COMPARISON OF THE PERFORMANCE OF THREE TYPES OF MULTIPLE REGRESSION FOR PHENOLOGY IN BAVARIA IN A DYNAMICAL-STATISTICAL MODEL APPROACH

Felix Pollinger, Katrin Ziegler and Heiko Paeth

With 7 figures and 1 table
Received 13 April 2017 · Accepted 19 October 2017

Summary: Some of the most obvious consequences of anthropogenic climate change are observed changes in the dates of the occurrence of phenological events. Most prominently, observations from the Northern Hemisphere's extratropics indicate an earlier occurrence of spring events. Recent climate models include land surface schemes that provide representation of the vegetation. However, they are limited in simulating the plants' response to climate change. In this study we present results of a dynamical-statistical modeling approach for phenology in southeastern Germany, combining climate change simulations provided by a high resolution, state-of-the-art regional climate model (RCM) with three different types of regression methods: ordinary least squares (OLS), least absolute deviation (LAD) and random forest (RFO). We focus on changes in the day of the year (DOY) of Forsythia suspensa flowering, the earliest phenophase of the growing season in Bavaria. Based on roughly 2600 observations, collected at 94 phenological and 26 meteorological stations between 1952 and 2013, we compare the regressions via a bootstrap, using once 13 and once 4 meteorological variables as predictors. Altogether, we find the regressions with less variables to be more robust, while the regression estimates are nearly identical. Explained variance and RMSE (root mean square error) are 54.8 % and 8.8 days for RFO and 51.2 % and 9.1 days for the other regressions. These trained and cross validated statistical models are used to estimate the effects of future climate change on the DOY by applying them to the RCM simulations. For OLS or LAD, under a low (high) greenhouse gas emission scenario, we find a mean advance of the DOY of 8 (15) days by the end of the 21th century compared to the base period from 1961 to 1990. The spatial pattern of the change resembles the topography, with the strongest trends in the DOY over mountainous regions as a consequence of a simultaneous rise in temperatures and reduction in snow depth. RFO is restricted to the range of the observations and hence the response to the simulated climate is damped, resulting in an advance of DOY of only 5 (8) days and a reduction in variance. There is no apparent spatial pattern identifiable. Altogether, we find OLS and LAD to be more suitable for dynamical-statistical modeling of phenology than RFO.

Zusammenfassung: Zu den augenfälligsten Folgen des anthropogenen Klimawandels gehören beobachtete Veränderungen im zeitlichen Auftreten von phänologischen Ereignissen. Am markantesten deuten Beobachtungen aus den Außertropen der Nordhalbkugel auf den früheren Eintritt von Frühlingsereignissen hin. Aktuelle Klimamodelle verfügen über Landoberflächenschemata zur Abbildung der Vegetationsdynamik, allerdings sind sie nur eingeschränkt dazu in der Lage die Reaktion von Pflanzen auf Klimaänderungen zu simulieren. In dieser Studie präsentieren wir Ergebnisse eines dynamisch-statistischen Modellierungsansatzes für Phänologie in Bayern. Dafür kombinieren wir hochaufgelöste Klimawandelsimulationen eines aktuellen, regionalen Klimamodells (RCM) mit drei verschiedenen Regressionsmethoden: Gewöhnliche-Kleinste-Quadrate (OLS), Geringste-Absolute-Abweichung (LAD) und Random Forest (RFO). Wir untersuchen Änderungen im Eintrittsdatum der Blüte von Forsythia suspensa, der frühesten Phänophase der Vegetationsperiode in Bayern. Bei einer Datengrundlage von etwa 2600 Beobachtungen, die an 94 phänologischen und 26 meteorologischen Stationen zwischen 1952 und 2013 erhoben wurden, nutzen wir Bootstraps um die Regressionen je mit 13 und 4 meteorologischen Variablen als Prädiktoren zu vergleichen. Insgesamt erweisen sich die Regressionen mit der geringeren Anzahl Variablen als robuster, während die Regressionsschätzer beinahe identisch sind. Für den RFO ergibt sich eine erklärte Varianz von 54.8 % und ein RMSE (die Wurzel des mittleren quadratischen Fehlers) von 8.8 Tagen, die anderen Regressionen erreichen 51.2 % bzw. 9.1 Tage. Mit diesen so trainierten und kreuzvalidierten statistischen Modellen schätzen wir die Effekte des künftigen Klimawandels auf das Eintrittsdatum der Blüte indem wir sie auf die RCM Daten übertragen. Für OLS und LAD finden wir für ein Szenario mit geringen (hohen) Treibhausgasemissionen eine mittlere Verfrühung des Blühdatums von etwa acht (15) Tagen bis zum Ende des 21.Jahrhunderts im Vergleich zur Referenzperiode 1961 bis 1990. Im räumlichen Muster zeichnet sich die Topographie ab. Dabei zeigen sich die stärksten Trends in hochgelegenen Regionen, bedingt durch die gleichzeitige Erhöhung der Temperatur und Reduktion der Schneetiefe. Die Schätzungen von RFO sind auf den Wertebereich der Beobachtungen beschränkt. Entsprechend fällt die Reaktion auf die simulierte Klimaänderung mit einer Verfrühung von nur 5 (8) Tagen und einer Reduktion der Varianz gedämpft aus. Ein offensichtliches räumliches Muster ist nicht zu erkennen. Insgesamt erscheinen uns OLS und LAD für die statistisch-dynamische Modellierung von Phänologie als besser geeignet als RFO.

Keywords: phenology, statistic methods, Bavaria, climate change, regional climate models, regression analysis

https://doi.org/10.3112/erdkunde.2017.04.01 ISSN 0014-0015 http://www.erdkunde.uni-bonn.de

1 Introduction

Phenology deals with recurrent biological events including the causes of their timing with regard to abiotic factors, amongst others (Lieth 1974). For a specific plant at one specific location, without the occurrence of natural disaster or human interference, the intraannual changes of climatic conditions are the main drivers of the plants' annual cycle, the most obvious in regions characterized by pronounced thermic or hygrid seasons as well as a dynamic vegetation (see Schwartz 2013). Also, the year-to-year variations of the onset of certain phenological phases is in large parts caused by variations of the regional climate. In the last decades, one special focus of phenological studies has been the shift in the well documented phenological phases as a consequence of climate change. In particular, there is overwhelming evidence from hundreds of species throughout the Northern Hemisphere for an earlier occurrence of spring events (see for example Field et al. 2014 and the reviews of Menzel et al. 2006; Cleland et al. 2007; Richardson et al. 2013 and references therein). These individual responses might result in unforeseen ecological consequences (Thackeray et al. 2016).

The most sophisticated tools for exploring the effects of climate change are three-dimensional circulation models (Taylor et al. 2012). A special group of these are regional climate models (RCMs), which, due to their high spatial resolution, are especially useful for investigating the environmental impacts of climate change (Giorgi et al. 2006). However, while these models combine numerous modules for incorporating physical and biological aspects of the climate system (Flato et al. 2013), current land surface schemes are not capable of modeling vegetation phenology satisfactorily, not to mention the response of individual phenological phases (Richardson et al. 2012).

Given these apparent deficiencies of climate models, to study the response of vegetation or specific plants to future climate change, empirical models are employed. A common way to do this is to use observational data to derive a statistical relationship between the target variable, e.g. the day of the year (DOY) of the occurrence of a phenological phase, and a set of observed meteorological variables as predictors (e. g. Menzel 2003; Ma and Zhou 2012). If the established relationship appears to be robust, it is possible to apply the empirical model to climate model simulations, e.g. to estimate the phenological response to projected climate change. This approach is usually called dynamical-statistical modeling (Paeth et al. 2008; Awoyé et al. 2017). A number of empirical, often called process-based, phenology models of different complexity exists (e.g. Hänninen et al. 1994; Schaber and Badeck 2003; Setiyono et al. 2007; Morin et al. 2009) and some of these have been used in a dynamical-statistical approach, e.g. by Migliavacca et al. (2012) and Morin et al. 2009 for studies in North America. However, those models are typically designed for specific plants and phenological phases and might even perform badly for validation data (see Richardson and O'Keefe 2009 for a discussion).

Thus, given the large number of regions, plants and phases for which information about their response to climate change signals might be crucial for adaption, purely statistical models (e.g. Estrella and Menzel 2006; Primack et al. 2009) should also be considered for exploring these questions. An appropriate general statistical tool for this is regression analysis. Albeit commonly associated with linear, ordinary least squares (OLS), there are a number of different approaches available. Typically, these address some of the apparent weaknesses of OLS, such as a lack of robustness in the presence of extreme values or outliers. Furthermore, modern methods of statistical learning seem to perform better in terms of prediction, often by avoiding to constrain the predictions by assuming a global, relatively simple mathematical function as OLS does (Hastie et al. 2008).

