Phenology in Germany in the 20th Century: Methods, Analyses and Models

Dipl. Syst.-Wiss. Jörg Schaber

April 2002

Dissertation zur Erlangung des akademischen Grades
Doktor der Naturwissenschaften (Dr. rer. nat.)
Eingereicht im Fachbereich Geoökologie
der Naturwissenschaftlich - Mathematischen Fakultät
der UNIVERSITÄT POTSDAM, DEUTSCHLAND

Dissertation submitted for the obtainment of the degree of
Doctor rerum naturalium
in the department of Geoecology
of the
UNIVERSITY of POTSDAM, GERMANY

Gutachter/Reviewers:

Professor Dr. Axel Bronstert
Professor Dr. Wolfgang Cramer
Dr. Franz-W. Badeck

Table of Contents

List of Figures

List of Tables

Danksagung

An erster Stelle danke ich herzlich Dr. Franz Badeck, der die vorliegende Arbeit initiiert, ständig neu motiviert und mit großer Aufopferung in jeglicher Hinsicht unterstützt und begleitet hat. Ich danke ihm für die mannigfaltigen Erkenntnisse und Ideen, die er mir in vielen Gesprächen hat zuteil werden lassen. Ich bedanke mich herzlich bei Prof. Dr. Axel Bronstert und Prof. Dr. Wolfgang Cramer, welche meine Arbeit betreut, begutachtet und überhaupt erst ermöglicht haben. Dank gebührt meinen Kolleginnen in der Arbeitsgruppe Petra Lasch, Ylva Hauf und Dr. Felicitas Suckow für Ihre Unterstützung durch die Schaffung von Freiräumen zur Erstellung der Dissertation sowie durch tatkräftige Hilfe.

Diese Arbeit wurde durch das Hochschulsonderprogram HSP3 des Landes Brandenburg und das EU-Projekt 'Advanced Terrestrial Ecosystems Analysis and Modelling' (ATEAM) Nr. EVK2-2000-00075 gefördert, wofür ich mich bedanke.

Ich danke dem Deutschen Wetterdienst für die Bereitstellung der Daten. Den Mitarbeitern des DWD Frau Polte-Rudolf und Herrn Wessel danke ich für die Unterstüzung bei der Datenanalyse bzw. beim Zugang zum phänologischen Archiv in Potsdam.

Ich danke den Kollegen, welche mit ihren phänologischen Beobachtungen wertvolle Daten zu dieser Arbeit beigetragen haben oder mir mit fachlichem Rat zur Seite standen: Wilfried Ahrens, Franz Badeck, Markus Erhard, Uta Fritsch, F.-W. Gerstengarbe, Alexa Griesel, Arnulf Günther, Andreas Güntner, Cesar Ionescu, Fred Hattermann, Ylva Hauf, Anne Jäger, Thomas Kleinen, Matze Lüdeke, Lucas Menzel, Peter Mohr, Dr. Pincus von der Uni Potsdam, Didier Roche, Steven Sitch, Feli Suckow, Martin Watti Wattenbach, Peter Werner und Kirsten Zickfeld.

Ferner danke ich Frau Uffrecht und Irina Krysanova, welche mir durch die prompte und freundliche Beschaffung von Literatur das Arbeiten wesentlich erleichtert haben. Für die angenehmen Rahmenbedingungen, welche ich sehr zu schätzen weiß und für die ich mich herzlich bedanke, sorgte darüber hinaus das PIK im Allgemeinen und im Besonderen Wolfgang Cramer, Werner von Bloh, Michael Flechsig, Dietmar Gibietz, Joachim Glauer, Karsten Kramer, Helmut Miethke, und Alison Schlums.

Meiner Mutter danke ich für Ihre moralische und finanzielle Unterstützung. Silke danke ich für Zuspruch und Unterstützung in einer stressigen Zeit.

In Gedenken an Alfred Schaber

Potsdam, den 1.4.2002 Jörg Schaber

Abstract

The length of the vegetation period (VP) plays a central role for the interannual variation of carbon fixation of terrestrial ecosystems. Observational data analysis has indicated that the length of the VP has increased in the last decades in the northern latitudes mainly due to an advancement of bud burst (BB). This phenomenon has been widely discussed in the context of Global Warming because phenology is correlated to temperatures.

Analyzing the patterns of spring phenology over the last century in Southern Germany provided two main findings:

- The strong advancement of spring phases especially in the decade before 1999 is not a singular event in the course of the 20th century. Similar trends were also observed in earlier decades. Distinct periods of varying trend behavior for important spring phases could be distinguished.
- Marked differences in trend behavior between the early and late spring phases were detected. Early spring phases changed as regards the magnitude of their negative trends from strong negative trends between 1931 and 1948 to moderate negative trends between 1948 and 1984 and back to strong negative trends between 1984 and 1999. Late spring phases showed a different behavior. Negative trends between 1931 and 1948 are followed by marked positive trends between 1948 and 1984 and then strong negative trends between 1984 and 1999.

This marked difference in trend development between early and late spring phases was also found all over Germany for the two periods 1951 to 1984 and 1984 to 1999.

The dominating influence of temperature on spring phenology and its modifying effect on autumn phenology was confirmed in this thesis. However,

- temperature functions determining spring phenology were not significantly correlated with a global annual CO_2 signal which was taken as a proxy for a Global Warming pattern.
- an index for large scale regional circulation patterns (NAO index) could only to a small part explain the observed phenological variability in spring.

The observed different trend behavior of early and late spring phases is explained by the differing behavior of mean March and April temperatures. Mean March temperatures have increased on average over the 20th century accompanied by an increasing variation in the last 50 years. April temperatures, however, decreased between the end of the 1940s and the mid-1980s, followed by a marked warming after the mid-1980s.

It can be concluded that the advancement of spring phenology in recent decades are part of multidecadal fluctuations over the 20th century that vary with the species and the relevant seasonal temperatures. Because of these fluctuations a correlation with an observed Global Warming signal could not be found.

On average all investigated spring phases advanced between 5 and 20 days between 1951 and 1999 for all Natural Regions in Germany. A marked difference between late and early spring phases is due to the above mentioned differing behavior before and after the mid-1980s. Leaf coloring (LC) was delayed between 1951 and 1984 for all tree species. However, after 1984 LC was advanced. Length of the VP increased between 1951 and 1999 for all considered tree species by an average of ten days throughout Germany.

It is predominately the change in spring phases which contributes to a change in the potentially absorbed radiation. Additionally, it is the late spring species that are relatively more favored by an advanced BB because they can additionally exploit longer days and higher temperatures per day advancement. To assess the relative change in potentially absorbed radiation among species, changes in both spring and autumn phenology have to be considered as well as where these changes are located in the year.

For the detection of the marked difference between early and late spring phenology a new time series construction method was developed. This method allowed the derivation of reliable time series that spanned over 100 years and the construction of locally combined time series increasing the available data for model development.

Apart from analyzed protocolling errors, microclimatic site influences, genetic variation and the observers were identified as sources of uncertainty of phenological observational data. It was concluded that 99% of all phenological observations at a certain site will vary within approximately 24 days around the parametric mean. This supports to the proposed 30-day rule to detect outliers.

New phenology models that predict local BB from daily temperature time series were developed. These models were based on simple interactions between inhibitory and promotory agents that are assumed to control the developmental status of a plant. Apart from the fact that, in general, the new models fitted and predicted the observations better than classical models, the main modeling results were:

- The bias of the classical models, i.e. overestimation of early observations and underestimation of late observations, could be reduced but not completely removed.
- The different favored model structures for each species indicated that for the late spring phases photoperiod played a more dominant role than for early spring phases.
- Chilling only plays a subordinate role for spring BB compared to temperatures directly preceding BB.

Keywords: phenology, combined time series, physiology-based models, uncertainty, variability, outliers, trends, Natural Regions, trend turning points, *Fagus sylvatica*, *Galanthus nivali*s, *Sambucus nigra*, *Quercus robur*, *Aesculus hippocastanum*, *Betula pendula*, *Syringa vulgaris*

Zusammenfassung

Die Länge der Vegetationsperiode (VP) spielt eine zentrale Rolle für die interannuelle Variation der Kohlenstoffspeicherung terrestrischer Ökosysteme. Die Analyse von Beobachtungsdaten hat gezeigt, dass sich die VP in den letzten Jahrzehnten in den nördlichen Breiten verlängert hat. Dieses Phänomen wurde oft im Zusammenhang mit der globalen Erwärmung diskutiert, da die Phänologie von der Temperatur beeinflusst wird.

Die Analyse der Pflanzenphänologie in Süddeutschland im 20. Jahrhundert zeigte:

- Die starke Verfrühung der Frühjahrsphasen in dem Jahrzehnt vor 1999 war kein singuläres Ereignis im 20. Jahrhundert. Schon in früheren Dekaden gab es ähnliche Trends. Es konnten Perioden mit unterschiedlichem Trendverhalten identifiziert werden.
- Es gab deutliche Unterschiede in den Trends von frühen und späten Frühjahrsphasen. Die frühen Frühjahrsphasen haben sich stetig verfrüht, mit deutlicher Verfrühung zwischen 1931 und 1948, moderater Verfrühung zwischen 1948 und 1984 und starker Verfrühung zwischen 1984 und 1999. Die späten Frühjahrsphasen hingegen, wechselten ihr Trendverhalten in diesen Perioden von einer Verfrühung zu einer deutlichen Verspätung wieder zu einer starken Verfrühung.

Dieser Unterschied in der Trendentwicklung zwischen frühen und späten Frühjahrsphasen konnte auch für ganz Deutschland in den Perioden 1951 bis 1984 und 1984 bis 1999 beobachtet werden.

Der bestimmende Einfluss der Temperatur auf die Frühjahrsphasen und ihr modifizierender Einfluss auf die Herbstphasen konnte bestätigt werden. Es zeigt sich jedoch, dass

- die Phänologie bestimmende Funktionen der Temperatur nicht mit einem globalen jährlichen CO_2 Signal korreliert waren, welches als Index für die globale Erwärmung verwendet wurde.
- ein Index für grossräumige regionale Zirkulationsmuster (NAO-Index) nur zu einem kleinen Teil die beobachtete phänologischen Variabilität erklären konnte.

Das beobachtete unterschiedliche Trendverhalten zwischen frühen und späten Frühjahrsphasen konnte auf die unterschiedliche Entwicklung von März- und Apriltemperaturen zurückgeführt werden. Während sich die Märztemperaturen im Laufe des 20. Jahrhunderts mit einer zunehmenden Variabilität in den letzten 50 Jahren stetig erhöht haben, haben sich die Apriltemperaturen zwischen dem Ende der 1940er und Mitte der 1980er merklich abgekühlt und dann wieder deutlich erwärmt.

Es wurde geschlussfolgert, dass die Verfrühungen in der Frühjahrsphänologie in den letzten Dekaden Teile multi-dekadischer Fluktuationen sind, welche sich nach Spezies und relevanter saisonaler Temperatur unterscheiden. Aufgrund dieser Fluktuationen konnte kein Zusammenhang mit einem globalen Erwärmungsignal gefunden werden.

Im Durchschnitt haben sich alle betrachteten Frühjahrsphasen zwischen 1951 und 1999 in Naturräumen in Deutschland zwischen 5 und 20 Tagen verfrüht. Ein starker Unterschied in der

Verfrühung zwischen frühen und späten Frühjahrsphasen liegt an deren erwähntem unterschiedlichen Verhalten. Die Blattverfärbung hat sich zwischen 1951 und 1999 für alle Spezies verspätet, aber nach 1984 im Durchschnitt verfrüht. Die VP hat sich in Deutschland zwischen 1951 und 1999 um ca. 10 Tage verlängert.

Es ist hauptsächlich die Änderung in den Frühjahrphasen, die zu einer Änderung in der potentiell absorbierten Strahlung (PAS) führt. Darüber hinaus sind es die späten Frühjahrsphasen, die pro Tag Verfrühung stärker profitieren, da die zusätzlichen Tage länger und wärmer sind als dies für die frühen Phasen der Fall ist. Um die relative Änderung in PAS im Vergleich der Spezies abzuschätzen, müssen allerdings auch die Veränderungen in den Herbstphasen berücksichtigt werden.

Der deutliche Unterschied zwischen frühen und späten Frühjahrsphasen konnte durch die Anwendung einer neuen Methode zur Konstruktion von Zeitreihen herausgearbeitet werden. Der neue methodische Ansatz erlaubte die Ableitung verlässlicher 100-jähriger Zeitreihen und die Konstruktion von lokalen kombinierten Zeitreihen, welche die Datenverfügbarkeit für die Modellentwicklung erhöhten.

Ausser analysierten Protokollierungsfehlern wurden mikroklimatische, genetische und Beobachtereinflüsse als Quellen von Unsicherheit in phänologischen Daten identifiziert. Phänologischen Beobachtungen eines Ortes können schätzungsweise 24 Tage um das parametrische Mittel schwanken. Dies unterstützt die 30-Tage Regel für die Detektion von Ausreissern.

Neue Phänologiemodelle, die den Blattaustrieb aus täglichen Temperaturreihen simulieren, wurden entwickelt. Diese Modelle basieren auf einfachen Interaktionen zwischen aktivierenden und hemmenden Substanzen, welche die Entwicklungsstadien einer Pflanze bestimmen. Im Allgemeinen konnten die neuen Modelle die Beobachtungsdaten besser simulieren als die klassischen Modelle. Weitere Hauptresultate waren:

- Der Bias der klassischen Modelle, d.h. Überschätzung von frühen und Unterschätzung von späten Beobachtungen, konnte reduziert, aber nicht vollständig eliminiert werden.
- Die besten Modellvarianten für verschiedene Spezies wiesen darauf hin, dass für die späten Frühjahrsphasen die Tageslänge eine wichtigere Rolle spielt als für die frühen Phasen.
- Die Vernalisation spielte gegenüber den Temperaturen kurz vor dem Blattaustrieb nur eine untergeordnete Rolle.

Stichworte: kombinierte Zeitreihen, Physiologie-basierte Modelle, Unsicherheit, Variabilität, Ausreisser, Trends, Naturräume, Trendwendepunkte, *Fagus sylvatica*, *Galanthus nivalis*, *Sambucus nigra*, *Quercus robur*, *Aesculus hippocastanum*, *Betula pendula*, *Syringa vulgaris*

1 Introduction

One of the main effects of a changing climate on vegetation is the modified triggering of developmental switches in the plant, leading specifically to an altered length of the vegetation period. This thesis analyzes observed trends of plant developmental stages during the 20th century in Germany and tries to come to a better understanding of the processes that trigger plant developmental stages and their dependence on environmental conditions through modeling studies.

1.1 Introduction to phenology

Phenology (Greek: phainesthai - to appear; logos - science) is the study of the timing of recurring life cycle events and their dependence on environmental conditions. (Schnelle 1955, Lieth 1974). For plants such events are developmental stages like, e.g., bud burst, leaf unfolding, blossoming, fruit ripening and leaf coloring.

The oldest known phenological time series is the record of blossoming of cherry in Kyoto, Japan, which starts in 812 AD (Daiga and Suzuki 1947, Suzuki 1951 after Schnelle 1955). These observations served cultural activities, namely the timing of the yearly cherry blossoming festival, rather than scientific purposes. In Europe, the oldest phenological observations were recorded by the Marsham family in Stratton, England, starting in 1736 (Margary 1926, Sparks and Carey 1995). The modern systematic approach originates in 1751 when the Swedish researcher Linné described the purpose and method of phenological observations. He also established the first regional phenological network. The first international phenological network was established by the Societas Meteorologica Palatina in Mannheim, Germany, that was active from 1781 until 1792 (unknown authors 1781-1792 and later Schnelle 1955) and included several stations across Europe. In 1882, Hoffmann and Ihne established the first broader phenological observation network in Europe with precise observation instructions and recordings on a regular basis (Hoffmann and Ihne 1884, Ihne 1883 - 1941, Ihne 1892). This network continued observations until 1941. The German data from this network will be used in this thesis (Chapter 4, “Data” on page 41, Figure 4-1). In Germany, systematic phenological observations started in 1922. This official national network continues to make observations to this day and its records are the main data source of the following studies (Chapter 4, “Data” on page 41, Figure 4-1). Starting in the 1960s several phenological networks were established in Europe. One important European network is the International Phenological Gardens (IGP) initiated by Schnelle and Volkert in 1957, using clones of several tree and shrub species at about 80 stations all over Europe (Schnelle and Volkert 1957, 1964, Chmielewski 1996). The most comprehensive presentation of the history and development of phenological networks and research can be found in Schnelle (1955).

The purpose and intention of phenological networks and observations has changed over the centuries. The early phenologists of the 19th century followed the tradition of Linné and were interested in characterizing natural or rather climatic regions by defining phenological differences between landscape units and recording them in phenological plant calendars. This was regarded as a cheap supplement to meteorological measurements. After the First World War, the national networks were established to use phenological observations and climatological information predominantly in agrometeorology (Schrepfer 1922, Härle 1938, Schnelle 1955). Phenological information was used to select regions that were best suited for cultivation of certain crops and fruits. Knowledge of the timing of phenological events like fruit ripening or maturity of grains was also used to facilitate coordination in the use of harvesting machines, for instance. New fruit plantations were set up in areas where phenological observations implied low risk of night or late frost damage. Knowledge of the timing of phenological events was and still is used in plant protection, horticulture and viticulture, and for the timing of pesticide and herbicide application.

1.2 Phenology and climate change

In the 1990s phenological research saw a renaissance in the context of global change research. Global warming, caused by radiative forcing of greenhouse gases, e.g. CO_2, CH_4, (Ramaswamy et al. 2001) emitted by human activities like fossil-fuel burning, deforestation and land use change (Houghton et al. 2001), has triggered widespread consequences for nature and society. In this context phenological research focuses on certain areas affected by global change like ecology, agriculture and human health.

The potential impacts of climate change on ecosystems include climate-induced changes in species performance and competition and consequently species distribution, composition and diversity (Gitay et al. 2001). Plants and animals react differently to changing environmental conditions, e.g., temperature and precipitation, through changing growth patterns (Roumet and Roy 1996, Kirschbaum 2000, Vanninen and Mäkela 2000), reproduction (Post et al. 2001) or migration patterns (Huntley 1991). Environmental change has consequences for the competitive balance between species (Zotz 2000) potentially leading to an altered species distribution and composition in ecosystems. Examples are shown in simulation studies at the ecosystem level (Walker 1994, Lasch and Lindner 1995, Bugmann and Solomon 2000) as well as at the global scale (Prentice et al. 1992, 1993, Sykes et al. 1996, Cramer et al. 1999). Even though species-specific process-based models for phenology are generally not incorporated in plant biogeography models, phenology, determining the length of the vegetation period, is one of the major determinants of interannual variability in plant production (Kramer 1995, Goulden et al. 1996, Kramer et al. 1996, Chen et al. 1999, White et al. 1999, Kramer et al. 2000) and reproduction (Chuine and Beaubien 2001) in natural and agroecosystems. In recent years the awareness of the central role of phenology for the interannual variation of gross and net carbon exchange of terrestrial ecosystems and related species performance and distribution issues have led to attempts at improving the description of phenology in ecosystem and global vegetation models (Foley et al. 2000, Kucharik et al. 2000).

Vegetation phenology is an important factor influencing atmospheric dynamics in the lower boundary layer because the drag and Bowen ratio differ considerably between leaved and non-leaved vegetation states (Schwartz and Karl 1990, Schwartz 1992, Fitzgerald et al. 2001). The drag and Bowen ratio play important roles for the assessment of vegetation-atmosphere feedbacks (Foley et al. 2000, Claussen et al. 2001). A better understanding of the variation of surface roughness, albedo and Bowen ratio with the density of photosynthesizing tissues in the vegetation cover leads to an increasing interest in a better description of the lower boundary layer conditions for weather forecast and climate models and therefore an explicit representation of plant phenology (Ciret et al. 1999, Kucharik et al. 2000, Bergengren 2001).

Improved insight in possible phenological changes also contributes to a better understanding of other ecosystem processes like climate-induced changes in nutrient cycles, water fluxes, and carbon sequestration at the global scale (Kindermann et al. 1996, Lloyd 1999, Jackson 2000, Schimel et al. 2001, Thornley 2001) as well as on the regional and local scale (Lindroth et al 1998, Chen et al. 1999, White et al. 1999, Black et al. 2000, Wilson et al. 2000). The importance of phenology for such ecosystem processes is due to its central role for the timing of the physiologically active and non-active states of the vegetation, especially for deciduous plants.

Phenology is also a major determinant in the timing of biotic interactions. The demography of grazers, browsers and predators varies with their own developmental responses to climatic condition as well as with the changes in forage and prey availability due to plant phenology (Gitay et al. 2001)

Phenological issues in agricultural research cover, e.g., the interaction between crops and pests and diseases, growing season length, frost damage and food quality (Parry et al. 1990, Walker 1991, Goudriaan and Zadoks 1995). The timing of phenological phases plays an important role in determining the allocation of assimilates to reproductive versus vegetative growth and above-ground versus below-ground biomass (Mitchel et al. 1996). Modeling the developmental stages of crops is one of the core elements in modeling production and yield as well as in models for the control or management of pests and diseases. There is a large amount of literature proving the importance of phenology for agroecosystems and giving a detailed overview is beyond the scope of this thesis (Goudriaan and van Laar 1994).

Phenological research in public and human health focuses on forecasting allergic pollen that cause hay-fever (Chuine 1999, Chuine et al. 2000) or vector-borne diseases like malaria (Walker 1991). In Europe, a pollen monitoring network exists that comprises about 300 stations. The prediction of mass proliferation of pollen and early warning for allergic persons requires the best achievable understanding of weather sensitivity of flowering phenology and models that build on this knowledge.

Several spatial levels of phenolgical observations play a role in climate change related phenology research. On the first level there are observations of individuals of plants or animals that comprise the above mentioned classical phenology. On this level various recent reports documented the influence of climate and weather on phenology of plants (Bradley et al. 1999, Menzel and Fabian 1999, Schwartz and Reiter 2000, Menzel 2000, Sparks et al. 2000, Chmielewski and Rötzer 2001, 2002) and animals like birds (Crick and Sparks 1999, Sparks 1999), fish (Ahas 1999) and insects (Parmesan 1999, Roy and Sparks 2000). On this level most studies reported a lengthening of the vegetation period mainly because of advanced spring phases. A trend to later egg-laying and shifts in bird and insect migrations patterns were also observed in the last decades. On the second level of phenological observations there are observations of interannual and intraannual mass and energy fluxes between whole ecosystems like grasslands (Schimel et al. 1991) and forests (Goulden et al. 1996, Chen et al. 1999, White et al. 1999) and the atmosphere. On this level observations mostly cover only a few years but there is substantial evidence that the length of the vegetation period plays a crucial role in determining the annual integral of the fluxes (Goulden et al. 1996, Chen et al. 1999, White et al. 1999). On the third, global, level, remote sensing studies strengthened evidence that the length of the vegetation period has increased in northern latitudes in the last decades (Myneni et al. 1997). These findings were also implied by seasonal atmospheric CO_2 measurements from which Keeling et al. (1996) inferred increased plant activity and a prolonged vegetation period in the northern latitudes. This resulted in various studies on the effect of climate change and the change in growing season length on growth and functioning of ecosystems (as documented above), specifically forest ecosystems. Forests are of special interest to climate change research because they contain the largest stock of carbon in terrestrial ecosystems and play a dominant role in land cover-atmosphere feedbacks.

1.3 Phenology and forests

Forest scientists have become increasingly interested in phenological research in recent years (Lechowicz and Koike 1995). The timing of phenological phases is an important factor in analyzing changes in net production of trees in response to interannual variation and long-term changes in climate (Lechowicz 1984, Goulden et al. 1996, Chen et al. 1999, White et al. 1999, Kramer et al. 1996, 2000, Chuine and Beaubien 2001). Kramer et al. (1996) showed that a difference of one week in bud break corresponds to a difference of about 5% in the radiation absorbed during the growing season. Therefore, the timing of phenological phases is an important factor in determining variations in net production of trees on interannual and climatic change related time scales. Tree species react differentially to climate change not only with respect to changed growth patterns but also through frost damage (Hänninen 1991, Murray et al. 1994, Kramer et al. 1996, Repo et al. 1996), resulting in altered competitive relationships between tree species (Kramer 1994, 1995). Models for the phenological development of trees were developed to assess the impact of climate change on forest ecosystems (Cannell and Smith 1983, Hunter and Lechowicz 1992, Hänninen 1994, Leinonen 1996, Leinonen et al. 1997, Menzel 1997, Chuine 2000).

1.4 Phenological modeling

For most fields where phenological models are of importance like dynamic global vegetation models, general circulation models, forest growth models, crop models and pollen forecast models, a daily time step of phenological plant development is appropriate. A single day also is the smallest temporal resolution of most available phenological observations. Simulation studies (Kramer et al. 1996) and experiments (Chen et al. 1999) showed that a one day change in phenological plant development can make a significant difference in terms of annually absorbed radiation and thus growth performance. Therefore, this thesis is restricted to the consideration of phenological models with this temporal resolution.

There is no model for plant phenological events that is equally well applicable to both the most important species and a wide geographic range. The most obvious reason for this is the lack of data, not only phenological data but also related weather information. Both types of information are needed to derive the relationship between them which is subsequently implemented into models. The deeper reason for shortcomings of current phenological models is the simple fact that the physiological processes within the plant that control its phenological status remain largely unknown (Hanover 1980, Lavender and Silim 1987, Powell 1987). Therefore most models describe empirical relationships between phenological observations and weather. For individual species on a local scale such models perform well. However, their performance drastically decreases when the models are applied to other conditions than the ones they were fitted to because the model parameters are site-specific. It is desirable to have a phenological model that is applicable to a wide range of conditions, i.e. a model whose parameters depend only on the species and not on the specific environmental conditions to which the model is fitted. This is especially useful for climate impact studies where conditions are prescribed to change. To develop such models it is reasonable to assume that species-specific parameters are obtained when the model concentrates on hypothesized species-specific processes that lead to phenological events rather than on simple empirical relationships between phenology and weather. There have been attempts to introduce some more physiology-based processes into phenological models (Hänninen 1994). Such attempts are promising but did not obtain the desired result of a wide applicability (Kramer 1994), implying that the nature of such processes has not been fully understood yet.

1.5 Motivation, aims and structure of the thesis

This thesis is divided into parts. In the first part, a methodology is introduced and further developed that allows phenological data to be processed to make them more consistent and reliable. In the second part, observed trends of phenology in the 20th century in Germany are analyzed and relationships with environmental conditions are derived. In the third part, more detailed phenological models are developed, that try to gain new insights into the physiological processes that trigger plant phenology.

Only a few studies about phenology cover more than a decade (Gitay et al. 2001). Longer time series are only available for ground phenological observations but even studies using ground observations rarely cover more than 30 years. For climate change related time scales this is rather a short period. Schönwiese and Rapp (1997) pointed out that for regional and seasonal temperature data trends can become random when time series of less than 30 years are analyzed. Menzel et al. (2001) showed that the trend of phenological time series is also not independent of the time period analyzed. This is especially critical for single phenological time series that are not exact measurements but rather observations that are prone to various kinds of phenology-specific uncertainties. Single phenological time series are mostly incomplete and contain data gaps. They exhibit an intrinsic variability caused by observational uncertainties due to unclear observation instructions, genetic variability and other sources of uncertainty that are difficult to quantify (see Chapter 3). Some authors have tried to overcome these shortcomings by analyzing average time series (Chmielewski and Rötzer 2001, 2002). However, the data gaps in single station time series can cause a bias in the calculated trends of the average time series only because of the averaging procedure (see Chapter 2; “Results” on page 18).

Therefore, when causes and patterns of plant phenological reactions to climate are to be analyzed, two conditions should be fulfilled. First, a sound methodology for the analysis of phenological ground observations that can cope with their intrinsic uncertainties should be applied, and secondly, time series of more than 30 years should be used.

As a basis for the following analyses, in Chapter 2 a method proposed by Häkkinen et al. (1995) is further developed to obtain reliable phenological time series. The influence of uncertain single station time series is reduced and at the same time the shortcomings of averaging are eliminated.

In Chapter 3 the intrinsic uncertainty and variability of phenological data and their causes are analysed. An attempt is made to roughly quantify the different sources of uncertainty. These analyses are helpful to refine the described methodologies in the future. Moreover, quantifying phenological uncertainty is a first step in assessing the quality of models of phenology that can be achieved facilitating model development.

Based on the developed methods, Chapter 4 analyses phenological time series including new data extending the time frame of earlier studies to over 100 years. The specific questions analyzed in this chapter are:

- What are the patterns of plant phenological development in the course of the last century in Germany?
- What are the main causes and determinants of the observed phenological patterns?
- How large is the influence of climatic change on phenology at the regional scale?
- Are the conclusions drawn from former studies of phenological time series that covered the last four to five decades still valid over a whole century?
- Can the random influence of varying observational periods be reduced?

As elaborated above, models for plant phenological events are needed to project future climate change impacts on vegetation and specifically on trees using climate and weather scenarios. When the impact of a changing climate on ecosystems is assessed with global vegetation models, climate models or, more specifically, forest growth models, a robust and reliable submodel of tree phenology is mandatory. Moreover, to improve understanding of the important factors that control plant phenology new hypotheses regarding the underlying physiological processes have to be formulated and tested.

Therefore, the purposes of the modeling studies in this thesis were twofold: firstly, to gain new insights into the underlying processes that determine plant phenology by setting up a new physiology-based modeling framework, and secondly, to develop an adequate and adjusted model of phenology that is applicable for climate change impact studies for forest ecosystems all over Germany. A new model based on physiological concepts was developed in Chapter 5 and compared to other classical models, fitting them to the largest data set ever used for phenological model development.

The results of the three main parts of this thesis;

- development of an adequate methodology for obtaining reliable phenological time series for further analysis and processing (Chapter 2),
- phenological data analysis (Chapter 3 and Chapter 4) and
- phenological modeling (Chapter 5)

are discussed in each chapter separately while an overall summary and conclusions on the achievements of the present thesis are given in Chapter 6.

2 Evaluation of methods for the combination of phenological times series and outlier detection

2.1 Introduction

As indicated in Chapter 1, when working with phenological data some specific limitations of their use have to be considered. It must be realized that phenological data origin from observations rather than from exact measurements. To obtain the data phenological observers use observation instructions that leave room for interpretation. Additionally, the exact location of the observation and therefore the environmental conditions as well as the plant individuals are mostly unknown. These various sources of uncertainty introduce an intrinsic variability to phenological observations that is difficult to quantify. In Chapter 3 the various sources that add to the uncertainty of phenological data are discussed in detail and a first attempt is undertaken to quantify them. Moreover, phenological time series are often incomplete and reveal large data gaps, further complicating their analysis.

In this chapter methods proposed by Häkkinen et al. (1995) to cope with the uncertainty of single station time series are evaluated and further developed. The main idea behind those methods is to use phenological observations from adjacent stations to supplement the phenological information that is available at a single stations to fill data gaps and reduce the uncertainty of a single station series. Phenological information from several stations is combined to obtain a common but more reliable information.

There are several areas of application where methods for combining phenological time series can be useful and where they have already been applied (Häkkinen et al. 1995, Linkosalo et al. 1996, 2000, Linkosalo 2000). One main application is to construct a combined phenological time series to obtain reliable long time series out of many messy time series for trend analysis, for instance. In this application data gap filling is of primary interest. We will apply combined time series for the generation of reliable long time series for trend analyses in Chapter 4.

When phenological times series are analyzed in relation to weather, for model development for instance (see Chapter 5), the problem is that phenological observation are rarely made in conjunction with weather measurements. It is rather the case that to relate phenological events to weather, weather data from adjacent stations has to be used. It is clear that the microclimate at the location of the weather station can expected to be systematically different from the microclimate of the location of the phenological observation due to a variety of factors such as exposition, vegetation cover, slope, aspect, etc. (see Chapter 3). As a consequence the observed phenological events are not directly related to the corresponding weather measurements of the station nearby. In the case, however, that the phenological observations from several stations around the weather station are available (Fig. 2-1), overlap and certain assumption are imposed (see Chapter 5, “Models, data and methods” on page 76), a combined phenological time series can be constructed that can be assumed to be more closely related to the corresponding weather measurements than the individual time series are.

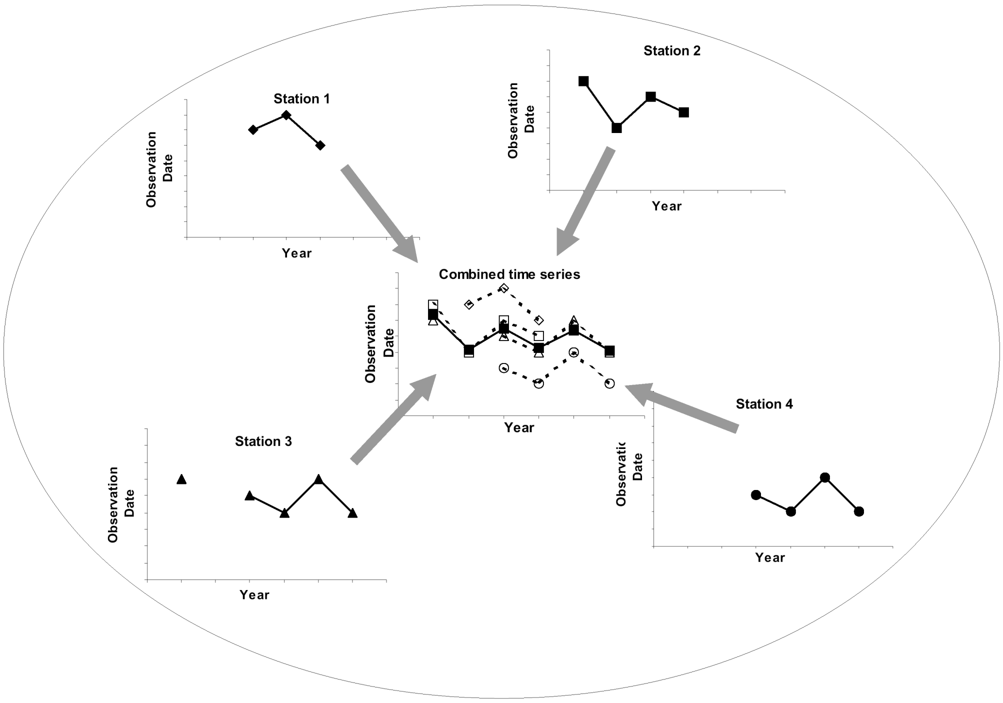


Figure 2-1. Merging several single station time series into one continuous combined time series

Additionally, combined time series can also be used to find outliers in individual time series (Linkosalo et al. 2000).

Häkkinen et al. (1995) proposed four methods to combine phenological time series but did not rank the best methods because they provided no comparable measure of performance.

The objective of this chapter was to find answers to the following questions:

- Are there theoretical reasons for selecting a certain method of combining fragmentary phenological data into one continuous time series in order to increase the reliability of the phenological information for subsequent analysis?
- Is there a particular technique that produces the most reliable estimates of the parameters of the combined phenological times series?
- What is the influence of different outlier detection methods on the resulting combined time series and what is an adequate way to detect and treat outliers?

The answer to the first question was sought by embedding all the methods proposed by Häkkinen et al. (1995) in a common theoretical framework that elucidates the underlying statistical assumptions of the methods. The second and third question were examined by applying several estimation techniques and two outlier detection methods in a Monte-Carlo simulation and to real data.

2.2 Methods and Data

2.2.1 A common theoretical background

Häkkinen et al. (1995) proposed four methods of combining phenological time series, in the following called method 1, 2, 3 and 4. We start by analyzing method 3, that was also the method used in later papers (Linkosalo et al. 1996, 2000).

The combined time series y_i, $i = 1,\ldots,M$ of method 3 was defined by Häkkinen et al. (1995) as

$$y_i = \frac{1}{n_i}\sum_{j=1}^{n_i}(x_{ij} - b_j),$$

Equation 2-1

The y_i can be interpreted as yearly averages of the observations x_{ij} of year i at station j, adjusted by b_j. The $n_i \le N$ is the number of observations, i.e. stations with observations, in year i. N is the total number of considered stations. Some x_{ij} are not defined because no observations were made in year i at station j. The b_j were found by an iterative optimization algorithm that minimizes

$$S = \sum_i\sum_j(x_{ij} - b_j - y_i)^2 \quad .$$

Equation 2-2

Defining n_{ij} as 1 if x_{ij} exists and 0 otherwise, we can redefine Eqn. 2-1 and Eqn. 2-2 as

$$y_i = \frac{1}{n_i}\sum_{j=1}^{N} n_{ij}(x_{ij} - b_j) = \overline{x_{i\circ}} - \frac{1}{n_{i\circ}}\sum_{j=1}^{N} n_{ij}b_j$$

Equation 2-3

and

$$S = \sum_{i=1}^{M}\sum_{j=1}^{N}(n_{ij}(x_{ij}-b_j-y_i))^2,$$ **Equation 2-4**

with $n_{i\circ} = \sum_{j=1}^{N} n_{ij}$ and $\overline{x_{i\circ}} = \frac{1}{n_{i\circ}}\sum_{j=1}^{N} n_{ij}x_{ij}$.

The minimum of *S* is found by differentiating *S* with respect to each of the parameters b_j and y_i and setting the resulting derivatives equal to zero. This results in Eqn. 2-3 and

$$b_j = \overline{x_{\circ j}} - \frac{1}{n_{\circ j}}\sum_{i=1}^{M} n_{ij}y_i$$ **Equation 2-5**

with $n_{\circ j} = \sum_{i=1}^{M} n_{ij}$ and $\overline{x_{\circ j}} = \frac{1}{n_{\circ j}}\sum_{i=1}^{M} n_{ij}x_{ij}$.

Inserting Eqn. 2-3 into Eqn. 2-5 and further impose that $\sum b_j = 0$ to obtain a unique solution we arrive, after several rearrangements, at linear equations for the b_j,

Equation 2-6

$$\left(n_{\circ j} + 1 - \sum_{i=1}^{M}\frac{n_{ij}}{n_{i\circ}}\right)b_j + \sum_{k \neq j}^{N}\left(1 - \left(\sum_{i=1}^{M}\frac{n_{ij}n_{ik}}{n_{i\circ}}\right)\right)b_k = x_{\circ j} - \sum_{i=1}^{M} n_{ij}\overline{x_{i\circ}} \quad for\; j,\, k = 1,...,N,$$

with $x_{\circ j} = \sum_{i=1}^{M} n_{ij}x_{ij}$.

