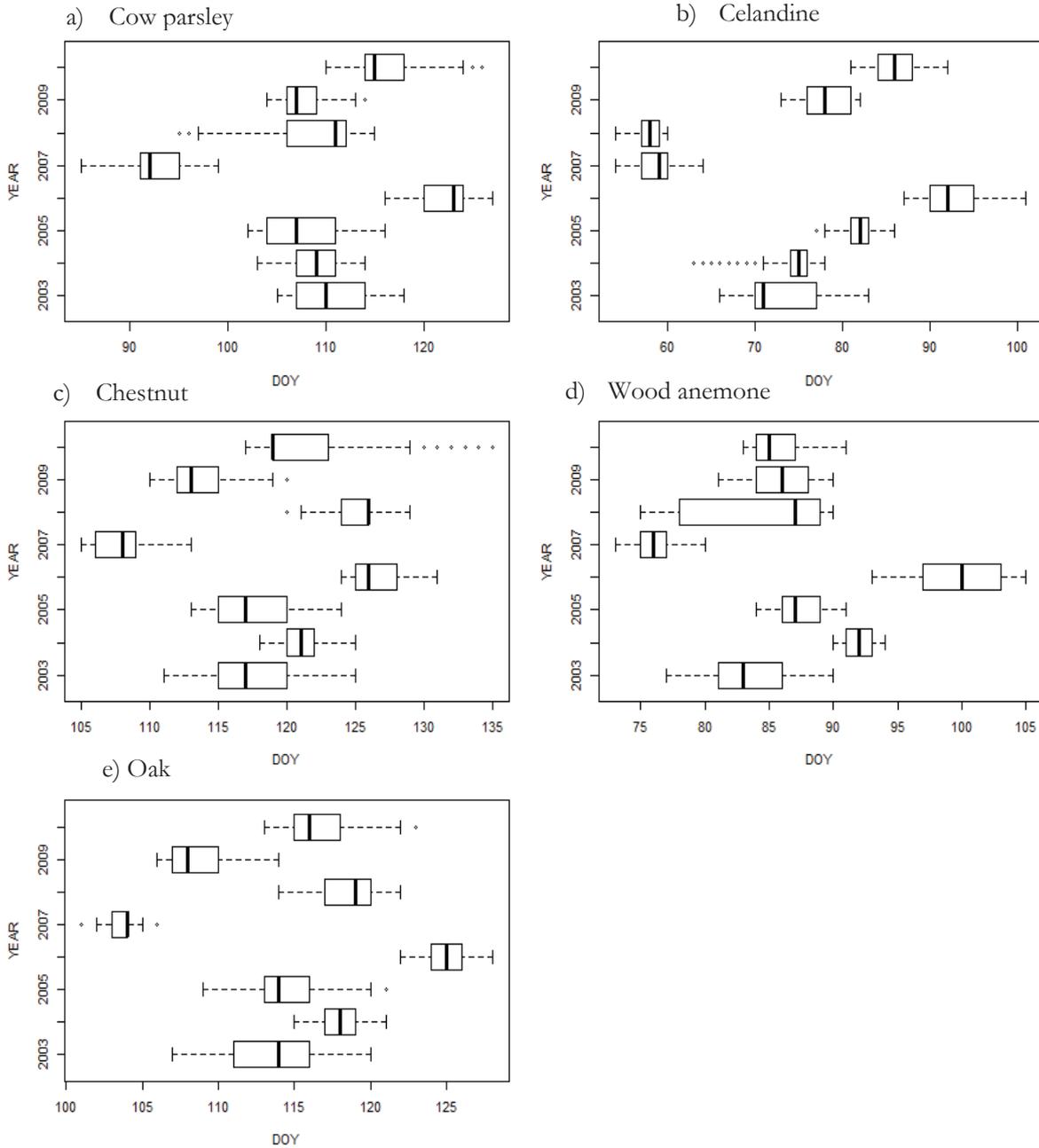


Figure 4.10. Box plots showing the duration of phenological event between the years 2003 and 2010.

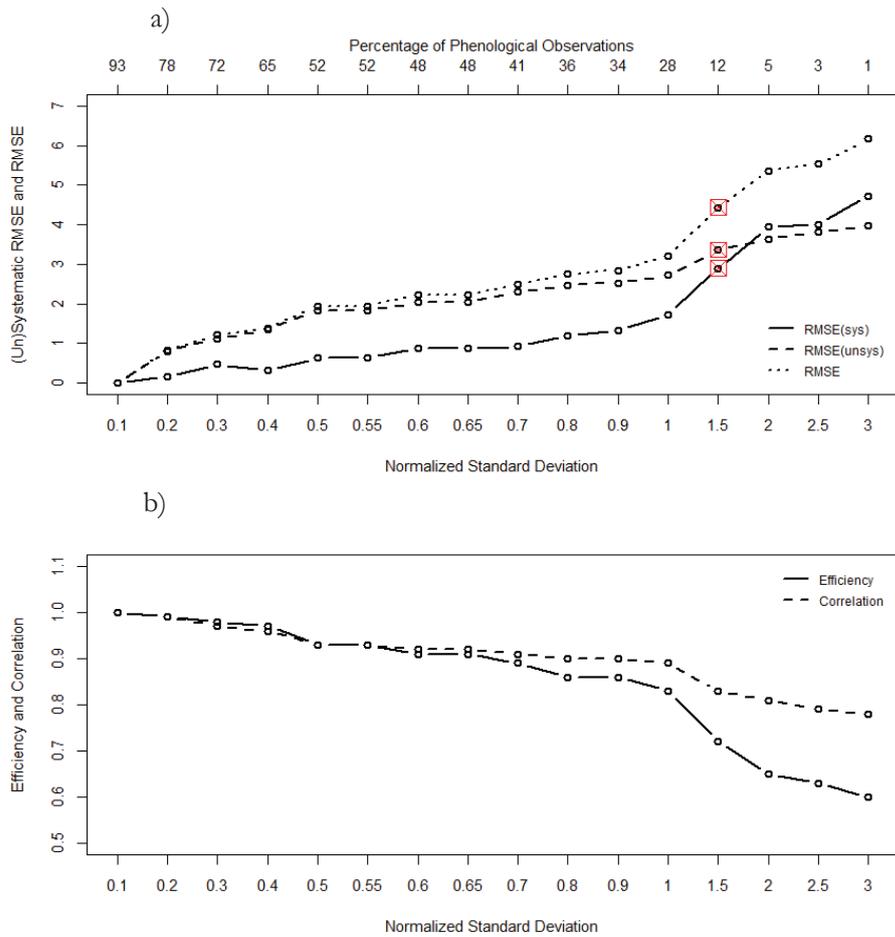


Figure 4.11. Systematic RMSE, unsystematic RMSE and RMSE (a) and efficiency (b) of the SW model after second round removal of temporally-biased observations of Oak

Furthermore, the box plots (Figure 4.10) could help determine the relation that exists between the DOY of the phenological event of one species and the other. For instance, cow parsley showed a relatively close DOY of phenological event with oak. The differences in the DOY (DOY of cow parsley – DOY of oak) of the phenological event between cow parsley and oak are shown in figure 4.12. These two species showed high variability in 2004, 2007 and 2008, which was caused by the varying sensitivity of the species to climatic variations.

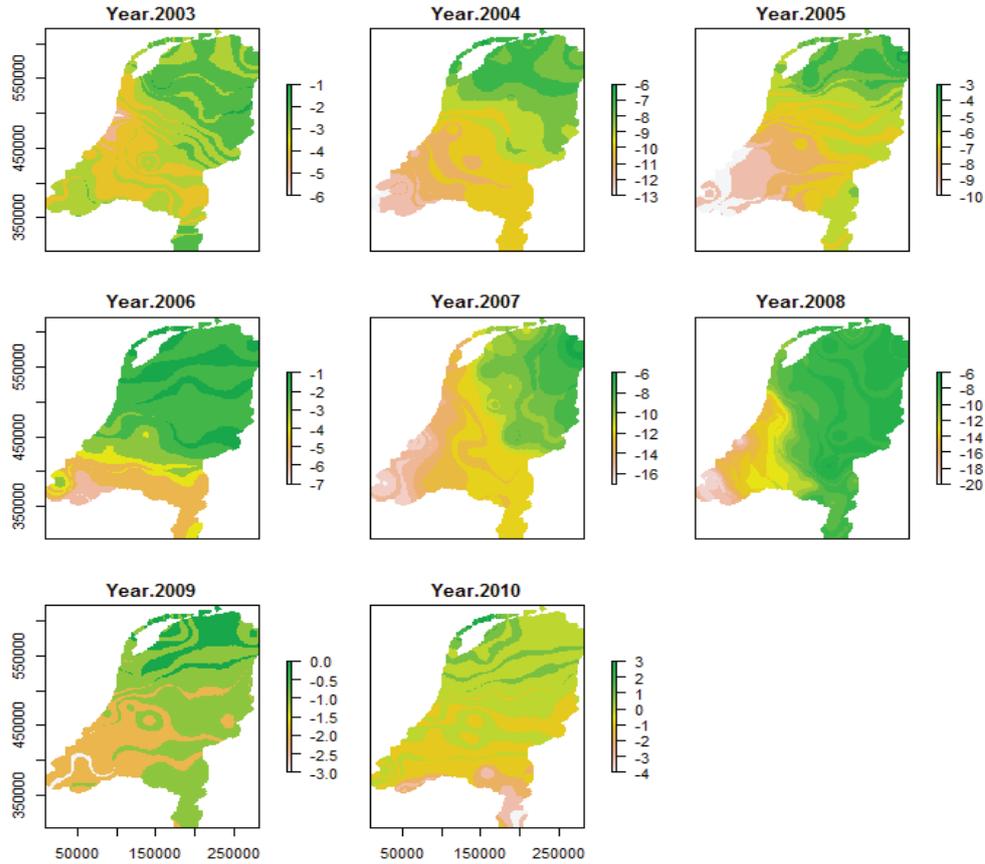


Figure 4.12. The variability in the DOY of the phenological event between Cow parsley and Oak between the years 2003 and 2010

4.4.3. Spatial bias

The new SW model was used to predict the phenological dates of the five plant species from the year 2003 until 2010 (Figure 4.13 (a)). The figure indicated the inter-annual variation in the prediction results caused by climatic variations. For instance, Cow parsley flowered between date 116 and 126 in 2006. However, the flowering in 2007 was between 86 and 98, which was earlier than that of 2006. Similarly, the flowering period was later in the northern part and earlier in the southern part of The Netherlands due to relatively high temperature in the southern part (Figure 4.14). The average annual temperature in figure 4.14 was calculated by using the daily average temperatures of the 365 days in a year. There was a higher temperature in the southern part than the northern part of The Netherlands. In addition, the spatial patterns of the DOY of flowering in the eight years period for Cow Parsley could be visualized in Figure 4.13 (b). The year 2007 showed earlier flowering than all the other years.

To determine areas with climatic variability, the SD of the prediction results in the 8 years period were computed for all species. The SD map (Figure 4.15) showed the variability in prediction results of species in the years between 2003 and 2010. The colours on the map indicated the amount of variability (SD) in days in the whole area of The Netherlands. Celandine had shown the highest variability in the flowering date ranging between 10.5 and 12.5 days. On the contrary, chestnut had the least variability in the leafing date ranging between 5 and 6.5. These inter-species differences in SD values were caused by the varying sensitivities of species to climatic variations.

Finally, we grouped the areas with climatic variability into highly, medium and low variable depending on the SD values (Figure 4.16). The percentages of the areas that belong to each group had been presented in tabular form (Table 4.10). The percentage of high variable areas ranged between 6 (cow parsley, chestnut, wood anemone) and 22 (Celandine). The percentage of medium variable areas ranged between 31 (cow parsley) and 76 (chestnut). Similarly, the percentage of low variable areas ranged between 14 (oak) and 63 (cow parsley).

Moreover, the distribution of non-temporally biased observations on highly, medium and low variable areas of The Netherlands is illustrated (Figure 4.16). In connection to this, we analysed the yearly volunteer phenological observations to determine their spatial belongingness in the high, medium and low variable areas (Figure 4.17). More than 50% of the volunteer phenological observations for Chestnut and Oak were made in areas with medium variability. Similarly, the higher proportion of phenological observations for Cow parsley and Celandine were reported in areas where there was low variability. Wood anemone had higher number of phenological observations in the medium and low variable areas. For all species, the smallest proportions of phenological observations were made in areas where there was high variability. Exceptionally, there were no observations made in the highly variable areas for Chestnut in 2004 and 2010, and for Wood anemone in 2006 and 2008, and for Oak in 2008 and 2010. In addition, the numbers of phenological observations in the high, medium and low variable areas were different from one year to the other depending on the number of volunteers who needed to report a phenological stage.

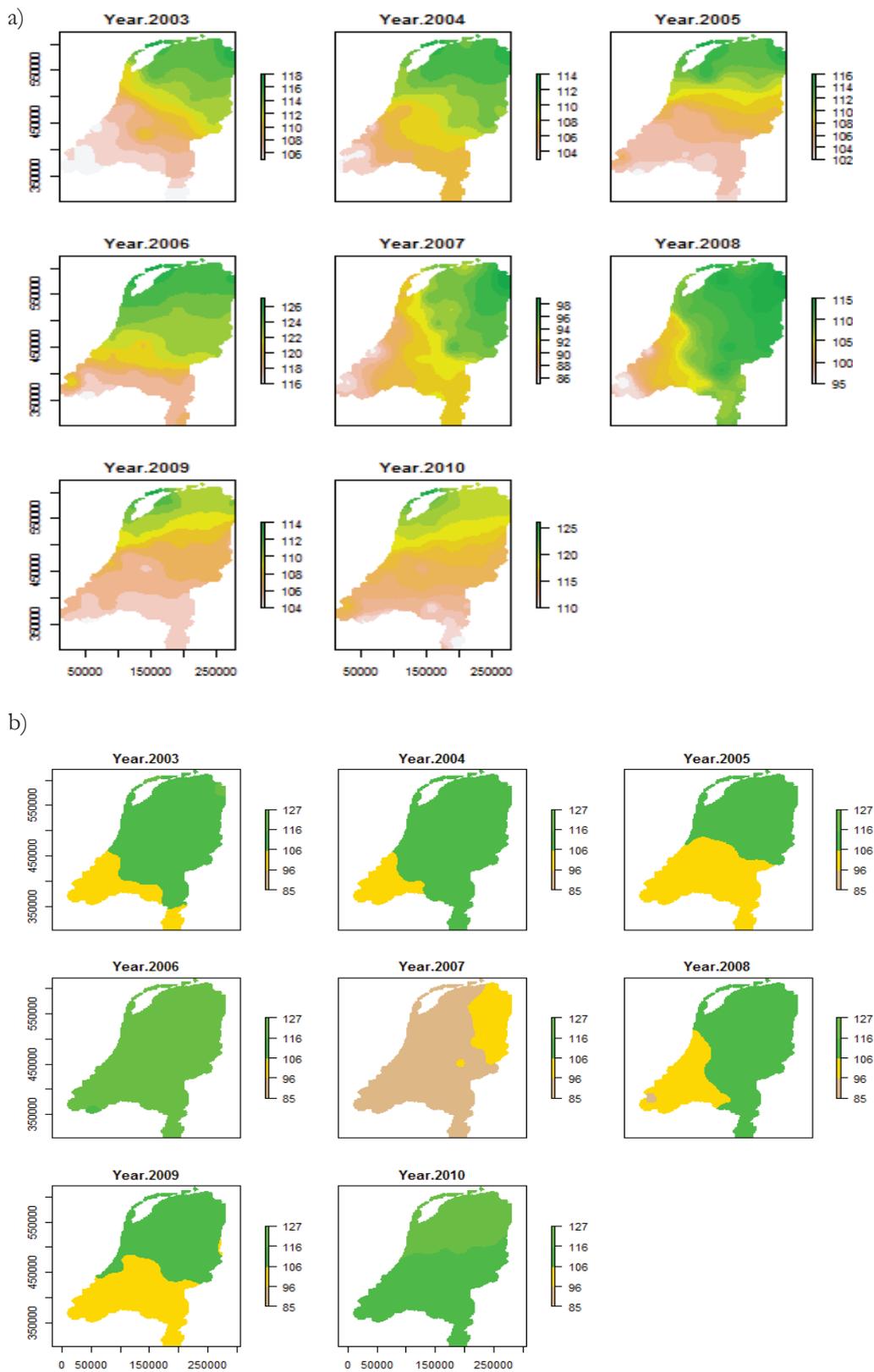


Figure 4.13. Prediction results of the SW model for Cow Parsley from the year 2003 to 2010 (a) with different scale depending on DOY of phenological event (b) with common scale ranging between the minimum and maximum DOY of phenological events in all years.