One aim of this paper is to demonstrate the usefulness and flexibility of the dynamical-statistical approach. We focus on changes in the DOY of *Forsythia suspensa* flowering for Bavaria, Germany, but the method could be transferred to other phenological features without large adaptions, especially as we demonstrate how to make use of spatially heterogeneous data. In this study, we employ and compare three types of regression for dynamical-statistical modeling: the parametric OLS and least absolute deviation regression (LAD) and the Random Forest algorithm (RFO), which has gained a lot of attention since its introduction (Breiman 2001) and has been widely adopted by data scientists.

We further use a high-resolution, state-of-the art RCM to estimate the effects of climate change. As we use two different emission scenarios we are able to estimate the potential for mitigation.

The paper is organized as follows: in section two, we introduce our study area, present our data and briefly discuss the principles of the three regressions. We further present the measures of model quality we use for model comparison and the bootstrap for estimating their robustness and reliability. Also, the pre-processing of the various datasets is explained. The next section contains the results of the study, which are discussed in section four before we draw our conclusions.

2 Material and methods

2.1 Study area

Bavaria is located in Central Europe and is, with an area of approximately 70500 km², the largest federal state of the Federal Republic of Germany. The climate is notably moderate with a slight gradient from more maritime conditions (Cfb climate in the Koeppen classification) in the north-western part to more continental climate (Dfb) in the east and south-east of the study area. Consequently, the prevalent natural vegetation type would be temperate forest without the, in reality vastly dominating, effects of human interference. Nowadays around 30 % of the area are covered with woods and forests. However, local conditions might differ largely from these averages, especially due to elevation. The southernmost part of Bavaria is part of the Alps and hence characterized by an Alpine climate.

2.2 Observational data

We focus on the DOY of *Forsythia suspensa's* flowering. This phase is considered the beginning of early spring as defined by the German Weather Service (Bruns et al. 2015). This phenophase is characterized by high variability, however, there are indications of an overall shift towards earlier dates for Germany (Menzel et al. 2001). *Forsythia suspensa* is considered a good proxy for the effect of climate variables like temperature. It is also part of the Global Phenological Monitoring network and the International Phenological Gardens of Europe Program (Chmielewski et al. 2013).

Observational data is provided by the German meteorological service and the Bavarian state office for the Environment. Phenological data collection is hereby carried out by a network of volunteers throughout Germany, whilst meteorological data is measured by operational weather stations. As the German meteorological service provides guidelines (Bruns et al. 2015) for the volunteers, the quality of the data can be considered quite high. Altogether, we can use 2592 observations of the DOY from 94 phenological stations, covering a time period from 1952 to 2013 (Fig. 1). We are therefore optimistic that we cover the range of environmental and climatic conditions in Bavaria quite well and that our statistical models can be reliably tuned. We only use a small portion of the more than 1000 phenological stations in Bavaria. This is due to the fact that we only consider stations for which there were at least 10 years of observations available, so as to get a realistic estimate of the average day of flowering per station. All DOYs are transformed into anomalies by subtracting this station mean.

The meteorological data consists of 13 variables, measured by 26 meteorological stations (Fig. 1) and provided in daily resolution. Please see the caption of Fig. 2 for a listing. Together, these variables allow for a rather complete assessment of each stations climate and represent all available climatic information possibly relevant for phenology.

While data availability for our study area can be considered quite positive, the predictors' preparation to obtain reliable results is not trivial. As the climatic information in general is gathered at different places than DOY, the pre-processing must ensure that it is representative for the phenological station. In addition, the optimal time span to consider during the year is a topic of discussion. Further, the statistical models must not be sensitive to location specific aspects that cannot be provided by climate models.

Altogether, we found our methods to produce the best results if the following is applied: For each phenological station we use climate data from the nearest meteorological station. We calculate for each variable accumulated anomalies over the 45 days before the station's mean DOY. This is an effective way to reduce the impact of climatic differences between the stations due to topography as well as the influence of different types of soils or even anthropogenic long term effects such as buildings. We found 45 days sufficient to ensure that the relationship between meteorological variables and DOY is not too strongly affected by the estimate of the mean DOY, while we can still make use of the available information. This is less the case when monthly or seasonal means are used (e.g. Menzel 2003; Chmielewski et al. 2004; Primack et al. 2009). Thus, we account for the individual characteristics of a location while preventing a situation where the predictors really are functions of the dependent variable, e.g. when yearly measures are calculated in reference to the actual DOY. Note that anomalies are also much more practical when transferring the statistical models to climate model data, as this removes potential systematic differences between the climatology of the model and the observed meteorological data.

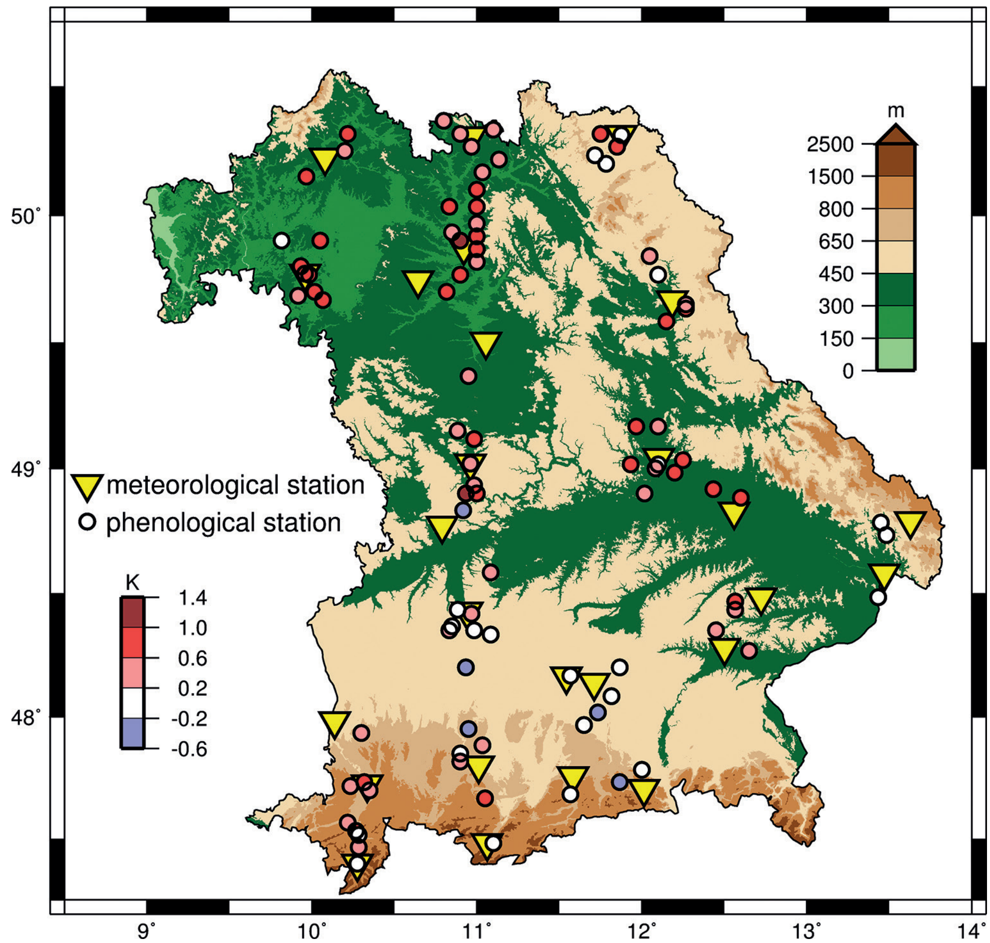

Fig. 1: Study area: elevation and locations of the phenological and meteorological stations used in this study. Filling indicates the available years' mean deviation in mean temperature from the long term mean of the 45-day period considered. See the text for further explanations.

To estimate the validity of this somewhat heuristic procedure, we address two aspects that could potentially affect our results. The first is the distance between phenological and meteorological station and hence the validity of the climate data. According to the provided meta data the maximum (mean) Euclidean distance is about 20 (9.5) km. We are confident that this doesn't endanger the expressiveness of the anomalies. However, the maximum difference in elevation is 1312 m and potentially harmful. The other issue is concerned with the robustness of the mean DOY. Considering that a maximum of 62 years of data is potentially available per station, it seems unlikely that a full 10-year record was gathered during years all characterized by strong climatic anomalies of the same sign (which could seriously affect the stations mean DOY). For mean temperature, Fig. 1 shows differences for each phenological station's mean and the 1961–1990 mean of the meteorological station during the considered 45 days. As none of these deviations exceeds half a standard deviation in either direction, they don't seem problematic in terms of representativity. Still, this may be the case for a small number of stations, whose impact on the regressions, however, should be small given the overall sample size and the fact that the possible range of the DOY is limited. Nonetheless, a far-off estimate for the mean DOY might affect the statistical-dynamical models' local performance.

We investigate the relevance of these effects on our results by redoing the analysis using three subsamples of our data. For that, we restrict our analysis to a) stations for which the difference in elevation to their next-neighbor meteorological station is less than 100 m, b) stations with 30 or more years of record and c) the intersection of a) and b).