Thus, the solution of Eqn. 2-4 with respect to $y = (y_1,...,y_M)$ and $b = (b_1,\ldots,b_N)$ is Eqn. 2-3 and

$$b = C^{-1}r$$ **Equation 2-7**

where $C=\{c_{jk}\}$ and $r=\{r_j\}$ for $j, k = 1, \ldots, N$ with

$$c_{jj} = n_{\circ j} + 1 - \sum_{i=1}^{M}\frac{n_{ij}}{n_{i\circ}} \text{ and } c_{jk} = 1 - \left(\sum_{i=1}^{M}\frac{n_{ij}n_{ik}}{n_{i\circ}}\right) \text{ for } k \neq j \text{ and } r_j = x_{\circ j} - \sum_{i=1}^{M} n_{ij}\overline{x_{i\circ}} \quad .$$

It can be shown (Searle 1971, Rencher 2000) that the same solutions Eqn. 2-3 and Eqn. 2-7 are obtained when we formulate the problem of finding y_i and b_j in Eqn. 2-4 as a linear model of a two-way crossed classification with fixed effects,

$$x_{ij} = m + a_i + b_j + e_{ij},$$ **Equation 2-8**

setting $y_i = m + a_i$ and apply the method of least squares to estimate the parameters, where *m* is a general mean, a_i is the effect of year *i*, *i=1,…,M* (accordingly called year effect in the following) and b_j is the effect of station *j*, *j=1,…,N* (called station effect in the following). The e_{ij} are independent identically distributed random errors with assumed expectancy $E(e_{ij}) = 0$ and common variance σ_e^2. To find a unique solution *m* is set to zero and it is assumed that $\sum b_j = 0$.

Thus, mathematically equivalent solutions are obtained when the problem of finding a combined phenological time series is formulated as in Eqn. 2-1 and Eqn. 2-2 or as a two-way crossed linear model with fixed effects (Eqn. 2-8).

Häkkinen et al. (1995) defined the combined time series according to method 1 as $y_i = \overline{x_{i\circ}}$, i.e. simply taking the average observation each year. We obtain this solution when we define our linear model as

$$x_{ij} = m + a_i + e_{ij}, \qquad \textbf{Equation 2-9}$$

i.e. a one-way linear model with fixed effects, and set $y_i = m + a_i$. The e_{ij} are independent identically distributed random errors with assumed expectancy $E(e_{ij}) = 0$ and common variance σ_e^2. To find a unique solution *m* is set to zero.

The combined times series according to Häkkinen et al.'s method 2 can also be defined within the framework of linear models. But this is not further considered here because it requires a long reference time series that is only rarely available.

Method 4 proposed by Häkkinen et al. (1995) formulates Eqn. 2-8 as a two-way crossed linear mixed model of randomized block design:

$$x_{ij} = m + a_i + b_j + e_{ij}, \qquad \textbf{Equation 2-10}$$

where b_j is the random block (station) effect, *m* is a constant and a_i is the fixed (year) effect. The random terms b_j and e_{ij} are by assumption independent identically distributed with expected zero mean. Moreover b_j is assumed to have the common variance σ_s^2 and e_{ij} have the common variance σ_e^2. The combined time series is again defined as $y_i = m + a_i$.

2.2.2 Parameter estimation

The error variance σ_e^2 is an important measure of both the reliability of the underlying data and the resulting combined time series because its magnitude is closely related to the range of the confidence intervals for the estimated parameters that were also used by Häkkinen et al. (1995) and Linkosalo et al. (1996) to estimate reliability.

The parameters *m*, a_i and b_j and the variance component σ_e^2 and σ_s^2 for the mixed model (Eqn. 2-10) can be estimated by various procedures that yield different results when data are missing, i.e. the data are unbalanced, and they also have different statistical properties (Searle 1987, Milliken and Johnson 1992). Here we explored the differences of the techniques of the Restricted Maximum Likelihood (REML) (Patterson and Thompson 1971, Corbeil and Searle 1976), Maximum Likelihood (ML) (Hartely and Rao 1967, Hemmerle and Hartley 1973), MIVQUE0 (Hartley et al. 1978, Searle 1987) and TYPE I or Henderson III (T1) (Searle 1971, Milliken and Johnson 1992). REML, ML and MIVQUE0 estimates were computed with the modules available in the SAS procedure MIXED (SAS Software Release 6.12). TYPE I was implemented following Thompson (1969) and Searle (1971) and the multiple linear least square regression (LS) for the fixed effect models (Eqn. 2-8) and (Eqn. 2-9) were calculated with the SAS procedure REG (SAS Software Release 6.12).

2.2.3 Robust estimation

The estimation techniques for linear models all use squared residuals to estimate the parameters and variance components. These techniques are susceptible to outliers (Dodge 1987, Rousseeuw and Leroy 1987, Barnett and Lewis 1996, Hubert and Rousseeuw 1997, Hubert 1997) because they emphasize extreme values. A variety of robust estimators have been proposed to identify and/or accommodate such outliers (Rousseeuw and Leroy 1987, Barnett and Lewis 1996). Hubert (1997) showed that, in the case of binary (dummy) regression variables, the L_1 regression based on minimizing the sum of absolute residuals $S = \sum |e_{ij}|$ rather than squared residuals has an optimal breakdown value although it is not robust in the presence of leverage points, i.e. outliers in the regressors (Hubert and Rousseeuw 1997). In linear models of designed experiments as in our case, however, leverage points do not occur because the regressors are not measured values but dummy regression variables that are prescribed by the design of the model. To detect outliers by robust regression, L_1 estimation of the linear model parameters was conducted using the algorithm of Barrodale and Roberts (1973, 1974). L_1 estimation was applied to the sum of absolute residuals of the two-way linear model with fixed effects (Eqn. 2-8). Calculation of the error variance component σ_e^2 of the L_1 estimation was done with the estimator proposed by McKean and Schrader (1987a, 1987b), where

$$\sigma_e = \sqrt{n}(\tilde{e}_{(n-k+1)} - \tilde{e}_{(k)})/(2z) \text{ and } k = (n+1)/2 - z\sqrt{n/4}. \qquad \textbf{Equation 2-11}$$

Here n is the number of non-zero residuals, the $\tilde{e}_{(i)}$, $i=1,...,n$ is the ordered set of non-zero residuals and z is an upper tailed standard normal critical value, here $z = 1.96$ after McKean and Schrader (1987a, 1987b).

2.2.4 Outlier detection

Biological variability of plant individuals, differences in microclimate, observational and protocolling errors add to the natural variability of phenological data that typically amounts to about one to two weeks (Baumgartner 1952, Schnelle 1955). A further attempt to quantify the natural variability of phenological data is made in Chapter 3. Observational errors result from individual interpretations of the observer's instruction, for instance. One mistake that can be detected despite the natural variability of phenological data is the month-mistake (MM) because of its strong deviation. A MM is a protocolling error that occurs when the observation date is noted in a wrong column or the observer uses the wrong column when transforming the actual date to the Julian date (day of year) using a conversion table or when transcribing the phenological information to a database (Schnelle 1955, Menzel 1997, Vasella 1997). Such events result in an observation date that is one or several months too early or too late. Most other mistakes are very difficult to detect without having access to the original observation reports or knowledge about the customs of the observers because they vary within the natural variability of phenological observations. Based on the above considerations and the fact that the most straight forward method of detecting outliers in linear models is by considering residuals (Barnett and Lewis 1996), those observation were considered as outliers whose estimated residuals of the linear models where larger than or equal to 30 days, i.e. where $|e_{ij}| \geq 30$. This method of detecting outliers is called the 30-day rule in the following.

The 30-day rule was compared with a classical method that was also applied by Linkosalo et al. (1996, 2000) proposed by Dixon (1950) and modified by King (1953). An extreme value of the ordered observations ($x_{(i,1)},\ldots,x_{(i,n)}$) of year i adjusted for the station effects b_j is considered an outlier when its test statistic T_i exceeds a critical value, where

$$T_i = max\left[\frac{x_{(i,n)} - x_{(i,n-1)}}{x_{(i,n)} - x_{(i,1)}}, \frac{x_{(i,2)} - x_{(i,1)}}{x_{(i,n)} - x_{(i,1)}}\right] .$$

In the literature only critical values for the 1% and 5% significance levels were found (Barnett and Lewis 1996).

2.2.5 Monte-Carlo analysis

The consequences of the different approaches for the estimation of the combined time series and outlier detection methods concerning the characteristic parameters were explored in a Monte-Carlo simulation study. Observations x_{ij} (in day of year DOY) were randomly generated according to a two-way linear model (Eqn. 2-8 and Eqn. 2-10) with prescribed parameters m, a_i, b_j, σ_s^2 and σ_e^2. The general mean m was set to DOY 120, the a_i were prescribed to vary according to a normal distribution with zero mean and a standard deviation of seven days. This reproduced the typical range of natural interannual variation of phenological time series. The other parameters depend on the Monte-Carlo simulation and are described below. Because phenological data are usually unbalanced 50% of the generated data were omitted. This is a representative value as the completeness of real data demonstrates (Tab. 2-5). To simulate outliers 1% of the resulting observations were made MMs by adding or subtracting 30 days (fifty-fifty chance) from the generated observations. This is the approximate amount of outliers that was found by Linkosalo et al. (1996, 2000).

Two simulation studies were conducted. In study one that resembles the case where a combined time series is sought for several stations around a weather station, we imposed variances similar to those found in the real data (Tab. 2-5). In study two that mimics a situation where many phenological stations within a region are combined for trend analysis, the variance components as well as the number of stations were doubled compared to those of study one. The main parameters of the two studies are summarized in Tab. 2-1.

Table 2-1. Setup of the Monte-Carlo simulations for comparing different methods of estimating effects and variance components of the linear models Eqn. 2-8, Eqn. 2-9 and Eqn. 2-10 with different estimation and outlier detection techniques. The observations x_{ij}, $i=1...M, j=1...N$ are generated according to the model $x_{ij} = 120 + a_i + b_j + e_{ij}$, where the a_i, b_j and e_{ij} are prescribed to vary according to a normal distribution N(x,y) with mean x and variance y. Per study the setup and subsequent estimations are repeated 500 times

Study	Average Completeness	Year effects a_i	Station effects b_j	Residual error e_{ij}	% Month mistakes	Number of stations (N)	Number of years (M)
Study 1	50%	N(0,49)	N(0,15)	N(0,30)	1	10	40
Study 2	50%	N(0,49)	N(0,30)	N(0,60)	1	20	40

For each study five runs were conducted. In the first run (a) the different estimation methods were applied to the generated data set without imposing MMs. In the second run (b) MMs were introduced but they were not detected and corrected. Finally, in the last three simulation runs (c, d and e) the generated time series were also corrected for the detected MMs by the 30-day rule and the Dixon test using 5% and 1% significance levels, respectively. Each of these ten simulation runs, in the following labeled as 1a, 1b,…,1e and 2a, 2b,…,2e according to study number and run, was repeated 500 times.

2.2.6 Phenological data

To analyze the effect of the different discordancy tests on real combined time series we applied them to data of the German Weather Service (DWD). Phenological times series that were recorded within in a radius of 10 km around a weather station and differed less than 50 m in elevation and had at least five years of observations were selected from the phenological and climatological database of the DWD and combined for each weather station. At these stations we used data for bud burst of four deciduous tree species, *Aesculus hippocastanum* L., *Fagus sylvatica* L., *Quercus robur* L. and *Betula pendula* Roth. Based on results from the Monte-Carlo simulation study the robust L_1 estimation was used for the outlier tests. After the removal of detected outliers we applied the T1 estimation procedure to the remaining data for variance component estimation.

Additionally the effect of the different estimation procedures on the estimated combined time series itself was studied using one example from the above derived combined time series. Nine phenological stations around weather station 2609 (Giessen, 8.7° N, 50.58° E, altitude 186 m, 1951-1999) were selected. The distance and altitude of the phenological stations and some characteristics are displayed in Tab. 2-2 and the graphs of the times series is shown in Fig. 2-2. The considered phase was chosen to be bud burst of horse chestnut (*Aesculus hippocastanum* L.)

Table 2-2. Phenological stations from the DWD phenological data base with a distance of less than 10 km and a difference in elevation of less than 50 m from the DWD weather station 2606 (Giessen, 8.7° N, 50.58° E, altitude 186 m, 1951-1999) that have at least 5 observations of bud burst of *A. hippocastanum*.

Station	Distance [km]	Altitude [m]	Observational time span	Number of observations
Station 1	2.12	160	1951 - 1983	31
Station 2	6.04	190	1951 - 1973	18
Station 3	4.94	160	1958 - 1991	33
Station 4	7.8	200	1951 - 1999	44
Station 5	7.15	190	1951 - 1978	28
Station 6	6.62	200	1987 - 1998	12
Station 7	9.79	200	1988 - 1995	7
Station 8	9.99	180	1952 - 1998	44
Station 9	5.98	180	1962 -1980	18

.

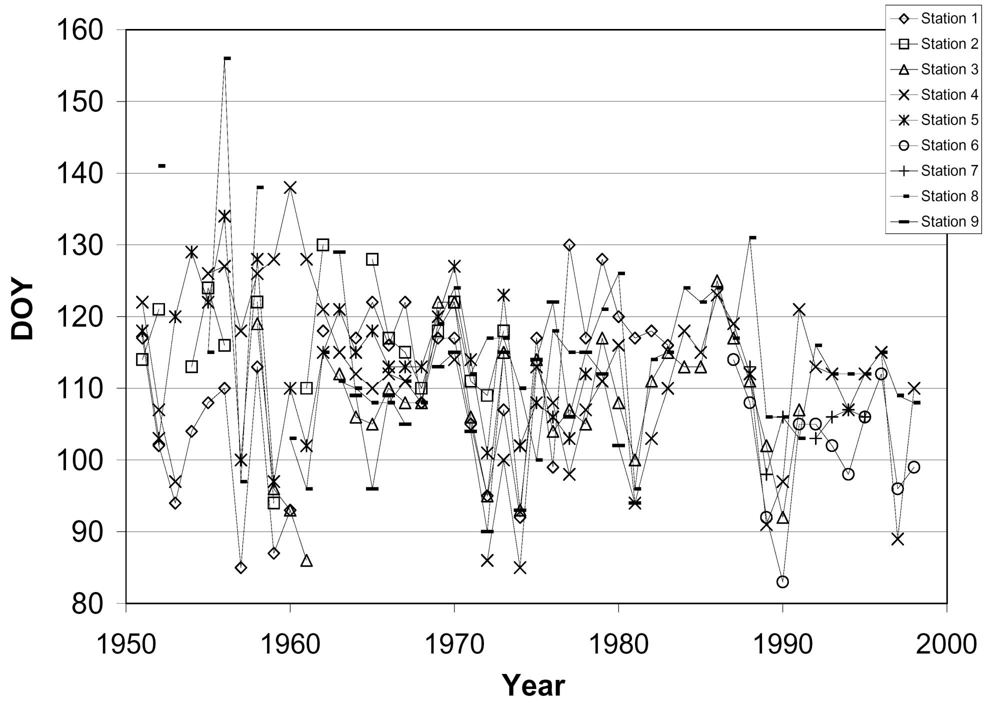


Figure 2-2. Phenological time series around weather station 2609 of the DWD in a radius of 10 km as described in Tab. 2-2. Observations in Julian day of year (DOY)

2.3 Results

2.3.1 Monte-Carlo analysis

For the Monte-Carlo analysis all effects/parameters and variance components had been prescribed and it could easily be checked whether the models and estimation methods reproduce those values for a large number of simulations. In Tab. 2-3 the mean absolute error MAE, the mean squared error MSE, the estimated residual variance σ_e^2 according to the models Eqn. 2-8, Eqn. 2-9 and Eqn. 2-10 and the mean deviation between the estimated and the imposed combined time series MDTS are displayed for the different simulation runs described in Tab. 2-1. MAE, MSE, MDTS and σ_e^2 are averaged over 500 simulation runs. MDTS is shown with its standard deviation σ. The differences among the methods concerning σ_s^2 are qualitatively the same as for σ_e^2 and are not shown. The robust L_1 estimation always had the lowest MAE, whereas the non-robust two-way model estimations performed better concerning the MSE. The one-way model estimation (the average times series) had the poorest performance concerning MAE and MSE. There were only small differences in terms of MAE and MSE between the different non-robust two-way models. The LS estimation for the fixed two-way model had slightly lower MSE in all scenarios. The estimation procedures showed, however, differences in their variance component estimation. It is especially noteworthy that the variance components estimated with the ML method were always lower than those estimated with the other methods. When no MMs were prescribed (simulation runs 1a and 2a) the variances determined with the ML method were clearly underestimated whereas results obtained with the other non-robust methods showed estimated variance components close to the imposed ones. LS, REML, MIVQUE0 and T1 agreed well in their variance components estimates. The MIVQUE0 variance estimates always had a higher variability (results not shown). Thus LS, REML and T1 seem to have the most reliable variance component estimates. The error variance of the L_1 estimation calculated by the method of McKean and Schrader strongly overestimated the imposed variances. The error variances varied between 15(16) and 45(72) in simulation run 1a(1b) and 32(45) and 74(93) in simulation run 2a(2b) for the non-robust two-way model estimations. In terms of the estimated combined time series (MDTS) the models did not differ a lot. 99% of the estimates for the combined time series did not differ more than about 10 days from the imposed values. The robust estimation L_1 was slightly less accurate than the other estimation methods. The fixed (Eqn. 2-8) and mixed (Eqn. 2-10) two-way models and their respective estimation methods all showed the same accuracy.

.

Table 2-3. Summary of the results of the Monte-Carlo simulations described in Tab. 2-1. The values are averages over 500 runs. MAE: mean absolute error, MSE: mean squared error, σ_e^2 : estimated error variance according to (8), (9), (10), MDTS: average deviation between estimated and prescribed combined time series with standard deviation, MM: number of month-mistakes over all 500 simulation runs.

Monte-Carlo simulation		**Model Eqn.**	**Estimation method**	**MAE**	**MSE**	σ_e^2	**MDTS± σ**
Study 1	Run a) No month-mistakes	Eqn. 2-8	L_1	3.5	26.8	88.3	-0.3±3.6
		Eqn. 2-9	LS 1-way	4.8	35.9	44.8	0.1±3.4
		Eqn. 2-8	LS-2-way	3.8	22.4	30.0	0.0±3.1
		Eqn. 2-10	ML	3.8	22.6	23.9	0.0±3.1
		Eqn. 2-10	REML	3.8	22.7	30.0	0.0±3.1
		Eqn. 2-10	MIVQUE0	3.8	22.7	30.0	0.0±3.1
		Eqn. 2-10	T1	3.8	22.7	30.0	0.0±3.1
	Run b) 736 MMs Not corrected for MMs	Eqn. 2-8	L_1	3.7	34.4	90.9	-0.2±3.7
		Eqn. 2-9	LS 1-way	5.1	43.9	54.7	0.1±3.7
		Eqn. 2-8	LS-2-way	4.1	28.8	38.7	0.1±3.4
		Eqn. 2-10	ML	4.1	29.2	30.8	0.1±3.4
		Eqn. 2-10	REML	4.1	29.4	38.8	0.1±3.4
		Eqn. 2-10	MIVQUE0	4.1	29.3	38.8	0.1±3.4
		Eqn. 2-10	T1	4.1	29.3	38.7	0.1±3.4
Study 2	Run a) No month-mistakes	Eqn. 2-8	L_1	5.4	56.0	178.9	-0.2±3.4
		Eqn. 2-9	LS 1-way	7.2	80.9	89.8	0.0±3.1
		Eqn. 2-8	LS-2-way	5.7	50.4	60.1	0.0±2.9
		Eqn. 2-10	ML	5.7	50.9	53.9	0.0±2.8
		Eqn. 2-10	REML	5.7	51.0	60.1	0.0±2.8
		Eqn. 2-10	MIVQUE0	5.7	51.0	60.2	0.0±2.8
		Eqn. 2-10	T1	5.7	51.0	60.1	0.0±2.8
	Run b) 1535 MMs Not corrected for MMs	Eqn. 2-8	L_1	5.6	64.3	187.2	-0.1±3.5
		Eqn. 2-9	LS 1-way	7.4	89.7	99.6	0.0±3.3
		Eqn. 2-8	LS-2-way	5.9	58.1	69.2	0.0±3.1
		Eqn. 2-10	ML	6.0	58.7	62.1	0.0±3.0
		Eqn. 2-10	REML	6.0	58.8	69.2	0.0±3.0
		Eqn. 2-10	MIVQUE0	6.0	58.8	69.1	0.0±3.0
		Eqn. 2-10	T1	6.0	58.8	69.2	0.0±3.0

The strategy for outlier detection had a substantial impact on the number of outliers detected. In Tab. 2-4 the results for the robust L_1 method, LS 2-way and T1 estimation are displayed. The results for the LS 1-way method are not mentioned because it proved to be inferior in terms of estimation accuracy (see above). All mixed model estimation methods performed similar in terms of outlier detection, thus, only results for the T1 methods are shown. In all simulation runs the highest number of outliers were detected after the robust L_1 estimation but the percentage of detected prescribed MMs was higher for the non-robust methods LS 2-way and T1.

Table 2-4. Summary of the results of the Monte-Carlo simulations described in Tab. 2-1. The values are averages over 500 runs. σ_e^2 : estimated error variance according to Eqn. 2-8, Eqn. 2-10, Detected outliers: in% of imposed month-mistakes (MMs), MM: detected imposed MMs in% of detected outliers.

	Monte-Carlo simulation	**Model Eqn.**	**Estimation method**	σ_e^2	**Detected outliers (% of imposed)**	**MM (% of detected)**
Study 1	Run c) 30-day residuals	Eqn. 2-8	L_1	90.9	41	99
		Eqn. 2-8	LS 2-way	37.0	11	100
		Eqn. 2-10	T1	36.8	12	100
	Run d) 5% Dixon test	Eqn. 2-8	L_1	86.4	205	22
		Eqn. 2-8	LS 2-way	33.3	127	34
		Eqn. 2-10	T1	33.3	127	33
	Run e) 1% Dixon test	Eqn. 2-8	L_1	88.7	74	26
		Eqn. 2-8	LS 2-way	36.8	28	51
		Eqn. 2-10	T1	36.8	30	53
Study 2	Run c) 30-day residuals	Eqn. 2-8	L_1	179.9	51	89
		Eqn. 2-8	LS 2-way	64.6	29	98
		Eqn. 2-10	T1	64.4	31	98
	Run d) 5% Dixon test	Eqn. 2-8	L_1	182.6	107	36
		Eqn. 2-8	LS 2-way	63.0	77	46
		Eqn. 2-10	T1	63.0	77	47
	Run e) 1% Dixon test	Eqn. 2-8	L_1	185.6	44	50
		Eqn. 2-8	LS 2-way	66.0	31	64
		Eqn. 2-10	T1	66.0	32	63

30-day rule (simulation run 1c, 2c): The L_1 fit detected about 41% and 51% of the number of imposed MMs in run 1c and 2c, respectively. Of these detected MMs 99% and 89% were indeed imposed MMs. The other estimation procedures detected a substantially lower number of MMs although the proportion of true MMs was higher.

Dixon tests (simulation run 1d,1e,2d,2e): The Dixon test found more outliers than the 30-day rule at both confidence levels in both studies and in combination with any estimation method. Using 5% confidence level results in detection of two to more than ten times more outliers compared to the 30-day rule. In simulation run 1d more outliers were found than the number of prescribed MM. Around 50% of the detected Dixon-outliers had not been imposed. Interestingly, the resulting estimated σ_e^2

was similar after removal of 30-day residuals and after removal of 1% Dixon-outliers in study 1, although about twice as much outliers were removed by the latter. In study 2 almost the same number of outliers was found with the non-robust methods by both the 30-day rule and the 1% Dixon test. However, the Dixon-test detected a substantially lower percentage of imposed MMs (63%) than the 30-day rule (98%). This correlates with the lower estimate of σ_e^2 for the latter.

The highest total number of the imposed MMs and at the same time a low percentage of non-MMs was found by the combination of the robust L_1-estimation method with the 30-day rule.

2.3.2 Phenological data

Based on the results of the Monte-Carlo simulation we applied the L_1 estimation to the described data sets for outlier detection and than the T1 method for variance component estimation.

Table 2-5. Application of 30-day rule and the Dixon tests after L_1 fit with subsequent T1 variance components estimation with data from the DWD. The values of the variance components are average values over all resulting combined time series. Outliers in percentage of observations in the resulting combined time series.

	Aesculus hippocastanum	***Fagus sylvatica***	***Quercus robur***	***Betula pendula***
No. of all original observations	9668	5909	7886	7886
No. of resulting time series (No. used weather stations)	69	47	55	57
No. of data points in resulting combined time series	3147	2069	2375	2684
Completeness including outliers (%)	50.9	46.7	46.9	50.4
av. T1 estimate of σ_e^2 including outliers	45.3	36.9	34.7	31.1
average estimates of σ_s^2 for all outlier procedures	17.8	13.6	15.1	13.3
30-day residuals (% obs. in comb. time series)	0.6	0.4	0.2	0.2
av. T1 estimate of σ_e^2 after removal of 30-day residuals	35.7	31.2	31.2	29.6
5% Dixon-type outliers (% obs. in comb. time series)	1.5	1.3	1.1	1.4
av. T1 estimate of σ_e^2 after removal of 5% Dixon-outliers	40.8	34.6	32.4	28.6
1% Dixon-type outliers (% obs. in comb. time series)	0.7	0.5	0.5	0.8
av. T1 estimate of σ_e^2 after removal of 1% Dixon-outliers	42.8	36.1	33.7	30.3
No.of 30-day residuals not detected by the 5% Dixon test	17	2	2	1
σ_e^2 after removal of 5% Dixon-type and 30-day residuals	33.7	29.6	29.5	27.9
No.of 30-day residuals not detected by the 1% Dixon test	22	4	5	5
σ_e^2 after removal of 1% Dixon-type and 30-day residuals	34.7	30.8	30.5	29.0

Even with the restriction that a combined time series was only calculated when there were more than five time series per weather station available, there were only about three observations available per year condensed to one on average (Tab. 2-5) such that the Dixon rule that is only defined for N > 3 could not be applied in many years. Average completeness was around 50% resulting in very unbalanced designs. The Dixon tests found more outliers than the 30-day rule for each species.The impact of the outlier detection procedures on the estimated station variance σ_s^2 was not very pronounced and is not further considered. The average values of σ_s^2 were 17.8, 13.6, 15.1 and 13.3 for *A. hippocastanum*, *F. sylvatica*, *Q. robur* and *B. pendula*, respectively. The 30-day rule found 0.2% to 0.6% outliers on average per species and combined time series decreasing the respective error variance σ_e^2 by 5% to 20%. The 1%(5%) Dixon test found 0.5%(1.1%) to 0.8%(1.5%) outliers decreasing the respective error variance σ_e^2 by 2%(6%) to 6%(10%). Although more values were deleted by the Dixon tests the resulting estimated σ_e^2 was lower after application of the 30-day rule, except for *B. pendula* in combination with the 5% Dixon test. The Dixon tests did not identify all observations with 30-day residuals as outlying. Removing these observations in addition to the Dixon-type outliers lead to a additional reduction of the average estimated σ_e^2. The percentage of outliers found by the 5% Dixon test was about the same magnitude as in former studies (Linkosalo et al. 1996, 2000). The variation of the error variances was higher than in the simulation study, ranging from 9 to over 120 for *A. hippocastanum* before and from 8 to 80 after deletion of outliers (results not shown).

For an individual combined time series (Tab. 2-2) the estimated yearly values differed up to ten days between the various procedures (results not shown). The one-way model as well as the robust fit that differed most from the two-way models fitted by classical non-robust procedures. This is not surprising because the one-way model does not correct for station effects, and the L_1 estimation is not as susceptible to extreme observations as the other methods. A difference of 1.2 days between the yearly estimates of the combined time series could be observed between the classical non-robust estimation procedures by two-way models. It was the two-way fixed effect model that differed maximally by one day from the estimates for the mixed models. This result differs slightly from what was found in the Monte-Carlo study, where the differences between the two-way models were negligible. In the Monte-Carlo study, however, averages over 500 simulation runs were used. The estimation procedures in the mixed model case did not have a significant influence on the values of the estimated parameters in this case. It can be stated, however, that all 95% confidence intervals for the year effects contained the estimated year effects of all other procedures.

The estimated error variances of the mixed model (estimated by REML, MIVQUE0, T1) and the fixed two-way model are all 73.6. Only ML and the fixed one-way model estimates differ from this value with 58.5 and 80.5, respectively. This large error variability was probably due to the large variability of the time series at individual stations between 1955 and 1960 (Fig. 2-2).

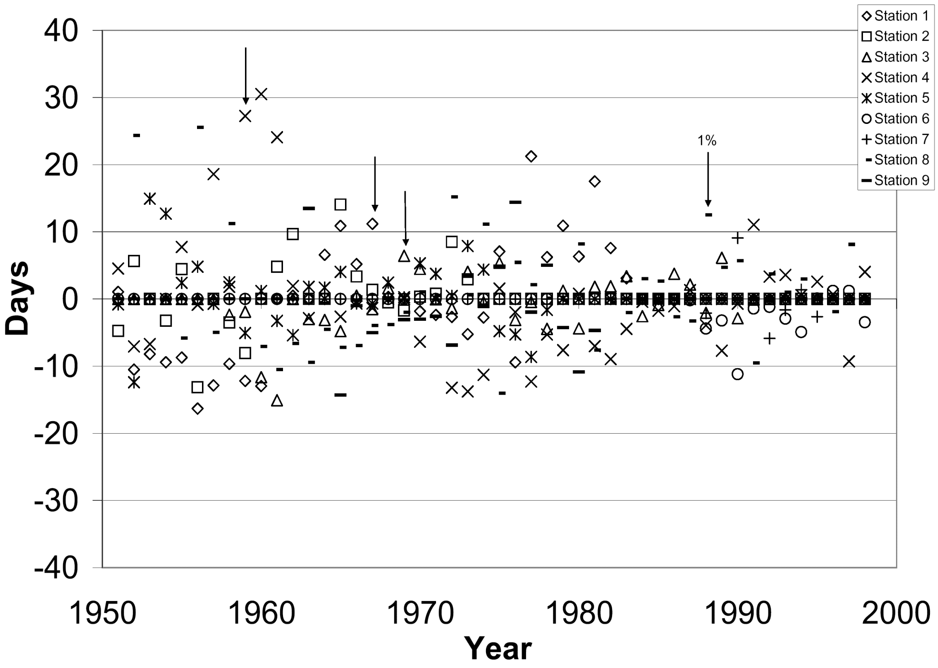


Figure 2-3. Residuals of the T1 fit for the combined time series at station 2906. Residuals marked with an arrow have been determined as outliers by the 5% Dixon test. Arrows labeled as '1%' mark outliers that have also been detected by the 1% Dixon test. The symbols are attributed to the same stations as in Fig. 2-2.

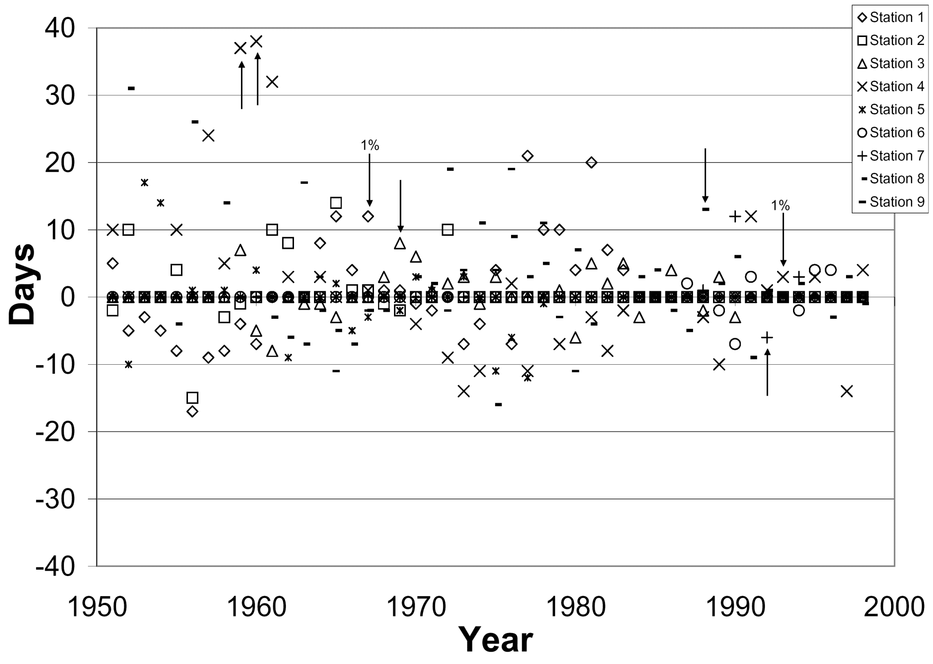


Figure 2-4. Residuals of the robust L_1 fit for the combined time series at station 2609. Residuals marked with an arrow have been determined as outliers by the 5% Dixon test. Arrows labeled as '1%' mark outliers that have also been detected by the 1% Dixon test. The symbols are attributed to the same stations as in Fig. 2-2.

The number of values that exceed the 30-day threshold increases from one to four when a robust L_1 fit (Fig. 2-4) is used instead of a T1 fit (Fig. 2-3). Applying the 5%(1%) Dixon test to the L_1 fit seven(two) outliers are found, opposed to four(one) outliers for the T1 fit. 5% Dixon test outliers are marked with an arrow in Fig. 2-3 and Fig. 2-4. Some of these were also detected by the 1% Dixon test and additionally labeled with '1%'. As seen in Tab. 2-5, not all residuals above the 30-day threshold were considered as outliers by the Dixon tests even though they represent undoubtedly extreme values, a typical masking effect (Barnett and Lewis 1996). Due to this masking effect the T1-estimated error variance of the combined time series was higher after removal of 5% Dixon outliers of the L_1 fit than after removal of 30-day residuals, namely 61.1 and 51.1, respectively. This corresponds to the reduction in estimated error variance by 17% through removal of seven Dixon outliers opposed to a reduction of 21% through a removal of four 30-day residuals. Removing the 30-day residuals of the L_1-fit also decreased the maximal absolute difference between the robust combined time series and the T1 time series from 9.3 to 6.2 days. After removal of 5% Dixon outliers the T1 time series had a maximal absolute difference of 7.5 days to the robust time series. Although more outlier were found by the 5% Dixon test, influential extreme observations are missed due to the masking effect, resulting in an unnecessarily high influence of extreme observations on the combined time series. The 1% Dixon test finds outliers with residuals of 3 and 12 days.

Phenological data are often subjected to trend analysis. We explore the effect of fitting a linear mixed model (T1), the robust model (L_1)and the LS 1-way model on the linear trend of the resulting combined time series of station 2609 from above (Tab. 2-2). As expected, the trend of the robust fit, as total change between 1951 and 1998 rounded to whole days, does not differ much between outlier detection methods. According to the robust fit the linear trend of the combined time series is delayed by one or two days in 48 years. The trend is not significant. The trends of the other two estimated time series show a net advancement between 1951 and 1998. Only the trend of the LS 1-way model including all observations is slightly significant ($P < 0.1$). The trends show marked differences between outlier detection procedures. Whereas including all observation the 2-way and the 1-way model show only a difference of one day, the difference increases to three days without 30-day residuals and falls back to one day again when we exclude Dixon-type outliers. This is because the Dixon test does not find the extreme observations in the beginning of the observation period (Fig. 2-3). But

Table 2-6. Total linear trend between 1951 and 1998, rounded to whole days, of the combined phenological time series of station 2609 using three estimation (T1,L_1,LS 1-way) and two outlier detecting procedures (30-day residuals, Dixon test).

Outlier procedure	Estimation procedure	Total linear trend [days]	P-values
Including all observations	Mixed T1	-5	0.15
	LS one-way	-6	0.09
	Robust L_1	2	0.67
Without 30-day residuals	Mixed T1	-1	0.74
	LS one-way	-4	0.32
	Robust L_1	2	0.68
Without Dixon-type outliers	Mixed T1	-4	0.36
	LS one-way	-5	0.18
	Robust L_1	1	0.78

even excluding those extreme observations by the 30-day rule the 1-way model still shows a stronger negative trend than the 2-way model. This is because the 1-way model is biased by the phenological stations 6 and 7 which only have observations at the end of the observation period (Tab. 2-2) and are in general earlier than the average (they have negative estimated station effects by the 2-way models).

2.4 Discussion

We showed that all methods proposed by Häkkinen et al. (1995) to combine phenological time series can be treated within the framework of linear models. Method 3 proposed by Häkkinen et al. (1995) is the same linear model as method 4, except that in method 3 the station effects are considered as fixed, whereas in method 4 the station effects are considered as random. Having defined this difference in the assumptions within the framework of linear models gives us theoretical arguments why to prefer one method or the other. The difference between the mixed model (method 4) and the fixed model (method 3) is of special interest here because the other methods perform poorly (method 1) or are of no relevance due to their application restrictions (method 2, see method section). It can be argued that given a certain region for that a representative phenology is to be estimated the distribution of phenological stations is random. Moreover, our interest is to estimate the year effects and their variance rather than single station effects, that are assumed to be arbitrarily chosen in that region. In such a situation the mixed model would be more appropriate. In another analysis we might be interested in the specific station effects to find stations that show extreme behavior, for instance, or check whether station differences are consistent throughout the years. In this case we would treat the station effect as fixed (Searle 1971). However, in terms of parameter estimation, error variances and confidence intervals the differences between fixed and mixed model are small when proper estimation methods are used like LS, REML, MIVQUE0 or T1.

It has been noted that phenological data are often not normally distributed (Schnelle 1955, Menzel 1997). Therefore the robust ANOVA in the L_1 norm might be more appropriate in the first place, even though L_1 estimates are not unequivocal (Bloomfield and Steiger 1983) and not as accurate as classical methods (Tab. 2-3). L_1 estimation is considered to be more appropriate for fatter tailed distributions (Dodge 1987). A better estimator for the variance component should be found than the one applied here (Eqn. 2-11). In Chapter 3, however, it is shown that on the one hand, the different sources adding to the station variance, i.e. microclimate, genetic variability and the observer, can be summarized under one source of variation and that on the other hand, genetic and observer influences can be assumed to be normally distributed. Thus, normally distributed error variances should be an acceptable assumption in most cases.