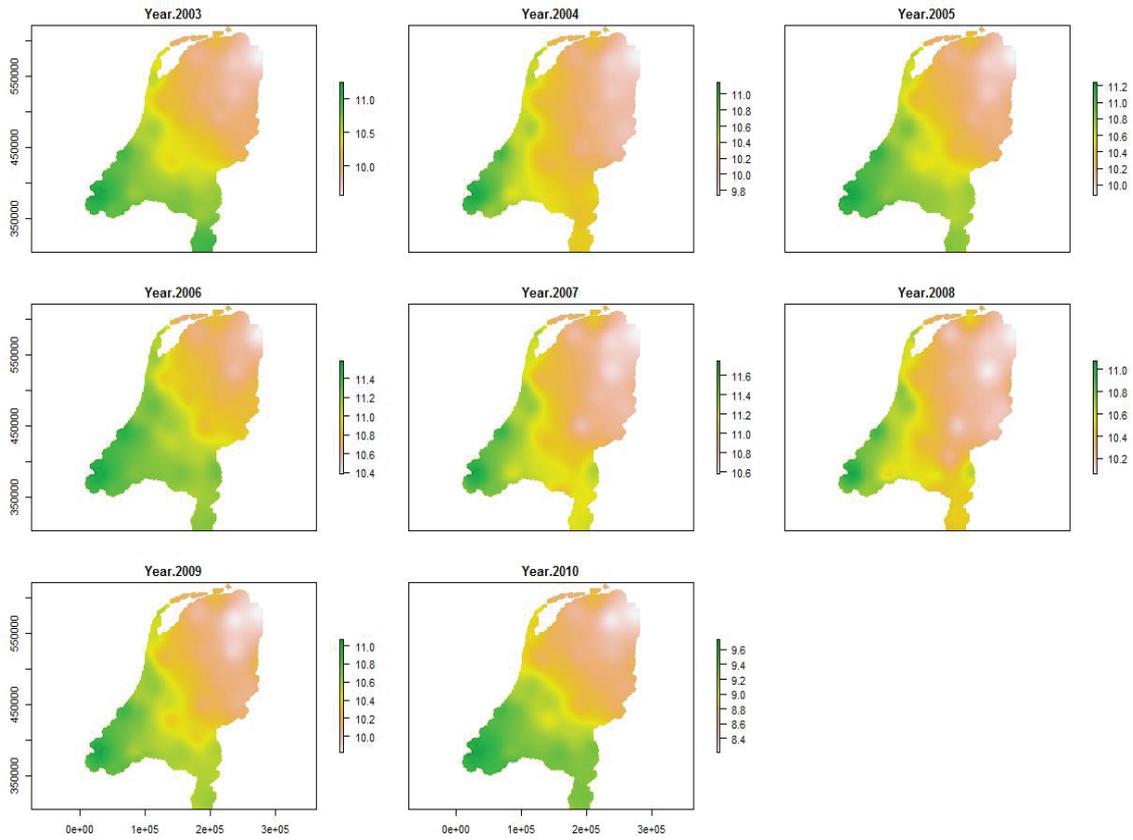


Figure 4.14. Average annual temperature of The Netherlands between the years 2003 and 2010.

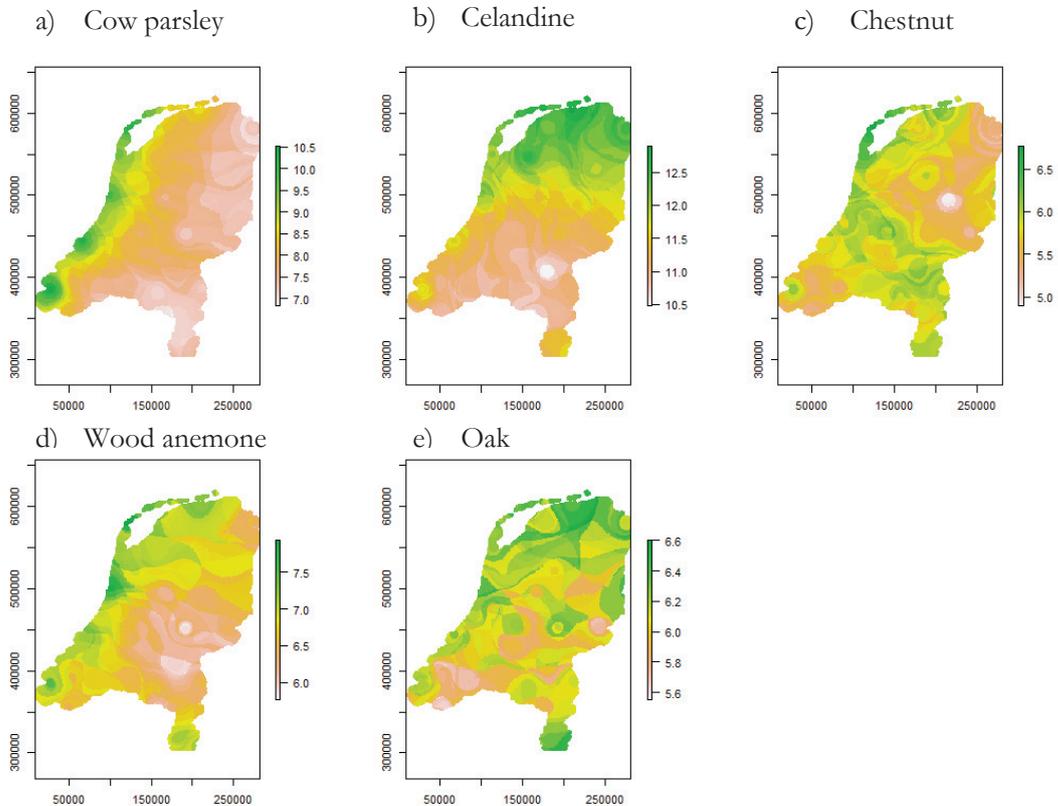


Figure 4.15. Standard deviation map of DOY of phenological event between the years 2003 and 2010

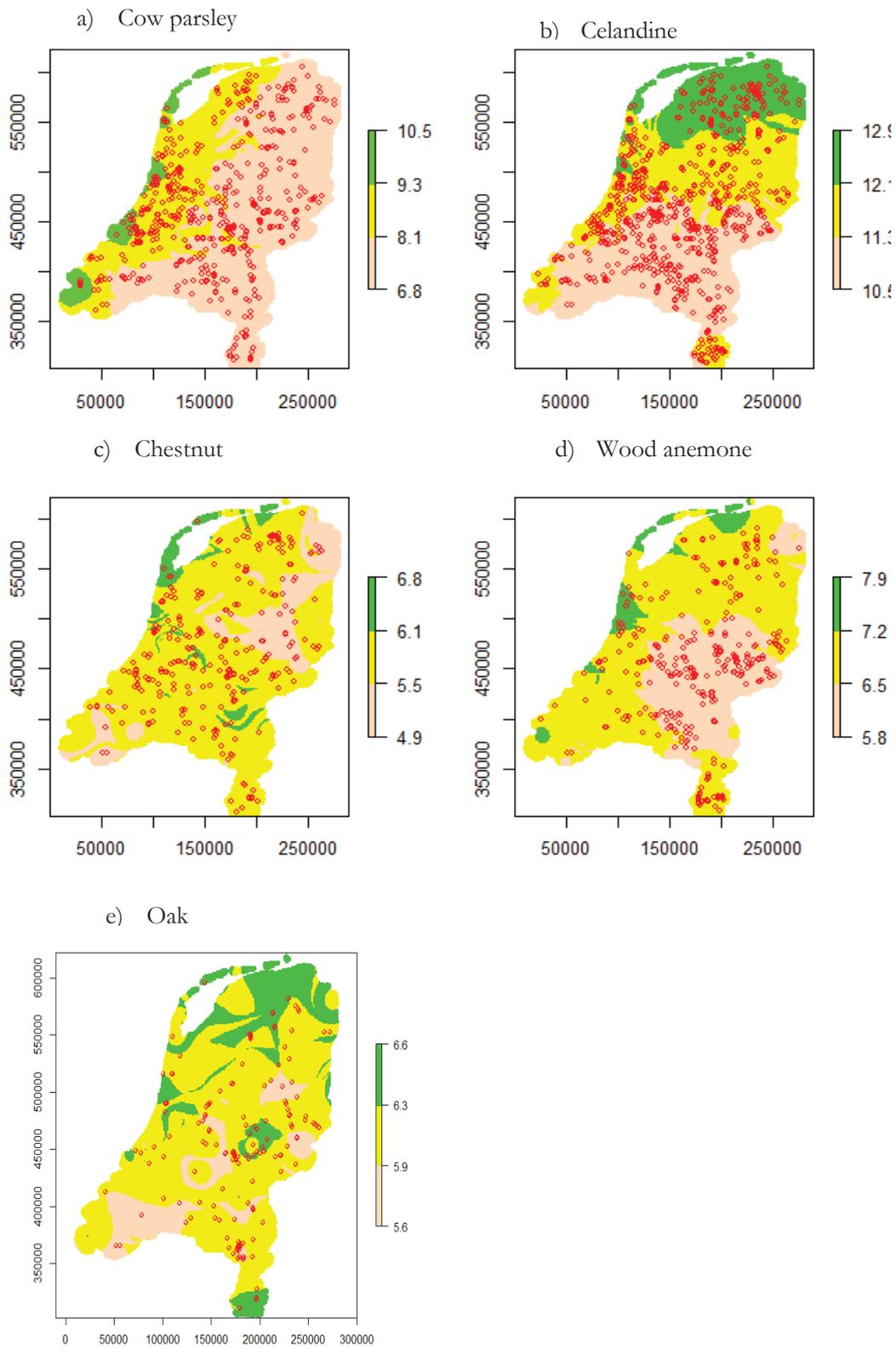


Figure 4.16. Volunteer phenological observations overlaid on areas showing high (green), medium (yellow) and low (pale colour) variability in the prediction of DOY of phenological stage.

Table 4.10 Percentages of high, medium and low variable areas in the whole area of The Netherlands

| Name of species | Highly variable (%) | Medium variable (%) | Low variable (%) |
|-----------------|---------------------|---------------------|------------------|
| Cow parsley | 6 | 31 | 63 |
| Celandine | 22 | 38 | 41 |
| Chestnut | 6 | 76 | 18 |
| Wood anemone | 6 | 63 | 31 |
| Oak | 19 | 68 | 14 |

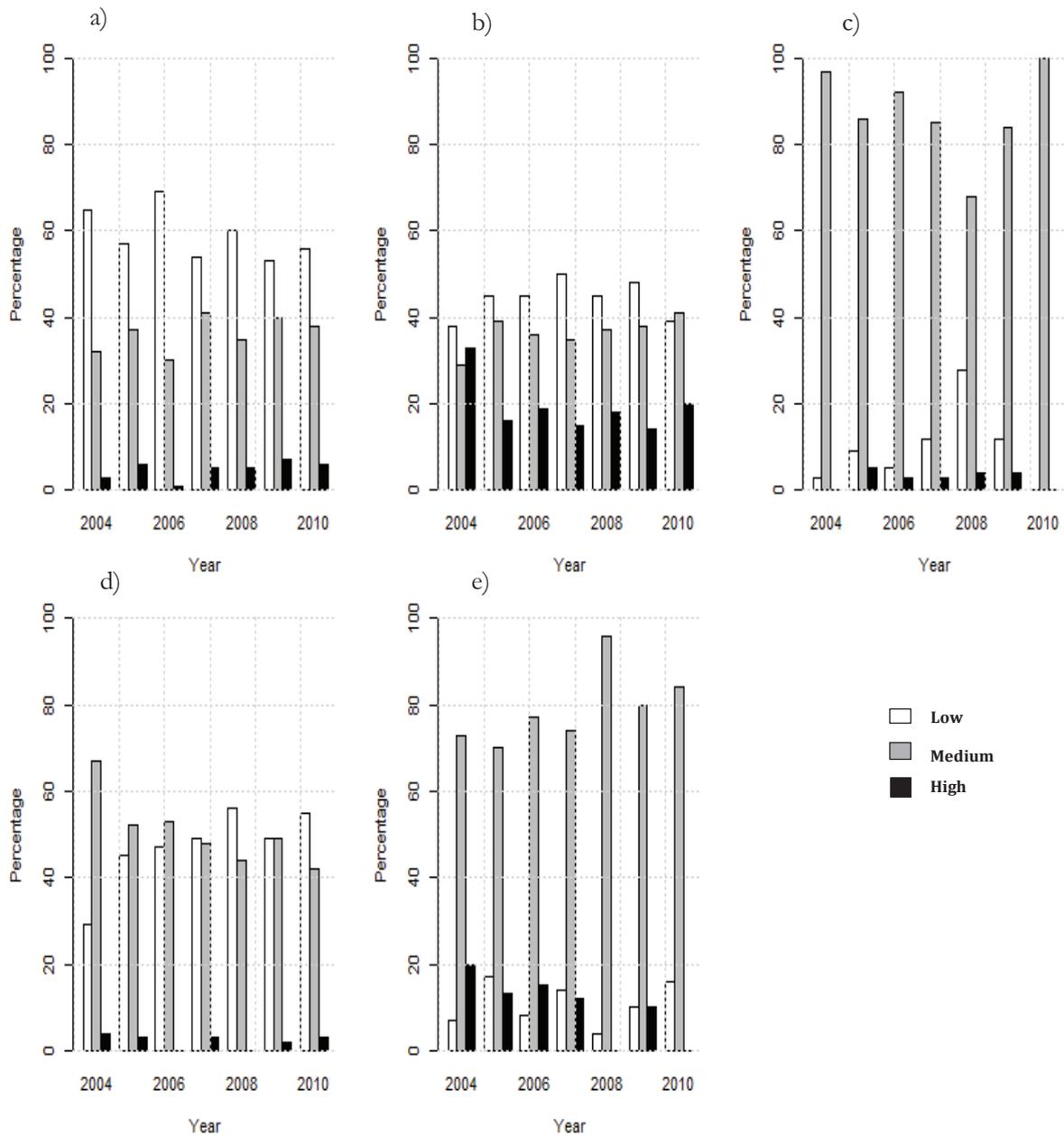


Figure 4.17. Percentages of volunteer phenological observations belonging to high, medium and low variable areas in The Netherlands (a) Cow Parsley (b) Celandine (c) Chestnut (d) Wood Anemone (e) Oak

For each species, the average number of existing observations belonging to low, medium and high variable areas were calculated from the yearly phenological observations (Table 4.11). Phenological observations in the medium variable areas were then used to determine the expected numbers of observations in low and high variable areas.