Fig. 2 displays the correlations of the processed variables. We intend to reduce the number of predictors as, in part due to our pre-processing, there

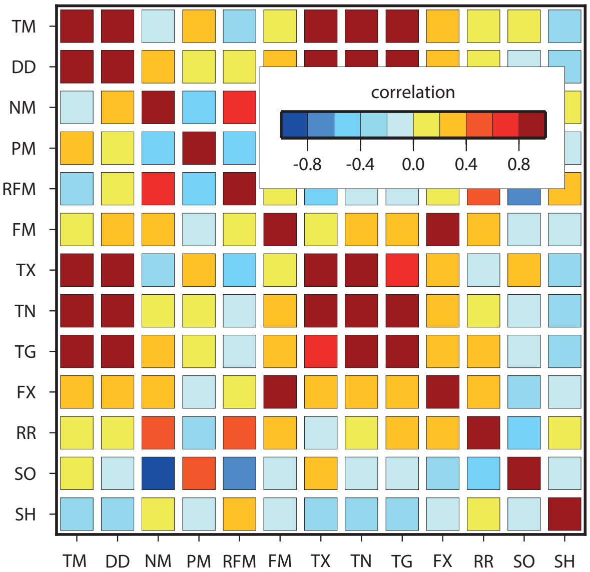

Fig. 2: Correlations of the predictors. See the text for the applied pre-processing. Abbreviations are due to the German meteorological service. TM: mean temperature; DD: vapor pressure; NM: cloudiness; PM: air pressure; RFM: relative humidity; FM: average wind speed; TX: maximum temperature; TN: minimum temperature; FX: maximum wind speed; RR: precipitation; SO: sunshine duration; SH: snow depth

are strong correlations between several of the predictor variables (Fig. 2). Furthermore, the adaptation of our procedure for other regions will be the easier the smaller the number of variables. As a subset, we choose mean temperature, mean wind speed, precipitation and snow depth, which represent different aspects of Bavaria's climate. Note that these tend to have small correlations with each other but strong correlations with several of the other variables.

2.3 Climate model data

The climate change simulations are part of the internationally coordinated EURO-CORDEX project (Jacob et al. 2014). We use daily data in a 0.11° x 0.11° horizontal resolution from the state-of-the-art RCM MPI-CSC-REMO2009 (addressed as REMO in the following) with boundary conditions provided by the global climate model MPI-ESM-LR. We consider two transient simulations, differing in the emitted amount of greenhouse gases and, hence, their atmospheric concentration. The Representative Concentration Pathway 4.5 scenario (RCP 4.5) assumes a rather small increase, which results in an average rise of radiative forcing of 4.5 W/m² by the end of the 21th century. The other considered scenario, RCP 8.5, is characterized by rather high emissions, which result in an anthropogenic radiative forcing of 8.5 W/m² (Moss et al. 2008). By utilizing both we can estimate the potential of consequent mitigation measures for the reduction of climate change impacts. Note that during the historical period (1950–2005) both runs share the same simulation driven by estimates of historical greenhouse gas concentrations and natural forcings. Hence, there are no differences due to the initial conditions and all differences between the simulations can be attributed to the applied forcing.

We calculated snow depth by accumulating snow depth changes, starting July 1st 1950, when the surface snow amount for all concerned gridboxes is 0.

The REMO data is processed analogously to the meteorological observations. We estimate the mean DOY of each gridbox via its next-neighbor of the phenological stations and calculate 45-day-sums relative to these dates for all variables and years. We normalize these time series with respect to 1961-1990.

2.4 Statistical methods

Here we introduce the statistical models used in this study. There are excellent references for all of them, so we restrict this discussion to some general features and foundations of each method to point out their differences.

In general, regression analysis aims at fitting a set of independent variables or predictors to a dependent variable. Let $\boldsymbol{y}$ be a vector that contains I independent realizations of Y and $\boldsymbol{X}$ be a matrix that contains I row vectors $\boldsymbol{x}_i$ of length J. Each $\boldsymbol{x}_i$ contains one realization x_{ij} per predictor X_j. Regression analysis aims at finding a prediction $\hat{y} = E(Y|\boldsymbol{x})$. In general, $\hat{y}_i$ won't meet y_i exactly. Instead, a prediction error $\varepsilon_i = y_i - \hat{y}_i$ occurs. OLS and LAD employ an explicitly stated function, whose coefficients are to be estimated, to minimize a quantity based on all ε_i. Both assume a parametric model that allows to express the prediction as a linear combination of the predictors. Using matrix notation these can be written as $\hat{\boldsymbol{y}} = \boldsymbol{Xb}$, if the first column of $\boldsymbol{X}$ contains a constant factor 1 and $\boldsymbol{b}$ the regression coefficients $b_0, b_1, \ldots, b_J$.

The most common way to do this is OLS, which minimizes $\Sigma_i \varepsilon_i^2$. The analytical solution to this problem is $\boldsymbol{b} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{Xy}$. OLS dates back to C.F. Gauss and can be considered a cornerstone of statistics for more than a century. However, there are well known problems associated with it, of which we tackle several in this paper: due to the derivation of its coefficients, OLS is heavily affected by outliers in the data. A relatively and absolutely small number of unusual values for *Y* might influence the fitted model in such a way that no useful application is possible. Furthermore, all deviations of the model are attributed to the dependent variable, while the predictors are considered true. Typically, this assumption is at least questionable when working with empirical data such as point-measurements which are taken to be representative for a nearby, but not identical, location.

LAD is quite similar to OLS, in terms that it assumes the same underlying function. The difference is that LAD attempts to minimize the sum of the absolute error of predictions $\Sigma_i|\varepsilon_i|$. Hence, LAD is significantly more robust towards numerical disturbances in the data than OLS, which is of advantage in many real world applications (Portnoy and Koenker 1997). There is no analytical solution for this, however, and iterative algorithms must be used. There might not even be a unique solution for the coefficients at all. See Dielman 2005 for a review.

The other type of regression considered uses a local, non-linear approach. Here, fitting is done without the constraints of an explicit model and hence more flexible. These approaches typically rely on a number of computationally expensive procedures like bootstrapping and random feature selection. The Random Forest (Breiman 2001) is a good example for this.

A random forest consists of a number (here: 500) of classification and regression trees (CART, Breiman et al. 1984), built by sequential binary splitting of an independent bootstrap sample of the data into *G* groups or nodes, attempting to minimize $\Sigma_g\Sigma_{i\in g}(y_i-\hat{y}_g)^2$. Here the prediction $\hat{y}_g$ is simply the mean of node *g*. The splits are based on a - at each node randomly selected - subset of the predictor variables. All nodes must contain at least five observations. When no more splitting is possible, the predictor space is divided into distinct and separated nodes. The tree estimates for the predictand are defined as the mean of the nodes. The estimates for additional data can be found by classifying it according to the splitting rules. RFO estimates of $\hat{\boldsymbol{y}}$ are then built as the mean of the tree estimates. Note that all predictions involved here aren't continuous but rather stepwise functions, limited to the range of the training data, a point that tends to get overlooked by users.

2.5 Model comparison and validation

We focus on two quantities to characterize and compare the quality of our statistical models: The coefficient of determination R^2 represents the explained variance, hence the linear association of the independent variable and the predictions. R^2 is defined as the squared correlation of the observed anomalies of the DOY and the statistical models' predictions. Further, to estimate the actual precision of the predictions, we use the root of the mean square error:

$$\mathrm{RMSE} = \sqrt{I^{-1}\sum_i \varepsilon_i^2}\,.$$

Also, statistical confirmation of a regression model is a crucial point if the estimated models are to be applied to an independent set of data. For OLS there is a huge number of well-designed procedures which allow - under some constraints - for statistical interference for each coefficient of a statistical model. However, there are by far less accepted inferential techniques for LAD or even for RFO. Therefore, in this paper, we prefer to use an identical bootstrap approach for all methods.

We divide our data into two subsets for the training of the statistical models and their validation, respectively. For validation, we randomly select at least 25 phenological stations, representing at least 500 samples. Then, the statistical models are fitted to the rest of the data, called training data. We apply these models to the validation data and calculate the measures of model quality. This is repeated 1000 times. Note that this is an independent bootstrap, not affecting the one of RFO, which is hence carried out 1000 times itself. This methodology allows us to give estimates for the uncertainty of the models' coefficients even without any prior knowledge or assumptions of their statistical distribution.

Note that each model is fitted using data from a great number of stations while other studies prefer to fit one regression per location (e.g. Primack et al. 2009) or to an areal mean (Menzel 2003). At least in our study, this would lead not only to a loss in generalization but also to obvious overfitting. Similar to regional frequency analysis (Hoskings and Wallis 1997), this is a way to capture the statistical relationships between climate variables and phenology in Bavaria rather than to model a number of specific time series of DOYs.