The influence of different outlier omitting strategies should be explored further. Month-mistakes are independent from other errors. Thus, it can be expected that 50% of the MMs go into the opposite direction than the errors of other origin and therefore are partly masked by these errors. This situation is mimicked in the structure of our Monte-Carlo simulation runs. Therefore, it can be expected that only about 50% of the MMs can be detected. Having this in mind the 30-day residual rule combined with a robust L_1 fit performs close to optimal. The non-robust procedures combined with the 30-day rule perform not as good because they decrease large residuals (Fig. 2-3, Fig. 2-4). In the case of normally distributed errors the Dixon test detects too many outliers especially at the 5% level. In general it can be expected that if more than 50% of MMs shall be detected this is only possible at the expense of dismissing values within the natural variability of the data. But even the 1% Dixon test that found a similar amount of outliers as the 30-day rule in study 2 rejected many false outliers. The Dixon test implies different variances for each year. On the one hand this is reasonable because the rate of phenological development varies from year to year. On the other hand this can result in the rejection of observations that only slightly deviate from the theoretical estimated value, e.g. year 1993 in Fig. 2-4, because the other residuals in this year are even smaller. The average error variance of about 30 found in the real data means that 95% of the residuals do not differ more than 10 days from the theoretical values. Even the maximal estimated error variance of about 100 means that 95% of the estimated variation is within 20 days. This evidence and experimental findings (Baumgartner 1952) give no biological background to discard phenological observations which are less than a week or even ten days off the estimated values because this is well within the natural unpredictable variability. Moreover, some extreme values can be missed because there might be other rather large residuals in this year, e.g. year 1952 in Fig. 2-4, that mask the outlier. For this reason it is not surprising that a distribution-free rule combined with a robust estimation performs better in the Monte-Carlo study and is more effective in removing large error variances due to extreme observations also when observational data are analyzed. The Dixon rule is not efficient in removing error variance because it misses some extreme observations that contribute significantly to σ_e^2 and thus the unreliability of the resulting combined time series.

There are types of mistakes that have a similar magnitude of impact as month-mistakes, like misprints and number rotations for instance, i.e. noting the julian date 163 instead of 136. The month mistake is a distinct feature, because it adds an error component of 30 days or a multiple of 30 days to the deviation of the observed value from the mean. This deviation is big relative to the variance due to biological variability. Therefore, a cut-off at 30 days from the expected value is an effective way to remove MMs and all mistakes that induce deviations above 30 days. Other observational and protocolling errors that cause error components smaller than 30 days can at the current stage of knowledge about the variance of phenophases in heterogeneous populations not be distinguished from the variance due to biological variability (see also Chapter 3). The Monte-Carlo analysis and inspection of observational time series especially demonstrate the danger of removal of correct observations when distribution-based methods are applied for outlier detection with usually small sample sizes available for phenological observations. However, a better characterization of the biological variability of phenology, that is urgently needed, can provide a basis for further refinement of the methods. A first approach to quantify the main sources of phenological variability and uncertainty is given in the next chapter. These studies are also needed to evaluate the danger of false identification of observations as MMs in cases where a distinct bimodal distribution of phenophases is produced by an intermittent occurrence of environmental conditions unfavorable to the phenological development, as e.g. a cold spill when the buds in a part of the population have already broken. In addition, studies on the statistics of observational and protocolling errors are needed. Such studies can serve to determine if a cut-off value lower than 30 days can be applied.

The presented results and the example of the trend analysis show that it is indeed important to correct for the station effects when trends of phenological time series are estimated. Otherwise we can have a bias in the resulting time series due to unequal distribution of observations in time in cases where station data are unbalanced. Chmielewski and Rötzer (2001) have analyzed trends of averaged (1-way model) time series of the International Phenological Gardens (IPG) between 1969 and 1998. The time series were averaged over several phases and stations within wide geographical regions. The data used by Chmielewski and Rötzer (2001) that were labelled to be within the Natural Region No. 8 (North Alpine Foreland, five stations, Bud burst of four species, totalling 455 observations) were digitalized from the journal of the IPG (Arboreta Phaenologica Nr. 1 - 43). Unfortunately the data of 1998 was not published yet and could not enter the comparison. The trend of the average bud burst time series (1-way model) amounted to a total of -9.5 days ($P < 0.1$) within the period 1969 through 1997. This is close to the value of -9.3 ($P < 0.1$) found by Chmielewski and Rötzer (2001)(3.1 days per decade) for the period 1969 through 1998. However, with a two-way mixed linear model (T1 estimation) the trends amounts to -7.8 days between 1969 and 1997 and is not significant ($P>0.1$).

We conclude this chapter with two recommendations. Firstly, we strongly recommend outlier detection when research is conducted with combined phenological data. In our view, one of the few mistakes that can be detected is the month-mistake because its deviation much is larger than the range of natural variability, even though the proportion of MMs to the total number of errors is not known. The non-parametric 30-day residual rule in combination with a robust L_1 fit is a stable and adequate procedure to detect month-mistakes and other extreme values. Secondly, after outlier removal, fitting a linear two-way mixed model using the TYPE I or REML estimation method to obtain a reliable continuous time series for further analysis is appropriate.

3 Variability and uncertainty of phenological data

3.1 Introduction

As previously mentioned, variability of phenological observations is composed of many kinds of uncertainties and errors (Chapter 2, ”2.2.4 Outlier detection” on page 14). In phenological data bases the occurrence of a certain phase is noted to have taken place on a fixed date at a certain site. This date is supposed to be representative for the area where the observation site is located (DWD 1991). It is difficult to determine how large the area or region is for that a phenological observation is representative. This depends on the region itself and also on the purpose the data is used for. The observation instructions of the DWD (DWD 1991) state that the site where an observer makes his observations should be characteristic for its community or district. Steep valleys, deep swales, exposed southern or northern slopes and other extreme locations should be avoided for phenological observations. In mountainous terrain the observation sites of the different plant individuals should not differ more than 50 m in elevation. With these restrictions microclimatic influences are supposed to be minimized. Nevertheless, apart from those obvious microclimatic influences it can be assumed that among several characteristic sites in a given community meteorological and other environmental conditions like soil, slope, aspect and vegetation cover have a modifying influence on phenology. Usually the exact location of the phenological observation is not known and this uncertainty transfers to the uncertainty and variability of the representative phenology in a given district.

Another prominent and well-known source of phenological variability is the genetic influence. There are various reports that document varying phenological behavior among provenances of the same species (see below).

Phenological data do not result from exact measurement but rather observations. The observation instructions try do define when a certain phase is reached. Bud burst, e.g., according to the instruction of the DWD (DWD 1991) is defined as the ’first day at which in at least three positions in a tree leaves have completely protruded the bud scales and have emerged to leaf base (petiole) such that the leaf shaft or leaf base is visible. The leaves are already in their final shape but have not yet attained full size’. It is clear that this definition still leaves room for interpretation and also adds to the uncertainty of phenological data apart from protocolling errors that have been studied in the previous chapter.

The purpose of this chapter is to roughly quantify these three distinct and independent types of errors or uncertainties and give a first approximation of the intrinsic uncertainty of phenological observations. As stated in the previous chapter (”2.4 Discussion” on page 25) this can be a helpful contribution to refine phenological data analysis. Secondly, when the intrinsic variability of phenological data is known, it can be assessed how well models can possibly approximate such data (see ”3.4.2 Results” on page 34).

3.2 Microclimatic and site influences

In most applications, phenological data are used in conjunction with meteorological measurements, mostly temperature. This has been indicated in the introduction (Chapter 1) and also becomes clear throughout this thesis. It is commonly agreed that temperature is the main determinant of phenological plant development (Schnelle 1955, Hickin and Vittum 1976, Thomson et al. 1982, Murray 1989, Worral 1993, Fitter at al. 1995, Diekmann 1996, Menzel 1997, Sparks et al. 2000). Many more references proving the crucial influence of temperature on phenology are given in this thesis. Thus, it can be hypothesized that the temperature variations among the characteristic sites of a community are a good measure for microclimatic and site influences on phenological observations.

There are numerous studies about the influence of soil, topography, relief and land cover on the radiation balance, temperature and other meteorological variables at various scales. A comprehensive review of all important aspects of microclimate is given in Geiger et al. (1995). Temperature variability within a region or district is difficult to assess. Temperature variability will not only vary among regions but also among scales within a certain region. There are temperature variation of over ten degrees in ridges of a crop field (Lessmann 1950) or within the bark of a tree (Haarlov and Petersen 1952), i.e. at a scale of a few centimeters. Such small scale variations are of minor interest to the phenologist except for very small plants like, e.g., *Galanthus nivalis*. There are studies about blossoming variations within a tree due to crown position and exposure (Scamoni 1938, Eisenhut 1957). This scale is also beyond the resolution of most phenological observations. Temperature variability at the scale of plant individuals are of interest in our context because phenology is mostly observed for individuals or groups of individuals. A resolution of 10 to 50 m² is appropriate for such an analysis.

The more heterogeneous the landscape is in terms of soil, relief and land cover, the larger the temperature variations that can be expected and accordingly the less representative a phenological observation will be within this community. To assess the microclimatic variations of a region either detailed models or intensive measurements or both are needed for this region. Such detailed studies have only been done for few regions. In one study, e.g., the climate of a region of 33 km² in the German Black Forest with a height range of about 200 meters has been investigated using 16 weather stations (Siegmund 1999). With a regression model that considered among topographic parameters also land cover types a standard deviation of mean March-May temperatures of 0.5 degrees within this region was calculated at a resolution of 50 m². Even though this region might be too large to identify just one representative phenological observation, this example shows that it is in principle possible to approximate microclimatic variability.

3.3 Genetic influences

There are many studies that show, mostly experimentally, the difference in phenological development between plant individuals of the same species but of different provenances or phenotype. It has been observed that there are late and early individuals of the same species, an information that has been used in forestry and agriculture for selection purposes.

Baumgartner (1952) observed the variability of the date of bud burst in 102 individually growing even aged mature beech trees (*F. sylvatica*) that were located in fairly homogeneous microclimatic and edaphic conditions. The trees all grew on the north-western side of a road and were subjected to little shading by the buildings on the opposite site of the road. Frequency distribution of the beginning of the green up, that was characterized by the date when ten percent of the buds were open is shown in Fig. 3-1 (based on data read from Fig. 1 in Baumgartner (1952); 98 trees, 4 excluded by Baumgartner). This frequency distribution has an average julian day of year (DOY) of 128 (127.9), a standard deviation of 2.7 days and a range of 17 days. The hypothesis of a normal distribution can not be rejected by the Kolmogoroff-Smirnoff (K-S) test and the χ^2-test at the 5% level.

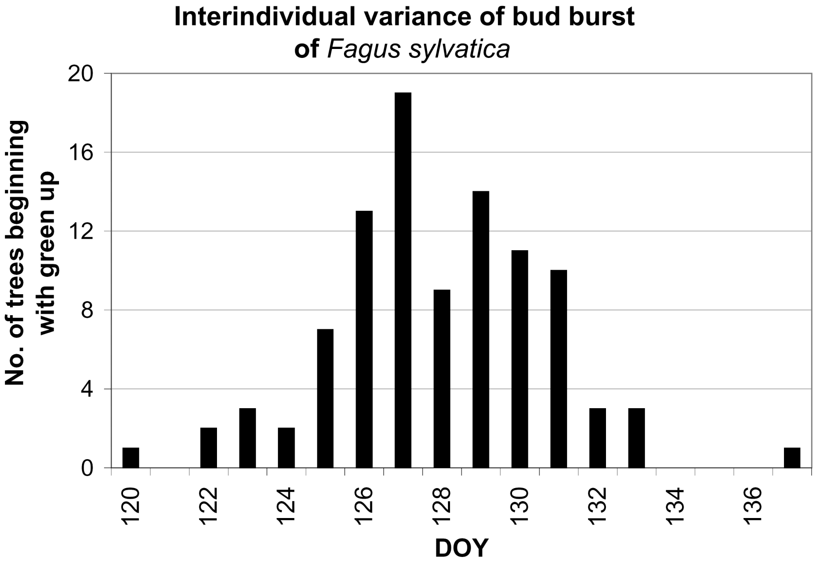


Figure 3-1. Number of trees beginning with green up, characterized by the date when ten percent of the buds were open. Data from Baumgartner (1952).

In Tab. 3-1 and Tab. 3-2 results from more experiments investigating differences in bud burst among individuals of one provenance and among clones, provenances or populations are compiled. For each experiment the individuals or groups of individuals were subjected to equal environmental conditions

Table 3-1. Variability of spring phases between tree individuals. N number of observed individuals. σ standard deviation of the observations. *a* and *b* indicate different years.

species	N	type of observation and source	σ [days]	range [days]	reference
Fagus sylvatica	127	observation of trees in a glass house CO_2-enrichment experiment, green tip on first bud visible source: one provenance	3.0*a*	12	Badeck (unpublished)
	148		2.7*a*	15	
	158		3.3*b*	14	
	154		3.5*b*	18	
	150		3.7*b*	24	
	163		3.8*b*	21	
Fagus sylvatica	131	observation of trees in a glass house CO_2-enrichment experiment, first buds open, source: one provenance	3.9	21	Badeck (unpublished)
	150		2.7	14	
Robinia pseudoacacia	56	observation of alley trees, first start of bud burst, source unknown	0.68	3	Badeck (unpublished)
Robinia pseudoacacia	56	observation of alley trees, all buds open, source unknown	0.76	3	Badeck (unpublished)

Bud burst of beech (*F. sylvatica*) trees in the experiment of Badeck (unpublished) show a similar variability as in the experiment of Baumgartner (1952). Even though the average standard deviation of roughly 3.3 days is slightly higher in the experiment of Badeck, the related range of 17.4 days is almost the same as in Baumgartner (1952). The observations of bud burst on locust trees (*R. pseudoacacia*) show a lower variability than *F. sylvatica*. Whereas the observations on *F. sylvatica* were conducted in two subsequent years the observations of *R. pseudoacacia* are from one year only. The two standard deviations of observations of first green tips on *F. sylvatica* in year *a* (Tab. 3-1) are lower than in year *b*. This gives a first hint that the variability of phenological phases between individuals is modified by external conditions.

For the comparison of phenological variability between provenances, clones or populations (Tab. 3-2), the data within these groups were averaged before comparison. The variability between provenances is in some cases remarkably higher than between individuals of one provenance. Some experiments, however, can be assumed to be outlying because their conditions were extreme and beyond the natural range of conditions usually experienced by the respective species. *A. saccharum* in Florida (Kriebel and Wang 1962) is an outlier in this compilation because Florida is not within the natural range of occurrence of *A. saccharum*. In the chilling experiments with *A. saccharum* the standard deviations of 13.0 and 13.9 days were obtained with incomplete chilling which is unlikely to occur under natural conditions. However, the other two chilling experiment of Kriebel and Wang (1962) with *A. saccharum* demonstrate that the variability between provenances also depends on environmental condition, here chilling.

Table 3-2. Variability of bud burst between clones and provenances. N number of observed individuals. σ standard deviation of the observations. *a* and *b* indicate different years. * extreme conditions. The data were scanned from graphs in the original papers.

species	N	type of experiment	origin of plants	σ [days]	range [days]	reference
Acer saccharum	4	observations outdoor at Wooster, Ohio, USA	Georgia, Tennessee, Ohio, Michigan	4.2	10	Kriebel & Wang (1962)
Acer saccharum	4	observations outdoor at Gainesville, Florida, USA	Georgia, Tennessee, Ohio, Michigan	19.0*	41	Kriebel & Wang (1962)
Acer saccharum	4	controlled conditions, constant temperature regime after different duration of chilling	Georgia, Tennessee, Ohio, Michigan	13.0*	29	Kriebel & Wang (1962)
				13.9*	33	
				4.7	11	
				2.9	6	
Betula pendula and B. pubescens	3	controlled conditions	56°N Denmark, 64°N and 69°N Norway, below 100 m	1.8	3.4	Myking & Heide (1995)
Abies amabilis	5	observations outdoor, 780 m	49°5’N to 51°10’N west of Rocky Mountains, 205-1404 m	5.4*a*	14	Worral (1983)
				10.2*b*	23	
Picea abies	6	comparison of clones in nursery 1965 and 1966	old trees in forest in Connecticut, probably one provenance	4.1a	9.7	Worral & Mergen (1967)
	17			2.7a	4.4	
	6			5.8b	12.9	
	17			6.2b	6.0	

For the two experiments with *Abies amabilis* and *Picea abies* that were conducted outdoor under natural conditions, it was tested whether the variances between two years were significantly different using the *F*-test (Sokal and Rohlf 1995). Only in the experiment with *Picea abies* the two standard deviations 2.7 and 6.2 were significantly different ($P<0.05$). These two standard deviations referred to the same clone. The significant difference in variability between two years of bud burst observations of *Picea abies* (Worral and Mergen, 1967) can be assumed to be caused by environmental conditions.

From the data compiled above, an average overall standard deviation between provenances of 4.8, ranging from 1.8 to 10.2 days, and an average overall range of 10.0 days, ranging from 3.4 to 23 days, can be deduced as a first rough guess of the provenance influence. Assuming that the variability among provenances is independent from variability within provenances, we can add an average interindividual variability of about 2.8 days (Tab. 3-1). Thus, from the data we considered, the magnitude of the genetic influence on spring phenology can be roughly approximated by a standard deviation of 7.6 days. Under the crude assumption that genetic variability is similar for the considered species, this means that for an average year the bud burst of 95% of observed tree individuals of several provenances at the same location can be assumed to range within 30 days.

3.4 The observer error

As indicated above, phenological observation instructions leave room for subjective interpretation. It can be assumed that when several observers make observations on the same plant individual, they will come to different results concerning the date of the occurrence of a phenological phase. Because I did not find studies that tries to quantify this type of error, an experiment was conducted to give a first approximation of the observer's influence on phenological observations.

3.4.1 The experiment

In spring 2000, 20 observers made phenological observations of bud burst on three individuals of birch (*Betula pendula*), beech (*Fagus sylvatica*), oak (*Quercus robur*) and may shoot of pine (*Pinus sylvestris*), respectively. The observers were given the observation instructions of the German Weather Service (DWD 1991). The definition of bud burst was already stated above and may shoot of pine is defined as the stage when at the 'top of the short shoots the first bud scales burst and the green needle tips become visible'. These definitions were accompanied by pictures of the respective phase. The observers were all natural scientists including one biologist but had no experience with phenological observations. Because the observers were all colleagues they were instructed not to exchange or discuss neither the observation instruction nor observation results to guarantee independent data. During the critical time observations should be made every day and it should not only be recorded the day when the respective phase occurred but also when observations were made at all.

3.4.2 Results

Not every observer observed all tree individuals. Furthermore, during the critical time weekends and holidays disrupted continuous daily observations for many observers. Thus, in the following only those observations were considered where the observer also made observations the previous day. In Tab. 3-3 the number of observers that recorded BB on a specific day of year (DOY) are noted for the deciduous species with the respective average, standard error and standard deviation with 95% confidence limits. In Tab. 3-4 only the summary statistics are noted for *P. sylvestris*. The distributions of the noted day of BB and may shoot were tested for normality per species. The hypothesis that the observations are normally distributed could not be rejected at the 5% level using both the K-S-test and the χ^2-test except for pine No. 3.

Table 3-3. Number of observers that recorded BB on the specific day of year (DOY). *B. p. Betula pendula, F. s. Fagus sylvatica, Q. r. Quercus robur.* Only those DOY of BB were considered where also observations the day previous to the noted DOY were made. Also noted: the average DOY with its standard error. l 95% cl: lower 95% confidence level of the standard deviation (std. dev.); u 95% cl: upper 95% confidence level of the standard deviation. N number of considered observations

DOY	***B. p.* 1**	***B. p.* 2**	***B.p.* 3**	***F. s.* 1**	***F. s.* 2**	***F. s.* 3**	***Q. r.* 1**	***Q. r.* 2**	***Q. r.* 3**
114	3	2							
115	6	6							1
116	3	4	6		1		1	2	5
117	1	2	2				1		4
118									
119			1				1		
120					1	1	1		1
121					1	1	2	2	
122				2	2	3		1	
123				7	1	1		2	
124				1		1			
average	115.2	115.4	116.6	122.9	120.7	122.0	119.0	120.3	116.6
std. err.	0.2	0.3	0.3	0.2	1.0	0.5	0.9	1.1	0.4
l 95% cl	0.6	0.7	0.7	0.4	1.6	0.8	1.3	2.0	0.9
std. dev.	0.9	0.9	1.0	0.6	2.5	1.3	2.1	3.0	1.3
u 95% cl	1.5	1.5	1.9	1.0	6.1	2.8	5.1	6.7	2.3
N	13	14	9	10	6	7	6	7	11

Thus, confidence limits could be calculated for the standard deviation of resulting observation distributions using the formula

$$\frac{s}{\sqrt{F_{n-1;\infty;\alpha/2}}} \le \sigma \le s\sqrt{F_{\infty;n-1;\alpha/2}}$$

with s estimated standard deviation, F F-distribution with n and ∞ degrees of freedom and $\alpha = 0.05$ rejection level. For *B. pendula*, *F. sylvatica* and *Q. robur* average standard deviation of the observations is 1.5 days. This signifies that for this year and for each of the nine individuals of *B. pendula*, *F. sylvatica*, and *Q. robur* 95% of the observations of BB varied within a range of roughly a week. This can also be directly deduced from Tab. 3-3. It is the one observer on DOY 116 of *F. sylvatica* No. 2 and the two observers on DOY 116 for *Q. robur* No. 2 that are not within the range of the week where the other observers noted the phase. These three observations are 4% of all observations. The individuals of *P. sylvestris* were very difficult to observe. This was because of the choice of individuals with high crown bases that were difficult to observe for observers without binoculars, but also because of the phase definition itself. It can be assumed that observations of may shoot are generally less reliable than observations of deciduous tree bud burst.

Table 3-4. Average DOY with its standard error. *P.s. Pinus sylvestris*. l 95% cl: lower 95% confidence level of the standard deviation (std. dev.); u 95% cl: upper 95% confidence level of the standard deviation. N number of considered observations

	P. s. 1	*P. s.* 2	*P. s.* 3
average	130.8	130.6	132.1
std. err.	1.8	1.7	1.9
l 95% cl	2.7	2.8	3.2
std. dev.	4.4	4.4	5.0
u 95% cl	10.8	9.7	11.0
N	6	6	5

3.5 Discussion

The microclimatic variability at the scale of interest between different representative sites in a certain district is extremely difficult to approximate. But it is possible and with the use of satellite observations such investigations might be facilitated. Moreover, it can be expected that temperature variability within a community varies from day to day depending on cloud cover, for instance. To translate such microclimatic uncertainties into related phenological variability, models of the temperature influence on phenology are needed. Such models again introduce an additional uncertainty because neither temperature is the only influencing factor nor the exact nature of the temperature influences on phenology is known. With some simple assumptions, however, it might be possible to deduce a lower bound of phenological variability due to the uncertainty of the microclimate in a given district, as a starting point.

Genetic variability is, in a first step, easier to approximate, even though in controlled experiments or field observations the whole range of genetic variability within a species can probably never be covered. The aim was to derive a first guess for this important influencing factor. The experiments with *Picea abies* showed that environmental conditions modify the observed variability. When the developmental speed of individuals is fast in one year, it can be expected that differences among individuals are less pronounced than when the environmental conditions favor a slow development. Year-to-year variation in observation uncertainty due to the genetic influence can be approximated with models that estimate this uncertainty as a function of the prevailing conditions around the date of the observation. To develop such models either more data on the genetic year-to-year variation or models for phenological phases that are parameterized for different individuals or provenances are needed. The later is possible with existing data. There are a lot of data of single station phenological observations of species individuals over a wide geographic range. Assuming different provenances among distant locations and fitting models (see Chapter 5) to each of these distant location, it can be assumed that the different parameterization of these models represent the behavior of different provenances. These models can be used for simulation experiments for provenance variation under varying weather conditions, which in turn can be used to develop models for the influence of weather conditions on genetic variability. However, the above estimated standard deviation of the genetic variability of bud burst in tree individuals is a first approximation of the magnitude of this uncertainty factor.

In the described experiment on the observer error the number of observations per tree individual was low. Therefore the stated standard deviations reveal a high standard deviation themselves. Moreover, as in the case of the genetic variability, the variation among observers will vary from year-to-year due to varying plant developmental speed, adding to the uncertainty of the derived variability of phenological data due to the observer. However, the results of the experiments allow to give a first approximation of the observers error of variational influence on phenological observations. The average standard deviation of the influence of the observer for observations of bud burst was estimated to be 1.5 days.

Because the observer error is independent from the genetic variability, the two standard deviations can be added, resulting in an approximate average standard deviation of the compound observers and genetic error of 9.1 days. This is more than the average estimated error standard deviation for real data of around 5.5 days (approximate average error variance of 30.0) in Table 2-5 on page 21 but below the maximal estimated error standard deviation of 10.0 days (see ”4.5 Discussion” on page 70). The error variance is a measure of variability of the phenological observations that can be assumed to include observer, genetic and site uncertainty. However, even a standard deviation of 9.1 days implies that 95%(99%) of the observations differ less than 18(24) days from the hypothetical mean. This is well below the 30 days that were assumed to be a month mistake (see ”2.2.4 Outlier detection” on page 14). This supports the month-mistake concept as being a safe method for the detection of outliers as long as there are no models for the change of the observation uncertainty with regard to the prevailing weather conditions around the date of the observations.

When a better quantification of the variance of climatic parameters, i.e. mainly temperature, becomes available either by models or by the restriction to a limited region, it is possible to estimate the variance of simulated phenological observations, too. Generating several temperature time series within the climatic uncertainty and feeding them into phenology models creates distributions of simulated phenological observations which in turn can be used to estimate the uncertainty of simulated observations. It can further be assumed that this estimated variability of the simulated observations is only caused by microclimatic uncertainty and not by model structure uncertainties. In connection with the above derived observational uncertainty it can then be derived how well the model can be expected to simulate the observations under the given uncertainties. It can be estimated, e.g., what the best fit of the model can possibly be given the uncertainty in observational and microclimatic data. Such knowledge is helpful for model development. This task, however, is left for further studies.

4 Regional phenology in Germany between 1880 and 1999

4.1 Introduction

As mentioned in the introduction, many phenological studies recorded an advancement of spring phases and a delay of autumn phases during the last decades that was widely discussed in the context of global warming (see Chapter 1). However, as Menzel (2001) pointed out, when looking at the local scale also a delay of spring phases and an advancement of autumn phases can be observed in the last decades. The analysis of trends in time series of phenological observations is rendered difficult because of incompleteness of the data and several sources of errors and uncertainty (see Chapter 2 and Chapter 3). The series have often been obtained by different observers and can have a large percentage of missing data, e.g. around 50% on average for the data set of the German Weather Service (DWD) (see Table 2-5 on page 21). Thus, when individual time series are compared, start and end year often differ, as well as intermediate years are often missing. Linear trend analysis is very susceptible to the values observed in start and end years. Changing start or end year by one year can already drastically change the resulting trend especially in highly variable phenological time series of a few decades (see Fig. 4-10). Thus, trends of time series with varying start and end dates are critical to compare. Reporting observed trends in days per year rather than absolute change (Menzel 2001) does not solve the problem of incomparability. Some authors solved this problem by averaging several phenological time series over the same period (Sparks 2000, Chmielewski 2001). In Chapter 2, I showed that when using yearly averages the resulting trends can be biased by unequal distribution of the stations in time as has been also demonstrated for a result in Chmielewski (2001) in Section 2.4 (see also Table 2-6 on page 24).

The method that overcomes drawbacks of analyzing both single time series and average time series is analyzing combined time series (Häkkinen et al. 1995, Chapter 2). Combining several fragmentary phenological time series for a certain region not only reduces the influence of individual time series and the noise in the time series but also gives the opportunity to apply proper outlier detection methods and normalize the combined time series to a common observation period.

A critical question is over what geographical extent phenological time series should be combined. Because in this chapter I am interested in regional differences of trends, the so-called Natural Regions (NRs) of Germany are a good choice. These Natural Regions are defined by the homogeneity of certain aspects of the landscape like vegetation, soil, climate and phenology (Meynen et al. 1992) (Fig. 4-1). Thus it can be assumed that combining phenological information over these NRs does not smooth too much of the natural phenological variability. It should be noted that the height distribution of the available phenological stations per NR will in most cases not represent the height distribution of the NR itself. Thus, it can be assumed that the combined phenological time series is not representative for average phenological times series of the NR in absolute terms. However, the combined phenological time series can be assumed to be representative for the characteristic trend of the NR because, by the definition of a NR, it can be assumed that trend development is approximately homogenous within a NR, i.e. independent from height.

It has been emphasized that a key to a better understanding of the change in phenology and its causes and future aspect is the availability of long time series of phenological observations (Peñuelas 2001). Schönwiese and Rapp (1997) point out that for climatological studies as far as the causal aspects are concerned it is reasonable to look at time series of about 100 years. For seasonal temperature data they showed that trends can become random when time series of less than 30 years are analyzed. Results from Menzel et al. (2001) and Gornik (1995) indicate that average trends stabilize for time periods over 20 years. Therefore, in this study long time series are analyzed avoiding the methodological shortcomings of both single time series and average time series through the use of combined time series. The general question was whether new aspects of the phenology in Central Europe evolve when we broaden our temporal window from a few decades to a whole century and more.

The key-questions analyzed in the following were:

- Can the trends in phenology that have been observed during the last decades also be found when the view upon phenology in Germany is broadened over a whole century.
- How do possible trends develop over the century and what determines them?
- What are the differences in trends between several phases and species and what are the reasons and consequences of those differences?
- Are there different trends in phenology among regions in Germany?
- What are the consequences of a changed phenology concerning the absorbed radiation?
- Can the random influence of varying observational periods be reduced?
- How large is the influence of global climatic change on phenology at the regional scale?

4.2 Data

The phenological data used in this study came from three distinct sources.

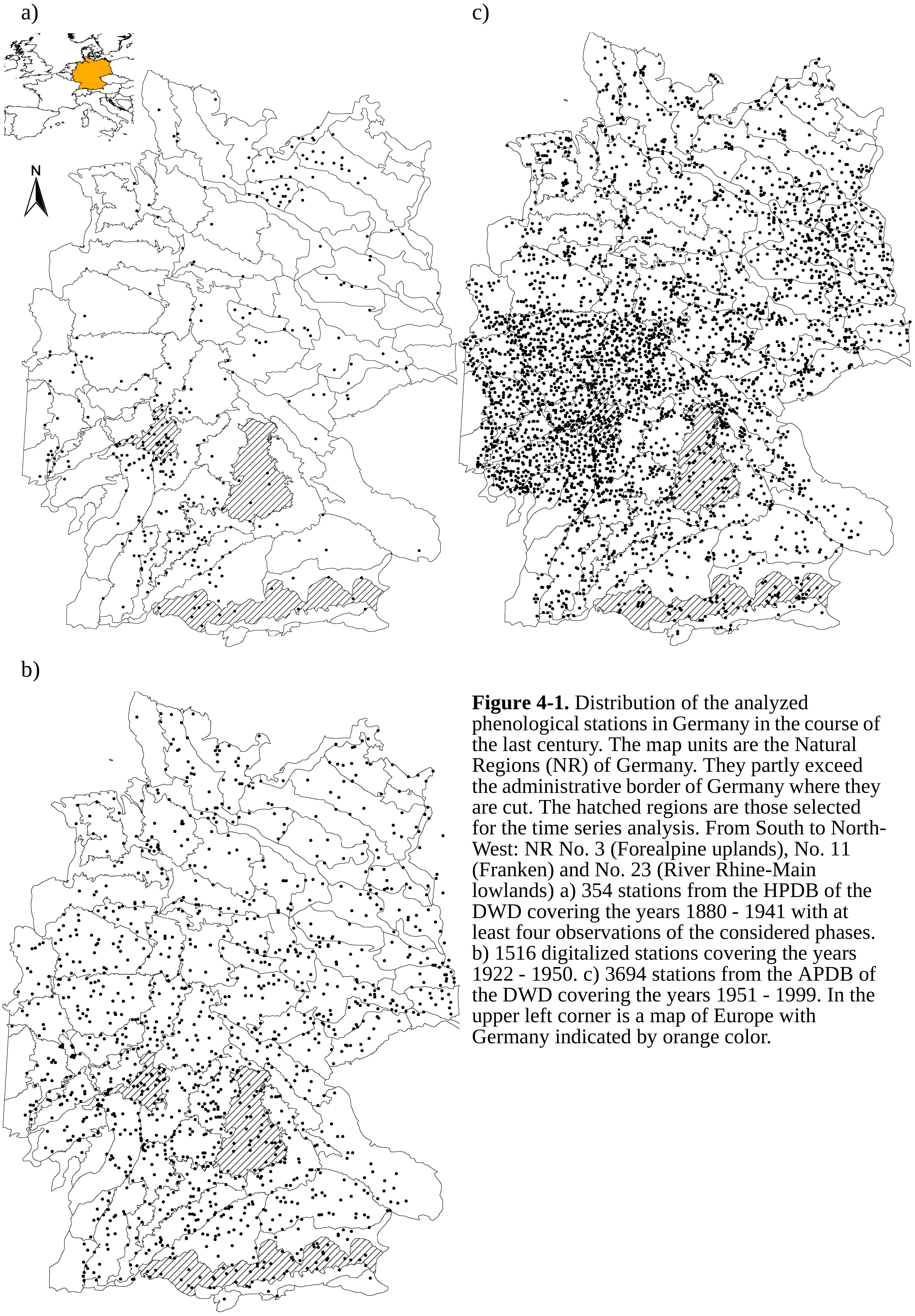


Figure 4-1. Distribution of the analyzed phenological stations in Germany in the course of the last century. The map units are the Natural Regions (NR) of Germany. They partly exceed the administrative border of Germany where they are cut. The hatched regions are those selected for the time series analysis. From South to North-West: NR No. 3 (Forealpine uplands), No. 11 (Franken) and No. 23 (River Rhine-Main lowlands) a) 354 stations from the HPDB of the DWD covering the years 1880 - 1941 with at least four observations of the considered phases. b) 1516 digitalized stations covering the years 1922 - 1950. c) 3694 stations from the APDB of the DWD covering the years 1951 - 1999. In the upper left corner is a map of Europe with Germany indicated by orange color.

The main data source was the DWD that provided digital data from its actual network that covers the years from 1951 until 1999. From this network 3694 stations were used that where distributed all over Germany (Fig. 4-1c)

Moreover, the historical phenological database (HPDB) of the DWD was used. This is also a digital database that comprises data from various sources and covers the years 1880 until 1941 (Plieninger et al. 1828 - 1920, Toepfer 1982 - 1910, Grossherzgl. statist. Bureau zu Schwerin 1860 - 1894, Ihne 1883 - 1941, Hoffman and Ihne 1884, Koepert 1891 - 1898, Ziegler 1891a, 1891b, 1904, Ihne 1892, Hertzer 1895, Schnelle 1950, Breitinger 1975, Frankhauser 1979 - 1983). From this database 354 stations were selected that were situated in Germany and had at least four observations of the phases being analyzed that are mentioned below. For the selected stations from the HPDB the geographical location was checked using a digital topological map of Germany.

To supplement the data for the time before 1951 and to fill the gap between 1941 and 1951 we digitalized phenological data that were available only in printed form. These data were collected by the volunteer network of the precursor of the DWD, the Deutscher Reichswetterdienst, and were published after the war (Schnelle and Witterstein 1952, 1964). These observations cover the years 1922 until 1944. Additionally, we digitalized the phenological data that were published between 1945 and 1950 in the meteorological yearbooks of the DWD and were not stored in the actual database of the DWD (DWD 1945- 1950). The position of the digitalized stations was also checked using a digital map of Germany. For the time from 1922 through 1950 data from 1516 phenological stations were digitalized.

Only the meteorological yearbooks of the former US-Zone in Germany covered the whole time span from 1945 until 1951, whereas the meteorological yearbooks of the other occupied zones started publishing later, e.g. the British-Zone started 1949. Thus continuous time series for the whole period from 1880 until 1999 could only be found for southern Germany.

The phases and species selected for the analysis cover most key phases of the different phenological seasons with a special emphasis on deciduous tree species. Moreover, only those NRs were selected that had the most complete and longest time series for several of these phases and species. As the optimal compromise between these two restrictions the NR No. 3 (Forealpine uplands), No. 11 (Franken) and No. 23 (River Rhine-Main lowlands) (Fig. 4-1, hatched regions) were selected with following species and phases:

- Early spring: Blossoming (B) of Snowdrop (*Galanthus nivalis* L.) and for the analysis for whole Germany also B Hazel (*Corylus avellana* L.) was included.
- Spring: Bud burst (BB) of European white birch (*Betula pendula* Roth.), common horse chestnut (*Aesculus hippocastanum* L.), European beech (*Fagus sylvatic*a L.) and English oak (*Quercus robur* L.).

- Early summer: B of *Aesculus hippocastanum*, common lilac (*Syringa vulgaris* L.), European black elder (*Sambucus nigra* L.) and for the analysis for whole Germany also mayshoot (M) of Scots pine (*Pinus sylvestris* L.) was included.
- Autumn: Leaf coloring (LC) of *Betula pendula*, *Aesculus hippocastanum*, *Fagus sylvatica* and *Quercus robur.*
- Length of vegetation period (VP) of *Betula pendula*, *Aesculus hippocastanum*, *Fagus sylvatica* and *Quercus robur.*

4.3 Methods

4.3.1 Combined time series

In the HPDB there were some phenological stations that also recorded observations in the APDB from the DWD. With the help of the DWD historical stations that were digitalized from paper could also be associated with actual phenological stations. Thus, time series could be produced with observations at the same location in the different data sets producing single station time series that spanned the whole century. These time series played a key role to connect the different phenological data bases. This interlocking of overlapping time series is a prerequisite to obtain long combined time series. This phenomenon of the data is called connectivity (Searle 1979) meaning that when two phenological time series are to be merged into one time series they must have at least one observation year in common to calculate station effects (Section 2.2.1 on page 11). Of course, it is unlikely that the same observer is continuously reporting phenological phases throughout the whole century. However, the data of the DWD does not specify when the observer has changed because station-ids are not associated with the observer within the database but rather with the location where the observations have been made. Systematical differences can be expected between observers. Nevertheless, when long time series over the whole century were reported under the same station-id, they were mathematically treated as observations of the same observer and the bias due to changing observers was neglected.