Table 4.11 Existing, expected and required number of volunteers in high, medium and low variable areas

| Species | Existing Number of Volunteers | | | Expected | | Required | |
|--------------|-------------------------------|--------|------|----------|------|----------|------|
| | Low | Medium | High | Low | High | Low | High |
| Cow parsley | 61 | 39 | 5 | 79 | 8 | 18 | 3 |
| Celandine | 76 | 62 | 29 | 67 | 36 | -9 | 7 |
| Chestnut | 6 | 47 | 2 | 11 | 4 | 5 | 2 |
| Wood anemone | 30 | 32 | 1 | 16 | 3 | -14 | 2 |
| Oak | 4 | 27 | 3 | 6 | 8 | 2 | 5 |

The results indicated the need for 18, 5 and 2 more volunteers for monitoring the phenological event of Cow parsley, Chestnut and Oak respectively, in low variable areas. Similarly, the required numbers of volunteers for high variable areas ranged between 2 (for Chestnut and wood anemone) and 7 (for Celandine). However, there were sufficient numbers of volunteer observations in low variable areas of Celandine (76) and Wood anemone (30). In general, this information could be used by phenological monitoring networks to give recommendations on the potential phenological monitoring areas of species.

4.5. PHASE 2.0 Model

4.5.1. Optimization

The maximum date of the phenological phase, which was used to determine the optimal starting DOY, was different for each plant species in the dataset. The maximum dates for Cow parsley, Celandine, Chestnut, Wood anemone and Oak were 360, 361, 274, 332 and 302, respectively. The optimization process that was used to find an optimal starting DOY of Oak is presented in Table 4.12. For each starting date, the efficiency values of model 2 (Efficiency2) and model 3 (Efficiency3) were computed and multiplied to choose the date with the maximum efficiency for model 2 and model 3.

First, daily temperature values from date 1 up to 302 were summed. The efficiencies of model 2 and model 3 with a starting date of 1 were 0.34 and 0.72, respectively. Secondly, daily temperature values were summed from date 2 up to 302. The efficiencies of model 2 and model 3 with a starting date of 2 were 0.36 and 0.75, respectively. In general, the same process was repeated until the starting date was equal to 302. Lastly, the starting date where there was a maximum value for Efficiency2 * Efficiency3, i.e. 42 for oak, was selected as the optimal starting DOY.

The optimal starting DOY that was determined by this iterative process is illustrated (Figure 4.18). The number of dots was equal to the number of iterations. That is, for each iteration, each dot have an efficiency value for model 2 (Efficiency2) and model 3 (Efficiency 3). The date where there was a maximum value of the product of Efficiency2 and Efficiency3 is shown in square box as an optimal starting DOY. The value shown on the y-axis was the efficiency of Model 2, which was used for predicting the phase-specific heat units from unspecific heat units. On the other hand, the value displayed on the x-axis was the efficiency of model 3, which predicted the DOY of phenological event from phase-

specific heat units. The optimal starting DOY for cow parsley, celandine and wood anemone was 1. Additionally, the optimal starting DOY of chestnut and oak were 100 and 42, respectively.

Table 4.12. Starting date and efficiencies of oak for model 2 and model 3 for iterations starting from $i=1\dots$ to $i=\text{max.ph}$

| Starting date | Efficiency2 | Efficiency3 | Efficiency2 * | Efficiency3 |
|---------------|-------------|-------------|---------------|-------------|
| 1 | 0.34 | 0.72 | 0.24 | |
| 2 | 0.36 | 0.75 | 0.27 | |
| 3 | 0.32 | 0.72 | 0.23 | |
| 4 | 0.36 | 0.74 | 0.27 | |
| . | . | . | . | . |
| . | . | . | . | . |
| 302 | . | . | . | . |

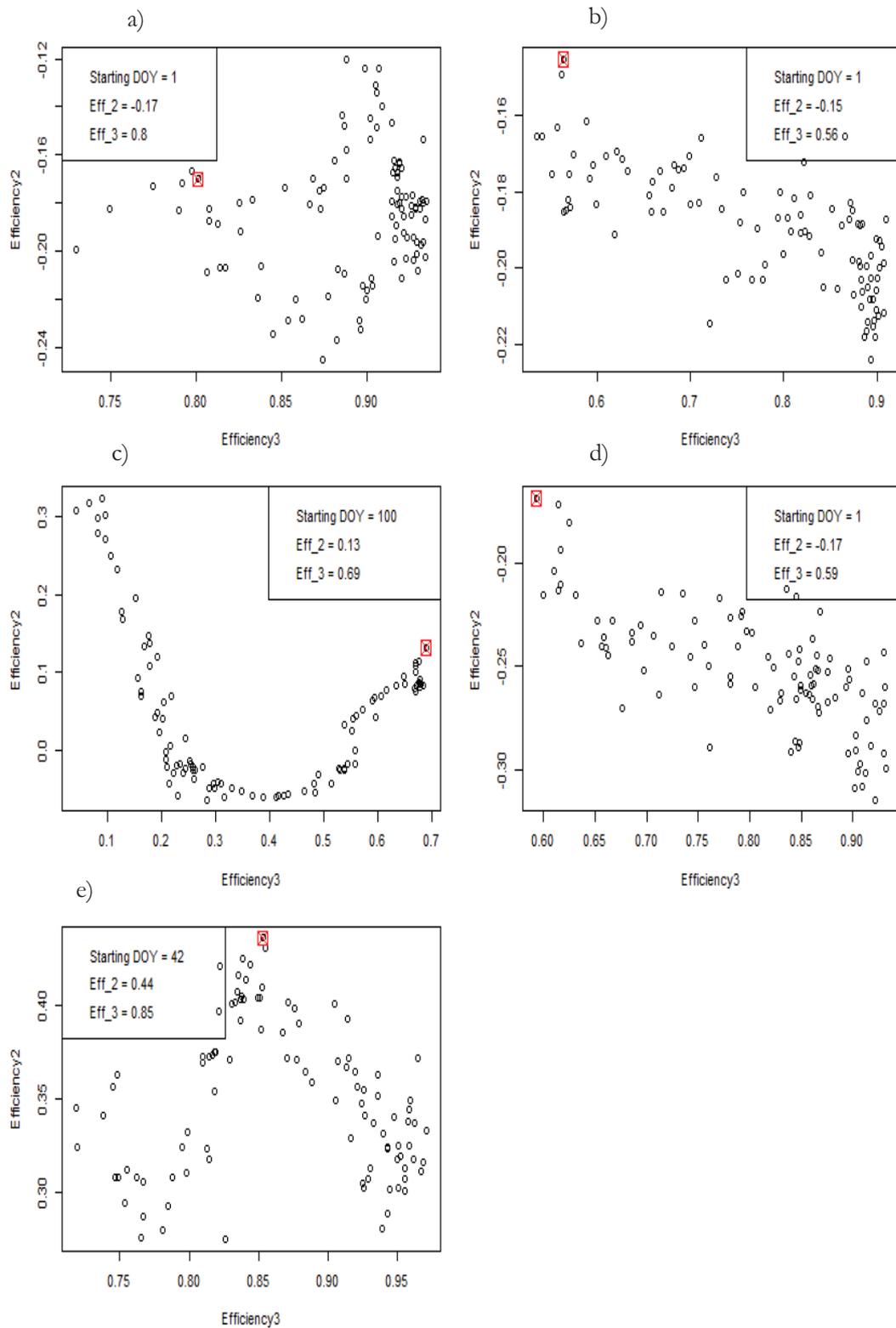


Figure 4.18. Optimization of the PHASE model for (a) Cow Parsley (b) Celandine (c) Chestnut (d) Wood Anemone (e) Oak. Each dot indicates an iteration, which was made for the choice of the starting DOY, from date 1 up to the date of maximum phase. Efficiencies of model 3 (Efficiency3) and model 2 (Efficiency 2) for every iteration are shown on the x and y-axis, respectively. The squared box indicates the iteration that resulted in a maximum Efficiency 2 * Efficiency3.

4.5.2. Model performance

The quality of PHASE model was assessed by using the RMSE and efficiency of both the calibration and validation dataset. The results of the quality measures on the calibration and validation dataset are

presented (Table 4.13). For the calibration dataset, the RMSE ranged between 7.47 (Chestnut) and 24.08 (Celandine) and the efficiency ranged between 0.49 (for Celandine) and 0.64 (for Chestnut). In addition, the RMSE of the validation dataset ranged between 11.27 (Chestnut) and 39.95 (Cow parsley) and the efficiency ranged between -2.27 (Chestnut) and -0.60 (Wood anemone). Hence, the model had shown better prediction results for chestnut and oak than the remaining species in the dataset.

Table 4.13. Statistical analysis of the PHASE model by using the calibration and validation dataset

| Name of Species | RMSE | Efficiency | RMSE | Efficiency |
|-----------------|-------------------|------------|------------------|------------|
| | Model Calibration | | Model Validation | |
| Cow parsley | 20.87 | 0.50 | 39.95 | -1.02 |
| Celandine | 24.08 | 0.49 | 33.26 | -1.11 |
| Chestnut | 7.47 | 0.64 | 11.27 | -2.27 |
| Wood Anemone | 19.93 | 0.50 | 20.23 | -0.60 |
| Oak | 14.33 | 0.63 | 16.06 | -0.99 |

4.6. Comparison of SW and PHASE 2.0 model

In order to compare the results of two different kinds of models (statistical and data-driven), the RMSE and efficiency of SW and PHASE model were compared (Smith et al., 2013). In the case of the calibration dataset, with temporally-biased observations, the PHASE model (Table 4.13) showed better prediction results than the SW model (Table 4.1). However, the PHASE model did not have a good predictive accuracy for the validation dataset. It was apparent that the RMSE and efficiency of the SW model (Table 4.4) was better than the PHASE model for the validation (independent) dataset.

In addition, we tested the performance of the PHASE model with a fixed starting DOY obtained from the SW model. The results of the PHASE model with a fixed starting DOY are presented in tabular form (Table 4.14). The RMSE of the model on the calibration dataset ranged between 7.54 and 20.77 and the efficiency ranged between 0.47 and 0.69. Hence, the optimal starting DOY determined by simulated annealing gave better prediction results for cow parsley, celandine, wood anemone and oak than the optimal starting DOY of PHASE model. Exceptionally, the efficiency and RMSE of the model on chestnut was slightly improved by the optimal starting DOY determined by the PHASE model.

On the other hand, the efficiency of the model on the validation dataset ranged between -1.64 and -1.2 and the RMSE ranged between 16.32 and 42.65 days. Therefore, the optimal starting DOY determined by the PHASE model had better prediction accuracy for the validation dataset than the fixed starting DOY. The overall results showed the varying sensitivity of the PHASE model on the five plant species.

Table 4.14. Statistical analysis of the PHASE model with a fixed starting DOY for calibration and validation dataset

| Name of Species | RMSE | Efficiency | RMSE | Efficiency |
|-----------------|-------------------|------------|------------------|------------|
| | Model Calibration | | Model Validation | |
| Cow parsley | 19.49 | 0.54 | 42.65 | -1.32 |
| Celandine | 20.77 | 0.63 | 37.87 | -1.64 |
| Chestnut | 7.54 | 0.47 | 10.15 | -1.60 |
| Wood Anemone | 16.88 | 0.60 | 24.51 | -1.47 |
| Oak | 11.84 | 0.69 | 16.32 | -1.20 |

Finally, the PHASE model was calibrated with non-temporally biased observations. That is, observations that were considered as temporally-biased (Section 4.4.1) were discarded and non-biased observations sets were used to calibrate the model. The results of the PHASE model for non-biased calibration and validation observation sets are presented (Table 4.15).

Table 4.15. Statistical analysis of the PHASE model after removal of temporally-biased phenological observations

| Name of Species | Optimal DOY | RMSE | Efficiency | RMSE | Efficiency |
|-----------------|-------------|-------------------|------------|------------------|------------|
| | | Model Calibration | | Model Validation | |
| Cow parsley | 100 | 6.81 | 0.68 | 25.67 | -0.34 |
| Celandine | 68 | 7.04 | 0.78 | 21.65 | -0.43 |
| Chestnut | 100 | 3.03 | 0.88 | 8.35 | -1.74 |
| Wood Anemone | 83 | 5.34 | 0.70 | 13.42 | -3.32 |
| Oak | 100 | 3.73 | 0.85 | 10.11 | -1.16 |

The removal of temporally-biased phenological observations improved the accuracy of the PHASE model (Table 4.15). The results showed an improvement in the RMSE and efficiency of the calibration and validation dataset. Similarly, new optimal starting DOY values were determined after removal of temporally-biased observations.

In general, the PHASE model showed better prediction accuracy for the calibration dataset (with and without temporally-biased observations) than the SW model. But, due to over-fitting of the model, the PHASE model showed worse prediction on the validation dataset than SW model.

5. DISCUSSION

5.1. Model Parameterization

5.1.1. Spring Warming

The rate of forcing in SW model increased with higher daily temperature values (Figure 4.1). These agrees with other findings (Olsson et al., 2013). In addition, the SW model did not show a good predictive accuracy in terms of RMSE and efficiency. This could be because of the effect of noise (i.e. temporally-biased observations) on parameterization of the model (Atkinson et al., 2012). For this reason, we assessed the performance of the SW model after removal of biased observations (Section 5.3.1)

5.1.2. UNIFORC

Our results for the UNIFORC model showed great difference in the optimal parameter values of T_0 and F^* from that of the SW model. The starting date (T_0) was earlier and, consequently, the heat requirement (F^*) became smaller.

5.1.3. UNICHILL

The UNICHILL model has more parameters than the other two models for accounting the chilling requirements. In line with other studies, the starting date of chill temperature accumulation was fixed to 1 September (Parker et al., 2011). The RMSE, efficiency and correlation of the UNICHILL model were worse than the SW and UNIFORC model. This model has more parameters which might not be fitted with small dataset. This in turn can result in over-fitting of the model. This finding is in agreement with other studies (Vitasse et al., 2011).

In general, we detected that all plant species had differing parameter values for the date of dormancy and phenological stage. This was because each species had differing sensitivity to climatic variation (Olsson et al., 2013). Similarly, the values of C^* and F^* were species-dependent and differed for the five plant species in the 'Natuurkalender' dataset.

5.2. Model Performance and Comparison

The model comparison in this study helped in choosing the modelling approaches that were suitable for the phenophases of the five plant species. The result of this research on the best performing model (i.e. SW) is in line with previous studies which discovered that forcing models perform better than models with a calculation of chilling requirement (Fu et al., 2012b; Richardson and O'Keefe, 2009). In addition, the SW model showed varying sensitivity among the five plant species. This is also supported by research findings of other scholars who mentioned the difficulties of building a model suitable for all species (Olsson et al., 2013). In connection to that, the RMSE, efficiency and correlation values of each phenological model were species dependent and differed for each species (Chuine et al., 1998). Overall, this study suggests that under the current climate conditions, chilling temperatures does not have a major effect on the investigated plant species and phenological events.

Although, smaller RMSE values were expected for the calibration dataset than the validation dataset (Fu et al., 2012b), we obtained higher RMSE, smaller efficiency and correlation values for the calibration dataset than the validation. This could be caused by noise (temporally-biased observations) in the data which could result in biased parameter values and lead to incorrect modelling results (Atkinson et al., 2012; Liu et al., 2004). However, these results are improved after the removal of temporally-biased observations from the calibration and validation dataset (Section 5.3). Moreover, the model showed large differences between the RMSE of the calibration and validation dataset. These differences could be attributed to

differences between the calibration and validation dataset. This was supported by the ANOVA test, which indicated that the means of the calibration and validation dataset were different.

5.3. Bias Identification

5.3.1. Temporal bias

The model showed better efficiency and correlation values as more observations were removed. But, phenological observations with valuable information could be lost as several observations were removed from the dataset. Therefore, there was a trade-off between the number of observations removed and the desired efficiency and correlation values of the model. In addition, as the model had different adaptation to the plant species, the systematic and unsystematic RMSE of the model was different depending on the species. For this reason, the optimal normalized SD threshold differed among plant species.