3 Results

3.1 Predictor subset

In Fig. 3 the results for out-of-the-box regressions with 4 and 13 variables are depicted. These fits use all available data. The scatterplots reveal that all six regression models result in very similar estimates for the anomalies of the DOY. As expected, the fits are better when more predictors are considered, however, the differences are small in terms of R^2 as well as of RMSE. The latter rises in all cases by about 0.3 days, while R^2 is reduced by approximately 3 %. Also the correlations between the model predictions $\hat{\boldsymbol{y}}$ are huge (OLS: 0.98, LAD: 0.96 and RFO: 0.97), hence we conclude that the fits for 13 and 4 predictors barely differ.

The estimates for LAD and OLS are nearly indistinguishable and RFO performs better than the two parametric models. However, it is interesting that a small number of potential outliers can easily be identified in each of the panels of Fig 3. Therefore, none of the fitted models is able to include these, no matter which mathematical approach is considered or whether or not we restrict our analysis to a subset of variables.

3.2 Bootstrap evaluation

The results of the bootstraps are used to estimate the consistency of our results. In the first step, we compare the overall performance of the statistical models using 4 as well as 13 predictors. Fig. 4 displays the measures of model quality, together with an estimate of their errors. In terms of R^2 as well as RMSE, Fig. 4 indicates better results for RFO than for OLS and LAD for both sets of predictors. As this is true for both the training and validation data, we are confident that this is not a consequence of overfitting. So, the nonparametric approach appears more successful in grasping the meteorological effects on the DOY than the linear parametric models. As a restriction to this finding, it should be noted that the RFO bootstrap estimates show the largest spread, indicating that they depend more on the training data than the other ones. However, there is a considerable overlap of the margins of error for all three regressions, so that the results of the statistical models are altogether not that different. We also can conclude that a simple linear model is not without justification. Again, for OLS and LAD the results show barely any differences.

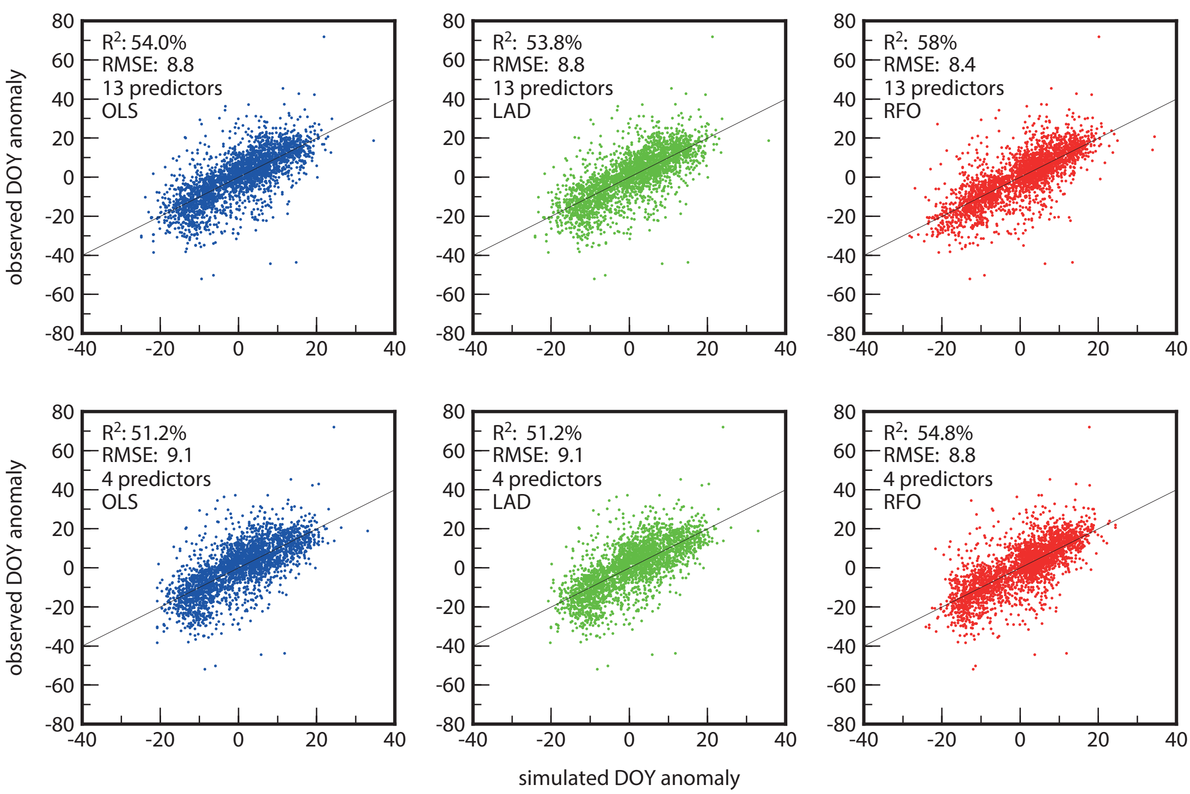

Fig. 3: Scatterplots of observed and simulated DOY anomalies of the flowering of Forsythia for three types of regressions, using 13 (top) and 4 (bottom) predictors

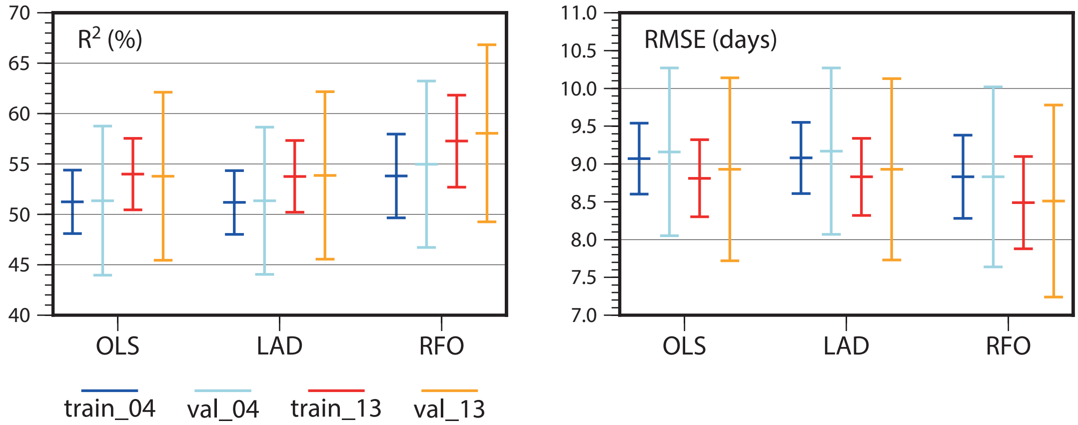

Fig. 4: Estimates for measures of model performance. Results for training datasets using 4 (train_04) and thirteen predictors (train_13) and validation data (val_04, val_13) for all considered regressions with 4 and 13 predictors, respectively. Means +/- two standard deviations of R^2 and RMSE from 1000 model fits.

On average, all regressions explain more than 50 % of the DOY's variance, both for training and validation data. The mean RMSE is approximately 9 days for all models, again RFO performs better than OLS and LAD. When only 4 predictors are taken into account, the performance of all regressions drops, but only to a minor extend. On average, R^2 declines by 3 % and the RMSE increases by 0.3 days, so this subset of climate variables obviously captures the bulk of relevant phenological information. Note that the RFO results in these cases are more similar to those of the parametric models, which we interpret as an indicator that the additional information from the 9 other predictors is mainly non-linear.

A somewhat puzzling feature of Fig. 4 is that in some cases the validation data seems to perform better than the training data. This is a quite uncommon outcome of a bootstrap analysis, but can be explained by the spurious data depicted in Fig. 3. Due to the selection algorithm, these are more often part of training datasets and hence affect those results more strongly. For RFO this is more pronounced than for OLS and LAD.

3.3 Intermodel comparison

Fig. 5 shows results of the intermodel comparison. The correlation between the predictions of all regressions is very high, for OLS and LAD nearly perfect. Furthermore, as the mean RMSE of these regression models is only about 0.5 days, they appear to produce effectively the same estimates for the DOY anomalies. Note, however, pairwise correlations between the estimates of RFO and OLS/LAD are also very high, resulting in a mean R^2 of about 87 % when 4 predictors are used. The RMSE in these cases is somewhat higher, around 3.5 days, but nonetheless the estimates of the different regressions are obviously quite consistent. Again, these findings hold true for both the training and the validation data. For 13 variables, the results are nearly identical, both in terms of the overall structure and the numerical values of R^2 and RMSE.

For the parametric models, we can further explore the predictors based on their standardized partial-regression coefficients. While for 4 considered predictors, all are found to be significant for both parametric models, temperature – as expected – has the strongest effect on DOY. Here, a change of one standard deviation of temperature results in a change in DOY of about 0.65 standard deviations. Wind speed is found to have the second strongest effect, but compared to the impact of temperature it clearly takes a backseat. The smallest effects are found for precipitation. An earlier occurrence of flowering is typically associated with positive anomalies for temperature and wind speed and negative ones for precipitation and snow depth, physical measures of a mild winter or an early onset of spring. These results are consistent for OLS and LAD and don't change much when 13 predictors are considered. However, due to the pronounced correlation especially of the temperature-based variables (see Fig. 2), the bootstrap analysis shows large variances for some key predictors. These models aren't used for the dynamical-statistical modelling. For RFO, scatterplots of the predictions and each predictor variable indicate qualitatively equivalent relationships.