The combined time series y_i, $i = 1,..., N$ (number of observation years) was derived from a two-way mixed linear model (Eqn. 2-10 ”2.2.1 A common theoretical background” on page 11). After removal of observations with 30-day residuals of the robust L_1 estimation (see ”2.2.3 Robust estimation” on page 14) the parameters y_i were estimated using the T1 method (see ”2.2.2 Parameter estimation” on page 13). The whole procedure is described in detail in Chapter 2.

4.3.2 Trends and linear regressions

Trends are either reported in days per year and defined as the slope a of a standard linear regression $y = ax + b$ of some phenological phase y against the year of observation x, or they are reported as absolute change in number of days between end and beginning of the observation period and defined as *a* times the range of the time series. Both ways the significance of the trend is equal to the significance of the slope *a* of the linear regression that is tested against zero using the *t*-test (Sokal and Rohlf 1995). The time series were tested for a normal distribution with the Shapiro-Wilk-test.

As a further test of the reliability of the calculated combined time series it was checked whether the variation of the phenological time series could be explained by variations of the most important driving force of phenology, i.e. temperature (Schnelle 1952, Murray 1989, Worral 1993). To analyze how much of the phenological interannual variation was explained by corresponding temperatures a stepwise linear regression of the phenological times series of the three selected NRs against mean monthly temperatures of the NRs was conducted. Mean monthly temperatures for the NRs were obtained from the Climate Research Unit (CRU) data set (Jones 1994, Jones et al. 1999) covering the years 1901 until 1998, where for each NR the average mean monthly temperatures of all grid cells inside minimum and maximum latitudinal and longitudinal extent of the NR was used. For spring and summer phases (B and BB) the mean monthly temperatures from October of the preceding year until the average month of the phase were considered. For autumn phases (LC) the months of average BB until the month of average LC were considered. Only those parameters entered and stayed in the model that were significant at the five percent level. The stepwise linear regression analysis was conducted with the subroutine REG of the SAS statistical software (Software Release 6.13).

Mean monthly temperatures of the NRs are influenced by various meteorological factors. On large temporal and spatial scales these factors are summarized under the term climate and on smaller scales they are called weather. Thus, the functions of mean monthly temperatures obtained by the preceding regression analysis are also influenced by climate and weather. To determine to what extent a global climate signal or a regional circulation pattern signal can be detected within these functions of temperature, that in turn determine regional phenology, a stepwise linear regression analysis of the temperature functions against a global radiative forcing signal (Jones et al. 1986, Wigley 1987) and a regional circulation index (NAO, normalized pressure index between Iceland and Azores) (Hurrell and van Loon 1997, Jones at al. 1997, Rogers 1997, Hurrel et al. 2001) was performed. Global warming has been mainly attributed to the anthropogenic greenhouse effect (Ramaswamy et al. 2001). Global atmospheric CO_2-equivalent concentrations as proxy for the anthropogenic radiative forcing have been statistically significantly correlated to the global warming signal (Grieser et al. 2000). Thus, we followed Grieser et al. (2000) and used the CO_2-equivalent concentrations normalized by the preindustrial value (279 ppm) as a global signal for the anthropogenic greenhouse effect (Fig. 4-2) that is also hypothesized to cause the recent observed trends in phenology through climate warming. As an index of regional circulation pattern different versions of the NAO index were used. The respective parameter only entered and stayed the model if it was significant at the 5% level.

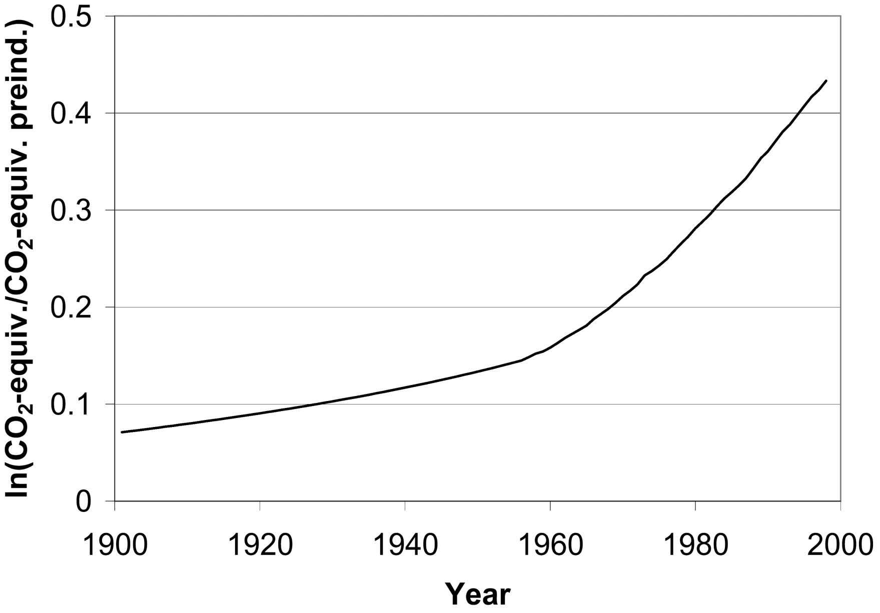


Figure 4-2. Yearly mean of the natural logarithm of the CO_2-equivalent concentrations normalized by the preindustrial value (279 ppm) (Wigley 1987, Grieser et al. 2000)

4.3.3 Trend turning points

Trend turning points were detected using a modified version of the sequential Mann-Kendall test (Sneyers 1990) that was also applied in climatological studies (Gerstengarbe and Werner 1999, Böhm et al. 2001). The normalized values of the Kendall's τ (Kendall and Gibbons 1990) were calculated progressively for N-1 subseries of the original time series $y_1, y_2, \ldots, y_N$ (thus, the first subseries is composed of the values y_1, y_2, the second of the values y_1, y_2, y_3, etc.). The progressive row τ_i, $i = 2,\ldots,N$, is calculated for each of these subseries as

$$\tau_i = (S_i - 1)/\sqrt{VAR(S_i)} \qquad \text{for } S_i > 0,$$

$$\tau_i = 0 \qquad \text{for } S_i = 0,$$

$$\tau_i = (S_i + 1)/\sqrt{VAR(S_i)} \qquad \text{for } S_i < 0,$$

with $S_i = \sum_{k=2}^{i-1} \sum_{j=k}^{i} sign(y_j - y_k)$ and $sign(x) = \begin{cases} 1 & \text{for } x > 0 \\ 0 & \text{for } x = 0 \\ -1 & \text{for } x < 0 \end{cases}$ and

$$VAR(S_i) = \frac{1}{18}\left[i(i-1)(2i+5) - \sum_{b_i}^{g_i} t_{b_i}(t_{b_i} - 1)(2t_{b_i} + 5) \right],$$

where g_i is the number of tied groups in the i-th subseries and t_{bi} the number of observations in the b_i-th tied group of subseries i. This results in a time series τ_i for $i = 2,\ldots,N$. Then a retrograde time series is calculated with the same procedure only for the reverse of the original rank series.

Sneyers (1990) points out that the intersections of the progressive and the retrograde time series mark approximate turning points of trend development within the original time series. These turning points are significant when they lie between two points where the progressive and the retrograde row exceed a predefined significance value. However, the sequential version of the Mann-Kendall test is susceptible to noise in the original time series, i.e. the more variation there is the less clearer become possible turning points by either finding too many or no intersections at all. Of course, the latter also happens when there are indeed no turning points within the time series.

After trend turning points were found the trends of several phenological phases between those points were analyzed over the last century in the three selected NRs.

Another focus was to see whether the trends and their change in time varied consistently not only over a long time span in the three selected NRs but also between all NRs in Germany. Because complete combined time series for all NRs in Germany were only available starting from 1951, differences among all NRs in Germany were only considered for the time frame 1951 until 1999.

4.3.4 Potentially absorbed radiation

The yearly gross primary production (GPP) per tree is an approximately linear function of the intercepted radiation over the vegetation period (Montheith 1972, 1977). The intercepted radiation is an approximately linear function of the incoming radiation (Ruimy et al. 1991). Thus, the average incoming daily radiation measured at 21 stations all over Germany over the period 1951 - 2000 can be assumed to approximate linearly the daily potentially intercepted radiation and also the potential daily GPP over that period on average (Fig. 4-3).

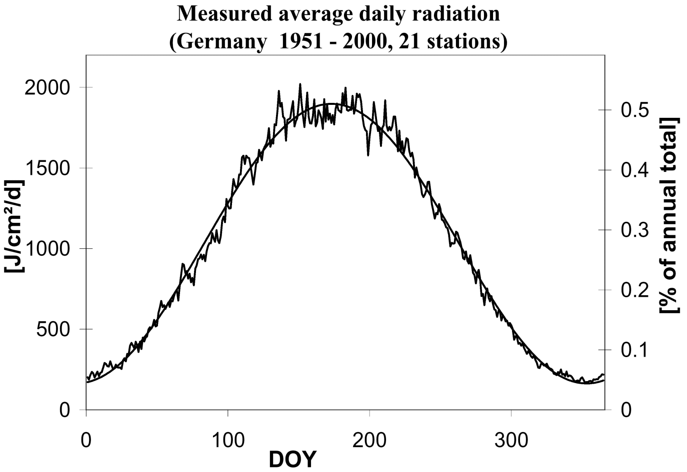


Figure 4-3. Average daily radiation measured at 21 stations all over Germany during the period 1951 through 2000 in J/cm^2/d and as % of the annual integral. DOY: Julian day of year. The smooth curve is a sine function fitted to the measured data (R^2=0.99).

The measured average incoming daily radiation was fitted to a sine curve that was used to calculate the daily potentially absorbed radiation over the vegetation period. The integral of this sine curve over the vegetation period was used as a proxy of the average yearly GPP an important determinant for the yearly carbon balance of the living tree. The impact of the change of the length of the vegetation period on average yearly GPP was investigated. The amount of potentially available light was also used by Kramer (1995) to evaluate the difference between phenological tree types.

4.4 Results

4.4.1 Combined time series

Even though gaps in individual time series (years without observation) are filled by the process of combination there still were some combined time series with remaining gaps. These gaps resided all before the year 1951 and mostly around 1945. These gaps were filled by the average of corresponding values from adjacent NRs corrected for their average deviation from the time series to be completed. Especially for the leaf coloring (LC) phases, gaps around 1945 were sometime large and could not be filled with data from adjacent NRs. To be able to follow the course of the length of the vegetation period over the century the LC time series remained in the analysis. For further analysis only those time series were considered where less than ten percent of the time series was generated by gap filling. For a summary of the resulting times series, their observational time span and their completeness before gap filling (see Tab. 4-1).

Some combined time series were not normally distributed according to the Shapiro-Wilk-test ($P < 0.05$). For the LC phases this is certainly due to a marked shift in the observation level in the middle of the 20th century. For the other phases visual inspection is more difficult (see Appendix A1 for figures of the combined time series). Applying the distribution-free Mann-Kendall trend test did not change the significance levels of the trends in most cases. Therefore the t-test was used throughout this chapter. As can be seen in Fig. 4-1, the further we go back in time the less phenological stations are available. This has consequences for the quality of the estimated combined time series per phase and NR. One measure for the reliability of the estimated combined times series are its estimated confidence intervals that are a function of the number of data available each year and the estimated variance components for the error and station variance (see ”2.2.1 A common theoretical background” on page 11).

Table 4-1. Phenological phases for the three selected Natural Regions No. 3, No. 11 and No. 23 (Fig. 4-1),their coverage during the observational time span and number of values within this time span (N). Average observation DOY with its standard deviation, Completeness before gap filling and number of filled data gaps. Averages marked with an asterix (*) indicate time series that were not normally distributed according to the Shapiro-Wilk-test ($P < 0.05$). For the LC phases the number of remaining gaps (years without observations) is given in brackets. For diagrams see appendix A1.

Natural Region	Phase	Coverage	N	Avg±σ	Complete-ness [%]	No. filled gaps
No. 3	B *G. nivalis*	1895-1999	105	65 ± 11.9	95.2	5
	BB *A. hippocastanum*	1890-1999	110	117 ± 7.2	100	0
	BB *B. pendula*	1890-1999	110	118 ± 8.1	93.6	7
	BB *F. sylvatica*	1890-1999	110	122 ± 5.4	98.2	2
	BB *Q. robur*	1890-1999	110	131 ± 6.7	91.8	9
	B *A. hippocastanum*	1890-1999	110	135 ± 6.9	100	0
	B *S. vulgaris*	1890-1999	110	135 ± 7.5	100	0
	B *S. nigra*	1890-1999	110	160 ± 7.7	100	0
	LC *A. hippocastanum*	1890-1999	110	281 ± 6.1	100	0
	LC *B. pendula*	1890-1924 1935-1999	99	279* ± 9.0	100	0 (10)
	LC *F. sylvatica*	1890-1941 1945-1999	107	285 ± 6.2	94.4	6 (3)
	LC *Q. robu*r	1890-1941 1945-1999	107	292* ± 8.2	94.4	6 (3)
No. 11	B *G. nivalis*	1935-1999	65	62 ± 11.9	100	0
	BB *A. hippocastanum*	1883-1999	117	111 ± 9.1	99.1	1
	BB *B. pendula*	1883-1999	117	113 ± 8.2	91.5	10
	BB *Q. robur*	1883-1999	117	129 ± 6.3	90.6	11
	B *A. hippocastanum*	1883-1999	117	132 ± 7.0	98.3	2
	B *S. vulgaris*	1883-1999	117	133 ± 6.9	100	0
	B *S. nigra*	1880-1999	120	159* ± 9.3	97.5	3
	LC *A. hippocastanum*	1883-1999	117	280 ± 6.2	94.9	6
	LC *B. pendula*	1883-1928 1936-1999	110	286 ± 7.5	94.5	6 (7)
	LC *Q. robu*r	1883-1941 1945-1999	114	290 ± 6.7	90.4	11 (3)

Table 4-1. (continued)

Natural Region	Phase	Coverage	N	Avg±σ	Complete-ness [%]	No. filled gaps
No. 23	B *G. nivalis*	1893-1999	108	51 ± 12.5	99.1	1
	BB *A. hippocastanum*	1882-1999	118	103* ± 8.0	99.2	1
	BB *B. pendula*	1880-1999	120	103 ± 7.8	99.2	1
	BB *F. sylvatica*	1880-1999	120	112 ± 6.7	96.7	4
	BB *Q. robur*	1882-1999	118	117 ± 7.0	96.6	4
	B *A. hippocastanum*	1880-1999	120	123* ± 6.9	99.2	1
	B *S. vulgaris*	1880-1999	120	123* ± 7.6	99.2	1
	B *S. nigra*	1880-1999	120	147* ± 6.6	99.2	1
	LC *A. hippocastanum*	1880-1999	120	280 ± 5.8	97.5	3
	LC *B. pendula*	1880-1999	120	285*± 11.2	95.8	5
	LC *F. sylvatica*	1880-1941 1945-1999	117	288 ± 6.9	96.6	4 (3)
	LC *Q. robu*r	1882-1941 1945-1999	115	290 ± 5.7	98.3	2 (3)

To obtain an overall picture of the reliability of the estimated combined time series the ranges of the 95% confidence intervals (RCIs) where pooled over all phases per year for spring (and early summer) phases and autumn phases, respectively. Fig. 4-4 and Fig. 4-5 show the median, the 5% and 95% quantiles of the RCIs for spring and autumn phases (Tab. 4-1), respectively. There is a clear distinction between the reliability of spring phases (B and BB) and autumn phases (LC) that is due to the lower amount of available data for autumn phases. The median of the RCIs of spring phases varies between ten and 25 days for the period from 1880 through 1922, decreases from 15 to five days between 1922 and 1951 and stays at around five days from 1951 through 1999. The RCIs for the autumn phases is almost twice as large at the beginning of the considered time period as for the spring phases. Between 1880 and 1922 the RCIs vary between 20 and 40 days. Between 1922 and 1950 the RCIs decrease from about 20 days to around 8 days, where they remain until 1999. The combined time series for LC and therefore also the derived time series for the length of the vegetation period (VP) for the three natural regions are unreliable for the period between 1880 and 1922.

Fig. 4-6, Fig. 4-7, Fig. 4-8 and Fig. 4-9 show examples of resulting combined time series with their respective 95% confidence intervals for B of *G. nivalis* in NR No. 23, BB of *A. hippocastanum* in NR No. 23, B of *S. vulgaris* in NR No. 3 and LC of *A. hippocastanum* in NR. No. 11, respectively. A compilation of diagrams of all derived and analyzed combined time series of the three NRs can be found in the appendix A1. From visual inspection it can be seen that there are no clear overall linear trends in the estimated combined time series of Fig. 4-6 to Fig. 4-9.

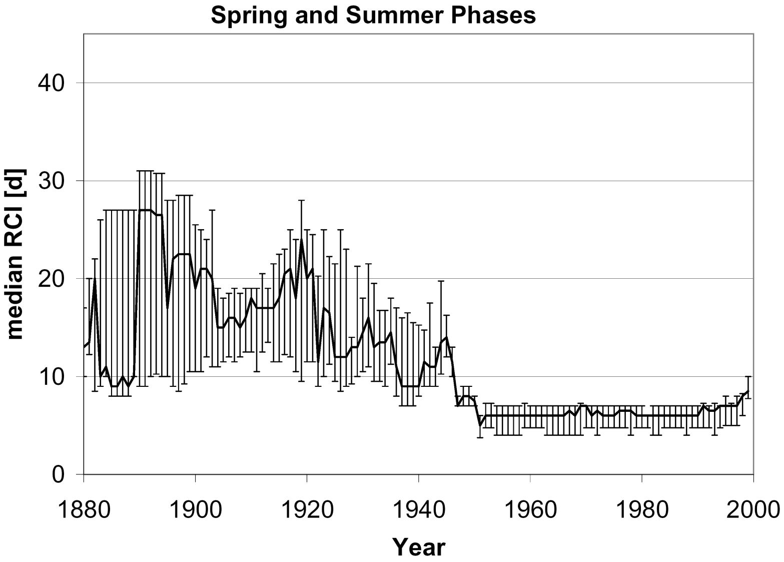


Figure 4-4. Median of the RCIs for the combined time series of Tab. 4-1 with 5% and 95% quantiles as error bars. Spring and early summer phases (B and BB).

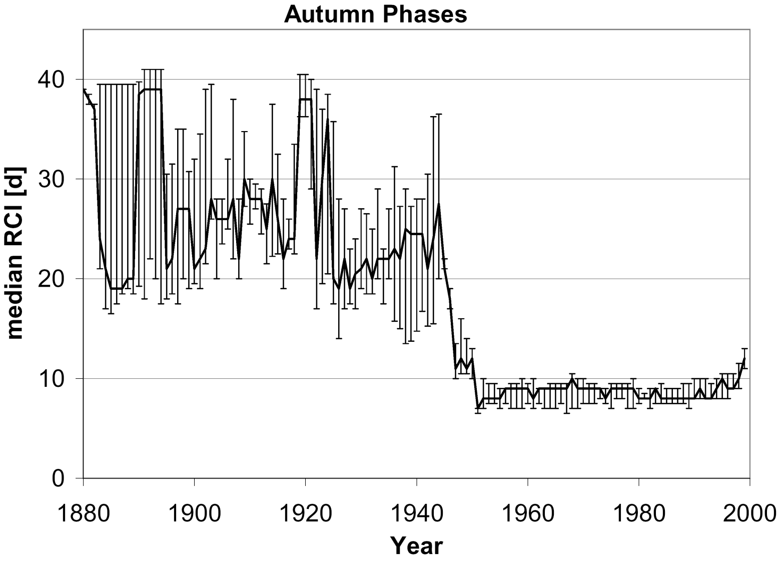


Figure 4-5. Median of the RCIs for the combined time series of Tab. 4-1 with 5% and 95% quantiles as error bars. Autumn phases (LC).

4.4.2 Overall trends

Trends as reported in recent literature (advancement of spring phases, delay of autumn phases) can be detected by visual inspection for the last decades but over the course of the whole century the trends of the last decades reveal themselves as parts of larger fluctuations. For B of *G. nivalis* in NR No. 23 (Fig. 4-6) the trends observed in the last decades can also be found between 1890 and 1918. Advancement of B of *S. vulgaris* in NR No. 3 (Fig. 4-8) in the last decades does also not seem to be extraordinary as it can be seen that around 1949 this phase was over a week earlier than in the 1990s.

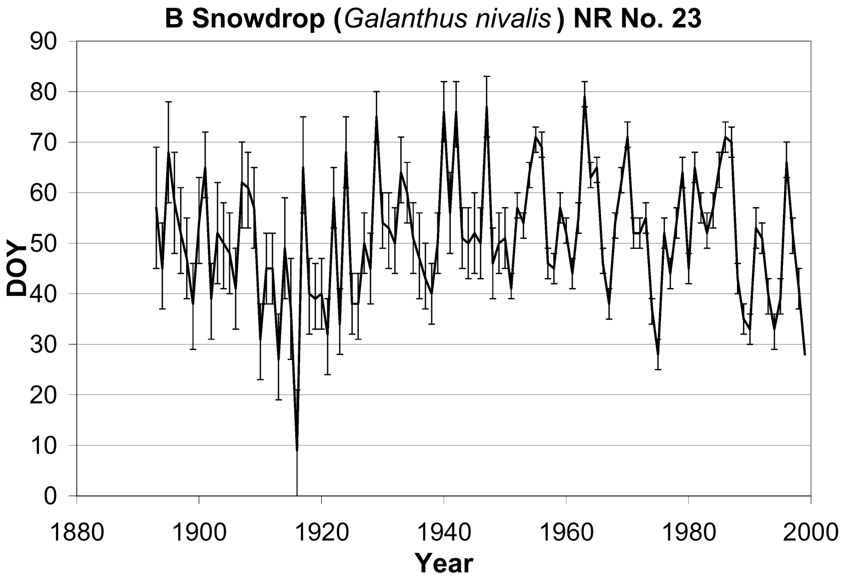


Figure 4-6. Combined time series for blossoming (B) of snowdrop (*Galanthus nivalis* L.) for the NR No. 23. Error bars show the estimated 95% confidence intervals. Values in Julian day of year (DOY). Early spring phase.

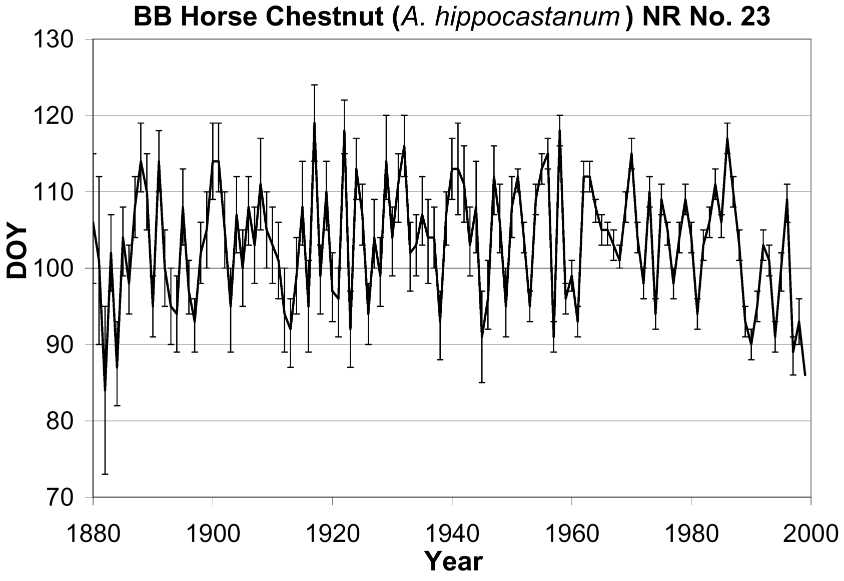


Figure 4-7. Combined time series for bud burst (BB) of Horse Chestnut (*Aesculus hippocastanum* L.) for the NR No. 23. Error bars show the estimated 95% confidence intervals. Values in Julian day of year (DOY). Spring phase.

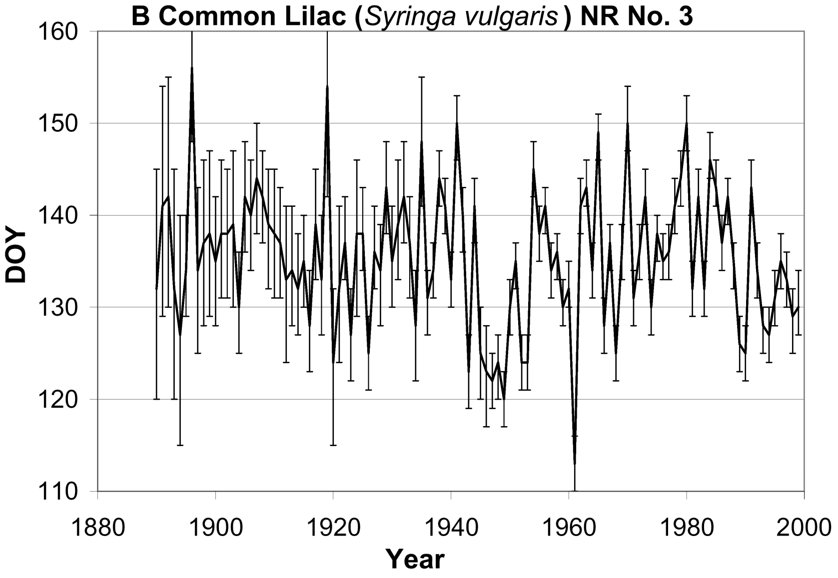


Figure 4-8. Combined time series for blossoming (B) of Lilac (*Syringa vulgaris* L.) for the NR No. 3. Error bars show the estimated 95% confidence intervals. Values in Julian day of year (DOY). Early summer phase.

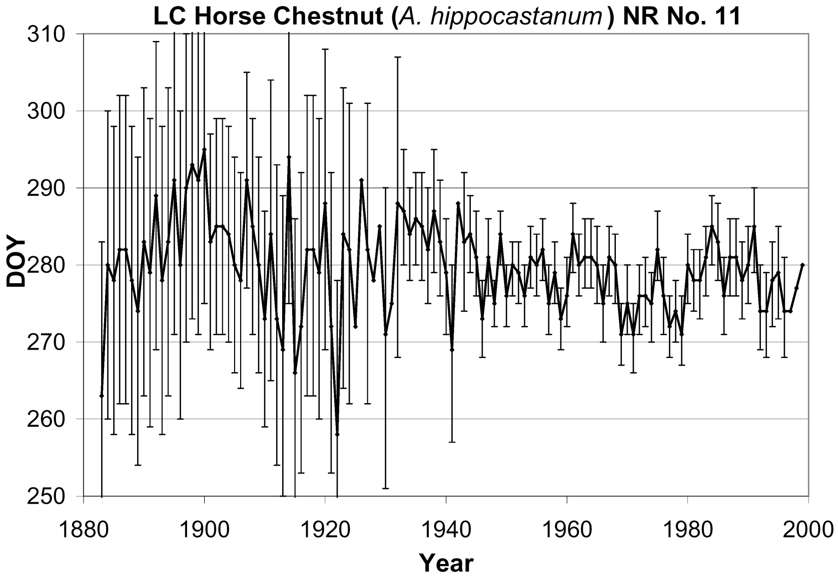


Figure 4-9. Combined time series for leaf coloring (LC) of Horse Chestnut (*Aesculus hippocastanum* L.) for the NR No. 11. Error bars show the estimated 95% confidence intervals. Values in Julian day of year (DOY). Autumn phase.

A similar picture can be seen for BB of *A. hippocastanum* in NR No. 23 (Fig. 4-7). Here the early BB of the 1990s was observed in other decades, too, even though reliability of the estimated time series decreases the further we go back in time. The latter notion is especially pronounced for the LC phase of *A. hippocastanum* in NR No. 11 (Fig. 4-9). In the period between 1880 and 1930 large confidence intervals coincide with large interannual variations such that statements over the whole time period between 1880 and 1999 are impossible. As can be seen in the appendix A1 not only confidence intervals of the autumn phases are large before 1950 but also in many LC time series a shift from a higher level before 1950 to a lower level after 1951 occurs.

Trends were calculated for all combined series of Tab. 4-1 for the whole period that was available (Tab. 4-2). Most time series showed negative trends. This was not only the case for spring and early summer phases (B and BB) but also for leaf coloring phases (LC). The LC phases even showed a more pronounced advancement than BB phases such that the length of the vegetation period (VP), calculated as the difference time series of BB and LC per tree species, also showed significant negative trends. For the more reliable B and BB phases the trends are not consistent among the NRs. Not only significances change but also the sign of the trends. Trends of B of *G. nivalis* is positive for NR 23 but negative for the NRs 3 and 11. Trends of BB of *B. pendula* are positive for NR 11 and negative for the NRs 3 and 23. Trends of BB of *F. sylvatica* are positive but not significant.

Table 4-2. Overall trends [days/year] of all analyzed phases (Tab. 4-1) for the three NRs for the whole observational period (for coverage of the observational period see Tab. 4-1). Significances of the trends are indicated by * (P < 0.1), ** (P < 0.05), *** (P < 0.01).

Phase	Trends NR No. 3	Trends NR No. 11	Trends NR No. 23
B *G. nivalis*	-0.097**	-0.24**	0.026
BB *A. hippocastanum*	-0.055**	-0.027	-0.026
BB *B. pendula*	-0.081**	0.034	-0.005
BB *F. sylvatica*	0.001	-	0.022
BB *Q. robur*	-0.103***	-0.006	-0.006
B *A. hippocastanum*	-0.03	-0.057**	-0.041**
B *S. vulgaris*	-0.024	-0.028	-0.026
B *S. nigra*	-0.102***	-0.087***	-0.04**
LC *A. hippocastanum*	-0.119***	0.044***	-0.077***
LC *B. pendula*	-0.166***	-0.103***	-0.094***
LC *F. sylvatica*	-0.093***	-	-0.056***
LC *Q. robu*r	-0.158***	-0.033*	-0.048***
VP *A. hippocastanum*	-0.067***	-0.07**	-0.073***
VP *B. pendula*	-0.097**	-0.082**	-0.069**
VP *F. sylvatica*	-0.094***	-	-0.077***
VP *Q. robu*r	0.055**	-0.026	-0.041

4.4.3 Stepwise regression analysis

As could be seen in Fig. 4-4 and Fig. 4-5 the reliability of the estimated combined time series increases substantially after 1951. To see whether the results of the stepwise linear regression analysis was influenced by the decreasing reliability of the time series before 1951, two analyses were conducted : a) taking all data into account starting 1901, when the temperature time series started and b) starting 1951. The values of the parameters of the stepwise linear regression changed only slightly when the analysis started in 1901 or 1951.

Table 4-3. : Summary of stepwise linear regression of phenological combined times series against mean monthly temperatures for each NR pooled over all NRs. R^2_{1901} is the overall coefficient of determination using the CRU temperature data from 1901 through 1998, R^2_{1951} is the overall coefficient of determination using the CRU temperature data from 1951 through 1998, for each month the parameters are given for the regression with temperature data from 1951. Bold values indicate those parameters that had the largest partial R^2. R^2_{NAO} and R^2_{CO2} are the partial regression coefficients of the regression of the function of mean monthly temperatures describing the phenology as indicated in the table against a NAO-index and global CO_2-equivalent concentration, respectively.

Phase	R^2_{NAO}	R^2_{CO2}	R^2_{1901}	R^2_{1951}	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep
B *G. nivalis*	0.32	-	0.74	0.76			-1.33	-1.73	**-2.76**							
BB *A. hippocastanum*	0.22	-	0.7	0.78					-0.8	**-2.51**	-2.29					
BB *B. pendula*	0.30	-	0.68	0.76					0.82	**-2.48**	-2.44					
BB *F. sylvatica*	0.25	-	0.75	0.82				-0.56	-0.55	-1.05	**-2.34**	-0.55				
BB *Q. robur*	0.23	-	0.66	0.75	-1.01			-0.55	-0.31	-1.14	**-2.84**	-0.96				
B *A. hippocastanum*	0.15	-	0.75	0.83	-0.5			-0.46	-0.33	-1.14	**-3.37**	-1.34				
B *S. vulgaris*	0.19	-	0.79	0.86	-0.5			-0.48	-0.56	-1.41	**-3.52**	-1.3				
B *S. nigra*	0.16	-	0.74	0.87	-0.73			-0.37	-0.71	-1.35	-2-23	**-2.69**				
LC *A. hippocastanum*			0.2	0.42								**-0.88**	-0.76			1.11
LC *B. pendula*			0.01	0.17											**1.03**	0.9
LC *F. sylvatica*			0.16	0.48								-0.54	-0.63		0.8	**1.79**
LC *Q. robur*			0.23	0.37							-0.52	-0.44	0.74		1.03	**1.13**
VP *A. hippocastanum*			0.52	0.75					0.72	**2.06**	2.3		-1.21		1.15	1.09
VP *B. pendula*			0.48	0.66					0.74	**2.07**	2.65				1.83	1.0
VP *F. sylvatica*			0.5	0.8	0.72	0.72		0.85	0.46	0.9	**2.24**		-0.84		0.73	2.08
VP *Q. robur*			0.46	0.72	1.22			0.62	0.33	0.66	**2.36**		-0.69		1.53	1.77

Therefore only parameter values of analysis b) are displayed in Tab. 4-3. The order of the partial regression coefficients (a measure for the percentage of explained variance) for the parameters for each regression did not change between the two analyses. Negative parameters indicate an advancement of the respective phase with increasing monthly temperature and positive parameters therefore model a delay in the respective phase with increasing monthly temperature. The parameters with the highest partial regression coefficient, i.e. the mean monthly temperatures that had the highest explanatory power, are indicated by bold font in Tab. 4-3. For the early spring phase B of *G. nivalis*, the monthly mean temperature that had the largest influence was February. Generally, the later the mean date of the phase the later was the most influential month. For BB of *B. pendula* and *A. hippocastanum* mean March temperatures were most influential whereas for BB of *Fagus sylvatica* and *Q. robur* it was April temperatures that explained the largest part of the variations. The same mean monthly temperatures that had the highest partial regression coefficients for the spring phases had also the highest partial regression coefficients when the length of the vegetation period (VP) was to be explained. For LC is was September and August temperatures that explained most of the phenological variation except for *A. hippocastanum* where it was May temperatures. For LC the August and September temperatures were associated with positive parameters indicating a delay of LC with increasing late summer and early fall temperatures. May and June temperatures had negative parameters.

In general, the spring and early summer phenology could well be explained by linear functions of mean monthly temperatures increasing their credibility. The coefficients of determination varied between 0.66 and 0.75 for the regression starting 1901 (R^2_{1901}) and 0.75 and 0.87 for regression starting 1951 (R^2_{1951}). When temperature data were fitted to phenological time series starting from 1951 the explanatory power of mean monthly temperatures increased probably due to the more reliable time series since 1951. For early spring and spring phases the percentage of explained variance (R^2) increased from around 71% to 77%, for early summer phases it increased by 10% from 76% to 86% on average. Also autumn phases could be much better explained by the temperatures of the last 50 years, indicated by an increase of the R^2 values from around 0.15 to about 0.38 and almost 0.5 for LC of *F. sylvatica*. This again, on the one hand, hints to the earlier mentioned notion that the estimated time series for autumn phases before 1950 might not be as reliable as spring phases (see Fig. 4-5), on the other hand, the timing of autumn phases is less understood than the timing of spring phases and is agreed to depend less on temperature than spring phases (Schnelle 1952). For these reasons and for the fact that changes in spring phases play a more dominant role for ecological consequences as will be shown below, the subsequent analysis of trend development over the century was restricted to spring and early summer phases.

Temperature in Europe is influenced by large scale weather patterns (Grieser et al. 2000). Thus, the linear functions of temperature, that in turn determine to a large part spring phenology, are influenced by large scale weather patterns. Moreover, a global warming signal has been detected for average Northern hemisphere temperatures (Grieser et al. 2000). Therefore it was tried via a simple regression analysis whether circulation patterns and a global warming signal could be detected in the linear functions of temperatures starting 1901. A stepwise linear regression was conducted with the described CO_2-equivalent concentrations (Fig. 4-2), that could enter the model as a regressor first, and of different versions of the NAO index as the second regressor. It is common practice to take a several months mean of the NAO to describe atmospheric circulation patterns. The regression analysis was conducted with mean NAO values of December until March, January until April, February until April and January until March. In all regressions the mean January until April NAO index was strongest correlated with the temperature functions except B of *G. nivalis* where it was mean January until March NAO index. A significant signal of the CO_2-equivalents could not be found. The partial regression coefficients for the NAO indices ranged from 0.15 to 0.32 with no clear distinction between early and late spring phases.

4.4.4 Trend turning points

Results from Menzel et al. (2001) and Gornik (1995) indicate that for phenological data more than 20 years are necessary to get a reliable trend. However, the trends also depend on where the time series are situated within the overall time frame. In Fig. 4-10 this is demonstrated for some example long-term time series. For time series of 20 years within 49 years time series the strength as well as the sign of the trend strongly depends on where the 20 years are situated (Fig. 4-10a). This is true for all four phases displayed in Fig. 4-10. For 30 years time frames of 49 years time series the same effect can be observed but with a more systematic influence of the end year. For 30-year time series that end before 1990 trends are mostly positive whereas for 30-year time series that end after 1990 trends are mostly negative. For the example of eight 49 years time series of *Q. robur* (Fig. 4-10c) this systematic influence of the analyzed time period is strong. All trends for 30-year time frames with end year later than 1989 are negative, whereas 70% of the trends for time series with end year before 1990 are positive. In total we find 40% positive trends in Fig. 4-10c. Menzel et al. (2001) found 33% positive trends and 67% negative trends for 775 time series of BB of *Q. robur* with 30 or more years in Germany. The above illustration indicates that part of the balance between positive and negative trends found by Menzel et al (2001) can be explained by the fact that they did not cover the same observational period. Unfortunately this analysis could not be done for the same data used by Menzel et al. (2001), because they were not available at the time this thesis was written.