Moreover, temporally-biased observations were randomly spread throughout the study area. That is, for all plant species, biased observations were not only located in areas where there was high climatic variability. Although the years 2006 and 2007 had the highest climatic variability between 1.3 and 1.8 and -2.3 and -1.9 respectively, there were more temporally-biased observations in the year 2005 where the SD ranged between -0.6 and 0.2. This indicated that the observations that were regarded as temporally-biased were not merely caused by the climatic variability in the area. This finding holds true for all of the plant species that were included in the investigation. Further, the biasedness could be linked to the quality of input data which might have been caused by the weekend (Żmihorski et al., 2012) or urban area effects (Bishr and Mantelas, 2008).

5.3.2. Parameterization of model after removal of temporal bias

As expected, the temporally-biased observations were outliers which could not be captured by the model. These observations affected the optimal parameter values that were determined by simulated annealing. Once these observations were removed, different parameter values were obtained. Unlike other species, oak showed a significant change of parameter values after removal of biased observations. This could be associated with the existence of 4% temporally-biased observations among the limited number of phenological observations (269) that was used to calibrate the SW model. Hence, the results of the optimization algorithm were affected by 4% of the noise in the data. In connection to this, oak showed a higher accuracy on the validation dataset than the calibration data after the removal of temporally-biased observations.

For this reason, we performed a second round analysis of systematic RMSE, unsystematic RMSE and RMSE. Four of the species in the dataset has a systematic RMSE smaller than the unsystematic RMSE but Oak had a higher systematic RMSE than the unsystematic. This indicated that our model had good prediction power for four of the species but needed reworking for Oak. Therefore, we did a second round removal of temporally-biased observations and determined new optimal parameter values. With these values, the RMSE, efficiency and correlation of the model were better for the calibration dataset than the validation. Additionally, the removal of temporally-biased observations from all species minimized the difference between the prediction results of the calibration and validation dataset.

We can conclude that the SW model had a better prediction precision for the calibration dataset of the five species than the validation dataset, as expected (Fu et al., 2012b). The results also confirmed the responsiveness of the phenological stages of the plants to temperature. In addition, the RMSE, efficiency and correlation of the model showed the ability of the models to adjust to variation in the data. Since there was no large difference between the RMSE of the calibration and validation dataset, we could assume that the parameters of the model were not dependent on the calibration dataset and the model was not over-fitted (Linkosalo et al., 2008).

The current parameter values of SW model could be different from parameter values obtained in other study areas depending on the varieties of the species that were studied (Leinonen and Hänninen, 2002). For instance, the parameter values that were obtained for Chestnut (after removal of temporally-biased observations) were compared with parameter values from literature. The parameter values of Chestnut, which were determined by SW model, with pollen data for Tb, F* and T0 in France were 13.9, 20.1 and 3, respectively (Chuine et al., 1998). Those values that were obtained in France were different from the current values, which were 3.27, 290 and 57 for Tb, F* and T0, respectively, in our study. We run simulated annealing for the same species a number of times to check if the algorithm was trapped in local minima and giving different parameter values at each run. However, the algorithm found similar results for the optimal parameter values. This indicated that the algorithm was not trapped into local minima rather it was finding the global minimum. The differences in the data used for model calibration could be one reason for the difference in parameter values. Unlike the pollen data, which was obtained from the weekly average of pollen shedding in the atmosphere, we used volunteer phenological observations to calibrate the model. For the pollen data, the middle day of the pollination week was considered as the flowering date of the species. In this study, the DOY of the phenological event (flowering, leafing) was the date of phenological observation reported by the volunteers. Similarly, the climatic and topographic variations across different regions and varietal differences could account for the different parameter values.

The model accounted the inter-annual variation but failed in elucidating good prediction for the annual variability. The differences in landscape and topography could account for climatic variations among the meteorological stations and the locations of the plants species that were explored (Fu et al., 2012b). In connection to that, the annual variability could be caused by the coarse temperature data, which was interpolated in 1km raster file, and used for model calibration. To this effect, the model couldn't handle annual variations. In addition, the inability of the model to consider other factors, like photoperiod, could have influenced the model from capturing the annual variability (Fu et al., 2012a; Menzel et al., 2006). Besides, the higher temperature in the southern part of The Netherlands than the northern part could affect the starting date of temperature accumulation. Consequently, an early starting date would enable the SW model to implicitly capture more chilling days in the north whilst the model would not enable to accumulate additional chilling days due to higher temperatures in the south (Olsson et al., 2013). This variation in the prediction results of the northern and southern part of The Netherlands could cause annual variations that could not be captured by the model.

5.3.3. Spatial bias

Plants of the same species, which were growing in different locations, could have local adaptations to the DOY of the phenological stage when exposed to varying climatic conditions (Fu et al., 2012b). In agreement with this, the DOY of phenological event of the five plant species were different between the years 2003 and 2010 because of climatic variations.

The SD of dates of phenological events among the years was used to detect climate change (Sparks et al., 2011; Vilhar et al., 2013). In line with such studies, we used the SD of phenological events between 8 years to determine areas with climatic variability. Hence, phenology was used as a climate change indicator across the whole area of The Netherlands. Since there was higher temperature in the southern part, there was an earlier onset of phenological stage for all species. This is in line with the trends that have been encountered in Europe (Menzel et al., 2006; Olsson et al., 2013). In addition, inter-species variability in DOY of phenological phases of the five species showed the varying sensitivity of species to different climatic conditions. Moreover, as the medium variable areas were the average (mid-point) values, the existing numbers of observations in the medium variable areas were used to determine the expected number of observations in low and high variable areas. This is in agreement with other studies (Shi et al.,

2012), which showed the importance of the number of phenological observations in an area and its effect on the DOY of the phenological event.

5.4. PHASE 2.0 Model

5.4.1. Optimization

The optimization of the PHASE model was not efficient because it was calculating psHU from date 1 up to the date of the maximum phase. Since there were outliers (temporally-biased observations) in the input phenological observations, the date of the maximum phase was not also accurate. Hence, the search space for the optimal starting DOY should be constrained based on the physiological behavior of the plants or the maximum date of phenophase of non-biased observation sets. For this reason, we have limited the search space between 1 and 100 by looking at the distribution of volunteer phenological observations.

Moreover, choosing the optimal starting DOY of the model based on the efficiency values computed from the OOB data could result in incorrect results. As discussed in section 5.4.2, the PHASE model showed bad performance on an independent dataset. Hence, in addition to the efficiency of the OOB data, the efficiency and RMSE on the training data should have been considered to determine the optimal starting DOY. Besides, the over-fitting of the model should be prevented to enable it to work on an independent dataset.

5.4.2. Model Performance

As the PHASE model does not make prior assumptions about the distribution of the data, it could learn from the data and make accurate predictions for the calibration data. However, the model showed higher RMSE and lower efficiency values for the validation dataset. This could be an "over fitting" problem which was caused by over-learning the training data and inability to make predications to an independent (test) data. In addition, the model might have fitted random noise in the data. Besides, there were large differences in the RMSE of the PHASE model between the calibration and validation dataset. These differences were caused by the over-fitting of the model on the calibration (training) dataset. This finding is in line with other studies (Linkosalo et al., 2008).

5.5. Comparison of SW and PHASE 2.0 Model

The optimal starting DOY of the PHASE model determined by the efficiency of the model on the OOB data resulted in incorrect values of optimal starting DOY. These in turn decreased the efficiency and RMSE of the PHASE model for the calibration and validation dataset. On the contrary, the optimal starting DOY determined by the SW model showed better results than the optimal DOY of the PHASE model for the calibration dataset. As the optimal starting DOY of PHASE model was determined by learning the training data, it showed better results for the validation dataset than the fixed starting DOY.

Moreover, the results indicated that the PHASE model had better prediction accuracy for the calibration dataset (with temporally biased observations) than the SW model. This indicated that the PHASE model was more robust to outliers than the SW model. On the contrary, the PHASE model showed low accuracy on the validation dataset. This might be caused by over-fitting of the training data, which might occur if the regression trees grown at each generation were complex. The robustness of random forest regression for outliers and the case of over-fitting was also supported in other findings (Hastie et al., 2003). In the case of this data set, PHASE model appeared to be superior to SW model in terms of RMSE and efficiency for the calibration dataset. In contrast, the SW model performed well for the validation (independent) dataset than the PHASE model.

6. CONCLUSIONS AND RECOMMENDATIONS

The main objective of this MSc thesis research was achieved by completing the sub-objectives. This section explains how the specific research objectives were achieved by answering the research questions.

6.1. Conclusions

1. Which statistical phenological modeling approach best fits the species/phenophase combinations of volunteered phenological observations?

In our study, we tested the performance of SW, UNIFORC and UNICHILL models on volunteer phenological observations. The models were fitted by using a global optimization algorithm called simulated annealing. Simulated annealing could find the global optimum without getting trapped into a local optimum. The SW model was found to be the best model in predicting the phenological stages of the five plant species in the Natuurkalender dataset.

The existence of temporally-biased observations caused biasedness in the parameter estimates as well as modelling results including RMSE, efficiency and correlation between observed and simulated values. Once these biased observations were removed, the quality (RMSE, efficiency and correlation) of the model was improved. On the other hand, the PHASE model used volunteer phenological observations to predict the phase-specific heat units and dates of phenological stages of the five plant species. This model was robust to outliers and could not easily be affected by the noise in the data. The performance of the model on the calibration dataset was better than the SW model. As a result of over-fitting, the prediction results of the PHASE model for the validation dataset were worse than the SW model. Eventually, the five plant species showed varying sensitivity to temperature and had different RMSE and efficiency. On the one hand, the SW and PHASE model showed better results in RMSE and efficiency for Chestnut and Oak, respectively. On the other hand, they showed worse results for Cow Parsley and Celandine.

2. How can we systematically exclude erroneous observations from phenological models based on volunteered geo-information?

The best statistical phenological model, SW, was used to detect and remove phenological observations that were temporally-biased. The SW model predicts the DOY of phenological event at locations of observations. The errors of each observation were computed and observations above a given normalized SD threshold were removed and the systematic and unsystematic RMSE of the model was checked. The normalized SD value by which the systematic RMSE of the model was smaller than the unsystematic RMSE and the smallest number of observations removed was selected as the threshold for removing temporally-biased observations. Hence, observations above the chosen species-specific threshold were regarded as erroneous (temporally-biased) observation sets.

3. How can the temporal windows of monitoring a phenophase be determined from the modeling results?

The DOY of the phenological event in the whole area of The Netherlands between the years 2003 and 2010 were different. In addition, a given year also showed variability in DOY at different locations. In connection to this, for each year, we determined the duration of phenological event (temporal windows). As there were climatic variations, the temporal windows had inter-annual variations. The general temporal windows that were exhibited between the 8 years period were between 85 and 127 for Cow parsley, 54 and 100 for Celandine, 106 and 135 for Chestnut, 73 and 104 for Wood anemone and 101 and 128 for Oak. In general, volunteers could be recommended to start monitoring the phenological events of these species during these periods. Specifically, the maps showing the DOY of the phenological event could help in determining the temporal windows at locations of interest.

4. How can phenological dynamics be used to determine sampling schemes?

The phenological dynamics on the DOY of the phenological events of the five plant species between the years 2003 and 2010 helped in determining areas that were spatially-biased. For every year and species, the SW model predicted the DOY of the phenological event based on daily temperature values. Then, areas showing variability on the DOY of each species during eight years were determined and grouped into high, medium and low variable areas. Hence, areas showing variability were regarded as spatially-biased and they were different for each species. In connection to that, areas that need to have more monitoring volunteers were determined. This information could be useful to give recommendations for volunteers on the potential phenological monitoring areas.

6.2. Recommendations

Based on the research data that were generated, the following recommendations can be made for future research works:

- Prevent the over-fitting of PHASE model by either using phenological observations of longer observation periods or by tuning parameters of the random forest regression. This could enable the model to make accurate predictions on an independent test dataset.
- Improve the phenological models to account for annual variability by using local temperature that are interpolated using fine grid temperature data.
- Incorporate other climatic factors (explanatory variables), which affect the timing of phenological events in the models. In connection to this, considering more climatic variables might improve modelling results and help determine accurate parameter values. In this case, volunteers should be made to provide details (metadata) about the location and phenological phase they are reporting.
- Test the results of this research in another study area or use different data, like pollen data and make comparison with VGI data to assess the value of VGI for environmental monitoring.
- The resulting maps showing the DOY of phenological event and SD maps showing climatic variations could be published in websites to be used by other users.

LIST OF REFERENCES

- Abdel-Rahman, E. M., Ahmed, F. B., and Ismail, R. (2012). Random forest regression and spectral band selection for estimating sugarcane leaf nitrogen concentration using EO-1 Hyperion hyperspectral data. *International Journal of Remote Sensing* **34**, 712-728.
- Alemu, G., Jing, M. C., and Chaoyang, W. (2013). Citizen Science: linking the recent rapid advances of plant flowering in Canada with climate variability. *Scientific Reports* **3**.
- Amorim, A. M. T., Gonçalves, A. B., Nunes, L. M., and Sousa, A. J. (2012). Optimizing the location of weather monitoring stations using estimation uncertainty. *International Journal of Climatology* **32**, 941-952.
- Atkinson, P. M., Jeganathan, C., Dash, J., and Atzberger, C. (2012). Inter-comparison of four models for smoothing satellite sensor time-series data to estimate vegetation phenology. *Remote Sensing of Environment* **123**, 400-417.
- Beaubien, E., and Hamann, A. (2011). Plant phenology networks of citizen scientists: recommendations from two decades of experience in Canada. *International Journal of Biometeorology* **55**, 833-841.
- Bellocchi, G., Rivington, M., Donatelli, M., and Matthews, K. (2010). Validation of biophysical models: issues and methodologies. A review. *Agronomy for Sustainable Development* **30**, 109-130.
- Betancourt, J. L., Schwartz, M. D., Breshears, D. D., Brewer, C. A., Frazer, G., Gross, J. E., Mazer, S. J., Reed, B. C., and Wilson, B. E. (2007). Evolving plans for the USA National Phenology Network. *Eos, Transactions American Geophysical Union* **88**, 211-211.
- Bird, T. J., Bates, A. E., Lefcheck, J. S., Hill, N. A., Thomson, R. J., Edgar, G. J., Stuart-Smith, R. D., Wotherspoon, S., Krkosek, M., Stuart-Smith, J. F., Pecl, G. T., Barrett, N., and Frusher, S. (2013). Statistical solutions for error and bias in global citizen science datasets. *Biological Conservation*.
- Bishr, M., and Mantelas, L. (2008). A trust and reputation model for filtering and classifying knowledge about urban growth. *GeoJournal* **72**, 229-237.
- Bluman, A. G. (2012). "Elementary statistics : a step by step approach," McGraw-Hill, New York.
- Blümel, K., and Chmielewski, F.-M. (2012). Shortcomings of classical phenological forcing models and a way to overcome them. *Agricultural and Forest Meteorology* **164**, 10-19.
- Bolmgren, K., Vanhoenacker, D., and Miller-Rushing, A. J. (2013). One man, 73 years, and 25 species. Evaluating phenological responses using a lifelong study of first flowering dates. *International Journal of Biometeorology* **57**, 367-375.
- Bonney, R., Cooper, C. B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K. V., and Shirk, J. (2009). Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy. *BioScience* **59**, 977-984.
- Bordogna, G., Carrara, P., Criscuolo, L., Pepe, M., and Rampini, A. (2013). A linguistic decision making approach to assess the quality of volunteer geographic information for citizen science. *Information Sciences*.
- Braunisch, V., and Suchant, R. (2010). Predicting species distributions based on incomplete survey data: the trade-off between precision and scale. *Ecography* **33**, 826-840.
- Breiman, L. (2001). Random Forests. *Machine Learning* **45**, 5-32.
- Brown, M., Sharples, S., Harding, J., Parker, C. J., Bearman, N., Maguire, M., Forrest, D., Haklay, M., and Jackson, M. (2013). Usability of Geographic Information: Current challenges and future directions. *Applied Ergonomics* **44**, 855-865.
- Cannell, and Smith, R. I. (1983). Thermal time, chill days and prediction of budburst in *Picea sitchensis*. *Journal of Applied Ecology* **20**, 951-963.
- Ceglar, A., Črepinšek, Z., Kajfež-Bogataj, L., and Pogačar, T. (2011). The simulation of phenological development in dynamic crop model: The Bayesian comparison of different methods. *Agricultural and Forest Meteorology* **151**, 101-115.
- Cheng, T., Haworth, J., and Manley, E. (2012). Advances in geocomputation (1996–2011). *Computers, Environment and Urban Systems* **36**, 481-487.
- Chmielewski, F.-M., Blümel, K., Scherbaum-Heberer, C., Koppmann-Rumpf, B., and Schmidt, K.-H. (2013). A model approach to project the start of egg laying of Great Tit (*Parus major* L.) in response to climate change. *International Journal of Biometeorology* **57**, 287-297.
- Chuine, I., and Cour, P. (1999). Climatic determinants of budburst seasonality in four temperate-zone tree species. *New Phytologist* **143**, 339-349.