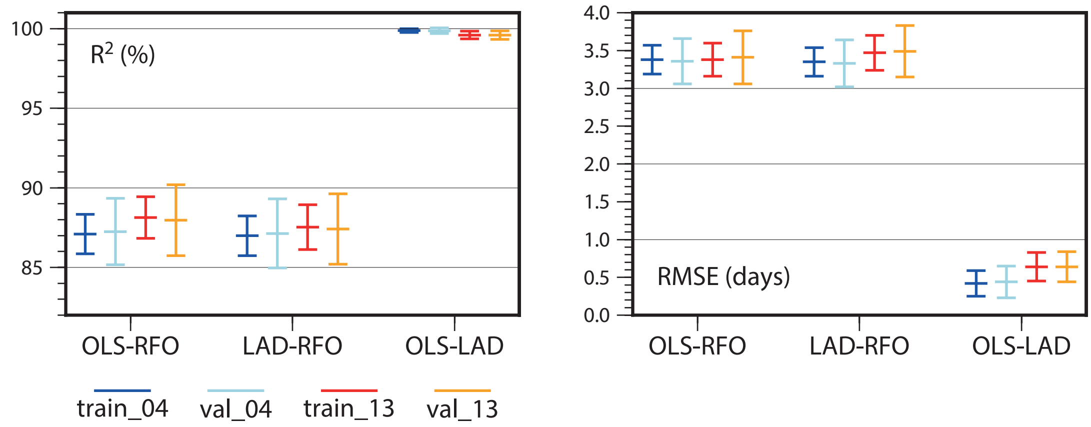

Fig. 5: Similar to Fig. 4 but estimates for model intercomparsion for all regressions.

3.4 RCM based estimates

Fig. 6 shows the mean development of the DOY in the considered model domain for 1950-2100. The gray zone in each panel marks the +/- 1 standard deviation of the time series of the DOY, calculated for the period 1961-1990, for which the means are all zero due to our normalization of the data. This area can be regarded as the range of typical differences in DOY between two subsequent years. In addition, Fig. 6 shows smoothed versions of each time series, which are less affected by REMOs interannual variability.

A common feature for all time series is a pronounced tendency towards earlier flowering dates until 2100 due to prescribed forcing. Also, no matter which regression is considered, the trend is stronger under RCP 8.5 than RCP 4.5. The segregation of the two scenarios becomes more obvious during the second half of the 21th century in accordance with the development of greenhouse gas concentrations. Additionally, the DOY time series for a scenario show very strong correlations between the three types of regression (correlation coefficients are at least 0.95), so short term fluctuations in the simulated climate result in homogeneous phenological responses as well. While these overall tendencies are identical for all regression types, the estimated change in DOY is of considerable range. OLS appears to be the most sensitive. During the last 30 years of the 21th century the DOY is projected to appear on average 14.8 (7.8) days earlier than during the reference period under the RCP 8.5 (RCP 4.5) scenario. It should be noted, however, that for LAD the changes are only marginally smaller. RFO on the other hand is obviously less affected by the forcing. Here, the changes in the DOY are only -7.4 (-4.5) days under RCP 8.5 (RCP 4.5) by the end of the 21 century. Thus, the ratio of the changes per scenario is well comparable to the one for OLS and LAD. Also these changes, albeit smaller than for the other regressions, are outside the estimated range of natural variability, since the standard deviation of RFO is clearly smaller than those of OLS and LAD. Also, as a response to the strong forcing of the RCP 8.5 scenario, RFO's standard deviation reduces significantly, while there are no significant changes in the variability for OLS and LAD.

In Fig. 7, the spatial pattern of the changes in DOY until the end of the 21th century are displayed. Of course the overall picture is consistent with the results shown in Fig. 6. In general, for OLS and LAD the local changes of the DOY are a function of the applied forcing and statistical model. Hence, deviations from a spatial mean response for one gridbox tend to be of the same sign for RCP 4.5 and RCP 8.5. For all combinations of these two factors, the mountainous areas appear to be subject to the most pronounced changes in DOY. The exceptions from this rule (notably in the southern part of Bavaria but north of the Alps) are most likely due to lakes situated in these gridbox. The differences between OLS and LAD as displayed in Fig. 7 are more or less negligible, and in general a consequence of the overall slightly stronger response of OLS. The changes in DOY for RFO are homogeneously smaller than for the other statistical models. Furthermore, especially for RCP 4.5, there occurs stronger scattering of the deviations from the spatial mean change, which might be due to the greater influence of natural variability in comparison to the forcing. RCP 4.5-RFO is also the only combi-

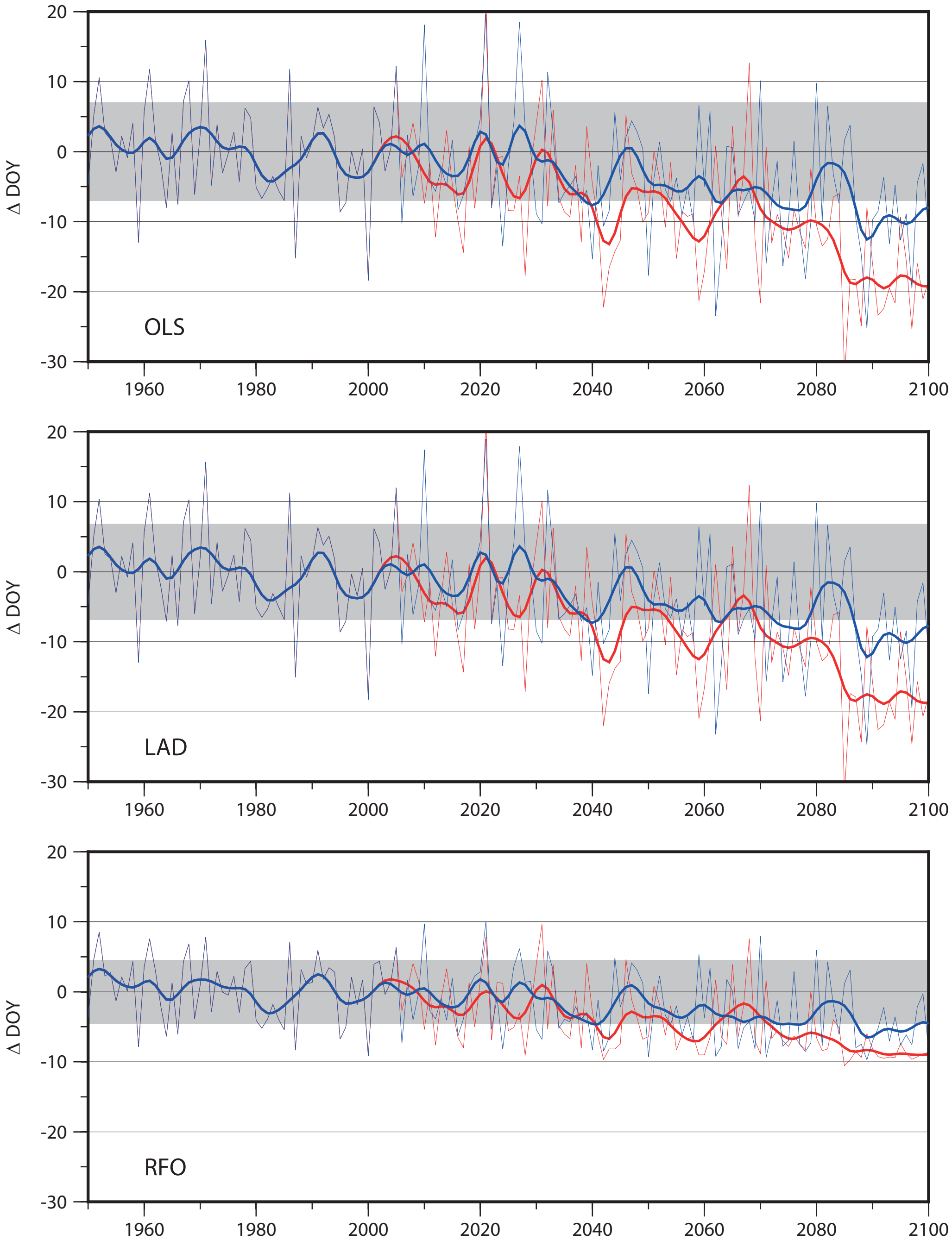

Fig. 6: Time series of variations of Bavaria's spatial mean DOY as derived from dynamical-statistical modeling. For the 21th century red lines show RCP8.5, blue lines RCP4.5. Bold lines are 10-year-lowpass-filtered versions of the time series. Gray area indicates +/- 1 standard deviation.

nation of forcing and statistical model for which a t-test doesn't indicate significant changes in mean DOY with $p < 0.01$ for a number (about 10 %) of gridboxes. Thus, for RCP 4.5-RFO, a dot indicates significant changes in the DOY in Fig. 7 while we don't use dots in the other panels. Altogether the interpretation of the spatial pattern of RFO based DOY changes is less straightforward and physically plausible than for the parametric models. Note that there is no focus on the mountainous areas. This probably rather unrealistic result is a consequence of the RFO's estimates restricted range.