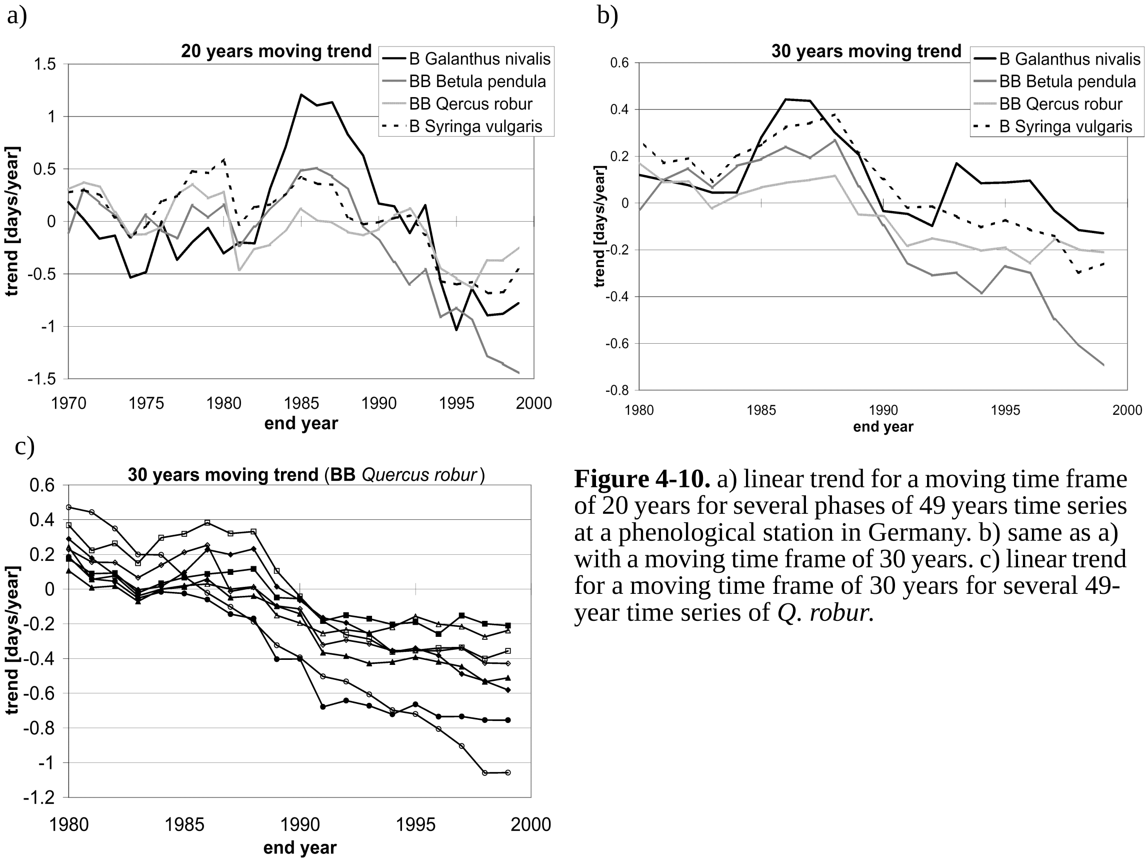


Figure 4-10. a) linear trend for a moving time frame of 20 years for several phases of 49 years time series at a phenological station in Germany. b) same as a) with a moving time frame of 30 years. c) linear trend for a moving time frame of 30 years for several 49-year time series of *Q. robur.*

It was concluded, however, that when phenological time series are compared they should cover the same observational period. Because the above derived century long time series had varying start years (they all had 1999 as end year) it was tried to detect common periods of marked trend behavior in which the trends could be easily compared among species and regions.

To determine common starting points for trend analysis for all considered phases the sequential version of the Mann-Kendall test was applied. From the regression analysis it became clear that not only spring and summer phases can be explained to a large part by linear functions of mean monthly temperatures but that it is March and April temperatures that largely determine spring phases, except for the early spring phase B of *G. nivalis* where mean February temperature has the largest influence. Because the sequential version of the Mann-Kendall test is susceptible to noise within the analyzed time series and the aim is to determine trend analysis periods for the ensemble of spring phenology the average of mean March and April temperatures was taken for the detection of possible trend turning points (Fig. 4-11). The sequential Mann-Kendall test only gives approximate trend turning points (Sneyers 1990) and is not reliable at the beginning of the time series due to small amount of available data in the first subseries. Thus, the intersections at the beginning and the end of the time series were neglected.

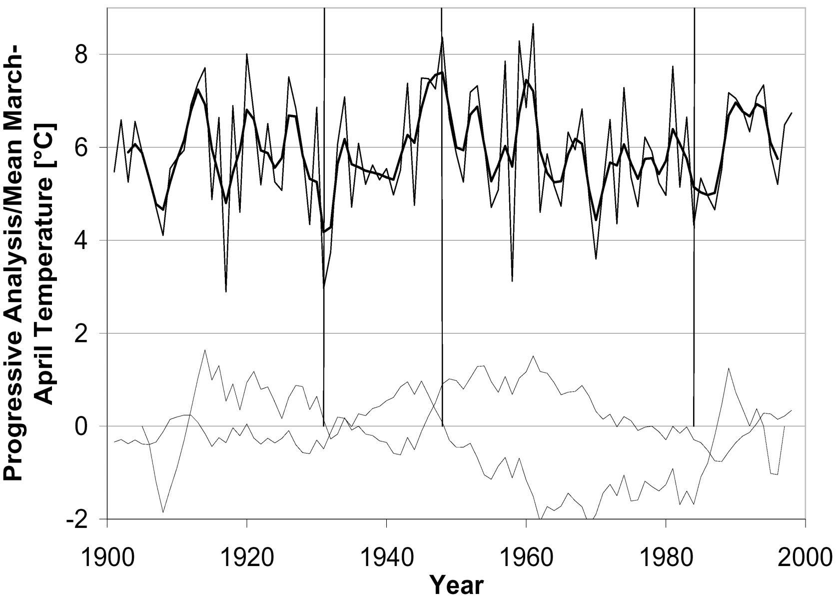


Figure 4-11. Average mean March and April temperature for the three selected NRs No. 3, 11 and 23. Bold line: time series smoothed with a 5 year gaussian filter. Vertical lines: detected trend turning points. Lower two lines: progressive and retrograde series from the sequential Mann-Kendall test

Three turning points were distinguished, i.e. 1931, 1948 and 1984. They were not or only weakly significant. Nevertheless, the three intersections of the progressive and retrograde rows seemed to mimic trend behavior of the mean March-April temperatures in the three NRs over the century. They were taken as clues for possible trend turning points of the associated phenology and to define normalized periods in which phenology can be analyzed and compared. Thus, four periods with possible different trend behavior were considered in the following: 1) until 1931, 2) 1931 - 1948, 3) 1948 - 1984 and 4) 1984 - 1999. The smoothed curve of Fig. 4-11 gives a visual support for these break points. Trends for spring and summer phases were calculated for these four periods for each of the three considered NRs (Fig. 4-12 - Fig. 4-14). Visual inspection of Fig. 4-11 can also give rise to the notion that already around 1970 a significant change of trend behavior took place. Therefore, an additional analysis was done with 1970 as trend turning point instead of 1984. In this analysis there were considerably less significant trends found in the period before as well as in the period after 1970 than when 1984 was taken as trend turning point (results not shown). This supports the results of the sequential Mann-Kendall test suggesting that the point in time when the developing trends became significant was in the mid 80s.

The comparison of the trends in the three NRs reveals that trends have significantly changed in the course of the last century (Fig. 4-12, Fig. 4-13 and Fig. 4-14). Trends for advanced spring and early summer phases as determined for the last 15 years have already been found in the first half of the last century in the 17 years between 1931 and 1948. This is obvious especially for the trends in NR No. 3 and No. 11. When we focus on trend developments of the last three periods we can see that the early spring phases B of *G. nivalis*, BB of *A. hippocastanum* and *B. pendula* always showed consistently negative trends that only changed in magnitude and significance.

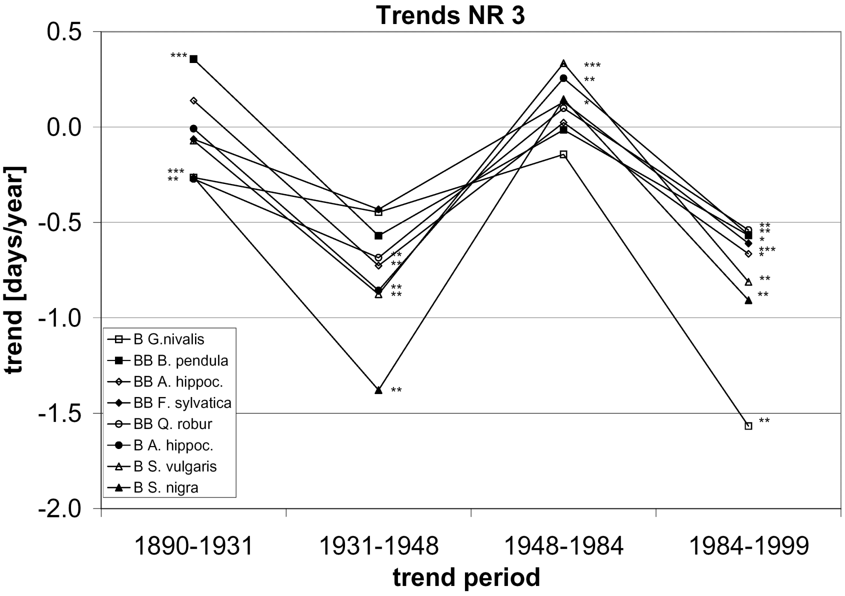


Figure 4-12. : Trends in NR No. 3 of spring and summer phases for the four periods detected by the sequential Mann-Kendall test. B Blossoming. BB Bud Burst. Stars indicate significance of trends: * (P < 0.1), ** (P < 0.05), *** (P < 0.01)

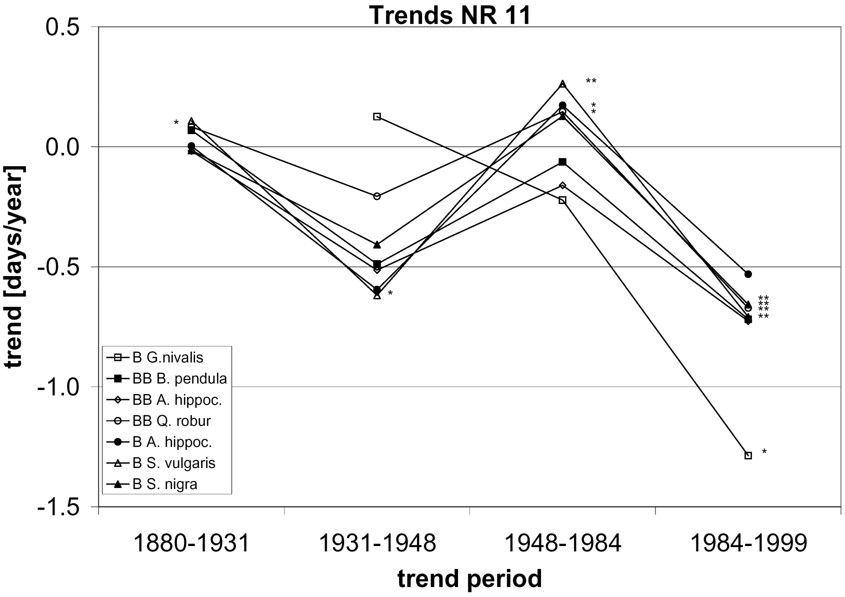


Figure 4-13. Trends in NR No. 11 of spring and summer phases for the four periods detected by the sequential Mann-Kendall test. B Blossoming. BB Bud Burst. Stars indicate significance of trends: * (P < 0.1), ** (P < 0.05), *** (P < 0.01)

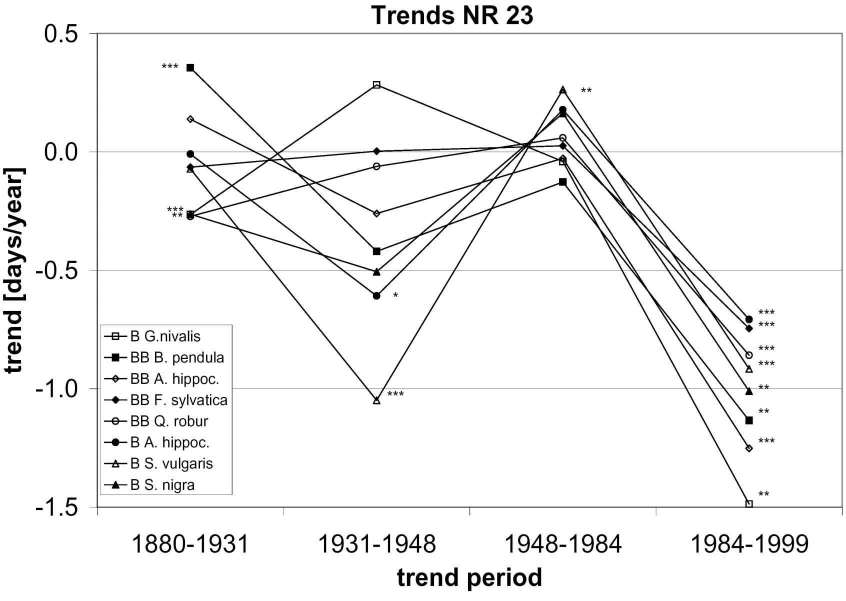


Figure 4-14. Trends in NR No. 23 of spring and summer phases for the four periods detected by the sequential Mann-Kendall test. B Blossoming. BB Bud Burst. Stars indicate significance of trends: * (P < 0.1), ** (P < 0.05), *** (P < 0.01).

The late spring and early summer phases BB of *F. sylvatica*, *Q. robur* and B of *A. hippocastanum*, *S. vulgaris* and *S. nigra* changed from significant negative trends to significant positive trends back to significant negative trends.

4.4.5 Regional differences

Similar trend turning points as in Fig. 4-11 were found when average March and April CRU temperatures for Germany were analyzed. Even though the picture of trend turning points for Germany was not as clear as for the three NRs analyzed above, a trend turning point could be determined in the mid 80s (results not shown). Because the earliest common starting point of phenological time series for all NRs in Germany was 1951, trends were calculated for the spring and summer and autumn phases mentioned in Section 4.3.1 for the two periods 1951 - 1984 and 1984 - 1999. In Fig. 4-15 to Fig. 4-18 the resulting trend maps for B *G. nivalis*, BB *B. pendula*, BB *Q. robur* and B *S. vulgaris* for all NRs in Germany are shown. Trends maps for all analyzed phases for the NRs in Germany (Tab. 4-4) are compiled in appendix A2. Blue colors indicate positive trends, red colors indicate negative trends. The significance of the trends is indicated by the density of white dots. NRs that were omitted because their combined time series did not cover the whole analyzed time span and where therefore not comparable are shown on white. For the trend analysis of for all NRs starting in year 1951 the VP was calculated as the combined times series of the differences series between BB and LC per year at each station due to the abundant data after 1951. Thus, in contrast to the time series of VP before confidence intervals were available.

The trend patterns of the three NRs (Fig. 4-12 - Fig. 4-14) were consistent all over Germany and mirror the general behavior also found for the three NRs above. The early spring phases B of *G. nivalis* and BB of *B. pendula* show consistently negative trends for most NRs that increase in magnitude and significance after the mid-1980ies. The late spring phase BB of *Q. robur* and the early summer phase B of *S. vulgaris* show significant positive trends before 1984 and significant negative trends after 1984 for most NRs. For all phases, however, there were a small number of exceptional NRs that did not fit into the general picture, especially before 1984. For the early spring phases B of *G. nivalis* and the late summer phase B of *S. vulgaris* there were only four out of 46 and one out of 52 NRs that showed trends of different sign than the average trend. For the more intermediate phases of BB of *B. pendula* and *Q. robur* there were 18 out of 52 and 11 out of 54 NRs that showed trends opposing to the mean. As can also be seen in appendix A2 for very early or very late spring phases the general notion is clearest that early phases have increased in magnitude and significance of negative trends in the last 15 year opposed to the period between 1951 and 1984, whereas late phases have flipped their trend behavior from significantly positive before 1984 to significantly negative after 1984. There was no clear regional pattern of NRs that did not follow the mean trend.

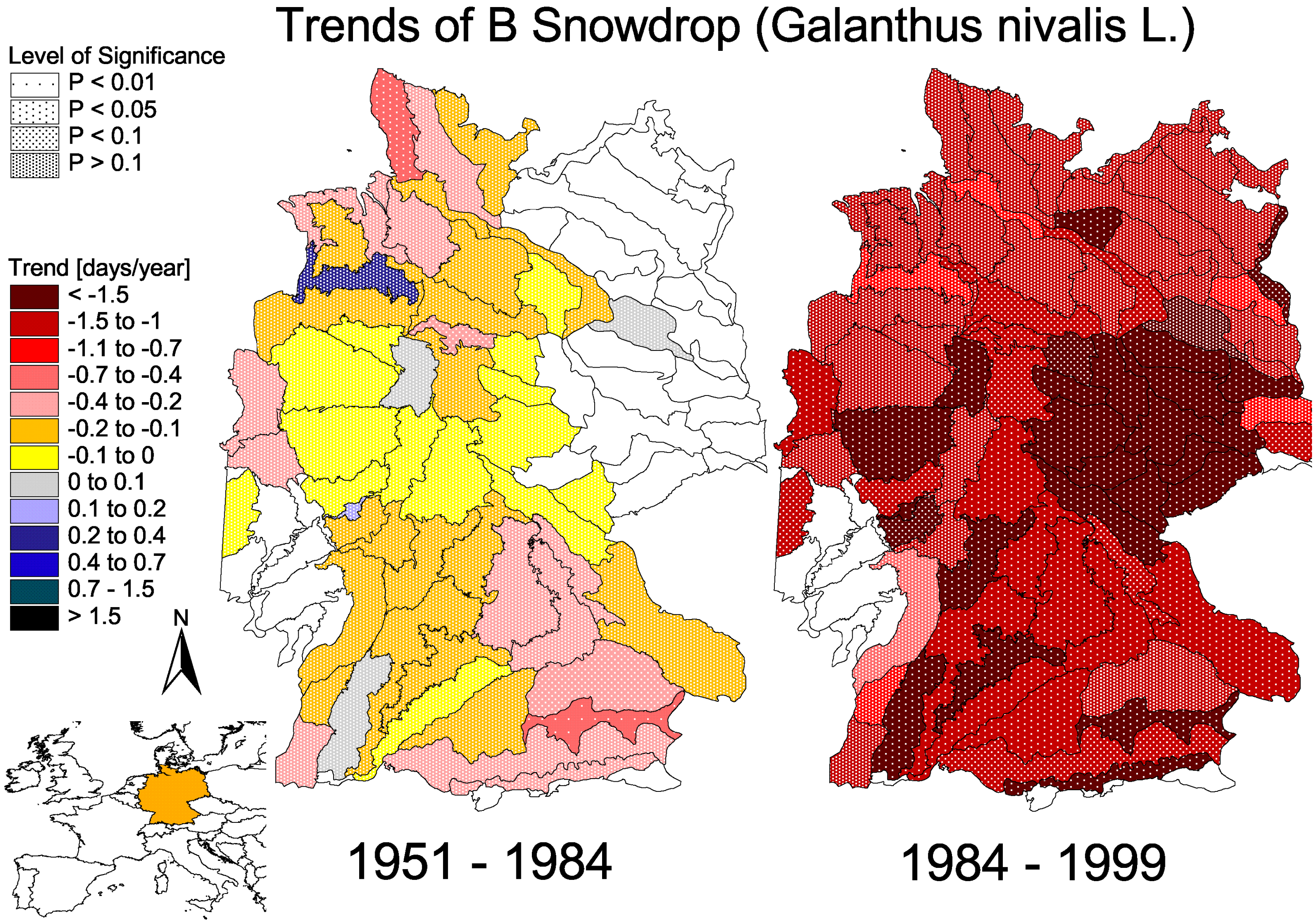


Figure 4-15. Trends of combined time series of blossoming of snowdrop (*Galanthus nivalis*) in Natural Regions in Germany that cover the period 1951-1984 or 1984-1999.

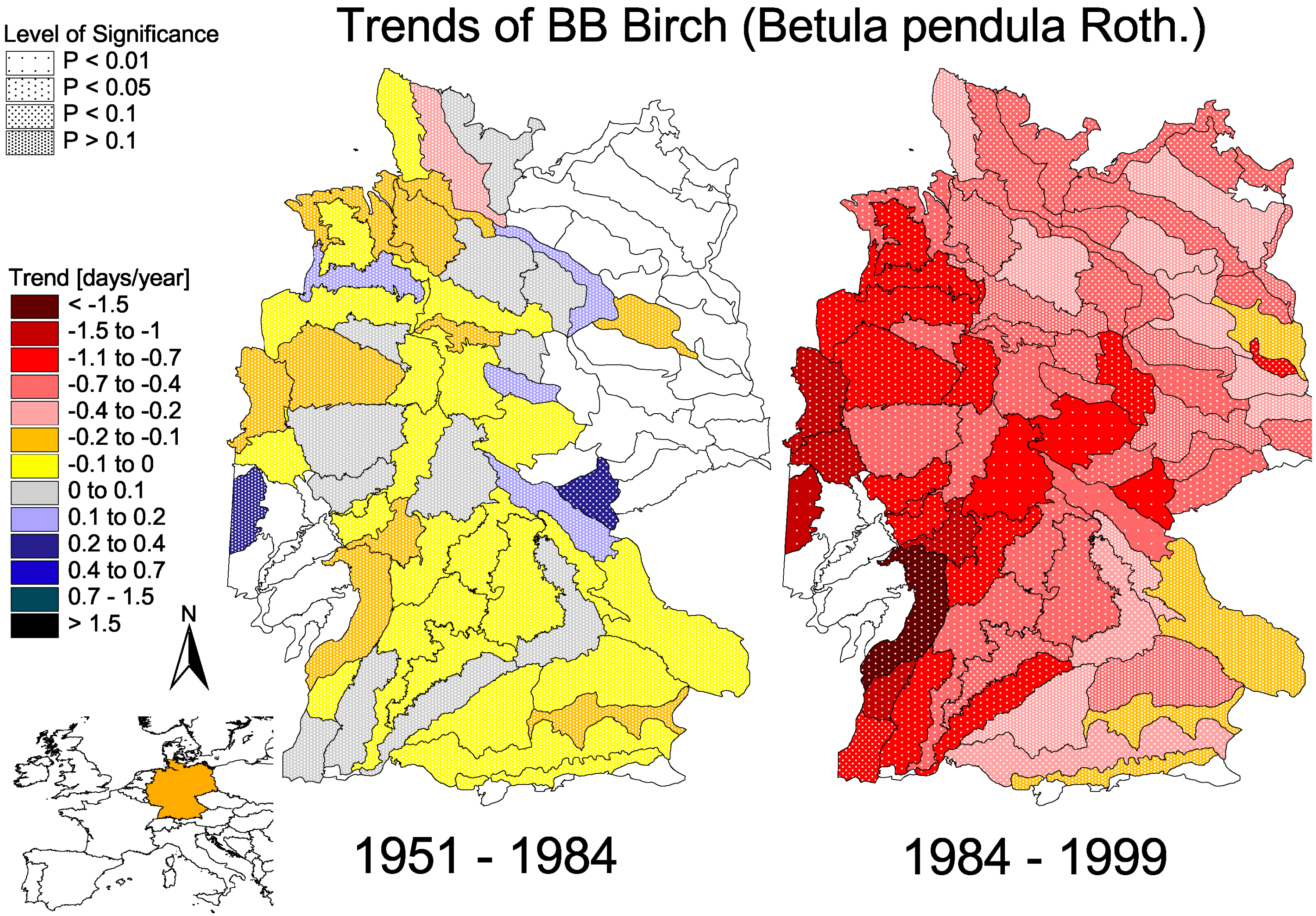


Figure 4-16. Trends of combined time series of bud burst of birch (*Betula pendula*) in Natural Regions in Germany that cover the period 1951-1984 or 1984-1999.

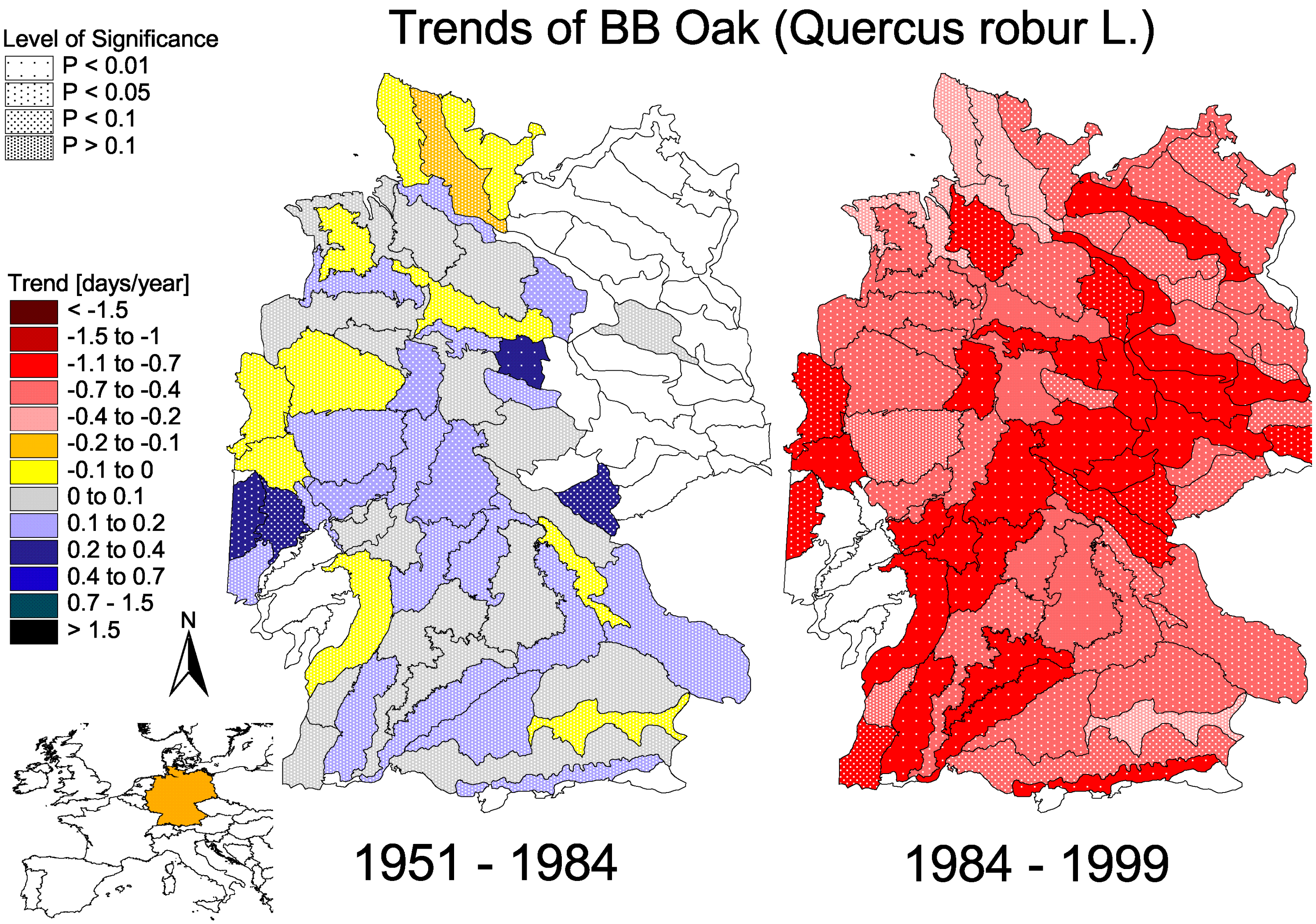


Figure 4-17. Trends of combined time series of bud burst of oak (*Quercus robur*) in Natural Regions in Germany that cover the period 1951-1984 or 1984-1999.

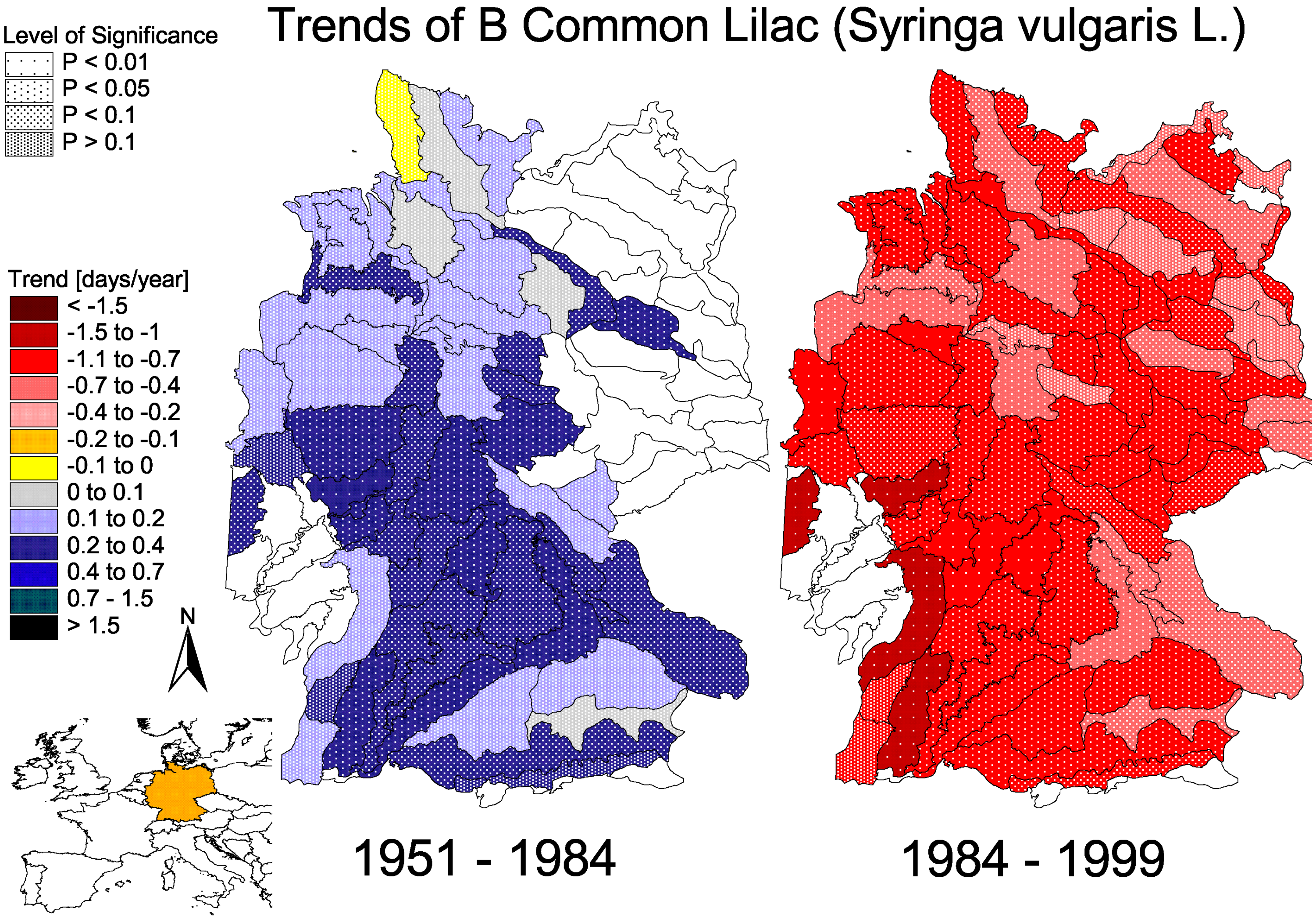


Figure 4-18. Trends of combined time series of blossoming of lilac (*Syringa vulgaris*) in Natural Regions in Germany that cover the period 1951-1984 or 1984-1999.

In Fig. 4-19 and Fig. 4-20 we see histograms of trends and their significances of BB, LC and VP of *A. hippocastanum* and *Q. robur,* respectively, for the NRs in Germany that cover the whole range of the periods 1951 through 1984 and 1984 through 1999 as shown in Fig. 4-16 and Fig. 4-17 and in the respective figures in appendix A2. Additional histograms for trends of BB, LC and VP of *B. pendula* and *F. sylvatica* for the NRs in Germany can be found in appendix A3.

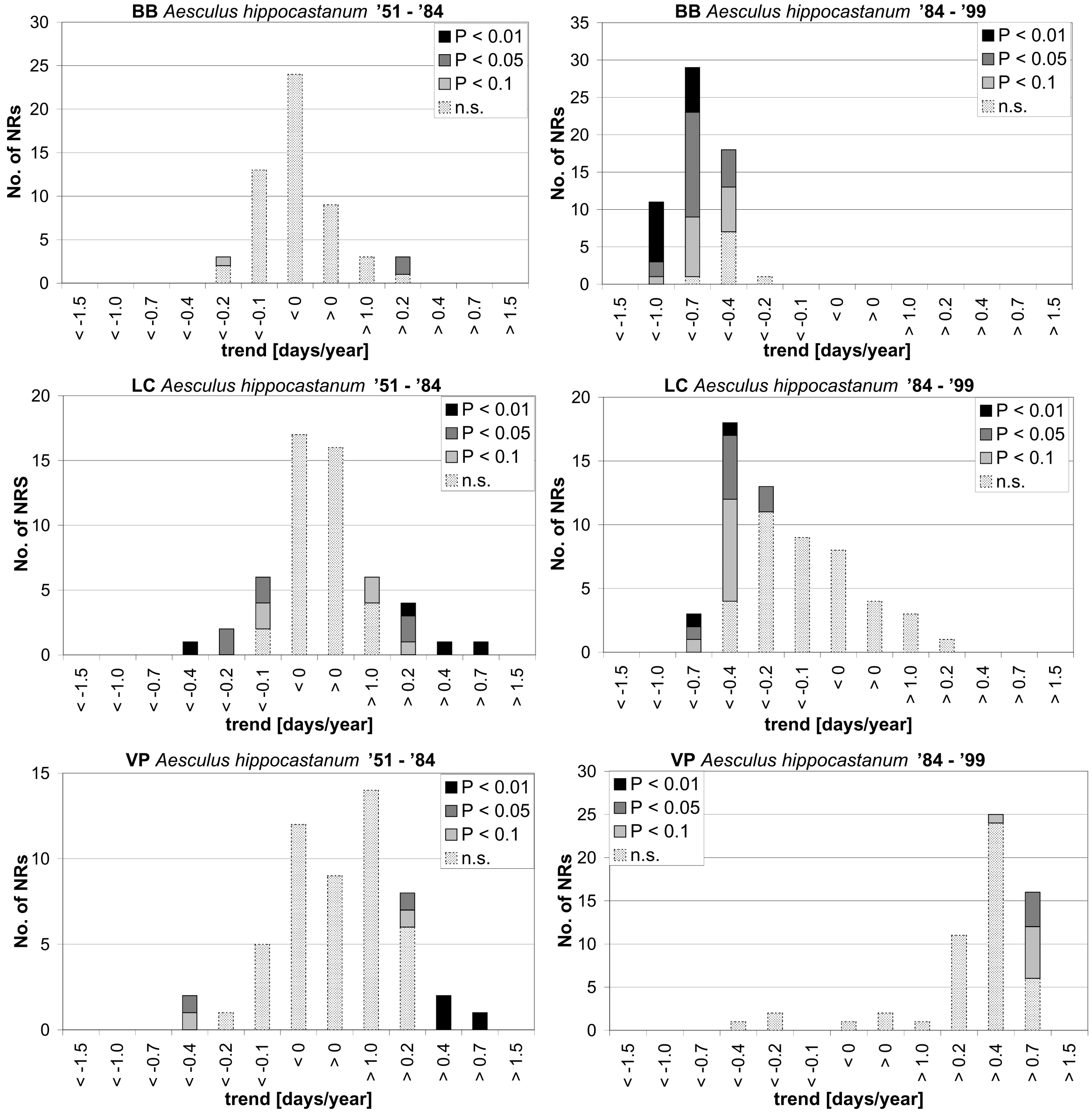


Figure 4-19. Histograms of trends and their significances over all NRs in Germany for BB, LC and VP of *A hippocastanum* as shown in Fig. 4-16 and in the respective figures in appendix A2.

Not only the BB phases show a distinct behavior between the two periods 1951 - 1984 and 1984 - 1999 as described above, but it is also the trends of the autumn phases and consequently the length of the vegetation period that change significantly in the mid-1980s.

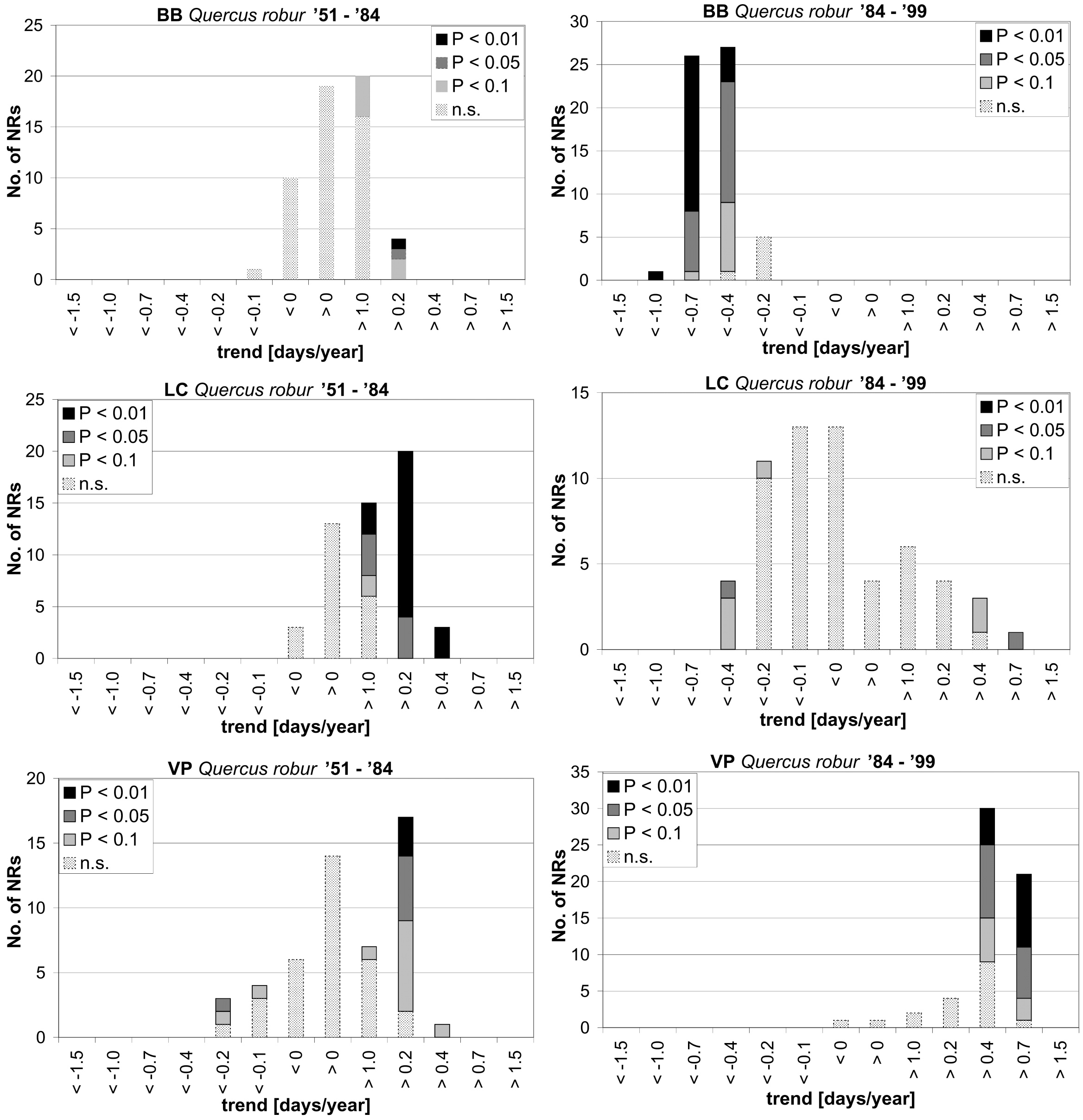


Figure 4-20. Histograms of trends and their significances over all NRs in Germany for BB, LC and VP of *Quercus robur* as shown in Fig. 4-16 and in the respective figures in appendix A2.