- Chuine, I., Cour, P., and Rousseau, D. D. (1998). Fitting models predicting dates of flowering of temperate-zone trees using simulated annealing. *Plant, Cell & Environment* **21**, 455-466.
- Chuine, I., Cour, P., and Rousseau, D. D. (1999). Selecting models to predict the timing of flowering of temperate trees: implications for tree phenology modelling. *Plant, Cell & Environment* **22**, 1-13.
- Chuine, I., Kramer, K., and Hänninen, H. (2003). Plant development models. *Phenology: An Integrative Environmental Science*, 217-235.
- Clark, R. M., and Thompson, R. (2010). Predicting the impact of global warming on the timing of spring flowering. *International Journal of Climatology* **30**, 1599-1613.
- Courter, J., Johnson, R., Stuyck, C., Lang, B., and Kaiser, E. (2013). Weekend bias in Citizen Science data reporting: implications for phenology studies. *International Journal of Biometeorology* **57**, 715-720.
- Cui, Y. (2013). A systematic approach to evaluate and validate the spatial accuracy of farmers market locations using multi-geocoding services. *Applied Geography* **41**, 87-95.
- Delaney, D., Sperling, C., Adams, C., and Leung, B. (2008). Marine invasive species: validation of citizen science and implications for national monitoring networks. *Biological Invasions* **10**, 117-128.
- Deligios, P. A., Farci, R., Sulas, L., Hoogenboom, G., and Ledda, L. (2013). Predicting growth and yield of winter rapeseed in a Mediterranean environment: Model adaptation at a field scale. *Field Crops Research* **144**, 100-112.
- Do, H. T., Lo, S.-L., Chiueh, P.-T., and Phan Thi, L. A. (2012). Design of sampling locations for mountainous river monitoring. *Environmental Modelling & Software* **27-28**, 62-70.
- Ferster, C. J., and Coops, N. C. (2013). A review of earth observation using mobile personal communication devices. *Computers & Geosciences* **51**, 339-349.
- Fink, D., Hochachka, W. M., Zuckerberg, B., Winkler, D. W., Shaby, B., Munson, M. A., Hooker, G., Riedewald, M., Sheldon, D., and Kelling, S. (2010). Spatiotemporal exploratory models for broad-scale survey data. *Ecological Applications* **20**, 2131-2147.
- Fishman, S., Erez, A., and Couvillon, G. A. (1987). The temperature dependence of dormancy breaking in plants: Mathematical analysis of a two-step model involving a cooperative transition. *Journal of Theoretical Biology* **124**, 473-483.
- Flanagan, A., and Metzger, M. (2008). The credibility of volunteered geographic information. *GeoJournal* **72**, 137-148.
- Fu, Y. H., Campioli, M., Demarée, G., Deckmyn, A., Hamdi, R., Janssens, I. A., and Deckmyn, G. (2012a). Bayesian calibration of the Unified budburst model in six temperate tree species. *International Journal of Biometeorology* **56**, 153-164.
- Fu, Y. H., Campioli, M., Van Oijen, M., Deckmyn, G., and Janssens, I. A. (2012b). Bayesian comparison of six different temperature-based budburst models for four temperate tree species. *Ecological Modelling* **230**, 92-100.
- Goodchild, M. F. (2011). Citizens as Sensors: The World of Volunteered Geography. In "The Map Reader", pp. 370-378. John Wiley & Sons, Ltd.
- Gouveia, C., Fonseca, A., Câmara, A., and Ferreira, F. (2004). Promoting the use of environmental data collected by concerned citizens through information and communication technologies. *Journal of Environmental Management* **71**, 135-154.
- Grunsky, E. C. (2002). R: a data analysis and statistical programming environment—an emerging tool for the geosciences. *Computers & Geosciences* **28**, 1219-1222.
- Haggett, P., Cliff, A. D., and Frey, A. (1977). "Locational analysis in human geography : 2. locational methods," Second edition/Ed. Edward Arnold, London.
- Harding, J., Sharples, S., Haklay, M., Burnett, G., Dadashi, Y., Forrest, D., Maguire, M., Parker, C. J., and Ratcliff, L. (2009). Usable geographic information—what does it mean to users? In "AGI GeoCommunity '09". AGI GeoCommunity.
- Hastie, T., Tibshirani, R., and Friedman, J. (2003). "The Elements of Statistical Learning: Data Mining, Inference, and Prediction," Springer.
- Hochachka, W. M., Caruana, R., Fink, D., Munson, A. R. T., Riedewald, M., Sorokina, D., and Kelling, S. (2007). Data-Mining Discovery of Pattern and Process in Ecological Systems. *Journal of Wildlife Management* **71**, 2427-2437.
- Hochachka, W. M., Fink, D., Hutchinson, R. A., Sheldon, D., Wong, W.-K., and Kelling, S. (2012a). Data-intensive science applied to broad-scale citizen science. *Trends in Ecology & Evolution* **27**, 130-137.
- Hochachka, W. M., Fink, D., Hutchinson, R. A., Sheldon, D., Wong, W.-K., and Kelling, S. (2012b). Data-intensive science applied to broad-scale citizen science. *Trends in ecology & evolution (Personal edition)* **27**, 130-137.

- Hunter, A. F., and Lechowicz, M. J. (1992). Predicting the timing of budburst in temperate trees. *Journal of Applied Ecology* **29**, 597-604.
- Jiguet, F., Devictor, V., Julliard, R., and Couvet, D. (2012). French citizens monitoring ordinary birds provide tools for conservation and ecological sciences. *Acta Oecologica* **44**, 58-66.
- Jun, Y., Kelling, S., Gerbracht, J., and Weng-Keen, W. (2012). Automated data verification in a large-scale citizen science project: A case study. In "E-Science (e-Science), 2012 IEEE 8th International Conference on", pp. 1-8.
- Júnez-Ferreira, H. E., and Herrera, G. S. (2013). A geostatistical methodology for the optimal design of space-time hydraulic head monitoring networks and its application to the Valle de Querétaro aquifer. *Environmental Monitoring and Assessment* **185**, 3527-3549.
- Kelling, S., Gerbracht, J., Fink, D., Lagoze, C., Wong, W.-K., Yu, J., Damoulas, T., and Gomes, C. (2012). "eBird: A Human/Computer Learning Network for Biodiversity Conservation and Research."
- Kim, C. (2009). Spatial Data Mining, Geovisualization. In "International Encyclopedia of Human Geography" (R. Kitchin and N. Thrift, eds.), pp. 332-336. Elsevier, Oxford.
- Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983). Optimization by Simulated Annealing. *Science* **220**, 671-680.
- KNMI (2013). Royal Netherlands Meteorological Institute. Vol. 2013.
- Kramer, K. (1994). SELECTING A MODEL TO PREDICT THE ONSET OF GROWTH OF FAGUS-SYLVAICA. *Journal of Applied Ecology* **31**, 172-181.
- Leinonen, I., and Hänninen, H. (2002). Adaptation of the timing of bud burst of Norway spruce to temperate and boreal climates. *Silva Fennica* **36**, 695-701.
- Levrel, H., Fontaine, B., Henry, P.-Y., Jiguet, F., Julliard, R., Kerbiriou, C., and Couvet, D. (2010). Balancing state and volunteer investment in biodiversity monitoring for the implementation of CBD indicators: A French example. *Ecological Economics* **69**, 1580-1586.
- Li, F., and Zhang, L. (2007). Comparison of point pattern analysis methods for classifying the spatial distributions of spruce-fir stands in the north-east USA. *Forestry* **80**, 337-349.
- Linkosalo, T., Lappalainen, H. K., and Hari, P. (2008). A comparison of phenological models of leaf bud burst and flowering of boreal trees using independent observations. *Tree Physiology* **28**, 1873-1882.
- Liu, H., Shah, S., and Jiang, W. (2004). On-line outlier detection and data cleaning. *Computers & Chemical Engineering* **28**, 1635-1647.
- Lovell, S., Hamer, M., Slotow, R., and Herbert, D. (2009). An assessment of the use of volunteers for terrestrial invertebrate biodiversity surveys. *Biodiversity and Conservation* **18**, 3295-3307.
- Luedeling, E. (2012). Climate change impacts on winter chill for temperate fruit and nut production: A review. *Scientia Horticulturae* **144**, 218-229.
- Luedeling, E., and Gassner, A. (2012). Partial Least Squares Regression for analyzing walnut phenology in California. *Agricultural and Forest Meteorology* **158-159**, 43-52.
- Luedeling, E., Zhang, M., McGranahan, G., and Leslie, C. (2009). Validation of winter chill models using historic records of walnut phenology. *Agricultural and Forest Meteorology* **149**, 1854-1864.
- Maisonneuve, N., Stevens, M., Niessen, M., and Steels, L. (2009). NoiseTube: Measuring and mapping noise pollution with mobile phones. In "Information Technologies in Environmental Engineering" (I. Athanasiadis, A. Rizzoli, P. Mitkas and J. Gómez, eds.), pp. 215-228. Springer Berlin Heidelberg.
- Malenovský, Z., Homolová, L., Zurita-Milla, R., Lukeš, P., Kaplan, V., Hanuš, J., Gastellu-Etchegorry, J.-P., and Schaepman, M. E. (2013). Retrieval of spruce leaf chlorophyll content from airborne image data using continuum removal and radiative transfer. *Remote Sensing of Environment* **131**, 85-102.
- Mendez, G., and Lohr, S. (2011). Estimating residual variance in random forest regression. *Computational Statistics & Data Analysis* **55**, 2937-2950.
- Mennis, J., and Guo, D. (2009). Spatial data mining and geographic knowledge discovery—An introduction. *Computers, Environment and Urban Systems* **33**, 403-408.
- Menzel, A., Sparks, T. H., Estrella, N., Koch, E., Aasa, A., Ahas, R., Alm-Kübler, K., Bissolli, P., Braslavská, O., Briede, A., Chmielewski, F. M., Crepinsek, Z., Curnel, Y., Dahl, Å., Defila, C., Donnelly, A., Filella, Y., Jatczak, K., Mâge, F., Mestre, A., Nordli, Ø., Peñuelas, J., Pirinen, P., Remišová, V., Scheifinger, H., Striz, M., Susnik, A., Van Vliet, A. J. H., Wielgolaski, F. E., Zach, S., and Züst, A. (2006). European phenological response to climate change matches the warming pattern. *Global Change Biology* **12**, 1969-1976.

- Miller, H. J., and Han, J. (2009). "Geographic data mining and knowledge discovery," Second edition/Ed. Taylor & Francis, Boca Raton.
- Miller, J. (1991). Short report: Reaction time analysis with outlier exclusion: Bias varies with sample size. *The Quarterly Journal of Experimental Psychology Section A* **43**, 907-912.
- Miranda, C., Santesteban, L. G., and Royo, J. B. (2013). Evaluation and fitting of models for determining peach phenological stages at a regional scale. *Agricultural and Forest Meteorology* **178–179**, 129-139.
- Møller, A. P., Rubolini, D., and 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* **105**, 16195-16200.
- Moller, M., Gla, x00Df, er, C., and Birger, J. (2011). Automatic interpolation of phenological phases in Germany. In "Analysis of Multi-temporal Remote Sensing Images (Multi-Temp), 2011 6th International Workshop on the", pp. 37-40.
- Natuurkalender (2013). The Nature. Vol. 2013.
- Olsson, C., Bolmgren, K., Lindström, J., and Jönsson, A. M. (2013). Performance of tree phenology models along a bioclimatic gradient in Sweden. *Ecological Modelling* **266**, 103-117.
- Parker, A. K., De CortÁZar-Atauri, I. G., Van Leeuwen, C., and Chuine, I. (2011). General phenological model to characterise the timing of flowering and veraison of *Vitis vinifera* L. *Australian Journal of Grape and Wine Research* **17**, 206-216.
- Plata, S. (2006). A note on Fisher's correlation coefficient. *Applied Mathematics Letters* **19**, 499-502.
- Polgar, C. A., Primack, R. B., Williams, E. H., Stichter, S., and Hitchcock, C. (2013). Climate effects on the flight period of Lycaenid butterflies in Massachusetts. *Biological Conservation* **160**, 25-31.
- Reynolds, J. H., Thompson, W. L., and Russell, B. (2011). Planning for success: Identifying effective and efficient survey designs for monitoring. *Biological Conservation* **144**, 1278-1284.
- Richardson, A., and O'Keefe, J. (2009). Phenological Differences Between Understory and Overstory. In "Phenology of Ecosystem Processes" (A. Noormets, ed.), pp. 87-117. Springer New York.
- Richardson, A. D., Keenan, T. F., Migliavacca, M., Ryu, Y., Sonnentag, O., and Toomey, M. (2013). Climate change, phenology, and phenological control of vegetation feedbacks to the climate system. *Agricultural and Forest Meteorology* **169**, 156-173.
- Roche, S., Propeck-Zimmermann, E., and Mericskay, B. (2013). GeoWeb and crisis management: issues and perspectives of volunteered geographic information. *GeoJournal* **78**, 21-40.
- Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., and Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing* **67**, 93-104.
- Sacks, W. J., and Kucharik, C. J. (2011). Crop management and phenology trends in the U.S. Corn Belt: Impacts on yields, evapotranspiration and energy balance. *Agricultural and Forest Meteorology* **151**, 882-894.
- Sáez, P. B., and Rittmann, B. E. (1992). Model-parameter estimation using least squares. *Water Research* **26**, 789-796.
- Saracco, J. F., Desante, D. F., Nott, M. P., HOCHACHKA, W. M., KELLING, S., and FINK, D. (2008). Integrated bird monitoring and the Avian Knowledge Network: Using multiple data resources to understand spatio-temporal variation in demographic processes and abundance. *Tundra to Tropics: Connecting Birds, Habitats and People. Proceedings of the 4th International Partners in Flight Conference*, 13-16.
- Scheifinger, H., Menzel, A., Koch, E., Peter, C., and Ahas, R. (2002). Atmospheric mechanisms governing the spatial and temporal variability of phenological phases in central Europe. *International Journal of Climatology* **22**, 1739-1755.
- Schwartz, M. D. (1999). Advancing to full bloom: planning phenological research for the 21st century. *International Journal of Biometeorology* **42**, 113-118.
- Shi, P., Chen, Z., Yang, Q., Harris, M. K., and Xiao, M. (2012). Influence of air temperature on the first flowering date of *Prunus yedoensis* Matsum. *Ecology and Evolution*, n/a-n/a.
- Siebert, S., and Ewert, F. (2012). Spatio-temporal patterns of phenological development in Germany in relation to temperature and day length. *Agricultural and Forest Meteorology* **152**, 44-57.
- Silvertown, J. (2009). A new dawn for citizen science. *Trends in Ecology & Evolution* **24**, 467-471.
- Smith, P. F., Ganesh, S., and Liu, P. (2013). A comparison of random forest regression and multiple linear regression for prediction in neuroscience. *Journal of Neuroscience Methods* **220**, 85-91.
- Solomatine, D., See, L. M., and Abrahart, R. J. (2008). Data-Driven Modelling: Concepts, Approaches and Experiences. In "Practical Hydroinformatics" (R. Abrahart, L. See and D. Solomatine, eds.), Vol. 68, pp. 17-30. Springer Berlin Heidelberg.