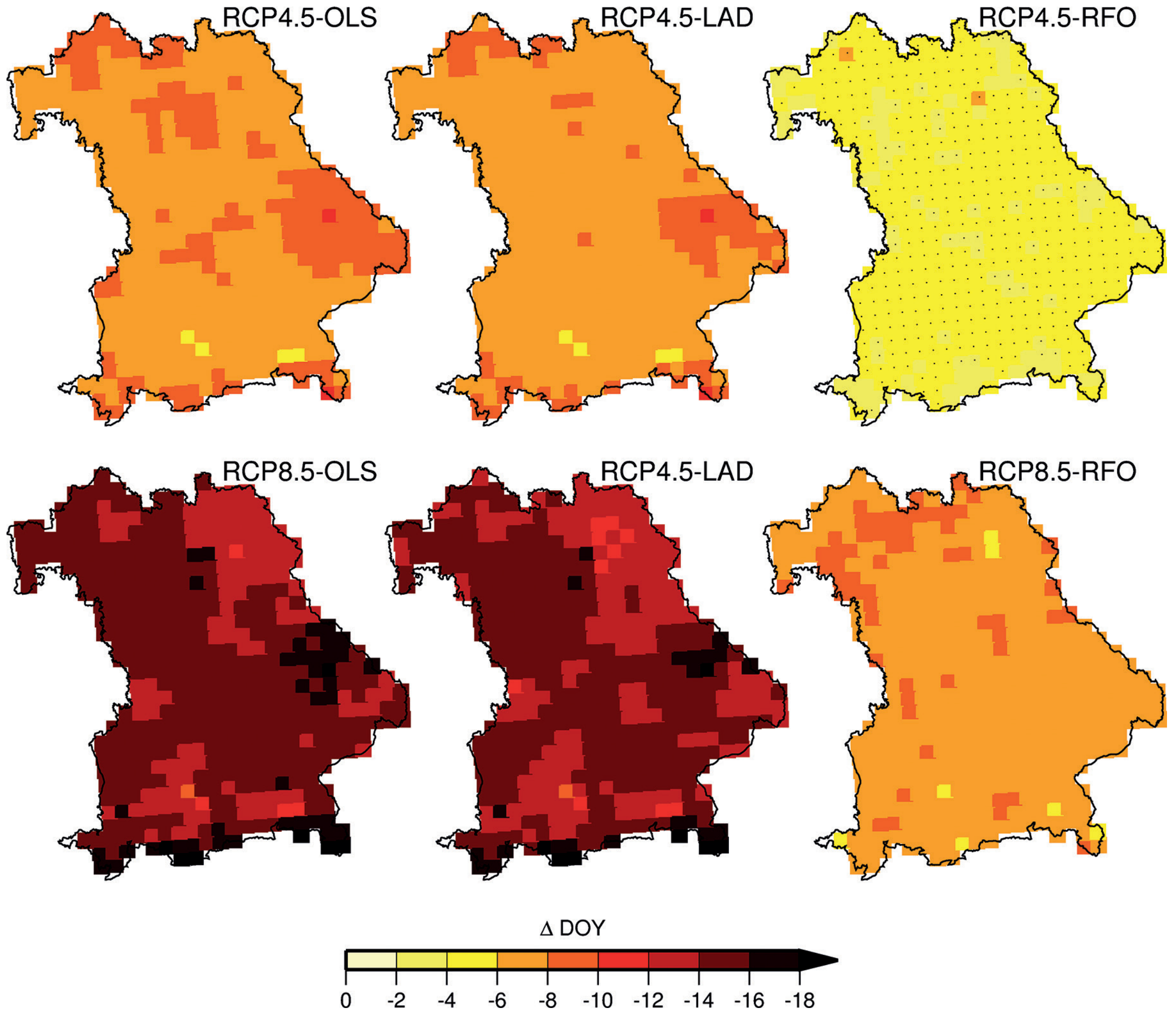

Fig. 7: Changes in DOY from 1961-1990 to 2071-2100 as simulated by the statistical-dynamical model. All changes are significant on the 1% level except for gridboxes in RCP4.5-RFO that are not marked by dot.

3.5 Effects of sample selection

Tab. 1 summarizes the effects of using more restrictive conditions for the selection of the data used for model fitting. We show results from the out-of-the-box regressions, analogous to those depicted in Fig. 3. The gain of restricting the analysis to stations for which at least 30 years of data are available is minor compared to the one that arises from limiting it to phenological stations that are within 100 m elevation difference to their next-neighbor meteorological station. When 4 predictors are considered, the model quality overwhelmingly declines. The smallest data set – containing only one third of the originally used stations – shows the highest values for R^2. For RMSE Tab. 1 indicates overall best results when 84 stations are taken into account. Note, that all estimates are inside the ranges of uncertainty depicted in Fig.4. Also, the pairwise correlations of the predictions are at least 0.95.

More important in terms of this study are effects that might occur due to different mean DOYs used as reference for REMO. Altogether, these are very minor. All possible versions of Fig. 6 are virtually identical. Concerning the regional change in DOY, there are essentially no effects under RCP 4.5 for OLS and LAD. For RCP 8.5, we find somewhat stronger effects when at least 30 years of data are required. For Bavaria's south-east and north-west we find differences up to 4 days, indicating that we might locally underestimate the trend in DOY as a consequence of calculating the mean DOY based on not enough observations. However, less than 4 % of the gridboxes show absolute changes of more than two days. For RFO regional change in DOY becomes even more homogeneous by excluding stations, but qualitative

Tab. 1: Measures of model quality for all three regressions with 13 (4) predictors restricting the sample to phenological stations within a vertical difference of 100 m of their next-neighbor meteorological station (first row), for which at least 30 years of data are available (second row), or both. The first (second) column indicates the number of stations (observations).

		R² in %			RMSE (days)		
stations	sample	OLS	LAD	RFO	OLS	LAD	RFO
84	2292	57.2 (53.8)	57.1 (53.7)	61.1 (58.0)	8.3 (8.6)	8.3 (8.7)	7.9 (8.2)
37	1396	55.0 (50.9)	54.8 (50.9)	59.3 (55.3)	8.9 (9.3)	9.0 (9.3)	8.5 (8.9)
31	1172	59.3 (53.9)	59.1 (53.9)	62.8 (58.4)	8.3 (8.8)	8.3 (8.8)	7.9 (8.4)

and quantitative effects are neglectable (less than 3 % of the gridboxes show changes of more than one day). Here, there are no apparent differences between RCP 4.5 and RCP 8.5.

4 Discussion

Given the overall results we conclude that our models succeed in their task to establish a robust statistical link between local climate conditions and the DOY of Forsythias flowering. As we are not aiming to predict the absolute DOY but rather model its variation, our results are ready to use along with modern high-quality climate simulations. We find a strong, overall tendency towards earlier flowering mainly in accordance with the effects of rising temperatures during the 21th century. As a consequence, the risk of late frost events should increase. Considering Forsythia, the economic effects won't be of much relevance. However, it is highly likely that other plants, including field crops and fruit trees, will respond to climate change in a similar way and hence become more vulnerable to frost damages (e.g. Chmielewski et al. 2004; Richardson et al. 2012).

Concerning the predictor selection, our bootstrap results indicate that the reduction to 4 predictors didn't affect the outcome of our study substantially. The partial-regression coefficients indicate that early occurrences of blossoming are typically associated with the positive phase of the North Atlantic Oscillation (NAO), which is known to affect phenophases all over Europe (Chemielewski and Rötzer 2001). In the study area, the NAO strongly affects temperature and wind but has minor effects on precipitation (Hurrell 1995). Of course, all of these affect snow depth, but the latter is more of a local aspect. However, our set of predictors might lack additional climatic information that may become more important under global warming conditions, such as the plants need for frosts before spring (e.g. Chuine et al. 1999; Chuine et al. 2016). In terms of the spatial pattern of changes in the DOY it should be considered that for mountainous areas not much observational data is available. However, the combined effects of changes in mean temperature and snow depth taken into account, the stronger signal in the DOY in these regions seems plausible.

For climate change studies, our approach is a useful complement or even alternative to highly specialized processed-based phenological models, especially considering the results of the cross-validating bootstrap that is useful for uncertainty assessment (Richardson et al 2013). Another advantage is that the effects of errors in the phenological observations (Larcher 2006) are reduced simply due to the large number of observations available and spread over a large area. It is capable to deal with relatively short time series and heterogeneous terrain, as the uncertainty induced by these issues clearly takes a back seat compared to the one associated with greenhouse gas forcing.