Trends in LC for *A. hippocastanum* (Fig. 4-19) and *B. pendula* (appendix A3) shift from an approximate normal distribution with zero mean to a left-skewed distribution with an emphasis on negative trends, whereas mostly positive trends for LC of *Q. robur* (Fig. 4-20) and *F. sylvatica* (appendix A3) detected for 1951-1984 broaden towards an inclusion of both significant negative and positive trends in the second period.

Trends in VP for *A. hippocastanum* (Fig. 4-19) and *B. pendula* (appendix A3) shift from an approximate normal distribution with zero mean to a right-skewed distribution with an emphasis on positive trends, whereas an emphasis on positive trends for LC of *Q. robur* (Fig. 4-20) and *F. sylvatica* (appendix A3) shift towards more significant positive trends between 1984 and 1999.

Table 4-4. Average trends over all Natural Regions (NRs) whose combined time series cover the periods 1951-1984, 1984-1999 and 1951-1999. Avg±σ: Overall mean date of the phase in Julian day of year ± standard deviation σ. N: total number of observations used for the analysis. total: total mean trend in the respective period rounded to days. No. of NRs: number of NRs that covered the respective period. Bold numbers: the mean trend over all used NRs is significantly different from zero at the 5% level (P<0.05).

Phase	Avg±σ	N	Trends ’51-’84		Trends ’84-’99		Trends ’51-’99	
			total [days]	**No. of NRs**	**total [days]**	**No. of NRs**	**total [days]**	**No. of NRs**
B *Corylus avellana*	59±25.7	49342	**-7**	46	**-31**	54	**-20**	42
B *Galanthus nivalis*	60±17.9	69527	**-5**	52	**-21**	59	**-12**	44
BB *Betula pendula*	112±11.6	64527	-1	52	**-11**	59	**-9**	44
BB *Aesculus hippocastanum*	113±12.4	75672	**-1**	55	**-12**	59	**-9**	44
BB *Fagus sylvatica*	120±9.9	60901	**2**	53	**-9**	44	**-5**	42
BB *Quercus robur*	127±11.1	62609	**3**	54	**-10**	59	**-6**	43
B *Aesculus hippocastanum*	131±11.0	63843	**6**	52	**-9**	59	**-2**	44
M *Pinus sylvestris*	133±12.0	44764	**5**	45	**-13**	59	**-4**	38
B *Syringa vulgaris*	133±11.1	66529	**7**	52	**-12**	59	**-4**	44
B *Sambucus nigra*	157±12.1	62975	**1**	50	**-14**	46	**-9**	43
LC *Aesculus hippocastanum*	277±12.5	71354	1	54	**-4**	59	-1	44
LC *Betula pendula*	278±14.0	60626	**3**	52	0	59	2	44
LC *Fagus sylvatica*	282±13.5	59663	**4**	53	-1	45	**3**	42
LC *Quercus robur*	287±12.9	62956	**6**	54	-1	59	**5**	43
VP *Quercus robur*	161±16.1	55157	**3**	52	**9**	59	**11**	41
VP *Fagus sylvatica*	162±15.8	54674	1	51	**7**	44	**8**	40
VP *Aesculus hippocastanum*	164±16.5	67612	**3**	54	**8**	59	**9**	44
VP *Betula pendula*	166±17.7	56912	**4**	52	**11**	59	**11**	44

In Tab. 4-4 the mean trends over all NRs that completely covered the respective periods 1951 - 1984 and 1984 - 1999, as well as mean trends over all NRs for the whole period 1951 - 1999 are displayed. On average our previous finding for the three selected NRs as well for the trend maps and trend histograms are confirmed for all NRs in Germany. There is a sharp increase in the magnitude of negative trends from the period before 1984 to the period after 1984 for the early spring phases (mean DOY of the phase < 115). Very early spring phases, B of *C. avellana* and *G. nivalis*, amplified their trends substantially from about a week advancement between 1951 and 1984 to a month advancement in the last 15 years, respectively, which again emphasizes the strong influence of the considered period upon trends. The early tree species *B. pendula* and *A. hippocastanum* have increased their small, partly non-significant, advancement of BB before the mid-1980s to an advancement of one and a half week in the last 15 years. Trends over the whole period from 1951 through 1999 are consequently also negative and range from a 20 day advancement for *C. avellana* to 9 day for BB of *B. pendula* and *A. hippocastanum*. BB of *F. sylvatica* and *Q. robur* as well as the B and M phases of early summer have switched from small but significant positive trends between 1951 and 1984 to significant negative trends between 1984 and 1999 of 9 to 14 days. However, trends over the whole period remain negative because the negative trends of the last 15 years outbalanced the positive trends before the mid-1980s and vary between -2 and -9 days for spring and early summer phases.

Autumn phases development reveals a less clear picture. Whereas before the mid-1980s autumn phases were mostly significantly delayed, they have advanced after 1984 but less pronounced except for *A. hippocastanum* that showed a significant advancement. This also why *A. hippocastanum* is the only species with a net advancement of LC between 1951 and 1999 whereas all other tree species experienced a net delay.

In terms of length of vegetation period of the four considered tree species the behavior of the spring phases dominates due to its larger magnitude. Thus, we can see an lengthening of the vegetation period throughout the whole time span 1951 until 1999 of around 10 days.

4.4.6 Potentially absorbed radiation

The measured average incoming radiation is a linear approximation of potentially absorbed radiation and thus average GPP. A relative change in the amount of absorbed radiation due to a change in the length of the vegetation period is equal to the relative change in GPP. In Tab. 4-5 the consequences of the change in length of the VP concerning potentially absorbed radiation are investigated. Only the results for the period from 1984 to 1999 are displayed. The measured average daily incident radiation over the period 1951 to 2000 and its corresponding fitted sine curve (Fig. 4-3) was almost equal to the one of the period 1984 to 1999. The trends of the phases given in Tab. 4-4 correlate with the respective average start and end days given in Tab. 4-5. The changes in potentially absorbed radiation are due to the change in BB and LC. The changes are stated in relation to the average radiation absorbed at the beginning of the investigation period as well as relative to incident radiation of the whole year.

Table 4-5. Change in relative average absorbed radiation due to change in the length of the vegetation period from 1984 to 1999. Averages over all NR in Germany whose combined time series covered the period 1984 to 1999. AR absorbed radiation. DOY day of year.

	B. pendula	***A. hippocastanum***	***F. sylvatica***	***Q. robur***
Av. start of VP 1984 (DOY)	114	116	122	129
Av. start of VP 1999 (DOY)	103	104	113	119
% change in AR due to BB relative to VP 1984	6.0	6.6	5.4	6.6
% change in AR due to BB relative to whole year	4.2	4.6	3.7	4.3
Av. end of VP 1984 (DOY)	278	279	284	289
Av. end of VP 1999 (DOY)	278	274	283	288
% change in AR due to LC relative to VP 1984	0.0	-1.6	0.3	-0.3
% change in AR due to LC relative to whole year	0.0	-1.1	0.2	-0.2
Av. length of VP 1984 (days)	164	163	162	160
Av. length of VP 1999 (days)	175	171	169	169
% av. AR 1984	69.8	69.2	67.8	65.6
% av. AR 1999	74.0	72.7	71.2	69.8
% change in AR from 1984-1999 relative to whole year	4.2	3.5	3.5	4.1

In Tab. 4-4 we saw that advancement of BB was largest for *A. hippocastanum*. This complies with the highest increase in absorbed radiation relative yearly incident radiation due to BB advancement. BB advancement of *B. pendula* was larger than for *Q. robur*, 11 and 10 days, respectively. However, the increase in absorbed radiation relative to 1984 was higher for *Q. robur* than for *B. pendula*, 6.6% and 6%, respectively. This effect can also be seen, but not as pronounced, when the change in absorbed radiation is stated relative to the total amount of yearly available radiation. In these terms *Q. robur* increases its radiation absorption by 4.3% and *B. pendula* by 4.2%. The change in LC is small and therefore it does not have much influence on the change in absorbed radiation except for *A. hippocastanum* where the significant advancement of LC of four days between 1884 and 1999 leads to a substantial decrease of potentially absorbed radiation. This is also why *A. hippocastanum* ranges last in terms of percent change in absorbed radiation together with *F. sylvatica*. Highest total increase in yearly potentially absorbed radiation had *B. pendula* followed by *Q. robur*.

4.5 Discussion

Visual inspection of the examples of the long-term combined time series given in Fig. 4-6 to Fig. 4-9 and the results of the trend analysis over the whole period in Tab. 4-2 show that linear regression analysis has to be handled with great care. We see in Tab. 4-2 that for some phases the trends over the whole period change among NRs or are not significant or contradict recent findings of other authors. Different trend directions among NRs can of course also be due to the different considered observational time spans. This again shows that trends of time series are difficult to compare as long as they do not cover the same period of time. The negative trends in LC and consequently VP give rise to the notion that trends might have changed over the century because most authors report the contrary concerning VP for the last decades. However, in Fig. 4-5 we saw that the estimations for the autumn phases were not reliable at the end of the 19th and the beginning of the 20th century making the results of the trend analyses for autumn phases questionable.

Analyzing combined time series rather than average or single time series is more reliable and adequate due to the reasons mentioned in the introduction and in Chapter 2. The large theoretical body coming with the theory of linear models also provides tools that measure the reliability of the obtained time series, e.g. confidence intervals for the estimated parameters. In recent publications where single or average time series have been studied (Sparks 2000, Menzel 2001, Chmielewski and Rötzer 2001) no such information was given. However, such an information can be useful. We saw that autumn phases might be obscured especially at the beginning of the century due to the small number of available data. In addition, changes in phase definitions can not be ruled out. Unfortunately, phase definition before 1922 were not available. Phase definitions starting 1922 were the same as in the actual observer’s instructions of the DWD (Schnelle and Witterstein 1952, 1964). Thus, analyzing trends of LC or VP before this time critical.

The regression analysis shows that a great deal of the regional phenological variance can be explained by linear functions of mean monthly temperatures. This gives further confidence in the obtained time series especially, for spring phases. Even autumn phases could rather well be explained for the time after 1951. It is interesting to note here that for all LC phases except birch negative parameters resulted for the summer months. This can be interpreted as a hint that drought stress in early summer might play a role for the timing of leaf coloring. Moreover, mean winter temperatures did not play a dominant role for spring phenology. When there was a significant effect of mean December or January temperature, then the corresponding parameter was negative, implying earlier spring phases with increasing temperature. Also October parameters were negative for spring phases. Thus, no effect of chilling could be found, i.e. earlier spring phases with lower temperatures (see also ”5.2 Models, data and methods” on page 76).

When we have long time series available we have the chance to look at a much broader picture of phenological development rather than analyzing the last decades and new methods to clarify this picture can be applied such that new insights can be obtained. Having long time series available statistical tests could be used that could detect trend turning points. It was shown that trends not only have changed throughout the century but that also there were marked differences between early and late spring phases. A shift in trend behavior was located between BB of *B. pendula* and *A. hippocastanum* that are dominated by March temperatures and BB of *F. sylvatica* and *Q. robur* that are dominated by April temperatures (Tab. 4-3). A clue for the reason of these differences can be obtained when we look at the course of March and April temperatures of our three NRs and Germany throughout the century (Fig. 4-21).

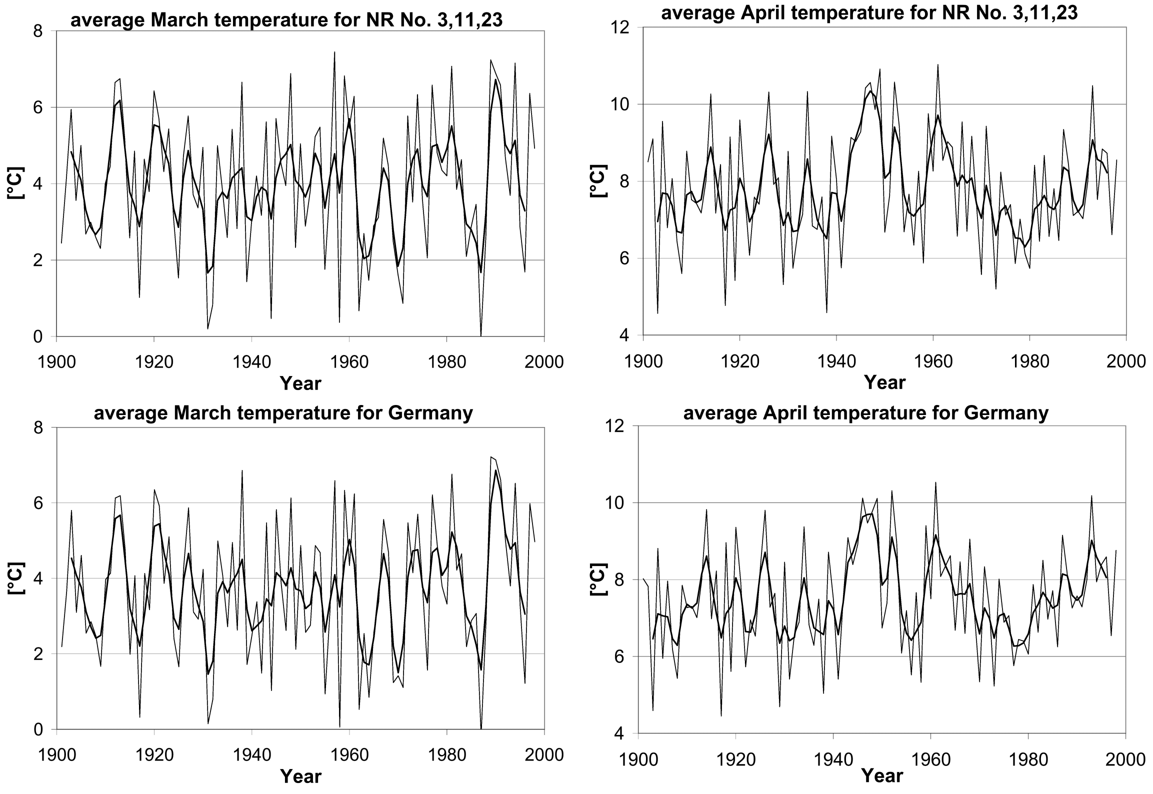


Figure 4-21. Mean March and April temperatures for the three NRs No. 3, 11 and 23 and for Germany (generated from the 1901-1998 CRU data set). Bold curves smoothed with a 5 five-year Gaussian filter.

There are obviously marked differences in temperature development in the three NRs No. 3, 11 and 23 as well as for whole Germany over the last century for the crucial months that determine spring phenology. In March, temperatures have increased on average accompanied with an increasing variation in the last 50 years, whereas April temperatures have clearly cooled between the end of the 1940s and beginning of the 1980s, followed by a marked warming. This structurally different behavior is mimicked by the phenological phases that are dominated by these monthly temperatures. Interestingly, April temperatures in the NRs No. 3,11 and 23 at the end of the 1940ies were higher than in the 1990ies (Fig. 4-21). This corresponds well with the finding that B for *S. vulgaris* was earlier end of the1940ies than in the last ten years (Fig. 4-9) in those NRs. These marked differences in phenology are not only valid for the three selected NRs but also apply to all NRs in Germany on average (Tab. 4-4). Here again we can find a hint in mean March and April temperatures that show for whole Germany basically the same picture (Fig. 4-21) only that mean April temperatures around 1949 are of about the same level as in the 90ies of the last century.

The time series over the whole century show that late spring phenology at the end of the first half of the 20th century was earlier than today in some regions in Germany. Moreover, similar trends towards advanced bud break have been observed in earlier decades with subsequent trend reversal (Fig. 4-12 - Fig. 4-14). This relates to the fact that no global anthropogenic radiative forcing signal could be detected in the long-term phenological time series (Tab. 4-3). This leads to the conclusion that the trends in phenology that have been observed in the last decades can not directly be attributed to an annual average Global Warming signal but are rather dominated by multi-decadal fluctuations in monthly temperatures. Such fluctuation have already been described for some NAO-indices (Hurrel 1995) that also play a role in the considered regional phenology (Tab. 4-3). It has to be emphasized that such multi-decadal fluctuations in phenology as well and in the NAO-indices are not only determined by regional but, more importantly, by seasonal temperatures. The global radiative forcing signal as well as the reported Global Warming trends at the global and European scale are mostly studied as yearly averages (Ramaswamy et al. 2001). The Global Warming signal in the analyzed phenological and temperature time series is therefore not only disguised by a higher variability due to a higher spatial resolution. The relation between Global Warming and phenology is further obscured by a different temporal reference. The radiative forcing signal is believed to cause the observed average yearly Global Warming, but how this signal influences differences in regional March and April temperatures remains unknown. As long as the climatologists have not revealed how Global Warming trends causally translate to such regional and seasonal temperature regimes the observed statistical relation between trends in phenology of the last decades and Global Warming trends can as well be a random event. That the correlation between phenology and a global radiative forcing signal can already be removed by simply extending the analyzed time frame was shown in this thesis. Thus, it is a critical statement that phenology of Central Europe is a good indicator of Global Warming as concluded by Chmielewski (2001).

Phenology is determined by seasonal temperatures which can show a different behavior than yearly averages. However, it was shown in this chapter that even such seasonal temperatures have to be analyzed in detail. Trend turning points in the mid-1980ies have been determined for spring temperatures (mean March-April-May) (MAM) in Europe before by climatologists (Böhm et al. 2001). For plant development and plant competition studies it is crucial, however, how temperatures develop in specifically this time. It was shown that March and April temperatures show a substantially different behavior with strong impact on the accompanying phenology. Schönwiese and Rapp (1997) analyzed trends of monthly temperatures for Europe. However, they considered the periods 1891 - 1990 and 1961 - 1990. Within these periods both March as well as April temperatures exhibited no significant trends for Germany. The determination of trends turning points, however, and the subsequent definition of periods where phenology showed a qualitatively different behavior revealed substantial differences in the development of March and April temperatures that are crucial for differential spring plant development.

Menzel et al. (2001), also analyzing data from the DWD, found both a lot significant positive and negative trends when analyzing 20 and 30 years time series within a time frame of 46 years. This was probably because the sign of the trend depends on the position of the 20 or 30 years within the last 50 years especially for late spring phases (Fig. 4-10). Menzel et al. (2001) found consistently average negative trends for spring phases and average positive trends for autumn phases when analyzing 20 and 30 year time series. The preceding analysis showed that it depends where such time intervals are selected and that depending on the time frame also consistent average positive trends for spring phases and consistent average negative trends for autumn phases can be detected in Germany (Tab. 4-4). Menzel et al. (2001) also analyzed mean phases for several subintervals. In their analyses mean BB of *F. sylvatica* and *Q. robur* did not change between the periods 1974-1996 and 1951-1973. This is not because they conveyed no trends but rather they shifted their trend behavior. Total trends were also different to the one found by Menzel (2001). Where Menzel found early spring advancement of eight to ten days I found advancements of about a month. Also absolute trends for BB, LC and VP were more pronounced in the present study. This can to a large part be attributed to the fact that Menzel et a. (2001) analyzed the period from 1951 to 1996, i.e. missing the phenologically early years from 1997 to 1999. This again emphasizes that trends can only be compared meaningfully when the same time intervals are analyzed. Similar to Menzel et al. (2001) no obvious regional patterns of trends were found.

As mentioned in Chapter 1 one main issue of phenological research is the investigation of phenology of forest ecosystems. For the aspects of forest growth and species composition it is not primarily the change in length of VP that is crucial but rather its effect on carbon balance of the living tree. When the impact of change in growing season length on yearly absorbed radiation is considered two factors are pivotal. The first is whether the change in VP takes place at the beginning or end of the VP and the second is its absolute position within the year. We saw in Tab. 4-5 that *B. pendula* profited the most from a prolonged VP. This was not only due to the fact that *B. pendula* had the strongest prolongation of VP but because this prolongation took only place in springtime where it is most profitable. *B. pendula* had the largest total amount of potentially absorbed radiation. This was also found by Kramer (1995). In Fig. 4-3 it can be seen that during the phase of BB, i.e. between DOY 100 and

130 (Tab. 4-5), average incident radiation is much higher with 1280 to 1660 J/cm^2/d than around LC, i.e. DOY 270 and 290, with 930 to 666 J/cm^2/d. *Q. robur* did also profit more from a longer VP than *A. hippocastanum* that had the largest advancement of BB. This was partly due to the fact that late spring species will generally profit more from a prolongation in spring time than early spring species because they exploit longer days. One day advancement around the DOY 100 adds 1280 J/cm^2 to the yearly carbon balance whereas an additional days around DOY 130 adds 1660 J/cm^2, i.e. 30% more. This again emphasizes the importance to determine as accurately as possible at what time of the year a possible climate change takes place. As we saw in Fig. 4-21 starting in the mid-1980s April temperatures seemed to increase more strongly than March temperatures favouring the late spring species like *Q. robur* as described above. This becomes an important factor when questions concerning species competition are considered. One day advancement in spring adds roughly 0.6% to the yearly absorbed radiation in this analysis. This in turn adds 0.4% to the yearly GPP (Tab. 4-5). This is roughly the value that was also found by Kramer (1995). Another result that is supported by the study of Kramer (1995) is that the relative decrease in yearly absorbed radiation by one day advancement of LC, i.e. roughly 0.3%, is half from what is gained by one day advancement of BB.

The analyzed potentially intercepted radiation is only a coarse measure of yearly GPP and to obtain better approximations phenology models have to be explicitly incorporated into forest growth models. Another factor that might become important is the change in the length of the period between bud burst and full leaf expansion. In this study is was assumed for simplicity that with BB trees fully exploit their interception capacity. Depending on weather conditions at BB trees need several days to weeks to have their leaves fully expanded. Again, this can only be analyzed by modeling studies because the phenological data give no information concerning timing of leaf expansion.

Despite all these shortcoming, an increase in GPP of about 3 - 4% in the period 1984 - 1999 was found as a first approximation. This identifies phenological changes as one source of the recently observed change in growth patterns of European forests (Spieker et al. 1996).

5 Physiology-based phenology models for forest tree species

5.1 Introduction

As mentioned before, the length of the vegetation period framed by the phenological events of bud burst and leaf coloring determines to a large part the annual carbon balance of deciduous trees (Kramer 1995, Goulden et al. 1996, Chen et al. 1999, White et al. 1999, Kramer et al. 2000). Goulden et al. (1996) and Chen et al. (1999) showed that it is especially the day of leaf emergence rather than the day of leaf coloring that controls annual carbon fixation by photosynthesis in forests. This is mainly because at the time of leaf emergence days are usually already long and temperatures high, so that a day advancement in spring consequently adds more to the annual carbon balance than an extra day in autumn where days are short and temperatures low. This has also been demonstrated in the previous chapter (see ”4.4.6 Potentially absorbed radiation” on page 68) and in a simulation study by Kramer (1995). It is also for this reason that phenological models concentrate on simulating the day of bud burst rather than leaf coloring or leaf fall.

The nature of the physiological processes within the plant that control its phenological status remain largely unknown (Hanover 1980, Lavender and Silim 1987, Powell 1987). However, it is reasonable to assume that, even if the processes are generally the same for most species, the balance between these processes is species-specific. Therefore, it can be expected that when such physiology-based processes are incorporated into a model, this model should exhibit a wide geographical applicability because the parameters only depend on the species and not on the special environmental conditions where it was developed. Such a model is especially useful for climate impact studies where environmental conditions are prescribed to change.

There exist several models that simulate bud burst (BB) of deciduous species and may shoot of evergreen species (Cannell and Smith 1983, Hänninen 1991, 1994, Kramer 1994, 1996, Linkosalo et al. 2000). All of these models, however, have only been applied to very special data sets like single species (Cannell and Smith 1983, Hannerz 1999), small data sets from a geographically limited region (Kramer 1994, 1996, Murray 1989) or clones (Diekmann 1996, Menzel 1997). When phenology of a wider geographical region including several species and provenances is to be modelled to simulate sensitivity of regional forest carbon budgets to climate change or validate satellite observations, these models with their parameterizations are no longer useful in many cases. This is probably because they are largely based on empirical relationships between phenology and weather.

The purpose of this modeling study was twofold. First, hypotheses about the physiology of plant developmental control were tested for their applicability in phenology modeling. Good modeling results strengthen these hypotheses and generate new insights into the dominating physiological processes that control plant phenology. Second, with this physiology-based approach a model should be developed that is able to simulate BB of several deciduous tree species over a wide geographical region.

5.2 Models, data and methods

5.2.1 Phenological models

First attempts to model the phenological development of plants originate in 1735 when Réaumur first described his observation that stages of plant development can be described by critical temperature sums (Wang 1960, Robertson 1968). This model of accumulated heat units has been used frequently by many authors in many different disciplines (Penning de Vries et al. 1989, Hunter 1992, Worrall 1993, Goudriaan and van Laar 1994, Diekmann 1996, Schaber 1997, Hannerz 1999). Temperature has also been identified as good predictor of regional phenology in Chapter 4.

A more elaborated model of phenology that has been developed for tree bud burst and may shoot is based on the observation that trees leaf out earlier the colder the preceding winter was. This finding has been confirmed experimentally. This means in mathematical terms that the critical temperatures sum is decreased by a prolonged period of low temperatures, called chilling (Cannel and Smith 1983, Murray et al. 1989, Hunter and Lechowicz 1992). Interestingly, this effect could not be found by the regression analyses between temperature and regional phenology conducted in Chapter 4. Mean temperature of winter months (December, January and February) did not play a dominant role for regional spring phenology (Table 4-3 on page 54). When an significant effect was found, however, the respective parameter was negative, implying earlier occurrence of the respective phase with increased temperature, i.e. the opposite effect to chilling. November mean temperature did not play a significant role at all, and October mean temperatures also showed negative parameters. It is possible, however, that mean monthly temperatures do not show a chilling effect due to their coarse temporal resolution.

Hänninen (1994) concentrates on the distinct phases of ontogenetic development of tree buds that were identified by many authors (Samish 1954, Smith and Kefford 1964, Vegis 1964, 1973, Sarvas 1974) introducing a more physiology-based aspect in phenology modeling. These distinct ontogenetic stages are difficult to distinguish but most authors agree that there is a phase, called rest by many authors, that is characterized by a period where buds remain dormant even under favorable environmental conditions. This phase is believed to be determined by growth-arresting conditions within the bud. It was observed that these conditions can be removed when the bud is exposed to a certain period of cold temperatures, i.e. chilling. After this period the bud is again able to react to positive environmental conditions. This stage was called quiescence by some authors. It is commonly agreed that the main factors defining positive or negative environmental conditions are temperature and photoperiod (Waring 1953, 1956, Sarvas 1972, Vegis 1973, Johnson and Thornley 1985).

The nature of these different phases and their physiological and morphological status is not fully understood, yet. Thus, the distinction of these phases or states and their transitions are not clearly defined. However, most investigators agree that the regulation of these phases and their transitions is controlled by the actions and interactions of growth-promoting compounds or hormones and growth inhibiting compounds (Samish 1954, Waring and Saunders 1971, Hanover 1980, Powell 1987). Thus, it is the hypothesized balance or ratio between promotory and inhibitory agents that determines the physiological state of development of the plant and its reaction to external driving forces. Several candidates of such substances were isolated, however, the nature of their interaction remains unclear and experimental evidence is often contradictory (Lavender and Silim 1987). Despite the lack of knowledge about the details of hormonal control of tree phenology, a first attempt was made to put the agreed first principles of hormonal aspects of phenology into a simple conceptual framework that can be easily translated into a mathematical model:

The abundance or concentration of certain enzymes in cells is determined by the rates of synthesis and breakdown. Control of these processes is subject to a lot of research, however, it is known that temperature and photoperiod play a prominent role. Temperature, for instance, can act through pure physical mechanisms like its influence on viscosity and diffusion. Moreover, synthesis of proteins usually has an activation energy or temperature and an optimal temperature beyond that synthesis rates decrease again (Vegis 1973, Johnson and Thornley 1984). Photoperiod has been observed to be the driving force of a biochemical trigger acting through the phytochrome system (Waring 1956, Nitsch 1957, Perry 1971, Heide 1993a, 1993b). From these simple but basic principles a model for the abundance or concentration of an inhibitory compound I and a promotory compound P is hypothesized and formulated as a system of two simple coupled difference equations:

$$\begin{aligned} \Delta I &= a_1 f_1(T) g_1(L) - a_2 f_2(T) g_2(L) I \\ \Delta P &= a_3 f_3(T) g_3(L)(1-I) - a_4 f_4(T) g_4(L) P \end{aligned}$$

Equation 5-1

where the a_i are scaling parameters and the f_i and g_i, $i = 1,...,4$, are functions of temperature T and photoperiod (day length) L, respectively. Temperature T and photoperiod L are themselves functions of time, in our case day of the year (DOY). Breakdown of the compounds *P* and *I*, indicated by the negative terms in Eqn. 5-1, is assumed to be a first order reaction, whereas the synthesis terms, indicated by the positive terms in Eqn. 5-1, are formulated as simple forcing terms. The synthesis term of the promotor *P* is damped by the presence of the inhibitor *I*. *P* and *I* are normalized arbitrarily to vary between one and zero.

To obtain an applicable model, assumptions about the parameters, functions, boundary and initial conditions have to be made: we let the model start at t_0, the observed date of leaf coloring (LC) of the previous year. This seems to be a natural starting date to calculate BB phenology because the preceding winter is known to play a role. For reasons of simplicity, on this date rest is assumed to be deepest. That means we assume the maximal concentration of inhibitory substances and minimal abundance of promotory compounds, setting $I(t_0) = 1$ and $P(t_0) = 0$, i.e. with the current family of models initial deepening of rest is not simulated. Bud burst takes place when *P* reaches the threshold $P = 1$.

Following the observation that almost all physiological processes have a minimum, optimal and maximum temperature, the temperature dependence of the functions f_i are modelled as a triangular function (Hänninen 1994) (Fig. 5-1) where

$$f_i(T) = \begin{cases} \dfrac{T - T_{min}}{T_{opt} - T_{min}} & \text{for } T_{min} \le T \le T_{opt} \\ \dfrac{T_{max} - T}{T_{max} - T_{opt}} & \text{for } T_{opt} \le T \le T_{max} \\ 0 & \text{else} \end{cases} \qquad \textbf{Equation 5-2}$$

for $i = 1,...,4$ with *T* being daily mean temperature and T_{min}, T_{opt}, T_{max} fixed temperature thresholds that are different for each f_i, respectively.

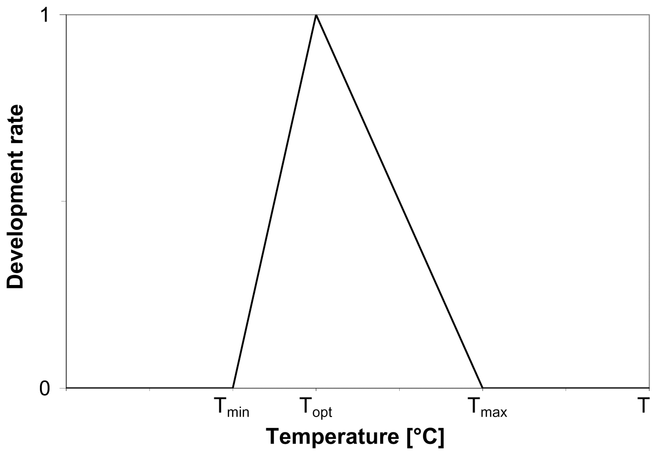


Figure 5-1. Triangular function of temperature (Eqn. 5-2) modifying the rate of development of promotor *P* and inhibitor *I* (Eqn. 5-1 and Tab. 5-1).

Because photoperiod is another important factor influencing tree phenology, the compounds *P* and *I* can be modified by multiplication of a day length (L) dependent function, i.e. either $g_i(L) = L/24$ or $g_i(L) = (24-L)/24$, depending on whether the process is assumed to be promoted by long days or long nights. Breakdown of the inhibitor *I* is assumed to be accelerated by long days, whereas build-up of the inhibitor is assumed to be supported by long nights. The contrary is assumed for the promotor *P*. Synthesis of the promotor *P* is favored by long days and breakdown of P is accelerated by long nights (Waring 1956, Nitsch 1957, Perry 1971).

To keep the number of parameters as low as possible some restrictions were introduced in the general framework of Eqn. 5-1. It was imposed that $f_1(T) = f_4(T) = 1$, i.e. the forcing term of the inhibitor *I* and the breakdown term of the promotor *P* only depend on photoperiod *L*. Further it was imposed that either a_1 or a_4 is zero. This finally resulted in 12 possible models each one having nine parameters. They are listed in Tab. 5-1 and are called Promotor-Inhibitor models (PIMs) in the following.

Table 5-1. Promotor-Inhibitor Models (PIMs) derived from the general framework in Eqn. 5-1 and the imposed restrictions described in the paragraph above. ΔI, ΔP rates of change of the inhibitor *I* and promotor *P*. f_I, f_P temperature dependent triangular functions according to Eqn. 5-2 for the inhibitor *I* and promotor *P*, respectively, each having three parameters T_{min}, T_{opt}, T_{max}. a_i, $i = 1,..., 4$ are scaling parameters. $I(t_0) = 1$, $P(t_0) = 0$. $0 \le I, P \le 1$. time step: one day. t_0 = day of leaf coloring the previous year. Bud burst: $P > 1$

Model No.	Inhibitor Model	Promotor Model
1	$\Delta I = -a_2 f_I I$	$\Delta P = a_3 f_P(1-I) - a_4 P$
2	$\Delta I = -a_2 f_I I$	$\Delta P = a_3 f_P(1-I) - a_4 P(24-L)/24$
3	$\Delta I = a_1(24-L)/24 - a_2 f_I I$	$\Delta P = a_3 f_P(1-I)$
4	$\Delta I = -a_2 f_I I$	$\Delta P = a_3 f_P(1-I)L/24 - a_4 P$
5	$\Delta I = -a_2 f_I I$	$\Delta P = a_3 f_P(1-I)L/24 - a_4 P(24-L)/24$
6	$\Delta I = a_1(24-L)/24 - a_2 f_I I$	$\Delta P = a_3 f_P(1-I)L/24$
7	$\Delta I = -a_2 f_I IL/24$	$\Delta P = a_3 f_P(1-I) - a_4 P$
8	$\Delta I = -a_2 f_I IL/24$	$\Delta P = a_3 f_P(1-I) - a_4 P(24-L)/24$
9	$\Delta I = a_1(24-L)/24 - a_2 f_I IL/24$	$\Delta P = a_3 f_P(1-I)$
10	$\Delta I = -a_2 f_I IL/24$	$\Delta P = a_3 f_P(1-I)L/24 - a_4 P$
11	$\Delta I = -a_2 f_I IL/24$	$\Delta P = a_3 f_P(1-I)L/24 - a_4 P(24-L)/24$
12	$\Delta I = a_1(24-L)/24 - a_2 f_I IL/24$	$\Delta P = a_3 f_P(1-I)L/24$

For each species it was determined which of these model formulation best suited the observed dates of BB and corresponding course of temperature and day length.

Three additional models from literature that have been successfully applied to different data sets and whose principles were describe above were also fitted to data. The mathematical details of these models are listed in ”Appendix A4 Model descriptions” on page 141.

The sequential model (SM) is also based on physiological considerations (Hänninen 1994) and was found to be best fitting among several other models for beech (*F. sylvatica*) bud burst (Kramer 1994, Kramer et al. 2000). It considers rest and quiescence as two distinct phases that follow sequentially in time. Transition from rest to quiescence does not take place unless a critical state of chilling is attained. Similarly, BB only takes place after a critical period of warm forcing temperatures. The model has seven parameters (”A4.1 The sequential model (SM)” on page 141).

The Cannel and Smith model (CSM) empirically models the observation that increased chilling in winter, decreases the required temperature sum for BB in spring. It was developed by Cannel and Smith (1983) and modified by Menzel (1997). Opposed to the SM in the CSM temperature sums can accumulate parallel to chilling. The model has four parameters (”A4.2 The Cannel and Smith model (CSM)” on page 142).

The linear temperature sum model (TSM) (Wang 1960, Robertson 1968, Kramer 1994, Menzel 1997) simply integrates temperatures above a certain threshold starting from a defined date up to a fixed critical value. The model has three parameters (”A4.3 The linear temperature sum model (TSM)” on page 142).

These three models are also compared to the NULL-model that is the mean date of BB.

5.2.2 Data

The phenological data used for this study come from the phenological network of the German Weather Service (DWD). For the purpose of modeling tree leaf emergence in relation to weather, phenological stations were selected for that corresponding weather measurements are available. Weather data were also provided by the DWD.

For the reason discussed in Chapter 2 combined phenological time series were calculated from all phenological stations available in the vicinity of a weather station. Assuming that the weather station measurements are characteristic for a certain region, it is assumed that the corresponding characteristic phenology is obtained when at least five phenological stations time series are combined that are not further away than 10 km and do not differ more than 50 m in elevation from the weather station. Phenological observations at stations with the latter characteristics were found to be highly correlated (Hense et al. 2002).