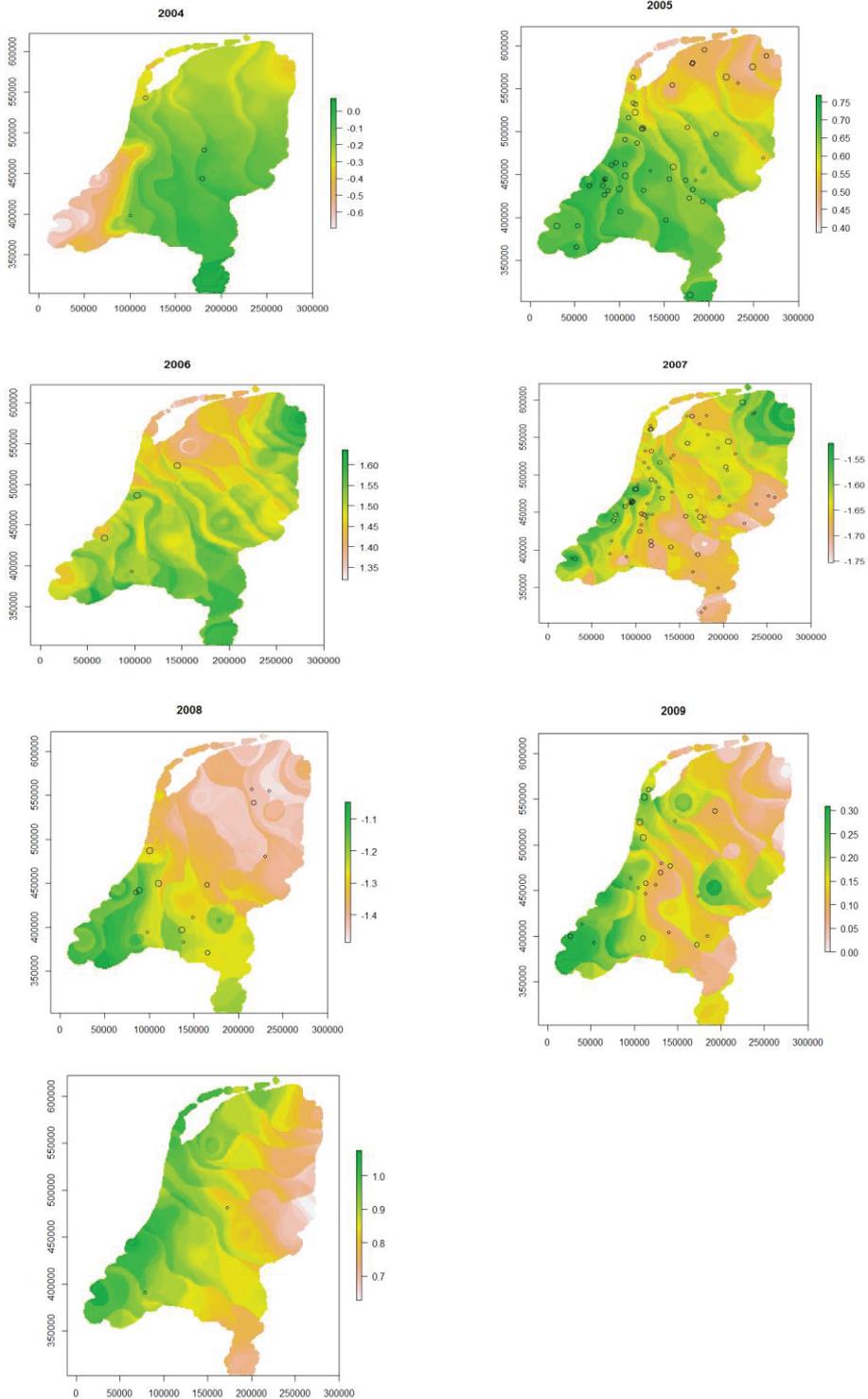
- Sparks, T., Górska-Zajączkowska, M., Wójtowicz, W., and Tryjanowski, P. (2011). Phenological changes and reduced seasonal synchrony in western Poland. *International Journal of Biometeorology* **55**, 447-453.
- Stein, A., and Georgiadis, N. (2006). Spatial Marked Point Patterns for Herd Dispersion in a Savanna Wildlife Herbivore Community in Kenya. In "Case Studies in Spatial Point Process Modeling" (A. Baddeley, P. Gregori, J. Mateu, R. Stoica and D. Stoyan, eds.), Vol. 185, pp. 261-273. Springer New York.
- Sullivan, B. L., Wood, C. L., Iliff, M. J., Bonney, R. E., Fink, D., and Kelling, S. (2009). eBird: A citizen-based bird observation network in the biological sciences. *Biological Conservation* **142**, 2282-2292.
- Tao, F., Zhang, S., and Zhang, Z. (2012). Spatiotemporal changes of wheat phenology in China under the effects of temperature, day length and cultivar thermal characteristics. *European Journal of Agronomy* **43**, 201-212.
- Tulloch, A. I. T., Possingham, H. P., Joseph, L. N., Szabo, J., and Martin, T. G. (2013). Realising the full potential of citizen science monitoring programs. *Biological Conservation* **165**, 128-138.
- Valentini, N., Me, G., Ferrero, R., and Spanna, F. (2001). Use of bioclimatic indexes to characterize phenological phases of apple varieties in Northern Italy. *International Journal of Biometeorology* **45**, 191-195.
- van Bussel, L. G. J., Ewert, F., and Leffelaar, P. A. (2011). Effects of data aggregation on simulations of crop phenology. *Agriculture, Ecosystems & Environment* **142**, 75-84.
- Vilhar, U., Beuker, E., Mizunuma, T., Skudnik, M., Lebourgeois, F., Soudani, K., and Wilkinson, M. (2013). Chapter 9 - Tree Phenology. In "Developments in Environmental Science" (F. Marco and F. Richard, eds.), Vol. Volume 12, pp. 169-182. Elsevier.
- Vitasse, Y., François, C., Delpierre, N., Dufrêne, E., Kremer, A., Chuine, I., and Delzon, S. (2011). Assessing the effects of climate change on the phenology of European temperate trees. *Agricultural and Forest Meteorology* **151**, 969-980.
- Willmott, C. J. (1981). ON THE VALIDATION OF MODELS. *Physical Geography* **2**, 184-194.
- Yanenko, O., and Schlieder, C. (2012). Enhancing the Quality of Volunteered Geographic Information: A Constraint-Based Approach. In "Bridging the Geographic Information Sciences" (J. Gensel, D. Josselin and D. Vandenbroucke, eds.), pp. 429-446. Springer Berlin Heidelberg.
- Yang Xiang, S. G., Brian Suomela, Julia Hoeng (2012). Generalized Simulated Annealing for Efficient Global Optimization: the GenSA Package for R. *The R Journal*.
- Zhang, S., and Tao, F. (2013). Modeling the response of rice phenology to climate change and variability in different climatic zones: Comparisons of five models. *European Journal of Agronomy* **45**, 165-176.
- Żmihorski, M., Sparks, T. H., and Tryjanowski, P. (2012). The Weekend Bias in Recording Rare Birds: Mechanisms and Consequences. *Acta Ornithologica* **47**, 87-94.

APPENDICIES

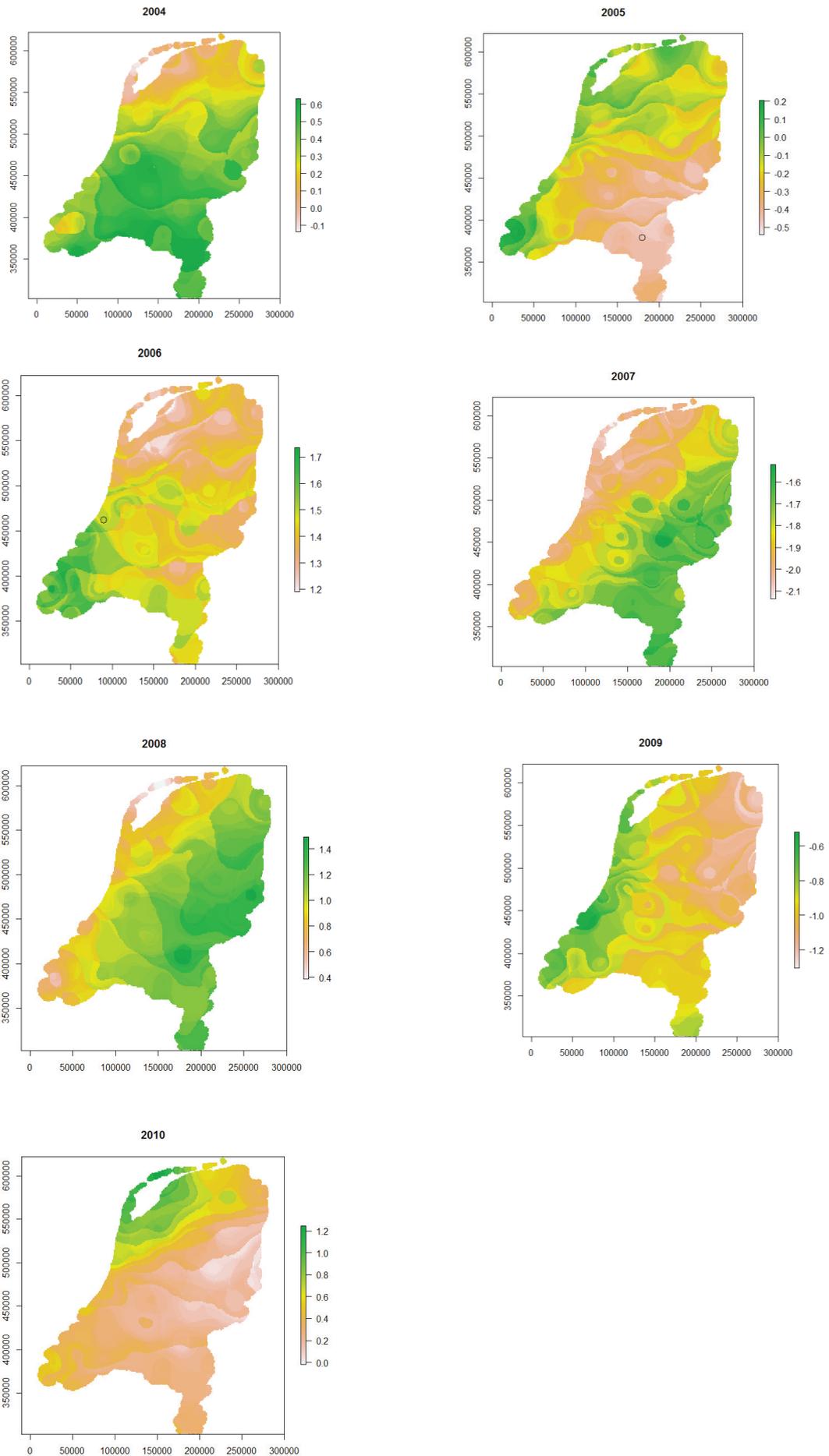
APPENDIX A

- (1) Spatial distribution of year based temporally-biased volunteer phenological observations (circle symbols). The colour of the map indicates the variability in DOY of the phenological event between one year and all the other years. For each phenological observation, the sizes of the dot indicate the amount of normalized SD value.