As a major methodological question, we were interested in the comparison of the three different regressions. Given the results for LAD and OLS, it seems that we could have restricted our analysis to one of these as the results appear to be virtual identical while RFO results show notable differences to the parametric regressions. However, given the scope of this study, we elected to show the comparable results for all regressions. While, our data didn't demand for a more robust approach than OLS, our results show that LAD is an alternative to OLS even if it is not required by the data. It could be argued that the spurious data (Fig. 4, Fig. 5) should have been excluded from the analysis. However, we didn't want to do so since we didn't want to make such a stringent prior assumption. We also believe that the finding that all three types of regression fail in a similar way is an interesting aspect of our study: none, including the RFO, can deal with all possible structures of variation in the data. Nonetheless, RFO was found superior over parametric regression in different fields (e.g. Svetnik et al 2003; Oliveira et al 2012) and we also found that it captured some minor climatic effects better. Our results, however, demonstrate that RFO, simply due to its definition, cannot respond to an appropriate extent to the applied forcing, but is limited to the observed range of the dependent variable. It is stuck in the extreme values observed in the past, resulting also in a statistically significant reduction of variance during the 2071-2100 period. Hence, its use in statistical-dynamical approaches should be considered with great care. Especially the spatial patterns of changes in the DOY simulated by OLS and LAD are considerably more in line with theoretical and empirical findings (e.g. Richardson et al. 2013) than those by RFO. Further, partial-regression coefficients allow for a relatively straightforward interpretation of the statistical model. RFO measures such as variable importance (see Breiman 2001) are hereby less useful. If a relatively small number of predictors that all affect the predictand is considered – a situation commonly regarded desirable – the results barley differ for each variable. Also, neither qualitative nor quantitative aspects of the relationships can be assessed directly. Regardless, parametric models might lose functionality as well when their predictors are outside the range of the observations used to estimate the regression coefficients. And the continuous advancement of flowering dates with increasing temperature is, of course, even physically impossible (e.g. Chuine et al 1999). Considering all results from the present study, however, parametric regression seems more useful for when it comes to analyzing the statistical model to time periods far beyond the one used for training the model.

5 Conclusions

All three types of regression were capable of detecting and modeling a robust statistical relationship between climate and the day of *Forsythia suspensa* flowering using 4 or 13 predictors and explaining well over 50 % of total variance. The best model fits were achieved by the RFO for both sets of predictors. In terms of dynamical-statistical modeling, however, it is less able to respond to future climate change than OLS and LAD, damping the simulated signal. Rising temperatures are the main driver of the advancement of the DOY. While the overall tendencies of the statistical-dynamical models are plausible, it would be a useful alternative to have powerful dynamical vegetation modeling included in climate models.

Acknowledgements

We thank the German meteorological service and the Bavarian state office for the Environment for providing the meteorological and phenological data used in this study. The elevation data used in Fig. 1 was kindly provided by the Bavarian state office for Digitizing, Broadband and Survey under creative common license CC BY 3.0DE. We further acknowledge the World Climate Research Programme's Working Group on Regional Climate, and the Working Group on Coupled Modeling, former coordinating body of CORDEX as well as the Helmholtz-Zentrum Geesthacht, Climate Service Center and the Max Planck Institute for Meteorology for producing the REMO data and making it available via the Earth System Grid Federation infrastructure. We are thankful to one anonymous reviewer for the helpful comments on the manuscript.

References

Awoyé, O. H. R.; Pollinger, F.; Agbossou, E. K. and Paeth, H. (2017): Dynamical-statistical projections of the climate change impact on agricultural production in Benin by means of a cross-validated linear model combined with Bayesian statistics. In: Agricultural and Forest Meteorology 234–235, 80–94. https://doi.org/10.1016/j.agrformet.2016.12.010

Breiman, L.; Friedman, J.; Stone, C. J. and Olshen, R. A. (1984): Classification and regression trees. The Wadsworth and Brooks-Cole statistics-probability series. London.

Breiman, L. (2001): Random forests. In: Machine Learning 45, 5–31. https://doi.org/10.1023/A:1010933404324

Bruns, E.; Elsässer, M.; Heitland, W.; Höfer, M.; Maier, H.; Meier, U.; Müller-Platz, W.; Raath, J. and Simon, W. (2015): Anleitung für die phänologischen Beobachter des Deutschen Wetterdienstes. Vorschriften und Betriebsunterlagen für die phänologischen Beobachter des Deutschen Wetterdienstes 17. Offenbach.

Chmielewski, F.-M and Rötzer, T. (2001): Response of tree phenology to climate change across Europe. In: Agricultural and Forest Meteorology 108, 101–112. https://doi.org/10.1016/S0168-1923(01)00233-7

Chmielewski, F.-M.; Müller, A. and Bruns, E. (2004): Climate changes and trends in phenology of fruit trees and field crop in Germany 1961-2000. In: Agricultural and Forest Meteorology 121 (1-2), 69–78. https://doi.org/10.1016/S0168-1923(03)00161-8

Chmielewski, F.-M.; Heider, S.; Moryson, S. and Bruns, E. (2013): International phenological observation networks: Concept of IPG and GPM. In: Schwartz, M. D. (ed): Phenology: an integrative environmental science. Heidelberg, 137–153. https://doi.org/10.1007/978-94-007-6925-0_8

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. In: Plant, Cell & Environment 22 (1), 1–13. https://doi.org/10.1046/j.1365-3040.1999.00395.x

Chuine, I.; Bonhomme, M.; Legave, J. M.; de Corátar-Atauri, G.; Charrier, G.; Lacointe, A. and Améglio (2016): Can phenological models predict tree phenology accurately in the future? The unrevealed hurdle of endodormancy break. In: Global Change Biology 22 (10), 3444–3460. https://doi.org/10.1111/gcb.13383

Cleland, E. E.; Chuine, I.; Menzel, A.; Mooney, H. A. and Schwartz, M. D. (2007): Shifting plant phenology in response to global change. In: Trends in Ecology & Evolution 22 (7), 357–365. https://doi.org/10.1016/j.tree.2007.04.003

Dielman, T. E. (2005): Least absolute value regression: recent contributions. In: Journal of Statistical Computation and Simulation 75, 263–286. https://doi.org/10.1080/0094965042000223680

Estrella, N. and Menzel, A. (2006): Responses of leaf colouring of four deciduous tree species to climate and weather in Germany. In: Climate Research 32, 253–267. https://doi.org/10.3354/cr032253

Flato, G.; Marotzke, J.; Abiodun, B.; Braconnot, P.; Chou, S. C.; Collins, W.; Cox, P.; Driouech, F.; Emori, S.; Eyring, V.; Forest, C.; Gleckler, P.; Guilyardi, E.; Jakob, C.; Kattsov, V.; Reason, C. and Rummukainen, M. (2013): Evaluation of climate models. In: Stocker, T. F.; Qin, D.; Plattner, G.-K.; Tignor, M.; Allen, S. K.; Boschung, J.; Nauels, A.; Xia, Y.; Bex, V. and Midgley, P. M. (eds.): Climate change 2013: the physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change. Cambridge, 741–866.

Giorgi, F.; Jones, C. and Asrar, G. R. (2006): Addressing climate information needs at the regional level: the CORDEX framework. In: World Meteorological Organization Bulletin 58, 175–183.

Häninnen, H. (1994): Effects of climatic change on trees from cool and temperate regions: an ecophysiological approach to modeling of bud burst phenology. In: Canadian Journal of Botany 73, 183–199. https://doi.org/10.1139/b95-022

Hastie, T.; Tibshirani, R. and Friedman, J. H. (2008): The elements of statistical learning: data mining, inference, and prediction. New York.

Hoskings, J. R. M. and Wallis, J. R. (1997): Regional frequency analysis: an approach based on L-Moments. Cambridge. https://doi.org/10.1017/CBO9780511529443

Hurrell, J. W. (1995): Decadal trends in the North Atlantic Oscillation: regional temperatures and precipitation. In: Science 269, 676–679. https://doi.org/10.1126/science.269.5224.676

Field, C. B.; Barros, V. R.; Dokken, D. J.; Mach, K. J.; Mastrandrea, M. D.; Bilir, T. E.; Chatterjee, M.; Ebi, K. L.; Estrada, Y. O.; Genova, R. C.; Girma, B.; Kissel, E. S.; Levy, A. N.; MacCracken, S.; Mastrandrea, P. R. and White, L. L. (eds.) (2014): Climate change 2014: impacts, adaptation, and vulnerability. Part A: global and sectoral aspects. Contribution of working group II to the fifth assessment report of the intergovernmental panel on climate change. Cambridge

Jacob, D.; Petersen, J.; Eggert, B.; Alias, A.; Christensen, O. B.; Bouwer, L.; Braun, A.; Colette, A.; Déqué, M.; Georgievski, G.; Georgopoulou, E.; Gobiet, A.; Menut, L.; Nikulin, G.; Haensler, A.; Hempelmann, N.; Jones, C.; Keuler, K.; Kovats, S.; Kröner, N.; Kotlarski, S.;Kriegsmann, A.; Martin, E.; Meijgaard, E.; Moseley, C.; Pfeifer, S.; Preuschmann, S.;Radermacher, C.; Radtke, K.; Rechid, D.; Rounsevell, M.; Samuelsson, P.; Somot, S.; Soussana, J.-F.; Teichmann, C.; Valentini, R.; Vautard, R.; Weber, B. and Yiou, P. (2014): EURO-CORDEX: new high-resolution climate change projections for European impact research. In: Regional Environmental Change 14, 563–578. https://doi.org/10.1007/s10113-013-0499-2

Larcher, W. (2006): Altitudinal variation in flowering time of Lilac (Syringa vulgaris L.) in the Alps in relation to temperatures. In: Sitzungsberichte und Anzeiger der mathematisch-naturwissenschaftliche Klasse, Abteilung I: Biologische Wissenschaften und Erdwissenschaften 212, 3–18.