With the above criteria a data set of 81 weather stations and 495 phenological stations was obtained (Fig. 5-2). In Fig. 5-3 a histogram of the distance distribution of the resulting phenological stations is displayed. The combined time series were obtained with the procedure recommended in Chapter 2. The available data per station were first fitted with the robust L_1 estimation (”2.2.3 Robust estimation” on page 14) and observations whose resulting residuals were larger than 30 days (30-day rule, ”2.2.4 Outlier detection” on page 14) were removed. Subsequently, the final combined time series for model fitting was calculated by the T1 method (”2.2.2 Parameter estimation” on page 13) assuming a two-way mixed model (Eqn. 2-10 ”2.2.1 A common theoretical background” on page 11). To calculate combined time series it has to be assumed that all influences that cause variation among stations (observers, microclimate, phenotypic and genetic traits) are approximately normally distributed and can be summarized under one source of variation. That this can be assumed to be the case is shown in Chapter 3.

The model fits were done for those deciduous tree species for which a high number of observations of BB were available. Only those observations were taken into account that had a corresponding observation of LC in the previous year because the calculations for the Promotor-Inhibitor model were assumed to start on this date. Except for *Larix decidua* for which calculations were allowed to start on the 1st November because there were not many LC dates available. The resulting combined time series had a total number of observations over all years and stations of

- 2475 data points for horse chestnut (*Aesculus hippocastanum* L.),
- 1937 data points for birch (*Betula pendula* Roth.)
- 1536 data points for beech (*Fagus sylvatica* L.)
- 1811 data points for oak (*Quercus robur* L.)
- 1441 data points for larch (*Larix decidua* Mill.)

Figure 5-2. Phenological and weather stations in Germany that were used for the modeling study. 495 phenological stations were selected indicated by the dots. These are phenological stations that are situated within a distance of 10 km from a weather stations and do not differ more than 50 m in elevation from the respective weather stations. The weather stations are situated in the center of the circles having a radius of 10 km. 81 weather stations were used.

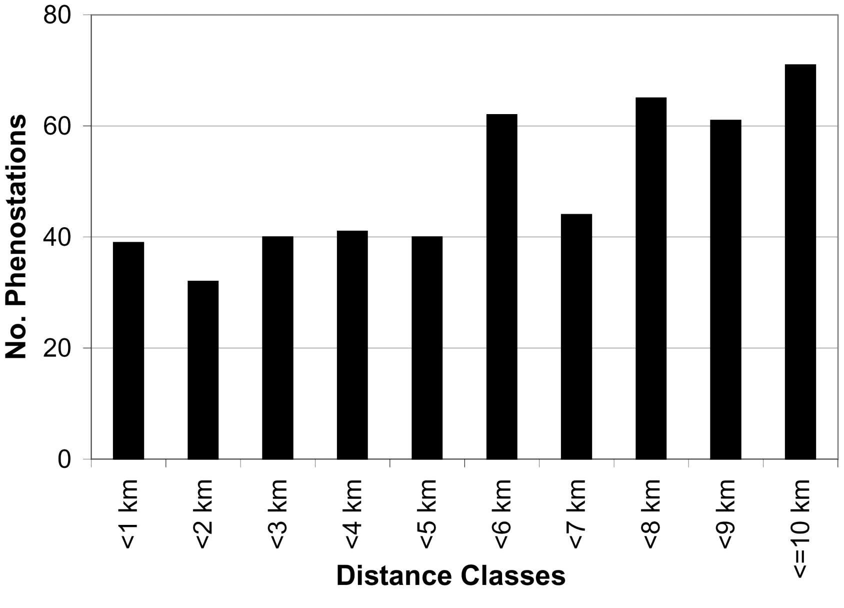


Figure 5-3. Histogram of the number of phenological stations in distance classes from the respective weather stations

5.2.3 Methods

The phenological models (except the Null-model) were fitted to the observations by minimizing the sum of squared residuals

$$f(x) = \sum_i r_i^2(x)$$

in the parameter space $x=(x_1,...,x_N)$ (N Parameters), where the residuals $r_i(x) = m_i(x) - o_i$, $m_i(x)$ is the day of BB according to the model with its respective parameter x and o_i is the observed day of BB in year i.

A simulated annealing algorithm (Metropolis et al. 1953, Kirkpatrick et al. 1983, Ingber 1989) was used for the optimization. This algorithm has been successfully applied earlier to phenological data (Chuine et al. 1998, 1999). Here a modified version of the implementation of Ingber (1989) was used that is available on the internet (www.ingber.com).

The method of simulated annealing is essentially a Monte-Carlo importance sampling of the parameter space in analogy with thermodynamics, specifically the way that liquids freeze and crystallize, or metals cool and anneal (Press et al. 1992). In a system in thermal equilibrium at temperature T the energy is probabilistically distributed among all different energy states E according to a Boltzmann-distribution $Prob(E) \sim \exp(-E/kT)$. Metropolis et al. (1953) first used this principle in numerical calculations in the following manner. When the set of parameters x is analogous to the thermodynamic state S of some system, then $f(x)$ can be viewed as the energy E of S. The system is defined to change its configuration from S_1 to S_2, or in terms of energy E_1 to E_2 if $E_2 < E_1$. If $E_2 > E_1$ then the transition takes place with the probability of $exp(-(E_2-E_1)/kT)$. Thus, in opposition to optimization algorithms that search for minima in direction of decreasing energy, e.g. steepest decent and Newton methods (Press et al. 1992), there is always the chance that the system configuration switches to a state of higher energy, meaning that it is able to jump out of local minima and resume its search for a global minimum. The simulated annealing algorithm uses five elements to work properly:

- A description of possible system configuration, i.e. a model and its parameters, possibly an additional probability density of the values attained by individual variables in the parameter space x.
- A generator of random steps in the parameter space.
- A cost or energy function $f(x)$ whose minimization is the goal of the procedure.
- A probability density for acceptance of a new cost-function or parameter set given the previous values, e.g., the Boltzmann-distribution, that is controlled by the temperature parameter T. Other distributions can also be used.
- A annealing schedule $T(k)$ of the control parameter T in k annealing time steps.

Ingber (1989, 1993) uses a different energy distribution than the original Boltzmann-distribution and showed that with his modifications a faster convergence of the algorithm can be expected. The convergence of the algorithm is very sensitive to its parameters, specifically the annealing schedule (Ingber 1989, Chuine 1998). There is no rule how to optimally adjust the parameters of the algorithm and a good configuration must be found by trial and error and adjusted individually to each problem.

An important step in the use of the algorithm is random sampling of the parameter space and calculating the corresponding state before subsequent decision whether the system is allowed to accept a new generated state or not. Calculation of states, i.e. $f(x)$, is in our case computing-time intensive compared to deciding whether to accept a new state or not. This makes the algorithm suitable for parallelization. The algorithm can be expected to be faster when several computers generate a set of states in parallel from which the algorithm accepts a new state, than when only one computer generates all the states. Thus, the algorithm using the code from Ingber (1986, www.ingber.com) was parallelized with a message passing interface (MPI) (www.mpi-forum.org, Gropp 1994) and migrated to a IBM parallel computer RS6000 SP.

For each optimization ten different initial parameter configurations were used to see wether the resulting minima are stable dependent on initial values. The quality of the convergence of the algorithm is characterized by the relative coefficient of variation rCV of the found minima over the ten optimization runs where $rCV = \frac{s/\bar{x}}{\sqrt{10}}100$ with as s standard deviation and $\bar{x}$ as the mean over the ten minima. The rCV characterizes the variation of the minima in percentage of the mean (Sachs 1997).

The available data were split into two sets of equal size for per each species. This was done by splitting each station series into even and uneven years and attributing them to the dependent or independent data set depending on the start year. One half of the data set was used to fit the model the other half was used for independent testing also called internal and external validity by Chuine (1998). Convergence of the algorithm with respect to the parameters was characterized by their standard error over the ten different runs.

Randomly selected initial parameter settings for the ten optimization runs per model may lead to different minima. Thus, obtaining these minima can also be viewed as a random process. Therefore it can be tested statistically whether the models do significantly differ with respect to their average minimum. This is done by the ANOVA method of Least Significant Difference (LSD) (Sokal and Rohlf 1995, Sachs 1997). The difference of the averages over the ten different optimization runs are tested against the critical value LSD, where

$$LSD = Q_{[\alpha, k, \nu]}\sqrt{\frac{MS_{within}}{n}}$$

with $Q_{[\alpha,k,\nu]}$ being the significance level of the studentized range at significance level α, for k normally distributed items and ν degrees of freedom of the estimated standard deviation of the items MS_{within}, also called 'mean squares within groups' in ANOVA terminology. n is the number of samples for each item. In this case the items are the 15 different models (12 PIMs + SM + CSM + TSM), $n = 10$, thus $\nu = 135$. α is set to 0.05. The tabulated value for Q in Sachs (1997) was used. When the difference between the average minimum of two models exceed LSD, then these averages are significantly ($P < 0.05$) different.

The quality of the model fits is measured by four different measures. The first measure is sum of squared residuals $f(x)$ that is used as the optimization criterion. A more descriptive version of the SSR is the root mean square error (RMSE). The second is the mean absolute error (MAE) and the third is the R^2 value of the linear regression of the simulated against the observed values. The latter can also be expressed as the percentage of variance explained by the model. The fourth measure of model performance is a function of the slope of the regression of the simulated against the observed values. It can be seen as a measure of the trend in the residuals of the model fit.

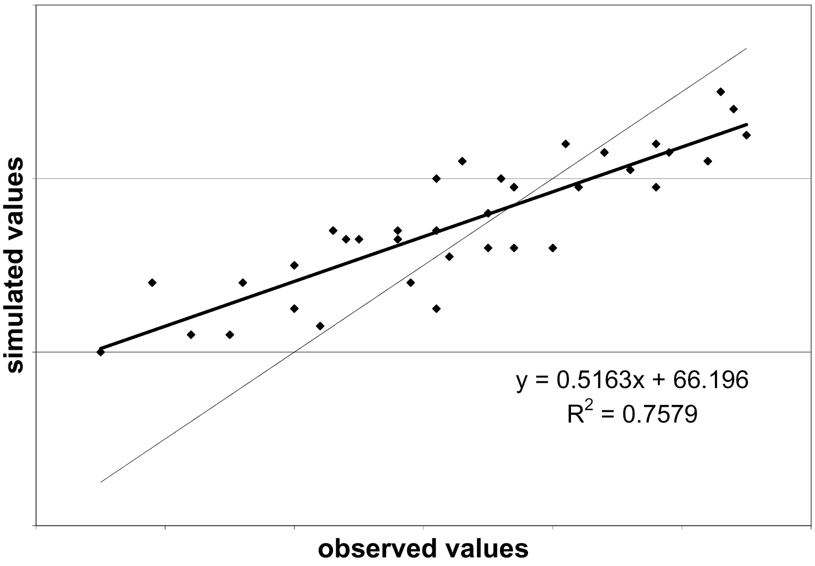


Figure 5-4. Example of a linear regression of observed values against simulated values of phenological observations. The thick line is the regression line. The thin line is the 1:1 line that would be the perfect fit.

Fig. 5-4 we gives a hypothetical example of a regression of observed values against simulated values of phenological observations in arbitrary units. It can be seen that the R^2 value with 0.76 is, at least as far as phenology is concerned, satisfactory. However, early observations are consistently overestimated whereas late observation are underestimated as can be seen by comparison to the perfect fit (indicated by the thin line in Fig. 5-4). Thus, the model is not able to reproduce the observed natural variability of the data. This is also indicated by a slope of the regression of less than unity. Such a situation could also occur with 100% explained variance, i.e. $R^2 = 1$. An additional measure of model performance is therefore defined as 1 - slope of the regression called bias, meaning that the smaller the bias, the less is the deviation of our model from the perfect fit.

5.3 Results

Parallelization significantly speeded up the optimization process depending on the number of processors used. For the optimization of the Promotor-Inhibitor Model (PIM) for 10 runs with different initial parameters for, e.g., *A. hippocastanum* computation time could be decreased from more than two weeks to three days using one and 40 processors, respectively. Computation time also depended on the amount of data and even more on the number of parameters of the model. For the SM model with seven parameters the optimization procedure took about two hours using 40 processors and for the CSM with four parameters computing time amounted to about one hour and ten minutes for the TSM. Only in the case of the TSM for *Q. robur* and the SM for *L. decidua* the ten different optimization runs converged to the same minimum in several cases, even though the annealing schedule was slowed down several times after test runs. However, the variation of the SSR of the different repetitions was small (Tab. 5-2).

Table 5-2. Optimization results. After the species name the number of observations is given that is used for model fitting. An equal number was used for model testing. For the PIMs the model with the lowest sum of squared residuals in a single optimization run (min SSR) is displayed. The mean SSR (average SSR of ten repetition runs among which is the min SSR) is given with its respective relative coefficient of variation (rCV). Root mean square error (RMSE) and mean absolute error (MAE) rounded to days.

Species (N)	**Model**	**fitting**							**testing**				
		mean SSR	**rCV [%]**	**min SSR**	**R^2**	**bias**	**RMSE**	**MAE**	**min SSR**	**R^2**	**bias**	**RMSE**	**MAE**
A. hippocas (1238)	PIM 8	36592	0.10	36477	0.695	0.3	5	4	35312	0.708	0.278	5	4
	SM	70386	0.02	70320	0.416	0.558	8	6	68440	0.434	0.548	7	6
	CSM	41795	0.29	41654	0.654	0.313	6	4	40275	0.671	0.29	6	4
	TSM	42731	0.08	42586	0.648	0.305	6	5	43396	0.648	0.304	6	5
	NULL	119645	-	-	-	-	10	8	120762	-	-	10	8
B. pendula (987)	PIM 2	24145	0.17	24036	0.722	0.27	5	4	31726	0.657	0.326	6	4
	SM	59622	0.32	58960	0.335	0.583	8	6	66701	0.298	0.621	8	7
	CSM	27246	2.24	25906	0.7	0.32	5	4	30580	0.668	0.356	6	4
	TSM	30619	0.11	30543	0.646	0.345	7	6	36229	0.605	0.389	6	5
	NULL	86093	-	-	-	-	9	7	91813	-	-	10	8
F. sylvatica (786)	PIM 11	21740	0.13	21583	0.486	0.504	5	4	20485	0.514	0.477	5	4
	SM	37263	0.36	36836	0.149	0.792	7	5	35457	0.172	0.781	7	5
	CSM	27390	2.03	26063	0.386	0.562	6	4	25315	0.407	0.54	6	4
	TSM	26141	0.04	26101	0.379	0.630	6	4	24050	0.433	0.599	6	4
	NULL	41956	-	-	-	-	7	6	42155	-	-	7	6

Table 5-2. (continued)

Species (N)	**Model**	**fitting**							**testing**				
		mean SSR	**rCV [%]**	**min SSR**	**R^2**	**bias**	**RMSE**	**MAE**	**min SSR**	**R^2**	**bias**	**RMSE**	**MAE**
Q. robur (906)	PIM 12	27698	0.17	27824	0.613	0.346	6	4	27304	0.618	0.35	5	4
	SM	51453	0.24	51025	0.313	0.59	8	6	50392	0.323	0.576	7	6
	CSM	29307	0.59	28994	0.593	0.426	6	4	28231	0.603	0.414	6	4
	TSM	30309	0.03	30291	0.576	0.401	6	4	30600	0.569	0.412	6	4
	NULL	71105	-	-	-	-	9	7	70934	-	-	9	7
L. decidua (721)	PIM 2	41279	0.05	41204	0.585	0.403	8	5	37044	0.60	0.353	7	5
	SM	62005	0.45	61633	0.381	0.628	9	7	56611	0.381	0.601	9	7
	CSM	41818	0.53	41541	0.583	0.405	8	5	37420	0.594	0.36	7	5
	TSM	42341	0.04	42259	0.575	0.403	8	5	36661	0.604	0.349	7	5
	NULL	99176	-	-	-	-	12	8	91384	-	-	11	9

In Tab. 5-2 the mean of the sums of squared residuals (SSR) over the ten optimization runs is displayed with its respective relative coefficients of variation (rCV). Among ten optimization repetitions the SSR varied always by less than 1% of the mean except for the CSM for *F. sylvatica* and *B. pendula* where the rCV was roughly 2%. The CSM is most sensitive to initial conditions.

However, some model fits were not significantly different with respect to their mean SSR over ten runs. In Tab. 5-3 the models whose average SSR over ten optimization runs was not significantly different and at the same time where in the group with the lowest SSR are shown per species.

Table 5-3. Models with lowest SSR but non significant ($P < 0.05$) difference among SSR after ten optimizations with randomly selected initial parameter values. Ascending order of mean SSR from the top.

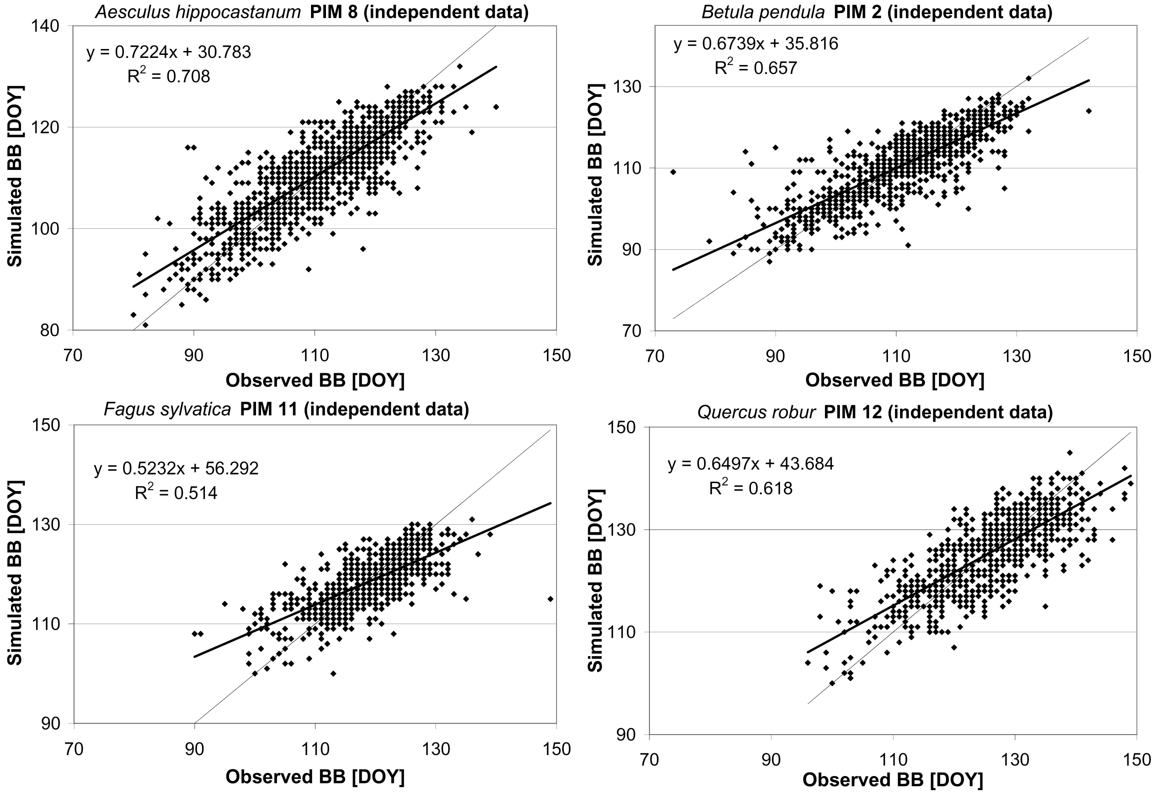
A. hippocastanum	***B. pendula***	***F. sylvatica***	***Q. robur***	***L. decidua***
PIM 8	PIM 2	PIM 4	PIM 11	PIM 5
PIM 2	PIM 5	PIM 11	PIM 12	PIM 4
PIM 5	PIM 8	PIM 5	PIM 10	PIM 2
PIM 10	PIM 11	PIM 10	PIM 6	PIM 1
PIM 11	PIM 10	PIM 8	PIM 5	PIM 11
PIM 4	PIM 4	PIM 2	PIM 4	CSM
			PIM 8	PIM 8
				PIM 10

In all species the group of models that are not significantly different and at the same time have the lowest average SSR have five models in common, i.e. PIM No. 4, 5, 8, 10, and 11. For each species the model with the lowest average SSR is not necessarily the model with the lowest overall SSR as can be seen by comparing the model column of Tab. 5-2 with the first row of Tab. 5-3. For *F. sylvatica*, e.g., the model with the lowest mean SSR is PIM 4 (Tab. 5-3), the lowest overall SSR, however, was found among the ten optimization runs of PIM 11 (Tab. 5-2). For *L. decidua* the CSM is within the group with the smallest SSR, for the other species this group only consists of PIMs.

For the PIMs the model with the minimum SSR and R^2 could differ for fitting and testing, respectively. For *B. pendula* and *Q. robur* the model with the lowest SSR and lowest R^2 when tested against independent data was PIM 5, for *F. sylvatica* it was PIM 10. In this case, however, the average SSR over ten optimization runs of these models was not significantly different (Tab. 5-3) and the R^2 values differed by 0.01. For *L. decidua* and *A. hippocastanum* is was the same PIM that found the lowest SSR for fitting and testing, respectively. For the fitting procedure the PIMs always converged to the lowest SSR and highest R^2 value compared to the classical models. In the case of testing the models with independent data, it was the TSM and the CSM for *L. decidua* and *B. pendula*, respectively, that had the lowest SSR and highest R^2 value. The PIMs had the lowest bias for all species, fitted and tested, except for testing of *L. decidua* where TSM had the lowest bias. The CSM performed better than the TSM, except for *F. sylvatica* and *L. decidua*. In terms of RMSE and MAE the PIM performed best for *A. hippocastanum*, *F. sylvatica* and *Q. robur*. For *B. pendula* and *L. decidua* the CSM and TSM performed equally well as the PIM concerning RMSE and MAE. The SM always had the lowest performance concerning all measures but still improved the NULL-model. On average the performance of the models decreased according to the ranking PIM, CSM, TSM, SM and NULL model with the above mentioned exceptions.

Best performance with respect to SSR also meant best performance concerning R^2, RMSE and MAE but not concerning the bias. In the case of *Q. robur* and *B. pendula* the PIM No. 5 performed best concerning SSR for model testing. With respect to the bias, however, it was PIM No. 10 and 11 that had the lowest bias for model testing of *Q. robur* and *B. pendula*, respectively. The total difference, however, was less than 0.01. Per species the two models had no significantly different average SSR over ten optimization runs (Tab. 5-3). Generally, the difference between all measures of model performance was small for the PIMs that were not significantly different concerning average SSR. Among those models the bias and R^2 values differed less than 0.03.

In general, the PIMs improved the classical model (SM, CSM, TSM). In terms of R^2 for model testing, the maximal improvement of an increase of 11% explained variance was achieved for *F. sylvatica*. For the other species this increase ranged from zero to 3% in the amount of explained variance for *B. pendula* and *A. hippocastanum*, respectively. For model fitting improvement of model performance for the PIMs in terms of R^2 ranged from 1% for *B. pendula* to 10% for *F. sylvatica* relative to the best classical model. In Fig. 5-5 plots of simulated against observed dates of BB are shown for the five species and best tested PIM models.



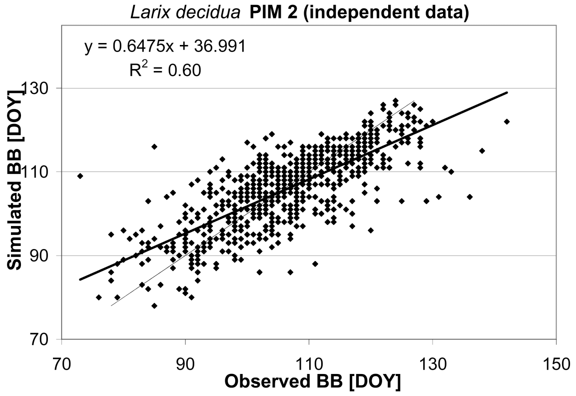


Figure 5-5. Simulated day of bud burst (BB) in Julian day of year (DOY) against observed values. PIM Promotor-Inhibitor Model. Results for the PIMs that best reproduced an independent data set. Thick line: Linear regression of simulated BB as a function of observed BB. Thin line: 1:1 line, indicating the perfect fit.

For the model results shown in Fig. 5-5 the respective parameter values and their allowed ranges during the optimization process are tabulated in Tab. 5-4. For the SM, CSM and TSM the parameters and ranges are stated in the ”Appendix A5 Parameters of the model fits for SM, CSM and TSM” on page 143. In general, the standard error of the parameters over the ten optimization runs is small. For *Q. robur* the variation of the parameters of the triangular temperature dependence function f_I is higher than for the other species. The limits of the parameters of f_I and f_P were adjusted after several test optimization runs such that the resulting optimal parameter set would not reach the limits of the allowed interval. Only for T_{opt} of f_I for *A. hippocastanum* the parameters still reach the allowed limits after several readjustments. The resulting temperature levels were lower for f_I than for f_P, for the allowed ranges as well as for the final values.

Table 5-4. Parameter values for the PIM models. For a description of the parameters see Eqn. 5-2 and Tab. 5-1. min, max: allowed ranges for the parameters during model fits. min SSR: Parameters for the PIM that performed best in terms of SSR for model fitting. mean: Average parameters after ten optimization runs. stderr: Standard error of the average parameters after ten optimization runs.

		f_I					f_P				
		T_{min}	T_{opt}	T_{max}	a_1	a_2	T_{min}	T_{opt}	T_{max}	a_3	a_4
allowed range	min	-25	-15	0	0	0	-20	0	5	0	0
	max	10	20	35	1	1	15	40	45	1	1
***A. hippo-castanum* PIM 8**	min SSR	-24.2	-15	23.13		0.095405	-10.94	23.55	38.57	0.053592	0.044243
	mean	-20.	-13.7	23.8		0.097702	-11.81	30.26	36.75	0.062753	0.045797
	stderr	1.01	0.49	0.36		0.002	0.3	1.83	1.71	0.002563	0.000837
***B. pendula* PIM 2**	min SSR	-24.96	-10.0	15.05		0.030619	-7.03	21.8	25.35	0.064803	0.045432
	mean	-22.32	-9.29	12.75		0.040222	-7.27	28.25	33.82	0.079167	0.048417
	stderr	0.79	0.16	0.29		0.002	0.48	1.72	2.06	0.004	0.001
***F. sylvatica* PIM 11**	min SSR	-10.34	-0.89	18.11		0.058326	-10.03	28.61	44.49	0.109494	0.039178
	mean	-9.15	-2.39	21.2		0.055358	-9.47	28.89	33.81	0.11551	0.039986
	stderr	0.38	0.32	0.97		0.002	0.25	1.24	1.77	0.004	0.000
***Q. robur* PIM 12**	min SSR	-23.05	-0.3	16.91	0.010379	0.055149	3.46	34.55	34.55	0.331253	
	mean	-24.29	-0.4	17.3	0.01165	0.059358	3.42	27.47	35.7	0.25418	
	stderr	0.22	0.22	0.28	0.001	0.003	0.06	1.87	1.49	0.02	
***L. decidua* PIM 2**	min SSR	-16.71	-13.60	34.98		0.005598	-3.24	18.38	37.02	0.112225	0.017149
	mean	-16.60	-14.18	31.36		0.006846	-3.32	26.24	32.46	0.206242	0.016441
	stderr	0.60	0.22	1.08		0.001	0.18	2.10	2.23	0.066568	0.001485

A sensitivity analysis for PIMs that best fitted the data (Tab. 5-3) was conducted with respect to the parameters for each species. Each parameter was changed by $\pm 10\%$ and the consequences in term of the relative change in the related SSR recorded. As a measure for sensitivity of the parameter with respect to the resulting SSR

$$S = \left| \frac{\Delta SSR}{SSR} \Big/ \frac{\Delta p}{p} \right|$$

with *p* being the respective parameter was taken. In Tab. 5-5 maximum *S* of either $+10\%$ or -10% is displayed for the best fitting PIM per species and parameter.

Table 5-5. Sensitivity *S* of the parameters of the best fitting PIM per species.

	T_{min}	T_{opt}	T_{max}	a_1	a_2	T_{min}	T_{opt}	T_{max}	a_3	a_4
A. hippocastanum PIM8	0.0	0.2	0.7		1.4	4.8	14.8	0.0	34.7	15.1
B. pendula PIM 2	0.0	0.2	0.9		1.3	2.5	12.8	0.0	19.6	9.0
F. sylvatica PIM 11	0.0	0.0	0.1		0.6	2.0	9.5	0.0	14.8	4.9
Q. robur PIM 12	0.0	0.0	1.3	2.0	3.3	1.1	6.0	0.0	3.1	
L. decidua PIM 2	0.0	0.5	0.7		3.6	0.6	5.1	0.0	5.8	0.8

The parameters of the temperature dependent breakdown function f_I show a low sensitivity for all models and species. The scaling parameter a_2 of f_I is more sensitive than the parameters of f_I concerning the resulting fit in terms of SSR. The parameters of the temperature dependent synthesis function f_P show a substantially higher sensitivity than the parameters of f_I. It is especially T_{opt} of f_P that plays a prominent role. T_{max} shows no marked influence on SSR in both f_I and f_P. The most sensitive are the scaling parameters a_3 and a_1/a_4. However, sensitivity does not only depend on the model structure but also on the other parameters. This shows the substantial difference in sensitivity of a_3 and a_4 for PIM 2 between *B. pendula* and *L. decidua*.

5.4 Discussion

The new promotor-inhibitor model (PIM) has similarities to other physiology-based models that have been proposed before, specifically the so-called parallel and deepening rest model (Hänninen 1994, Kramer 1994). Landsberg (1974) stated that for dormancy release it is essential that response to forcing temperatures is taken into account, even when the critical state of chilling has not yet been attained. The buds potential to respond to forcing temperatures increases with the time spent under chilling conditions. Thus, chilling and the accumulation of heat units parallel each other in contrast to successive completion of chilling and forcing as prescribed in the sequential model (SM). The concept of decreased accumulation of forcing temperatures through incomplete chilling is also incorporated in the PIM by the dampening factor (1-*I*) in the forcing term for the promotor *P* (Eqn. 5-1, Tab. 5-1). Promotor increment through temperature is modelled by a triangular temperature dependent function as for the inhibitor whereas in the parallel and other models of Hänninen (1994) this was realized by a logistic function. The deepening rest model (Kobayashi et al. 1982) stipulates that the state of chilling must increase before it can loosen its block on assimilation of heat units again. This notion is also allowed for in the PIM through the forcing term for the inhibitor *I* but it depends on the special model structure whether increasing *I* is allowed and if so then increasing *I* depends on the day length. For the other synthesis and breakdown terms the PIM also consequently features the influence of day length which has not been done in this form before. Thus, on the one hand, there are some elements in the new model framework that can also be discovered in other physiology-based and classical phenology models, even though model development started from other premises, on the other hand, the new approach resulted in new combinations and an extension of old concepts that seem promising. Especially the breakdown of the promotor is a new feature in the PIM. This relates to a decrease of state of forcing that has not been allowed before in the classical models.

Hänninen (1994) has combined several mathematical formulations of classical processes that are agreed to play a role in bud burst phenology, incorporated them into a common framework and came up with 96 different models. Two of them, the SM and the parallel model have also been used in this study. There has been no attempt so far to thoroughly test all these model with a consistent data set. Hänninen (1994) has tested all of his proposed methods but used parameter sets for the submodels as described in the literature. Kramer (1994) and Menzel (1997) fitted a limited number of these models to a limited data set. With the simulated annealing optimization algorithm in combination with parallel computing, it now becomes possible not only to test but also to fit all available phenology models to extensive data sets with a reasonable effort in modeling and computation time. A complete testing of all available models promises new insights into the important processes determining phenology of trees.

The simulated annealing algorithm is not only substantially accelerated through parallelization but also reveals satisfying convergence characteristics. Whereas Kramer (1994) had convergence problems using standard differential methods, Chuine (1998) and this study suggest that simulated annealing is an adequate method for fitting phenological models. Even though it was rarely the case that among ten optimization runs the same final minimum was found, the fact that all the resulting minima differed mostly less than one percent (Tab. 5-2) gives confidence that a global minimum is among the ten minima or at least close to them. Because the parameters belonging to the ten optimization runs also differed only little in most cases (Tab. 5-4), trust in the model fits is further increased. It would be interesting the apply different annealing algorithms or other restricted random search methods like genetic algorithms to compare results and gain further confidence in the resulting optimal fits.

The fact that several model structures yielded similar or not significantly different results when fitted to the data gives rise to the notion that the model might be overparameterized or that some model elements or processes do not add to the explanatory power of the model (see below). However, those model that were always among the best, i.e. PIM No. 4, 5, 8, 10, 11 and mostly also 2, had some interesting features in common (Tab. 5-1). Concerning the inhibitor part, those models exhibited only breakdown of the inhibitor, i.e. allowed only a continuous release of growth arresting internal conditions. Consequently, the promotor part showed a contrasting reaction and included breakdown of promotory compounds. Because the models only decided to either include synthesis of inhibitor or breakdown of promotor a direct comparison of the two processes was not possible. However, because breakdown of promotor was favoured, deepening of rest after partial rest release did not seem to play a dominant role. This suggests that indeed the state of forcing decreases in spring. This means that under unfavorable conditions in spring time a once attained physiological state of development is not only retained but set back and has to resume before it can further develop. PIM No. 1 and No. 7 included the promotor breakdown term but were not among the best fitted models. These two models, however, did not include any day length modification function showing another important result; day length seems to be an important factor for spring phenology. Even though this is known for a long time and often proved by experiments, day length was only seldomly introduced in phenological models (Hänninen 1994, Menzel 1997). Partly because former modeling studies did not find a relevant effect of day length or temperature was thought to be sufficient. However, in this study it improved the models.

The temperature response functions adjusted to the expected levels. This was also partly due to the predefined model structure, but the fact that breakdown function of inhibitor f_I had optimal temperatures below or around zero supports the chilling concept, meaning that low temperatures do enhance subsequent development.

There was a distinction of model structure among species. For the best fitted models only the late spring deciduous species, *F. sylvatica* and *Q. robur*, included the day length factor in the promotor synthesis term (PIM11 and PIM 12). The early spring and the needle leafed species *A. hippocastanum*, *B. pendula* and *L. decidua* did not include the day length factor in the promotor synthesis term of the model that fitted best to the data (PIM8 and PIM2). Based on these finding it can be speculated that the models support experimental evidence that for late spring species day length plays a more pronounced role than for early species. This has also been shown in experiments with *F. sylvatica* (Wareing 1953, Heide 1993a, 1993b). Among those models whose average SSR was not significantly different (Tab. 5-3) no such distinction could be made, therefore this evidence for different species featuring different processes is weak. However, it were those models that were always and almost exclusively in the group of the best models that had following feature in common: no breakdown of inhibitor, synthesis of promotor and day length influence on the promotor (PIMs No. 2,4,5,8,10,11). Exception was *Q. robur* where PIM 6 and 12 where substituted for PIM 2 and *L. decidua* where additionally PIM 1 and CSM where among the best models.

One of the possible model refinements to obtain a better discrimination between the different model structures can be derived from the results shown in Tab. 5-4 and Tab. 5-5. It has been mentioned above that the model might be overparameterized. The optimal temperature of the triangular response functions are for most species at a level that is close to the maximal temperatures observed, meaning that one half of the triangular functions becomes obsolete. For f_I the temperature allowing for maximal rate of inhibitor breakdown T_{opt} is below -10°C except for *F. sylvatica* and *Q. robur* where it is below 0°C. In the data used here daily mean temperatures only seldomly reach levels below -10°C. The same is the case for T_{opt} of f_P. Here optimal temperature is always above 18°C which is a level that is only seldomly reached in spring time in Germany. The sensitivity analysis supports the notion that the lower(upper) part of the triangular function $f_I(f_P)$ is obsolete because T_{min} of f_I and T_{max} of f_P have no marked influence on model performance concerning SSR (Tab. 5-5). Thus, the parameters T_{min} of f_I and T_{max} of f_P can be dropped from the model without substantial impact on model performance. Thus, a two-parametric saw-tooth function or just a two-parametric linear function could have also been used. To obtain an optimal parameter set for only seven parameters the models would have to be fitted again which is an exercise for further studies. Other functions could also be used. Using a triangular function might not be optimal. The triangular functions seem to adjust themselves to wide ranges to have a broad range of optimal temperatures. A function with a broad range of optimal temperatures and a faster (nonlinear) decrease to zero outside the optimal range might be more appropriate here.

Even though the SM was found to be best in earlier simulation studies (Kramer 1994), it was outperformed by the classical models TS and CSM in this study. The convergence criteria for the SM were as good as for the other models implying that this was not due to an optimization problem. It also shows that a high number of parameters does not necessarily improve model results. However, the parameters of the SM sometimes reached the imposed limits, especially for the temperature parameters. Allowing wider ranges as Kramer (1994) could improve the results. Substantial improval of BB simulation of the PIM compared to the classical models was achieved for *F. sylvatica*, implying that the TSM and the CSM lack one or more important processes for this species, e.g. inclusion of day length (Heide 1993a). Among the best fitting models only for *L. decidua* a classical model (CSM) could not be distinguished from other best fitting models (Tab. 5-3), even TSM performed better concerning the bias and SSR when tested against independent data. The improvement in model performance was sometimes low in terms of R^2 but as indicated in Chapter 3 the maximal achievable R^2 is below unity because of the uncertainty of both observational data and model driving parameters, i.e. mainly temperature. When more data on the uncertainty of phenological and temperature data become available a maximal achievable R^2 can be estimated (see Chapter 3). On this basis of a maximal achievable R^2 or other measure of performance a better judgment can be made about how large the improval of model performance has been.

For the other modelled species both the TSM or the CSM perform satisfyingly well and for the sake of parsimony they seem to be sufficient for most applications.

For model simulations over longer time periods, like e.g. several decades, possibly in connection with tree growth models there is one factor that might become important and favours the PIM again, i.e. the bias. The general notion of phenological models is that they tend to overestimate early observations and underestimate late observations. The models are not flexible enough to map extreme conditions and are rather suited to simulated average observations. A measure for this systematic mistake is the introduced bias. For *B. pendula* the CSM was better than the PIM concerning SSR and R^2. However, the CSM underestimated the observed values on DOY 130 by four days and observed values on DOY 140 by nine days on average whereas the PIM only by five days. For *F. sylvatica* improvement of the bias also had significant effects. As mentioned earlier, the day of BB in late spring has a substantial impact on the annual carbon balance of trees due to the already long days and favorable temperatures (see ”4.4.6 Potentially absorbed radiation” on page 68). When incorporating models of phenology in forest growth models a reduction in the bias of BB for late or early years of two or more days will have significant long term effects on tree growth simulations. The fact that the bias could not be entirely removed even with the better PIMs, indicates that the models are still not flexible enough with respect to extreme weather conditions. The growth promoting influence of high temperatures, for instance, might increase more stronger in time than it was already modelled by the influence of day length. Introduction of more non-linearities into the model could further improve the bias. On visual inspections no systematic station bias could be found, i.e. stations that were systematically overestimated or underestimated which could be a hint to provenance influences. However, this should be tested more extensively in future studies.