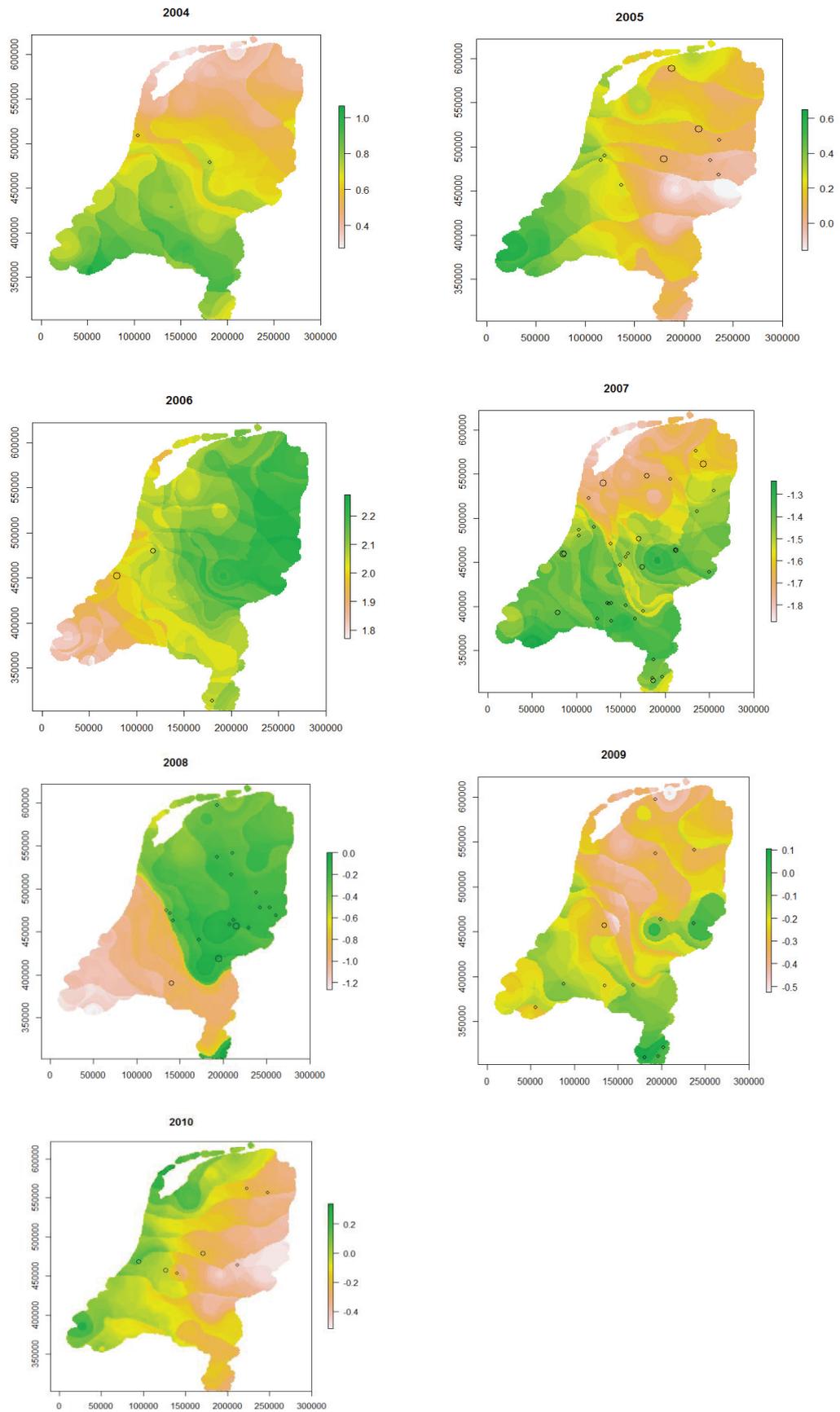
(a) Celandine

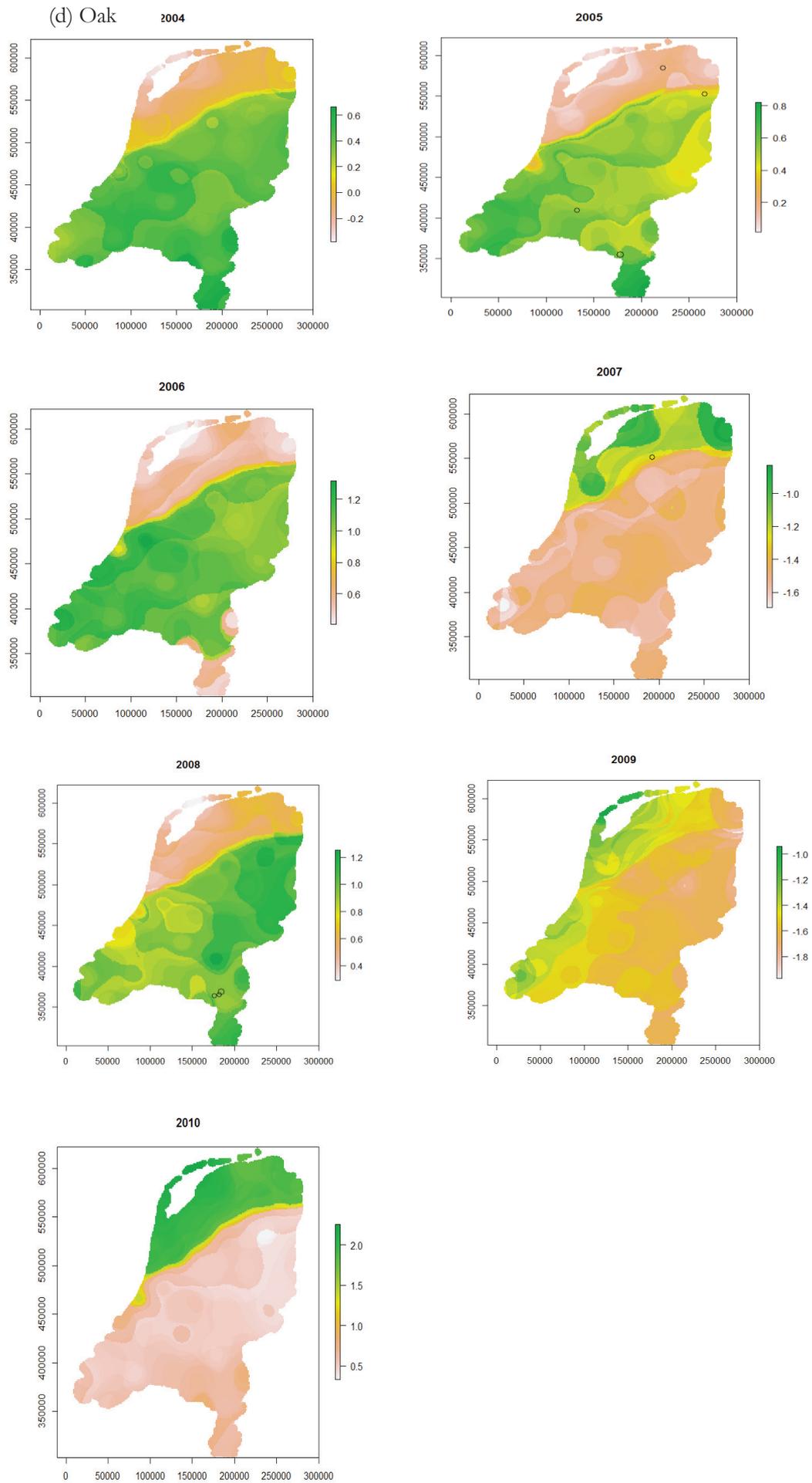


(b) Chestnut



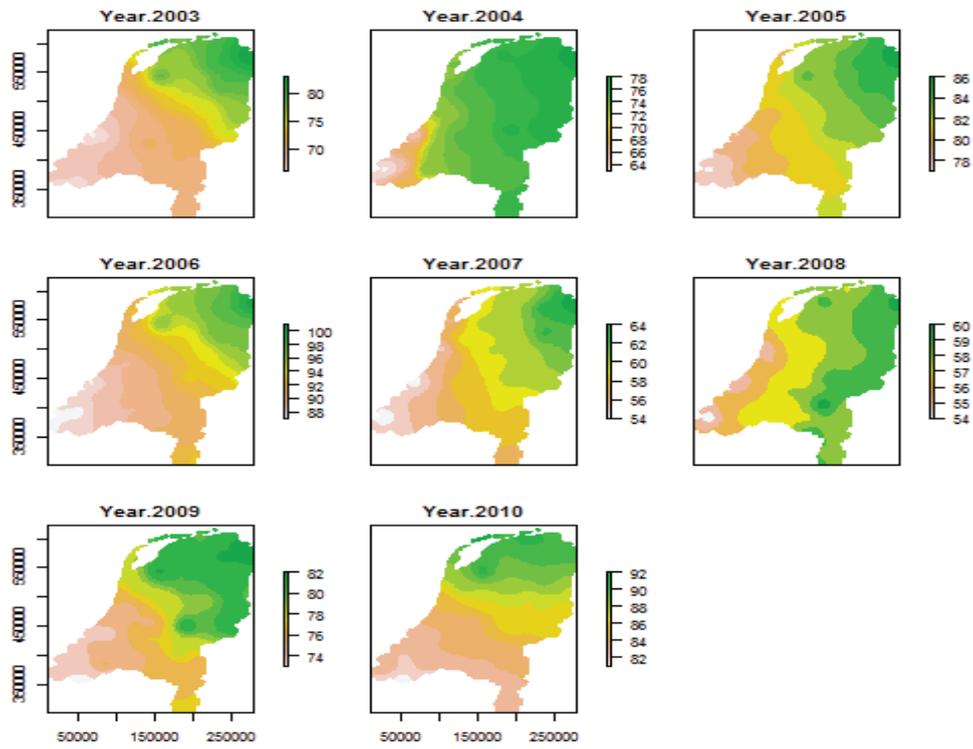
(c) Wood anemone



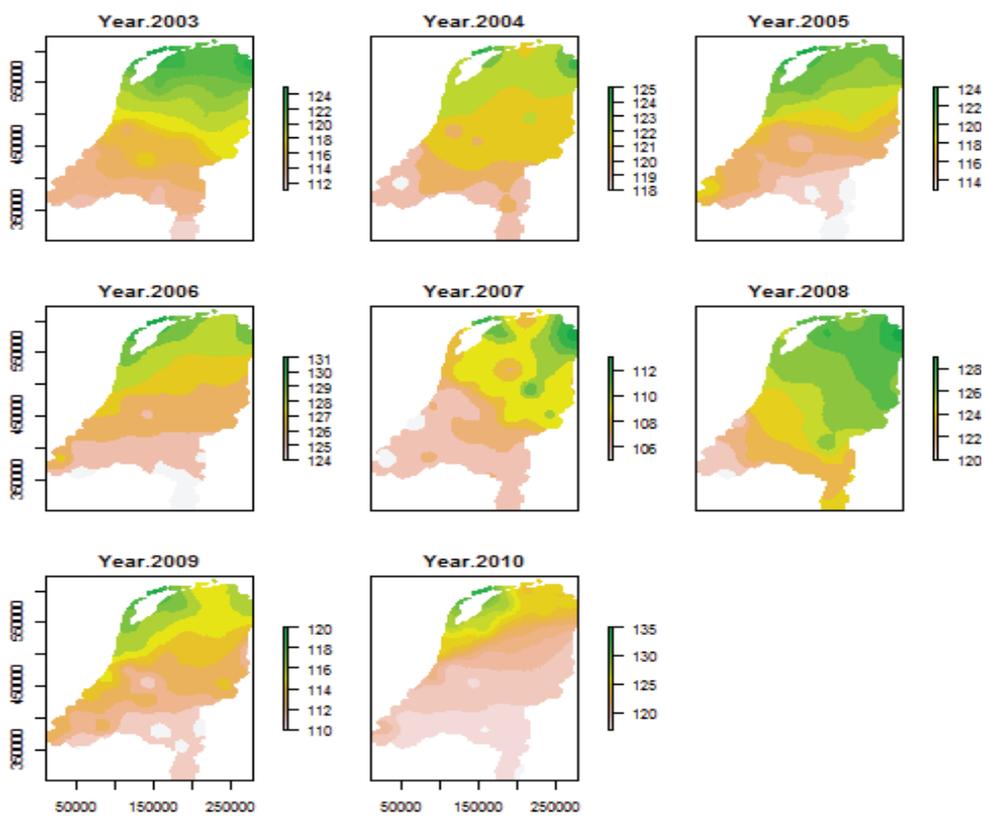


(2) Prediction results of the SW model from the year 2003 to 2010 in the whole area of The Netherlands

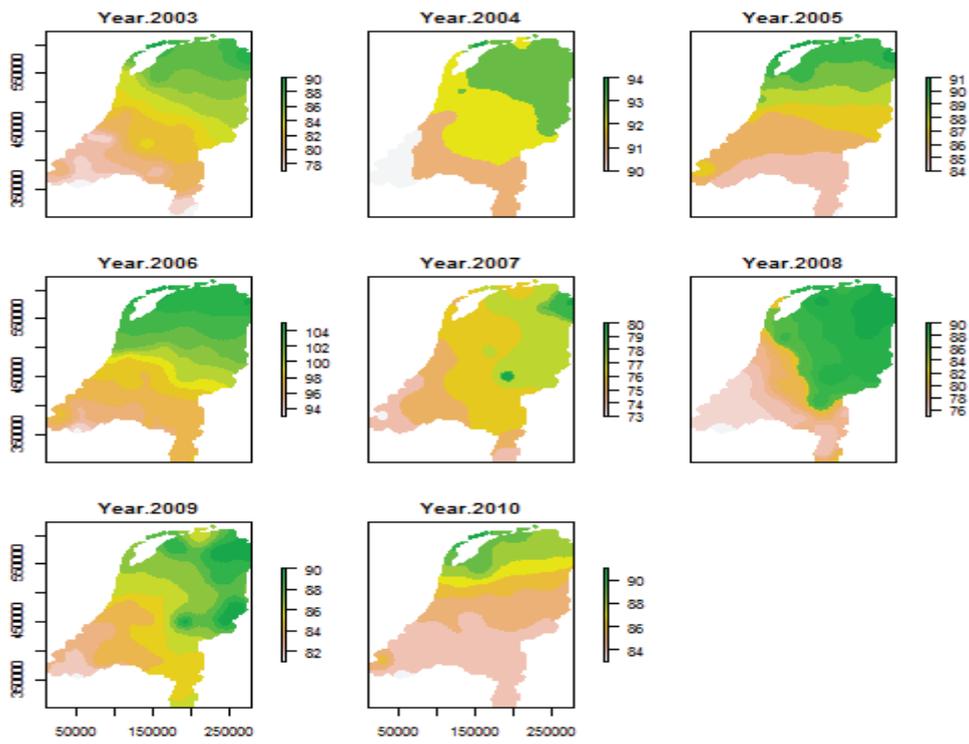
(a) Celandine



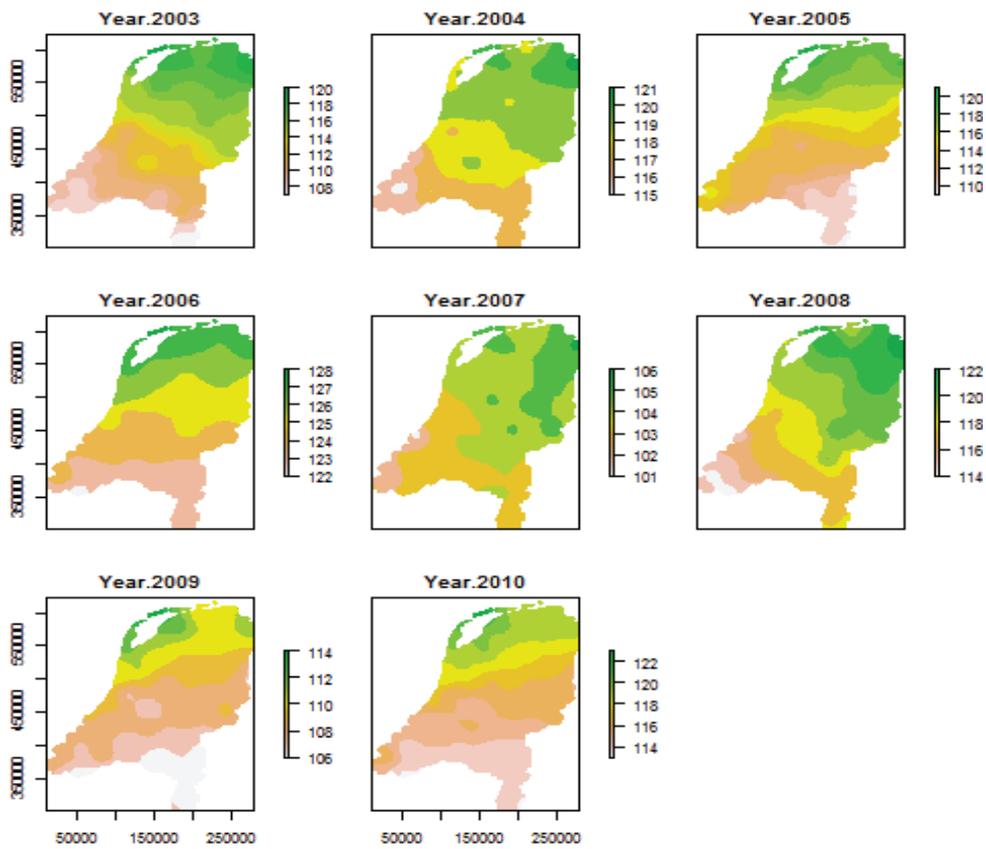
(b) Chestnut



(c) Wood anemone



(e) Oak



APPENDIX B

R Scripts used

1. Scripts used for SW model

```
# optimized Parameters are: xx[1]=Tb, xx[2]=t0, xx[3]=F*
library(GenSA)
library(raster)
library(rgdal)
#Read csv file
inputdata = read.csv("Celandine_Calibration.csv")
#Creates a matrix named data
data <-inputdata[,1:4]
data[,1] <-inputdata$jaar #year
data[,2] <-inputdata$x # x
data[,3] <-inputdata$y # y
data[,4] <-inputdata$Date # Date of observation

#set a lower and upper bounds for GenSA
lower <- c(0,365-quantile(data[,4],0.50),50)
upper <- c(10,quantile(data[,4],0.50)+365,2000)

#Load Raster file of temperatures
for(year in 2003:2010) {
  Tnam<-paste("T_", year, sep = "")
  assign(Tnam, brick(paste("T_", year, "_all_warp.bsq", sep=""))) }

#Extract temperatures for locations year of observation and year before
for(year in 2003:2009) {
  XY<-cbind(data[which(data[,1]==year+1),2], data[which(data[,1]==year+1),3])
  d1all = extract(get(paste("T_",year,sep="")), XY)[,1:365]
  d2all = extract(get(paste("T_",year+1,sep="")), XY)[,1:365]
  dall<-cbind(d1all,d2all)
  if (year>2003) {alld<-rbind(alld,dall)}
  else {alld<-dall}
}

funcSWModel_new <-function(xx) {
  xx[2]<-round(xx[2])
  # subtract base temperature and make negative values equal to 0
  a<-alld-xx[1]
  a[a<0]<-0
  # cumulative sum of temperatures from starting date (xx[2]) to end
  csa<-t(apply(a[,xx[2]:730], 1, cumsum))
  # find first date with tempsum above threshold (xx[3])
  dop<-apply(csa>=xx[3]*1, 1,which.max)+xx[2]-1
  dop[which(csa[,length(xx[2]:730)]< xx[3])<- NA
  if (sum(is.na(dop))> 10) {
```

```

    #print("F is too large for the selected starting day")
    RMSE<- 10000
    return (RMSE)
  }

# if there are less than 9 NA in the dataset
dop[is.na(dop)]=999
# if the dop is in the previous year
if (min(dop,na.rm=T)>365) {
  dop<-(dop-365)}
else {
  #print("Wrong combination; the predicted dop is too small")
  RMSE<- 10000
  return (RMSE)}
EFF = 1-((sum((dop-data[,4])^2))/(sum((data[,4]-mean(data[,4]))^2)))
RMSE<-sqrt(sum((dop-data[,4])^2)/length(dop))

}
#Optimization
out1 <- GenSA(lower = lower, upper = upper, fn = funcSWModel_new,control = list(smooth =
FALSE, temperature=200000, max.time=1200))
out1[c("value","par","counts")]