Lieth, H. (1974): Phenology and seasonality modeling. Berlin. https://doi.org/10.1007/978-3-642-51863-8

Ma, T. and Zhou, C. (2012): Climate-associated changes in spring plant phenology in China. In: International Journal of Biometeorology 56 (2), 269–275. https://doi.org/10.1007/s00484-011-0428-3

Menzel, A.; Estrella, N. and Fabian, P. (2001): Spatial and temporal variability of the phenological seasons in Germany from 1951 to 1996. In: Global Change Biology 7 (6), 657–666. https://doi.org/10.1046/j.1365-2486.2001.00430.x

Menzel, A. (2003): Plant phenological anomalies in Germany and their relation to air temperature and NAO. In: Climatic Change 57 (3), 243–263. https://doi.org/10.1023/A:1022880418362

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.; Måge, 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 Zust, A. (2006): European phenological response to climate change matches the warming pattern. In: Global Change Biology 12, 1969–1976. https://doi.org/ 10.1111/j.1365-2486.2006.01193.x

Migliavacca, M.; Sonnentag, O.; Keenan, T.F.; Cescatti, A.; O'Keefe, J. and Richardson, A. D. (2012): On the uncertainty of phenological responses to climate change, and implications for a terrestrial biosphere model. In: Biogeosciences 9, 2063–2083. https://doi.org/10.5194/bg-9-2063-2012

MORIN, X.; LECHOWICZ, M.J.; AUGSPURGER, C.; O' KEEFE, J.; VINER, D. and CHUINE, I. (2009): Leaf phenology in 22 North American tree species during the 21st century. In: Global Change Biology 15, 961–975. https://doi.org/10.1111/j.1365-2486.2008.01735.x

MOSS, R.; BABIKER, M.; BRINKMAN, S; CALVO, E.; CARTER, T.; EDMONDS, J.; ELGIZOULI, I.; EMORI, S.; ERDA, L.; HIBBARD, K.; JONES, R.; KAINUMA, M.; KELLEHER, J.; LAMARQUE, J. F.; MANNING, M.; MATTHEWS, B.; MEEHL, J.; MEYER, L.; MITCHELL, J.; NAKICENOVIC, N.; O'NEILL, B.; PICHS, R.; RIAHI, K.; ROSE, S.; RUNCI, P.; STOUFFER, R.; VAN VUUREN, D.; WEYANT, J.; WILBANKS, T.; VAN YPERSELE, J. P. and ZUREK, M. (2008):Towards new scenarios for analysis of emissions, climate change, impacts, and response strategies. Technical summary. Intergovernmental panel on climate change. Geneva.

OLIVEIRA, S.; OEHLER, F.; SAN-MIGUEL-AYANZ, J.; CAMIA, A. and PEREIRA, J. M. (2012): Modeling spatial patterns of fire occurrence in Mediterranean Europe using multiple regression and random forest. In: Forest Ecology and Management 275, 117–129. https://doi.org/10.1016/j.foreco.2012.03.003

PAETH, H.; CAPO-CHICHI, A. and ENDLICHER, W. (2008): Climate change and food security in tropical West Africa. In: Erdkunde 62 (2), 101–115. https://doi.org/10.3112/erdkunde.2008.02.01

PRIMACK, R. B.; IBANEZ, I.; HIGUCHI, H.; LEE, S. D.; MILLER-RUSHING, A. J.; WILSON, A. M. AND SILANDER, J. A. (2009): Spatial and interspecific variability in phenological responses to warming temperatures. In: Biological Conservation 142 (11), 2569–2577. https://doi.org/10.1016/j.biocon.2009.06.003

PORTNOY, S. and KOENKER, R. (1997): The Gaussian hare and the Laplacian tortoise: computability of squared-error versus absolute-error estimators. In Statistical Science 12 (4), 279–300. https://doi.org/10.1214/ss/1030037960

RICHARDSON, A. D. and O'KEEFE, J. (2009): Phenological differences between understory and overstory: a case study using the longterm Harvard Forest records. In: NOORMETS, A. (ed.): Phenology of ecosystem processes. New York, 87–117. https://doi.org/10.1007/978-1-4419-0026-5_4

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. In: Agricultural and Forest Meteorology 169, 156–173. https://doi.org/10.1016/j.agrformet.2012.09.012

RICHARDSON, A. D.; ANDERSON, R. S.; ARAIN, M. A.; BARR, A. G.; BOHRER, G.; CHEN, G. S.; CHEN, J. M.; CIAIS, P.; DAVIS, K. J.; DESAI, A. R.; DIETZE, M. C.; DRAGONI, D.; GARRITY, S. R.; GOUGH, C. M.; GRANT, R.; HOLLINGER, D. Y.; MARGOLIS, H. A.; MCCAUGHEY, H.; MIGLIAVACCA, M.; MONSON, R. K.; MUNGER, J. W.; POULTER, B.; RACZKA, B. M.; RICCIUTO, D. M.; SAHOO, A. K.; SCHAEFER, K.; TIAN, H. Q.; VARGAS, R.; VERBEECK, H.; XIAO, J. F. and XUE, Y. K. (2012): Terrestrial biosphere models need better representation of vegetation phenology: results from the North American Carbon Program Site Synthesis. In: Global Change Biology 18, 566–584. https://doi.org/10.1111/j.1365-2486.2011.02562.x

SCHABER, J. and BADECK, F. W. (2003): Physiology-based phenology models for forest tree species in Germany. In: International Journal of Biometeorology 47, 193–201. https://doi.org/10.1007/s00484-003-0171-5

SCHWARTZ, M. D. (ed) (2013): Phenology: an integrative environmental science. Heidelberg. https://doi.org/10.1007/978-94-007-6925-0

SETIYONO, T. D; WEISS, A.; SPECHT, J.; BASTIDAS, A. M.; CASSMAN, K. G. and DOBERMANN, A. (2007): Understanding and modeling the effect of temperature and daylength on soybean phenology under high-yield conditions. In: Field Crops Research 100 (2-3), 257–271. https://doi.org/10.1016/j.fcr.2006.07.011

SVETNIK, V.; LIAW, A.; TONG, C.; CULBERSON, J. C.; SHERIDAN, R. P. and FEUSTON, B. P. (2003): Random forest: a classification and regression tool for compound classification and QSAR modeling. In: Journal of chemical information and computer sciences 43 (6), 1947–1958. https://doi.org/10.1021/ci034160g

TAYLOR, K. E.; STOUFFER, R. J. and MEEHL, G. A. (2012): An overview of CMIP5 and the experiment design. In: Bulletin of the American Meteorological Society 93, 485–498. https://doi.org/10.1175/BAMS-D-11-00094.1

THACKERAY, S. J.; HENRYS, P. A.; HEMMING, D.; BELL, J. R.; BOTHAM, M. S.; BURTHE, S.; HELAOUET, P.; JOHNS, D. G.; JONES, I. D.; LEECH, D. I.; MACKAY, E. B.; MASSIMINO, D.; ATKINSON, S.; BACON, P. J.; BRERETON, T. M.; CARVALHO, L.; CLUTTON-BROCK, T. H.; DUCK, C.; EDWARDS, M.; ELLIOTT, J. M.; HALL, S. J. G.; HARRINGTON, R; PEARCE-HIGGINS, J. W.; HØYE, T. T.; KRUUK, L. E. B.; PEMBERTON, J. M.; SPARKS, T. H.; THOMPSON, P. M.; WHITE, I.; WINFIELD, I. J. and WANLESS, S. (2016): Phenological sensitivity to climate across taxa and trophic levels. In: Nature 535, 241–245. https://doi.org/10.1038/nature18608

Authors

Dr. Felix Pollinger
Katrin Ziegler, M.Sc.
Prof. Dr. Heiko Paeth
University of Wuerzburg
Institute of Geography and Geology
Am Hubland
97074 Wuerzburg
Germany
felix.pollinger@uni-wuerzburg.de
katrin.ziegler@stud-mail.uni-wuerzburg.de
heiko.paeth@uni-wuerzburg.de