```

2) Scripts for UNIFORC model

```

#optimized parameters are: xx[1]=d, xx[2]=e, xx[3]=T0, xx[4]=F*
#set a lower and upper bounds for GenSA
lower <- c(rep(-6),rep(1),365-quantile(data[,4],0.50),rep(50)) # satisfy the constraint d<0 and e>0
upper <- c(rep(-1), rep(20),quantile(data[,4],0.50)+365,rep(2000))
funcUNIFORCModel_new <-function(xx) {
  xx[3]<-round(xx[3])
  # Calculate the sigmoidal dependency between daily temperature and rate of forcing
  a<- 1/(1 + (exp(xx[1]*(alld-xx[2])))
  a[a<0]<-0
  csa<-t(apply(a,xx[3]:730), 1, cumsum))
  # find first date with tempsum above threshold (xx[3])
  dop<-apply(csa>=xx[4]*1, 1,which.max)+xx[3]-1
  dop[which(csa[,length(xx[3]:730)]< xx[4])<- NA
  if (sum(is.na(dop))> 10 ) {
    #print("F is too large for the selected starting day")
    RMSE<- 10000
    return (RMSE) }
  # if there are less than 9 NA in the dataset
  dop[is.na(dop)]=999
  # if the dop is in the previous year
  if (min(dop,na.rm=T)>400) {
    dop<-(dop-365)}
  else {

```

```

#print("Wrong combination; the predicted dop is too small")
RMSE<- 10000
return (RMSE)}
EFF = 1-((sum((dop-data[,4])^2))/(sum((data[,4]-mean(data[,4]))^2)))
RMSE<-sqrt(sum((dop-data[,4])^2)/length(dop))}

```

3. Scripts for UNICHILL model

```

#parameters: a,b,c, C(state of chilling)
#d, e, F(State of forcing), t0: fixed
#xx1=a, xx2=c, xx3=b, xx4=d, xx5=e, xx6=C*,xx7=F*
lower <- c(rep(-6),rep(1),rep(-6),rep(-6),rep(1),rep(1),rep(10))
upper <- c(rep(-1), rep(10),rep(-1),rep(-1),rep(20),rep(200),rep(1000))
funcUNICHILL_new <-function(xx){
  # sigmoidal dependency for rate of chilling (a) and rate of forcing (b)
  a<-1/(1+exp((xx[1]*((alld-xx[2])^2)) + (xx[3]*(alld -xx[2]))))
  b <- 1/(1 + (exp(xx[4]*(alld-xx[5]))))
  a[a<0]<-0
  b[b<0]<-0
  t0= 243 #starting DOY fixed to 1 September
  # cumulative sum of temperatures from starting date (t0) to date of dormancy
  csa<-t(apply(a[,t0:365], 1, cumsum))
  dod<-apply(csa>=xx[6]*1, 1,which.max)+t0-1
  dod[which(csa[,length(t0:365)]< xx[6])<- NA
  if (sum(is.na(dod))> 0 ){
  RMSE<- 10000
  return (RMSE) }
  cc<-b
  for(i in 1:length(dod))
  {
  cc[i,1:(dod[i]-1)] <- 0
  }
  csa2<-t(apply(cc[,1:730], 1, cumsum))
  dop<-apply(csa2>=xx[7]*1, 1,which.max)
  dop[which(csa2[,length(1:730)]< xx[7])<- NA
  if (sum(is.na(dop)>10)){
  RMSE<- 10000
  return (RMSE) }
  dop[is.na(dop)]=999
  if (min(dop,na.rm=T)>365) {
  dop<-(dop-365) }
  else {
  RMSE<- 10000
  return (RMSE) }
  if (length(which(dop>quantile(data[,4],0.75)))>10){
  RMSE<- 10000
  return (RMSE) }
  RMSE<-sqrt(sum((dop-data[,4])^2)/length(dop))}

```

4. Scripts for removal of observations above a given normalized SD

```
data2= read.csv("Celandine_Predicted.csv") #reads a data that is predicted by the model
#dev_col: is the column containing the normalized SD values, dev is the normalized SD value
SD_filter <- function(data1,dev_col,dev)
{
  # filt --- |sdev| >=dev
  # left --- |sdev| <dev
  filtdata = matrix(nrow=0,ncol=7)
  leftdata = matrix(nrow=0,ncol=7)
  tmparr = matrix()
  for(i in 1:length(data1[,1]))
  {
    tmparr[1:4] = data1[i,1:4]
    tmparr[5:7] = data1[i,(dev_col-2):dev_col]
    if ((abs(data1[i,dev_col]))>= dev)
      filtdata = rbind(filtdata, tmparr)
    else
      leftdata = rbind(leftdata, tmparr)
  }
  datapair <- list("Nfiltdata"=filtdata,"Nleftdata"=leftdata )
  return(datapair)
}
filt_data <-SD_filter(data2,7,0.7)
colnames(filt_data$Nfiltdata) <- c("Year","X", "Y", "Date","Predicted","Error", "S_Deviation")
colnames(filt_data$Nleftdata) <- c("Year","X", "Y", "Date","Predicted","Error", "S_Deviation")
write.csv(filt_data$Nfiltdata,"Removed_data.csv",row.names=FALSE)
write.csv(filt_data$Nleftdata,"Left_data.csv",row.names=FALSE)
```

5. Scripts used for calculating systematic and unsystematic RMSE

```
Predicted_data=read.csv("Celandine_Predicted.csv") # reads a csv file that contains all the records
including predicted and observed values
systematic.lm=lm(Predicted_data$Predicted ~ Predicted_data$Date) # makes a linear model between
the predicted values and observed values
obar = mean(Predicted_data$Date) #mean of observed values
n=length(Predicted_data$Date) # length of observed
a=systematic.lm$coefficients[[1]] # reads the value of the intercept from linear model
b= systematic.lm$coefficients[[2]] # reads slope from linear model
p=a+(b*Predicted_data$Date) #Calculates the pi, obtained by the linear relation between observed
and predicted
systematic_RMSE=sqrt(sum((p-Predicted_data$Date)^2)/n)
unsystematic_RMSE=sqrt(sum((Predicted_data$PredictedDOY-p)^2)/n)
RMSE=sqrt( sum((Predicted_data$PredictedDOY-Predicted_data$Date)^2)/n)
```

6. Scripts for predicting the DOY of the phenological event (for SW model) in the whole area of the Netherlands and calculating the SD of prediction between the years and plotting the SD map.

```

#Load Raster file of temperatures
for(year in 2003:2010) {
Tnam<-paste("T_", year, sep = "")
assign(Tnam, brick(paste("T_", year, "_all_warp.bsq", sep="")))
#Extract temperatures for locations year of observation and year before
spname <- 'Celandine'
getYearLocs <- function(year)
{
  Tnam<-paste("T_", year, sep = "")
  assign(Tnam, brick(paste("T_", year, "_all_warp.bsq", sep="")))
  ts <- get(Tnam)[["T_2003_all_warp.1"]]
  yrRows <- dim(ts)[1]
  locs <- matrix(nrow=0,ncol=2)
  for(i in 1:yrRows){
    notna <- which(!is.na(ts[i,]))
    len <- length(notna)
    ycoord <- yFromRow(ts, i)
    cat("At row ",i," (with ",len,")\n")
    for(j in 1:len) {
      locs<-rbind(locs,c(xFromCol(ts, notna[j]),ycoord))
    }
  }
  return(locs)
}
XY <- getYearLocs(2003)
funcSWModel2 <-function(spname,year,XY,xx1,xx2,xx3) {
  data = extract(get(paste("T_",year,sep="")), XY)[1:365]
  XY <-cbind(XY,c(rep(0,dim(XY)[1])))
  xx2<-round(xx2)
  a<-data-xx1
  a[a<0]<-0
  csa<-t(apply(a[,xx2:365], 1, cumsum))
  dop<-apply(csa>=xx3*1, 1,which.max)+xx2-1
  dop[which(csa[,length(xx2:365)]< xx3)]<- NA
  XY[,3]<-dop
  write.csv(XY,file=paste(spname,'_DOY_pred_',year,'.csv'),row.names=FALSE)
  return(XY)
}

pop.sd <- function(x)(sqrt(var(x)*(length(x)-1)/length(x)))

#calling...funcSWModel2(year,XY,xx1,xx2,xx3) for each year
for(year in 2003:2010){
  yearXYZ <- funcSWModel2(spname,year,XY,0.7328072,20,183.3402861)#celandine

```

```

png(paste(spname,'_DOY_plot_',year))
yearRas <- rasterFromXYZ(yearXYZ)
plot(yearRas)
dev.off()
if (year==2003){
  allyearsRas <- yearRas
  toSDVals <- yearXYZ[,3]
}
else {
  allyearsRas <- addLayer(allyearsRas,yearRas)
  toSDVals <- cbind(toSDVals,yearXYZ[,3])
}
}
names(allyearsRas) <- c('Year 2003','Year 2004','Year 2005','Year 2006','Year 2007','Year 2008','Year
2009','Year 2010')
png(paste(spname,'_DOY_plot_all'))
plot(allyearsRas)
dev.off()
allSDs<-apply(toSDVals, 1, pop.sd)
SDLocs <- yearXYZ
SDLocs[,3] <- allSDs
allMns<-apply(toSDVals, 1, mean)
NormSD <- toSDVals
for(j in 1:length(allSDs)){
  NormSD[j,] <- (toSDVals[j,]-allMns[j])/allSDs[j]
}
SDnNormSD <- cbind(SDLocs, NormSD)
write.csv(SDnNormSD,file=paste(spname,'_SDnNormSD_Vals.csv'),row.names=FALSE)
png(paste(spname,'_SD_Vals_plot'))
plot(rasterFromXYZ(SDLocs))
dev.off()

```

7. Scripts for partitioning the areas showing climatic variability in to low, medium and high and overlaying the distribution of volunteers on the top of these areas

```

spCals = read.csv("Celaninde_nonbiased_observations.csv")
allSpXYs <- spCals[,2:3] # reads the coordinates (x,y) of non-biased observations
spSDVals = read.csv("Celandine_SD_Vals.csv")
#computing and plotting SD distribution
minSD <- min(spSDVals[,3])
maxSD <- max(spSDVals[,3])
diff <- (maxSD-minSD)/3
allSDs <- length(spSDVals[,3])
lowSD <- length(which(spSDVals[,3]<=(minSD+diff)))
medSD <- length(which(spSDVals[,3]>(minSD+diff)&spSDVals[,3]<=(minSD+2*diff)))
highSD <- length(which(spSDVals[,3]>(minSD+2*diff)))
p1 <- round(lowSD*100/allSDs)
p2 <- round(medSD*100/allSDs)

```

```

p3 <- round(highSD*100/allSDs)
arg<-list(at=c(minSD,(minSD+diff),(minSD+2*diff),maxSD),labels=c(round(minSD,digits=1),
+ round((minSD+diff),digits=1),round((minSD+2*diff),digits=1),round(maxSD,digits=1)))
brk <- c(minSD,(minSD+diff),(minSD+2*diff),maxSD)
plot(rasterFromXYZ(spSDVals),col=c("peachpuff","yellow1","limegreen"),axis.arg=arg,breaks=brk)
points(allSpXYs, pch=21,col="red",cex=0.6)

```

8. Boxplots showing the DOY of phenological event vs the year

```

boxplot(Date~Year,
data=all_species,main=paste(""),ylim=c(range((all_species$V3))),xlab="DOY",ylab="YEAR",
horizontal=TRUE)

```

9. Scripts for drawing the graph between normalized SD (x-axis) and RMSE (systematic and unsystematic), efficiency and correlation (y-axis). The point where the optimal threshold value was obtained could also be seen on the graph

```

# Read a csv file with SD, RMSE (systematic), RMSE (unsystematic), efficiency, correlation and n for
each species

```

```

Book3<-read.csv("Book3.csv")
sp1 <- Book3[1:15,1:8] #Bosanemoon
sp2 <- Book3[16:29,1:8] #Oak
sp3 <- Book3[30:43,1:8] #Celandine
sp4 <- Book3[44:57,1:8]#Cow parsley
sp5 <- Book3[58:73,1:8]#chestnut
gr1 <- function(sp,Nr)
{
  par(ps = 11, cex = 1, cex.main = 1)

```

```

plot(sp$RMSE.sys.,type="b",lwd=2,xaxt="n",ylim=c(floor(min(sp[,3:5])),ceiling(max(sp[,3:5]))),col="
black",xlab="Normalized Standard Deviation",
  ylab="(Un)Systematic RMSE and RMSE",lty=1)
#title(main=paste("SD versus RMSE and for species: ',sp[1,1]',line=3)
axis(1,at=1:length(sp$SD),labels=sp$SD)
axis(1,at=1:length(sp$n),labels=round(((Nr-sp$n)*100)/Nr,0),side=3)
mtext("Percentage of Phenological Observations (Removed)", side=3, line=2)
lines(sp$RMSE.unsys.,type="b",lwd=2,lty=2)
lines(sp$RMSE,type="b",lwd=2,lty=3)
reqi<-max(which(sp$RMSE.sys.<sp$RMSE.unsys.))
points(c(reqi,reqi,reqi), c(sp[reqi,3],sp[reqi,4],sp[reqi,5]), pch=7,col="red",cex=2)

```

```

legend("bottomright",legend=c("RMSE(sys)","RMSE(unsys)","RMSE"),lty=c(1,2,3),lwd=2,ncol=1,bt
y="n",cex=0.8,inset=0.01)
}
gr2 <- function(sp)
{
  plot(sp$Efficiency,type="b",lwd=2,xaxt="n",ylim=c(round((min(sp[,6:7])-
0.1),digits=1),round((max(sp[,6:7])+0.1),digits=1)),col="black",xlab="Normalized Standard
Deviation",

```

```

ylab="Efficiency and Correlation")
axis(1,at=1:length(sp$$SD),labels=sp$$SD)
lines(sp$Correlation,type="b",lwd=2,lty=2)

legend("topright",legend=c("Efficiency","Correlation"),lty=c(1,2),lwd=2,ncol=1,bty="n",cex=0.8,inset=0.01)
}
# Nr values: Wood anomale(Bosanemoon) = 532, Eik = 249, celandine = 1341, Cow_Parsley = 775,
White_Horse_Chestnut = 390
par( mfrow = c( 3, 2), mar=c(4,4,4,4))
gr1(sp4,776)#Cow parsley
gr1(sp3,1341)#Celandine
gr1(sp5,390)#Chestnut
gr1(sp1,532) #Bosanemoon
gr1(sp2,249) #Eik
par( mfrow = c( 3, 2), mar=c(4,4,4,4))
gr2(sp4)#Cow parsley
gr2(sp3)#Celandine
gr2(sp5)#Chestnut
gr2(sp1) #Bosanemoon
gr2(sp2) #Eik

```

- 10. Scripts for drawing scatter diagrams between the predicted and observed values. Temporally-biased observations are shown in red dots and non-temporally biased are shown with black dots.**

```

data1= read.csv("Celandine_Predicted_non_biased.csv")
data2=read.csv("Celandine_predicted_biased.csv")
cor.tr <- cor.test(data1$Predicted,data1$Date, method="spearman")
plot(data1$Date, data1$Predicted, xlab="Observation(DOY)", ylab="Prediction(DOY)",
xlim=c(range(40,150)),ylim=c(range(40,150)),pch=1)
abline(lsfite(data1$Date,data1$Predicted), lwd=1)
legend("topleft",legend=c(paste("Correlation (rho) =",round(cor.tr$estimate,2), " RMSE =
",round(sqrt(mean((data1$Date-data1$Predicted)^2)),2))))
points(data2$Date,data2$Predicted,pch=24,col='red',bg='red')

```