6 Summary, conclusions and perspectives

The length of the vegetation period (VP) plays a central role for the interannual variation of carbon exchange of terrestrial ecosystems. This has been demonstrated by experiments as well as in simulation studies (see Chapter 1). Observational data analysis at various spatial resolutions has indicated that the length of the VP has increased in the last decades in the northern latitudes largely due to an advancement of bud burst (BB) (see Chapter 1 and Chapter 4). This phenomenon has been widely discussed in the context of Global Warming because phenology is correlated to temperatures.

6.1 Secular trends and trend reversal

To come to a better recognition and understanding of the patterns of phenology and their causes longer phenological time series are needed than those used in former studies that covered only the last decades (Menzel et al. 2001, Peñuelas and Filella 2001, Chmielewski and Rötzer 2002). By merging digitalized historical phenological observations with data from existing data bases it was possible to construct unique observational phenological time series for regions in Southern Germany that spanned over a hundred years. Analyzing the patterns of spring phenology over the last century in Southern Germany revealed that

- the strong advancement of spring phases that has been observed in former studies especially in the decade before 1999 are not singular events in the course of the 20th century but similar trends were observed in the decade before the 1950s. These trends occurred also at the same absolute level. Distinct periods of varying trend behavior for important spring phases could be distinguished.
- there are marked differences in trend behavior between the early and late spring phases. Early spring phases changed as regards the magnitude of their negative trends from strong negative trends between 1931 and 1948 to moderate negative trends between 1948 and 1984 and back to significant negative trends between 1984 and 1999. Late spring phases showed a different behavior. Negative trends between 1931 and 1948 are followed by significant positive trends between 1948 and 1984 and then significant negative trends between 1984 and 1999.

This marked difference in trend development between early and late spring phases especially in the last two analyzed periods was also found all over Germany for the two periods 1951 to 1984 and 1984 to 1999 (see Chapter 4). The curvilinear trend development found for late spring phases has not been described for plant phenology before and its detection was made possible by the application of new methodologies to derive reliable long-term phenological time series (see below and Chapter 2, Chapter 4). A similar trend behavior was described for bird egg-laying in the UK by Crick and Sparks (1999). They also noted that laying date became later in the 1960s and 1970s, and then earlier through the 1980s and 1990s. These trends were correlated to changes in mean March and April temperatures.

There are more historical data available for other regions in Germany that were not digitalized and used for this thesis. The patterns found in the secular trend analysis for Southern Germany were confirmed for the rest of Germany, starting from 1951. It is an interesting task for future studies to investigate secular phenology patterns all over Germany by including additional historical data.

6.2 Spring phenology and temperatures

Investigating the causes of the observed spring phenology patterns, the dominating influence of temperature on spring phenology and its modifying effect on autumn phenology was confirmed in this thesis. At the regional scale, spring phenology between 1901 and 1999 was shown to be correlated to linear functions of mean monthly temperatures ($R^2 \approx 0.7$) whereas autumn phenology was only weakly correlated to such functions ($R^2 \approx 0.15$) These correlations could be enhanced to $R^2 \approx 0.8$ and $R^2 \approx 0.36$ for spring and autumn phenology, respectively, using temperature data starting in 1951 (Chapter 4; Table 4-3 on page 54).

- None of the temperature functions determining spring phenology was significantly correlated with a global annual CO_2 signal.
- An index for large scale regional circulation patterns (NAO-index) could only to a small part ($R^2 \approx 0.23$) explain the observed phenological variability in spring.

The observed different trend behavior of early and late spring phases was mapped to the changing trend of mean March and April temperatures. Mean March temperatures increased on average over the 20th century, accompanied by an increasing variation in the last 50 years. April temperatures, however, decreased between the end of the 1940s and the mid-1980s, followed by a marked warming after the mid-1980s. In the middle of the 20th century mean April temperatures were higher than or at the same level as in the late 1990s for Southern Germany and the whole of Germany, respectively.

There are reports stating that mean annual temperature development on the global and European scales showed no trends between 1950 and 1990 followed by a marked increase (Ramaswamy et al.

2001). But even studies about temperature trends in Germany on a monthly basis (Schönwiese and Rapp 1997) did not report a substantial decrease in April temperatures between 1950 and the mid-1980s. This is because Schönwiese and Rapp (1997) analyzed 30-year climatic normal periods. In these periods this marked trend does not show.

It can be concluded that the negative trends of spring phenology in the last decades found in this and in former studies are part of multi-decadal fluctuations over the 20th century that vary with the species and the relevant seasonal temperatures. Because of these fluctuations a correlation with an observed Global Warming signal could not be found. In order to significantly attribute the observed advancement in spring phenology of the last decades to Global Warming, it has to be further investigated to what extent the observed Global Warming, which is in most cases an annually integrated signal, relates to distinct seasonal temperature regimes at a monthly resolution on a regional level.

6.3 Consequences for production and competition

On average all investigated spring phases have advanced between 5 and 20 days between 1951 and 1999 for all Natural Regions in Germany. A marked difference between late and early spring phases is due to the above mentioned differing behavior before and after the mid-1980s. Leaf coloring (LC) was delayed between 1951 and 1984 for all tree species. After 1984 LC was advanced. Length of the VP has increased between 1951 and 1999 for all considered tree species by an average of ten days throughout Germany (see Chapter 4; Table 4-4 on page 67).

Changes in the length of the vegetation period do not directly relate to changed potentially absorbed radiation because of the sinusoidal course of average daily radiation over the year. This thesis and earlier studies showed that it is predominately the change in spring phases which contributes to a change in the potentially absorbed radiation. Additionally, it is the late spring species that are relatively more favored by an advanced BB because they can additionally exploit longer days and higher temperatures per day advancement. However, to assess the relative change in potentially absorbed radiation among species it is important to consider changes in both spring and autumn phenology as well as where these changes are located in the year.

To assess the impact of the change in growing season length on species competition, models of the length of the VP have to be incorporated into more complex forest ecosystem models. This has been done by Kramer (1995). This study, however, considered a period of only 14 years with a simple representation of phenology using monospecific forest models. With now available multi-species forest growth models also representing such important processes as nutrient cycling and competition for light, water and nutrients, e.g. 4C (Schaber et al. 1999, Suckow et al. 2001), it should be possible to better assess climate change impacts on species competition and composition. With such models it is also possible to differentiate the effects of changed growing season length, nitrogen imissions and changed forest management on the recently observed growth trends of European forests (Spieker et al. 1996). A ranking of the sensitivities of these three important processes to the performance of forest ecosystems would be a valuable contribution to assess possibilities to mitigate the adverse effects of a changing climate for the forest sector. This will be done in the future using the multi-species process-based forest model 4C and including the species-specific parameterizations of the phenology models that have been developed in this thesis (see below and Chapter 5).

6.4 Combined time series and uncertainties

The above-mentioned marked difference between early and late spring phases, predominantly caused by the curvilinear trend development of late spring temperatures, has not been detected in earlier studies, even with similar data (Menzel et al. 2001, Chmielewski and Rötzer 2001, 2002). Prerequisite for these new findings was a methodology that could cope with the intrinsic uncertainties and shortcomings of the available phenological data that obscured and biased earlier studies (Menzel et al. 2001, Chmielewski and Rötzer 2001, 2002). It could be shown that the use of average time series (Chmielewski and Rötzer 2001, 2002) is likely to introduce a bias in the derived time series (see Chapter 2). The analysis of single station time series covering different periods (Menzel et al. 2001) is also likely to be biased because trends are not independent of the period being analyzed such that trends of time series covering different periods are difficult to compare (see Chapter 4; ”4.4.4 Trend turning points” on page 56). The method that overcomes both drawbacks is the analysis of so-called combined time series, first proposed by Häkkinen et al. (1995). In this thesis a conclusion was drawn about how to best estimate such combined time series (see ”2.4 Discussion” on page 25). The method was extended through the introduction of a robust estimation method allowing a more reliable detection of outliers, which are also a prominent source of bias in phenological data sets (see Chapter 2). The proposed methods allowed reliable phenological time series to be derived through the combination of several messy single station time series. For the analysis of regional phenology presented above, this methodology allowed the construction of time series that spanned over 100 years. With such long time series it was possible to define standard periods that were covered by all combined time series making proper trend comparison feasible and removing the bias of earlier studies. Detecting these periods of marked trend behavior was facilitated by the application of statistical methods to much longer time scales than used in earlier studies. The proposed methodology further allowed derivation of locally combined time series that can be assumed to better match local weather measurements than single station data. This increased the available data for model development (see Chapter 5).

For future refinement of data processing methods an attempt was made to approximate the contribution of the most important sources of variability to the uncertainty of phenological observations (see Chapter 3). Apart from the analyzed protocolling errors (see Chapter 2), the microclimatic site influences, genetic variation and the observers themselves were identified as sources of uncertainty. The uncertainty of phenological observations varies from year to year depending on the prevailing environmental conditions. Because of the lack of data a quantitative relationship between environmental conditions and this year-to-year variation could not be established. In a first assessment, however, using the available data it was concluded that 99% of the phenological observations at a certain site will vary within 24 days around a hypothetical mean. This gives further support to the proposed 30-day rule to detect protocolling outliers (see Chapter 2; ”2.2.4 Outlier detection” on page 14). The genetic uncertainty was estimated to be fourfold the uncertainty introduced by different observers. The microclimatic site variability could not be quantified and it can be expected to strongly vary among regions. Further refinement of such uncertainty analysis using more data will also facilitate model development. Given the intrinsic uncertainty of phenological data it is not possible even for a perfect phenological model to perfectly fit the observational data. When the different error sources can be quantified it will be possible to assess how well a model can be expected to perform.

6.5 Phenology models and parameterization

Models are needed for the projection of effects of a changing climate on phenology. In Chapter 1 the research areas are described in which phenological models are useful and have already been applied. Therefore a new phenology model for BB was developed that could not only better predict local phenology over a wide geographical range (Germany) than the classical models but also allowed new insights into the possible mechanisms controlling phenology in plant species.

The new model was based on simple interactions between inhibitory and promotory agents that are assumed to control the developmental status of a plant. Several alternative model structures were formulated emphasizing different hypothetical physiological processes. The parameter optimization process revealed that for most species a different model structure suited best. Driving variables were temperature and day length. The main modeling results were:

- The new models fitted the observations better than classical models.
- The bias of the classical models, i.e. overestimation of early observations and underestimation of late observations, could be reduced but not completely removed.
- The different favoured model structures for each species indicated that for the late spring phases (BB of *F. sylvatica* and *Q. robur*) the photoperiod played a more dominant role than for early spring phases (BB *B. pendula* and *A. hippocastanum*).

- Chilling only plays a subordinate role for spring BB compared to temperatures preceding BB.

The presented modeling approach generally performed better than classical models and allowed for a species-specific weighting of the dominating processes. Experimental studies supported the finding that day length plays an important role for late spring BB. However, potentials for model improvement have been shown. Reducing the number of parameters speeding up model optimization and studying a wider range of nonlinear temperature responses can further reduce the bias in BB prediction.

It has been shown that usually a change in BB has more effect on the yearly absorbed radiation than change in LC. However, in some cases a change in timing of LC can have significant effects on the amount of absorbed radiation (see above and Chapter 4; Table 4-5 on page 69). Thus, to come to the desired better representation of the length of the growing season in long-term simulations of forest ecosystems, models for the end of the VP should be developed. No model that tried to predict LC from environmental conditions was found in any of the studies cited here. Even though it was not the aim of this thesis to develop a LC model, first hints for important aspect of such a model were found. In the regression analysis in Chapter 4 (”4.4.3 Stepwise regression analysis” on page 53) almost half of the regional variation of LC phases could be predicted by linear functions of mean monthly temperatures (Table 4-3 on page 54). The months August and September had positive parameters indicating delayed LC with increasing temperature in these months. However, the late spring and early summer months April, May and June showed negative parameters indicating advanced LC with higher early summer temperatures. High temperatures, especially in June, are often connected with drought which is known to stress trees. Severe drought events are known to lead to advanced LC. A simple empirical approach combining the beneficial effect of high temperatures in early autumn with the adverse effect of high temperature in summer by a climatic water balance, for instance, seems to be a promising starting point for modeling LC.

7 Bibliography

References marked by an asterix (*) indicate sources of the historical phenological database (HPDB) from the German Weather Service (DWD). These references are sometimes not unequivocal and lack information that could not be obtained from standard sources. The references are then just stated as obtained from the DWD.

Ahas, R. 1999. Long-term phyto-,ornitho-, and ichthyophenological time series analyses in Estonia. Int. J. Biometeorol. 42:119-123.

Barrodale I., F.D.K. Roberts. 1973. An improved algorithm for discrete L_1 linear approximation. SIAM J. Numer. Anal. 10(5):839-848.

Barrodale I., F.D.K. Roberts. 1974. Algorithm 478 : Solution of an overdetermined system of equations in the L_1 norm. Comm. ACM 17:319-320.

Barnett V., T. Lewis. 1996. Outliers in Statistical Data. New York, John Wiley & Sons. 584p.

Baumgartner A. 1952. Zur Phänologie von Laubhölzern und ihre Anwendung bei lokalklimatischen Untersuchungen. Berichte des DWD in der US-Zone 42:69-73.

Bergengren, J. C., S. L. Thompson, D. Pollard, and R. M. DeConto. 2001. Modeling global climate-vegetation interactions in a doubled CO_2 world. Climatic Change 50:31-75.

Black, T. A., W. J. Chen, A. G. Barr, M. A. Arain, Z. Chen, Z. Nesic, E. H. Hogg, H. H. Neumann, and P. C. Yang. 2000. Increased carbon sequestration by a boreal deciduous forest in years with a warm spring. Geophysical Research Letters 27:1271-1274.

Bloomfield, P., and W. L. Steiger. 1983. Least absolute deviations: theory, applications and algorithms. Birkhäuser, Boston. 349p.

Böhm, R., I. Auer, M. Brunetti, M. Maugeri, T. Nanni, and W. Schoner. 2001. Regional temperature variability in the European Alps: 1760- 1998 from homogenized instrumental time series. International Journal of Climatology 21:1779-1801.

Bradley, N. L., A. C. Leopold, J. Ross, and W. Huffaker. 1999. Phenological changes reflect climate change in Wisconsin. Proceedings of the National Academy of Sciences of the United States of America 96:9701-9704.

*Breitinger, E. 1975. Die Schäfer'sche Chronik. Oberdingen.

Bugmann, H., R. Grote, P. Lasch, M. Lindner, and F. Suckow. 1997. A new forest gap model to study the effects of environmental change on forest structure and functioning. Pages 255-261 in G. M. J. Mohren, K. Kramer, and S. Sabaté, editors. Impacts of Global Change on Tree Physiology and Forest Ecosystems. Kluwer Academic Publishers, Dordrecht.

Bugmann, H. K. M., and A. M. Solomon. 2000. Explaining forest composition and biomass across multiple biogeographical regions. Ecological Applications 10:95-114.

Cannell M.G.R., R. Smith 1983. Thermal time, chill days and prediction of budburst in *Picea sitchensis*. J. Appl. Ecol. 20:951-963.

Chen W.J., Black T.A., Yang P.C., Barr A.G., Neumann H.H., Nesic Z., Blanken P.D., Novak M.D., Eley J., Ketler R:J., and C. R. 1999. Effects of climatic variability on the annual carbon sequestration by a boreal aspen forest. Global Change Biol. 4:41-53.

Chmielewski, F.-M. 1996. The International Phenological Gardens across Europe. Present state and perspectives. Phenology and Seasonality 1:19-23.

Chmielewski F.-M., and T. Rötzer. 2001. Response of tree phenology to climate change across Europe. Agricultural and Forest Meteorology 108:101-112.

Chmielewski F.-M., and T. Rötzer. 2002. Annual and spatial variability of the beginning of growing season in Europe in relation to air temperature changes. Clim. Res. 19:257-264.

Chuine, I., P. Cour, and D. D. Rousseau. 1998. Fitting models predicting dates of flowering of temperate-zone trees using simulated annealing. Plant, Cell & Environ. 21:455-466.

Chuine, I., P. Cour, and D. D. Rousseau. 1999. Selecting model to predict the timing of flowering of temperate trees: implications for tree phenology modelling. Plant, Cell & Environ. 22:1-13.

Chuine, I. 2000. A unified model for bud burst of trees. J. Theor. Biol.207(3):337-347.

Chuine, I., J. Belmonte, and A. Mignot. 2000. A modelling analysis of the genetic variation of phenology between tree populations. Journal of Ecology 88:561-570.

Chuine, I., and E. G. Beaubien. 2001. Phenology is a major determinant of tree species range. Ecology Letters 4:500-510.

Ciret, C., J. Polcher, and X. Le Roux. 1999. An approach for simulating the phenology of savanna ecosystems in the Laboratoire de Meteorologie Dynamique general circulation model. Global Biogeochemical Cycles 13:603-621.

Claussen, M., V. Brovkin, and A. Ganopolski. 2001. Biogeophysical versus biogeochemical feedbacks of large-scale land cover change. Geophysical Research Letters 28:1011-1014.

Cramer, W., H. H. Shugart, I. R. Noble, F. I. Woodward, H. Bugmann, A. Bondeau, J. A. Foley, R. H. Gardner, B. Lauenroth, L. F. Pitelka, O. Sala, and R. W. Sutherst. 1999. Ecosystem composition and structure. Pages 190-228 in B. H. Walker, W. L. Steffen, J. Canadell, and J. S. I. Ingram, editors. The Terrestrial Biosphere and Global Change: Implications for Natural and Managed Ecosystems. Cambridge University Press, Cambridge.

Crick H., T. Sparks. 1999. Climatic change related to egg-laying trends. Nature 399:423-424.

Corbeil R.R., and S. R. Searle. 1976. Restricted maximum likelihood estimation of variance components in the mixed model. Technometrics 18(1):31-38.

Daigo, Y., and M. Suzuki. 1974. Phenology in Japan. 217p.

Diekmann, M. 1996. Relationship between phenology of perennial herbs and meteorological data in deciduous forests of Sweden. Can. J. Bot. 74:528-537.

Dixon W.J. 1950. Analysis of extreme values. Ann. Math. Statist. 21:488-506.

Dodge Y. 1987. Statistical Data Analysis based on the L_1-Norm and Related Methods. Amsterdam, Elsevier Science Publishers. 464p.

DWD 1945 - 1950. Deutsches Meteorologisches Jahrbuch der US-Zone 1945 - 1950, DWD (German Weather Service).

Eisenhut, G. 1957. Blühen, Fruchten und Keimen in der Gattung *Tilia*. Thesis Universität München.

Fitzjarrald, D. R., O. C. Acevedo, and K. E. Moore. 2001. Climatic consequences of leaf presence in the eastern United States. Journal of Climate 14:598-614.

Foley, J. A., S. Levis, M. H. Costa, W. Cramer, and D. Pollard. 2000. Incorporating dynamic vegetation cover within global climate models. Ecological Applications 10:1620-1632.

*Frankhauser 1979-1983. Klimatologische und phänologische Beobachtungen im Canton Bern im Jahr 1869 bis 1882. Bern, Buchdruckerei Stampfli, Lack & Scheim.

Geiger, R., R. H. Aron, and P. Todhunter. 1995. The Climate near the Ground. Harvard University Press, Cambridge, Mass. 528p.

Gerstengarbe, F.-W., and P. C. Werner. 1999. Estimation of the beginning and end of recurrent events within a climate regime. Clim Res 11:97-107.

Gitay, H., S. Brown, W. Easterling, B. Jallow, J. Antle, M. Apps, R. Beamish, T. Chapin, W. Cramer, J. Frangi, J. Laine, L. Erda, J. Magnuson, I. Noble, J. Price, T. Prowse, T. Root, E.-D. Schulze, O. Sirotenko, B. Sohngen, and J. Soussana. 2001. Ecosystems and their goods and services. Pages 641-692. Chapter 5 in J. J. McCarthy, O. F. Canziani, N. A. Leary, D. J. Dokken, and K. S. White. Climate Change 2001: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Third Assessment Report of the Intergovernmental Panel on Climate Climate Change. Cambridge University Press.

Gornik, W. 1995. Further investigations into the problems of finding a mean value in phenological date series. Arboreta Phaenologica 40:26-27.

Goudriaan, J., and van Laar H.H. 1994. Modelling potential crop growth processes. Kluwer Academic Press, Dortrecht, The Netherlands. 238p.

Goudriaan, J., and J. C. Zadoks. 1995. Global climate change: modelling the potential responses of agro-ecosystems with special reference to crop protection. Environmental Pollution 87:215-224.

Goulden M.L., J.W. Munger, S.-M. Fan, B.C. Daube, S.C. Wofsky. 1996. Exchange of carbon dioxide by a deciduous forest: response to interannual climate variability. Science 271:1576-1578.

Grieser, J., T. Staeger, and C.-D. Schönwiese. 2000. Statistische Analyse zur Früherkennung globaler und regionaler Klimaänderungen aufgrund des anthropogene Treibhauseffektes. Berichte des Institutes für Meteorologie und Geophysik der Universität Frankfurt/Main 103. Eigenverlag. Frankfurt. 228p.

Gropp, W., E. Lusk, and A. Skjellum. 1994. Using MPI: Portable Parallel Programming with the Message-Passing Interface. MIT Press, Cambridge, MA, USA. 307p.

*Grossherzgl. statist. Bureau zu Schwerin. (Publisher) 1860-1894. Die Beoachtungen über die Entwicklung der Pflanzen in Mecklenburg (1860 - 1894).

Haarlov, N., and B. B. Petersen. 1952. Measurements of temperature in bark and wood of Sitka spruce, Kopenhagen.

Häkkinen R., T. Linkosalo, P. Hari. 1995. Methods for combining phenological time series: application to bud burst in birch (Betula pendula) in Central Finland for the period 1896-1955. Tree Physiol. 15:721-726.

Hänninen, H. 1991. Does climatic warming increase the risk of frost damage in northern trees? Plant, Cell and Environment 14:449-454.

Häninnen, H. 1994. Effects of climatic change on trees from cool and temperate regions: an ecophysiological approach to modelling of bud burst phenology. Can. J. Bot. 73:183-199.

Härle, A. 1938. Blüte- und Erntezeiten von Winterroggen and Winterweizen in Deutschland. Mitt. Biol. Reichsanst. f. Land- u. Forstwirtsch. 57.

Hannerz, M. 1999. Evaluation of temperature models for predicting bud burst in Norway spruce. Can. J. For. Res. 29:9-19.

Hanover, J. W. 1980. Control of tree growth. BioScience 30:756-762.

Hartley H.O., J.N.K. Rao. 1967. Maximum likelihood estimation for the mixed analysis of variance model. Biometrika 54:93-108.

Hartley H.O., J.N.K. Rao, L.R. Lamotte. 1978. A simple synthesis-based method of variance component estimation. Biometrics 34:233-242.

Heide, O. M. 1993a. Daylength and thermal time responses of bud burst during dormancy release in some northern deciduous trees. Physiologia plantarum 88:531-540.

Heide, O. M. 1993b. Dormancy release in beech buds (*Fagus sylvatica* L.) requires both chilling and long days. Physiologia plantarum 89:187-191.

Hemmerle W.J., H.O. Hartley. 1973. Computing Maximum Likelihood Estimates for the Mixed A.O.V. Model Using the W-Transformation. Technometrics 15(4):819-831.

Hense, A., R. Glowienka-Hense, M. Müller, and P. Braun. 2002. Spatial model of phenological observations to analyze their interannual variation in Germany. Agricultural and Forest Meteorology. accepted for publication.

*Hertzer, H. W. 1895. Grenzmarken der Pflanzenentwicklung bei Wernigerode. Schriften d. Naturwiss. Vereins d. Harzes i. Wernigerode 10.

*Hoffman H. and E. Ihne 1884. Phänologischer Aufruf. In: Beiträge zur Phänologie. Giessen.

Houghton, J. T., Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, X. Dai, K. Maskell, and C. A. Johnson, editors. 2001. Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press. 881p.

Hubert M. 1997. The breakdown value of the L_1 estimator in contingency tables. Statist. and Probab. Lett. 33:419-425.

Hubert M., P.J. Rousseeuw 1997. Robust estimation with both continuous and binary regressors. J. Statist. Plann. Interference 57:153-163.

Hunter, A. F., and M. J. Lechowicz. 1992. Predicting the timing of bud burst in temperate trees. J. Appl. Ecol. 29:597-604.

Huntley, B. 1991. How plants respond to climate change: migration rates, individualism and the consequences for plant communities. Annals of Botany 67:15-22.

Hurrell, J. W., and H. VanLoon. 1997. Decadal variations in climate associated with the North Atlantic oscillation. Climatic Change 36:301-326.

Hurrell, J. W., Y. Kushnir, and M. Visbeck. 2001. Climate - The North Atlantic oscillation. Science 291:603-604.

*Ihne, E. 1883-1941. Phänologische Mitteilungen. Arbeiten der Landwirtschaftskammer Hessen and Self Publishing. Darmstadt. 1-59.

*Ihne, E. 1892. Die ältesten pflanzenphänologischen Beobachtungen in Deutschland. Ber. d. Oberhess. Ges. f. Natur- u. Heilkunden 28:53 - 56.

Ingber, L. 1989. Very Fast Simulated Re-Annealing. Mathematical and Computer Modelling 12:967-973.

Ingber, L. 1993. Simulated Annealing - Practice Versus Theory. Mathematical and Computer Modelling 18:29-57.

Jackson, R. B. 2000. Belowground processes and global change. Ecological Applications 10:397-398.

Johnson, I. R., and J. H. M. Thornley. 1985. Temperature dependance of plant and crop processes. Ann. Bot. 55:7-24.

Jones, P. D. 1994. Hemispheric surface air-temperature variations - a reanalysis and an update to 1993. Journal of Climate 7:1794-1802.

Jones, P. D., T. M. L. Wigley, and P. B. Wright. 1986. Global Temperature-Variations between 1861 and 1984. Nature 322:430-434.

Jones, P. D., T. Jónsson, and D. Wheeler. 1997. Extension to the North Atlantic oscillation using early instrumental pressure observations from Gibraltar and South West Iceland. International Journal of Climatology 17:1433-1450.

Jones, P. D., M. New, D. E. Parker, S. Martin, and I. G. Rigor. 1999. Surface air temperature and its changes over the past 150 years. Reviews of Geophysics 37:173-199.

Keeling C.D., Chin. J. F. S., Whorf T.P., 1996. Increased activity of northern vegetation inferred from atmospheric CO_2 measurements. Nature 382:146-149.

Kendall, M. and J.D. Gibbons 1990. Rank Correlation Methods. E. Arnold, London. 260p.

Kindermann, J., F.-W. Badeck, G. Würth, and G. H. Kohlmaier. 1996. Interannual variation of carbon exchange fluxes in terrestrial ecosystems. Global Biogeochemical Cycles 10:737-746.

King E.P. 1953. On some procedures for the rejection of suspected data. J. Amer. Statist. Assoc. 48:531-533.

Kirkpatrick S., C. D. Gelatt, Jr. and M. P. Vecchi. 1983. Optimization by Simulated Annealing. Science 220:671-680.

Kirschbaum, M. U. F. 2000. Forest growth and species distribution in a changing climate. Tree Physiology 20:309-322.

Kobayashi, K., L. Fuchigami, and M. J. English. 1982. Modeling temperature requirements for rest development in Cornus sericea. Journal of American Society of Horticultural Science 107:914-918.

*Koepert, O. 1891-1898. Phänologische Beoachtungen aus dem Ostkreise des Herzogtums Sachsen-Anhalt aus dem Jahre 1890 bis 1895. Mitt. d. Vereins f. Erdk. z. Halle.

Kramer K. 1994. Selecting a model to predict the onset of growth of *Fagus sylvatica*. J. Appl. Ecol. 31:172-181.

Kramer K. 1995. Modelling comparison to evaluate the importance of phenology for the effects of climate change in growth of temperate-zone deciduous trees. Clim. Res. 5:119-130.

Kramer, K. and G.M.J. Mohren 1996. Sensitivity of FORGRO to climatic change scenarios: A case study of *Betula pubescens*, *Fagus sylvativa* and *Quercus robur* in the Netherlands. Climatic Change 34:231-237.

Kramer K., A. Friend, I. Leinonen. 1996. Modelling comparison to evaluate the importance of phenology and spring frost damage for the effects of climate change on growth of mixed temperate-zone deciduous forests. Clim. Res. 7:31-41.

Kramer K., I. Leinonen, D. Loustau. 2000. The importance of phenology for the evaluation of impact of climate change on growth of boreal, temperate and Mediterranean forests ecosystems: an overview. Int. J. Biometeorol. 44(2):67-75.

Kriebel, H. B., and C. W. Wang. 1962. The interaction between provenance and degree of chilling in bud-break of sugar maple. Silvae Genet. 11:125-130.

Kucharik, C. J., J. A. Foley, C. Delire, V. A. Fisher, M. T. Coe, J. D. Lenters, C. Young-Molling, N. Ramankutty, J. M. Norman, and S. T. Gower. 2000. Testing the performance of a Dynamic Global Ecosystem Model: Water balance, carbon balance, and vegetation structure. Global Biogeochemical Cycles 14:795-825.

Landsberg, J. J. 1974. Apple fruit bud development and growth; analysis and an empirical model. Annals of Botany 38:1013-1023.

Lasch, P., and M. Lindner. 1995. Application of two forest succession models at sites in north east Germany. Journal of Biogeography 22:485-492.

Lavender, D. P., and S. N. Silim. 1987. The role of plant growth regulators in dormancy in forest trees. Plant Growth Regulation 6:171-191.

Lechowicz, M. J. 1984. Why do temperate trees leaf out at different times? Adaptation and ecology of forest communities. The American Naturalist 126:821-842.

Lechowicz, M. J., and T. Koike. 1995. Phenology and seasonality of woody plants: an unappreciated element in global change research? Can. J. Bot. 73:147-148.

Leinonen, I. 1996. Dependence of dormancy release on temperature in different origins of *Pinus sylvestris* and *Betula pendula* seedlings. Scandinavian Journal of Forest Research 11:122-128.

Leinonen, I., T. Repo, and H. Hanninen. 1997. Changing environmental effects on frost hardiness of Scots pine during dehardening. Annals of Botany 79:133-138.

Lessmann, H. 1950. Temperaturverhältnisse in Häufelreihen. Jahr. Ber. d. Bad. Landeswetterd.:35-45.

Lieth, H. 1974. Phenology and seasonality modelling. Springer Verlag, Berlin.

Lindroth, A., A. Grelle, A. S. Moren, Anders. 1998. Long-term measurements of boreal forest carbon balance reveal large temperature sensitivity. Global Change Biology 4:443-450.

Linkosalo T., R. Häkkinen, P. Hari. 1996. Improving the reliability of a combined phenological time series by analysing observation quality. Tree Physiol. 16:661-664.

Linkosalo T. 2000. Analyses of the spring phenology of boreal trees and its response to climate change. PhD Thesis: University of Helsinki. Department of Forest Ecology Publications No. 22.

Linkosalo T., T.R. Carter, R. Häkkinnen, P. Hari. 2000. Predicting spring phenology and frost damage risk of Betula spp. under climatic warming: a comparison of two models. Tree Physiol. 20:1175-1182.

Lloyd J. 1999. The CO_2 dependance of photosynthesis, plant growth responses to elevated CO_2 concentrations and their interaction with soil nutrient status. II. Temperate and boreal forest productivity and the combined effects of increasing CO_2 concentrations and increased nitrogen deposition at a global scale. Functional Ecology 13:439-459.

McCarthy, J. J., O. F. Canziani, N. A. Leary, D. J. Dokken, and K. S. White, editors. 2001. Climate Change 2001: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Third Assessment Report of the Intergovernmental Panel on Climate Climate. Cambridge University Press. 1032p.

McKean J.W., R.M. Schrader. 1987a. Least Absolute Error Analysis of Variance. In: Statistical Data Analysis Based on the L_1-Norm and Related Methods. Ed. Dodge Y. Amsterdam, Elsevier Science Publishers B.V.:297-305

McKean J.W., R.M. Schrader. 1987b. Coefficients of determination for least absolute deviation analysis. Statist. and Probab. Lett. 5:49-54.

Melillo, J. M., T. V. Callaghan, F. I. Woodward, and E. Salati. 1990. Effects on ecosystems. Pages 283-310 in J. T. Houghton, G. J. Jenkins, and J. J. Ephraums, editors. Climate Change: The IPCC Scientific Assessment. Cambridge University Press, Cambridge.

Menzel A. 1997. Phänologie von Waldbäumen unter sich ändernden Klimabedingungen - Auswertung der Beobachtungen in den Internationalen Phänologischen Gärten und Möglichkeiten der Modellierung von Phänodaten. Forstwissenschaftliche Fakultät der Universität München. München. 150p.

Menzel A. and P. Fabian. 1999. Growing season extended in Europe. Nature 397:659.

Menzel A. 2000. Trends in phenological phases in Europe between 1951 and 1996. Int. J. Biometeorol. 44:76-81.

Menzel, A., N. Estrella, and P. Fabian. 2001. Spatial and temporal variability of the phenological seasons in Germany from 1951 to 1996. Global Change Biology 7:657-666.

Mergen, F. 1963. Ecotypic variation in *Pinus strobus*. Ecology 44:716-727.

Metropolis, N., A. W. Rosenbluth, M. Rosenbluth, and A. H. Teller. 1953. Equation of state calculations by fast computing machines. J. Chem. Phys. 21:1087-1092.

Milliken G.A., D.E. Johnson. 1992. Analysis of Messy Data. Volume I: Designed Experiments. New York, Chapman & Hall, 365p.

Mitchel, J. H., S. Fukai, and M. Cooper. 1996. Influence of phenology on grain yield variation among barley cultivars grown under terminal drought. Aust. J. Agric. Res. 47:757-774.

Monteith, J. L. 1972. Solar radiation and productivity in tropical ecosystems. J. Appl. Ecol. 9:744-766.

Monteith, J. L. 1977. Climate and the efficiency of crop production in Britain. Royal Society of London, Philosophical Transaction 281:277-294.

Murray, M. B., M. G. R. Cannel, and R. I. Smith. 1989. Date of bud burst of fifteen tree species in britain following climatic warming. J. Appl. Ecol. 26:693-700.

Murray, M. B., R. I. Smith, I. D. Leith, D. Fowler, H. S. J. Lee, A. D. Friend, and P. G. Jarvis. 1994. Effects of elevated CO2, nutrition and climatic warming on bud phenology in sitka spruce (*Picea sitchensis*) and their impact on the risk of frost damage. Tree Physiology 14:691-706.

Myking, T., and O. M. Heide. 1995. Dormancy release and chilling requirement of buds of latitudinal ecotypes of *Betula pendula* and *Betula pubescens*. Tree Physiology 15:697-704.

Myneni R.B., C.D. Keeling, C.J. Tucker, G. Asrar, R.R. Nemani. 1997. Increased plant growth in the northern latitudes from 1981 to 1991. Nature 386:698-702.

Nitsch, J. P. 1957. Photoperiodism in woody plants. Proceedings of the Am. Soc. for Horticultural Sciences 79:526-544.

Parmesan C., N. Ryrholm, C. Stefanescu, J.K. Hill, C.D. Thomas, W.J. Descimon, J.A. Thomas , and M. Warren. 1999. Polewards shifts in geographical ranges of butterfly species associated with regional warming. Nature 399:579-583.

Parry, M. L., J. H. Porter, and T. R. Carter. 1990. Agriculture: climatic change and its implications. Trends in Ecology and Evolution 5:318-322.

Patterson H.D., R. Thompson. 1971. Recovery of inter-block information when block sizes are unequal. Biometrika 58(3):545-554.

Peñuelas, J. and I. Filella. 2001. Phenology - Responses to a warming world. Science **294**:793-794.

Penning de Vries, F. W. T., D. M. Jansen, H. F. M. ten Berge, and A. Bakema. 1989. Simulation of ecophysiological processes of growth in severeal annual crops. Pudoc, Wageningen. 270p.

Perry, T. O. 1971. Dormancy of trees in winter. Science 171:29-36.

*Plieninger, Schober, et al. 1828 - 1920. Erscheinungen im Tier- Und Pflanzenleben / -reich. Jahresber. ü.d. Witterung i. Württemb. and Dtsch. Meteorol. Jb. f. Württemb.

Post, E., M. C. Forchhammer, N. C. Stenseth, and T. V. Callaghan. 2001. The timing of life-history events in a changing climate. Proceedings of the Royal Society of London Series B-Biological Sciences 268:15-23.

Powell, L. E. 1987. Hormonal aspects of bud and seed dormancy in temperate-zone woody plants. Hortscience 22:845-850.

Prentice, I. C., W. Cramer, S. P. Harrison, R. Leemans, R. A. Monserud, and A. M. Solomon. 1992. A global biome model based on plant physiology and dominance, soil properties and climate. Journal of Biogeography 19:117-134.

Prentice, I. C., M. T. Sykes, and W. Cramer. 1993. A simulation model for the transient effects of climate change on forest landscapes. Ecological Modelling 65:51-70.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. 1992. Numerical recipes in C. Cambridge University Press, New York. 994p.

Ramaswamy, V., O. Boucher, J. Haigh, D. Hauglustaine, J. Haywood, G. Myhre, T. Nakajima, G. Y. Shi, and S. Solomon. 2001. Radiative Forcing of Climate Change. Pages 349-416 in J. T.

Houghton, Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, X. Dai, K. Maskell, and C. A. Johnson, editors. Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press.

Rencher A.C. 2000. Linear Models in Statistics. New York, Wiley. 578p.

Repo, T., H. Hänninen, and S. Kellomäki. 1996. The effects of long-term elevation of air temperature and CO_2 on the frost hardiness of Scots pine. Plant, Cell and Environment 19:209-216.

Robertson, G. W. 1968. A biometeorological time scale for a cereal crop involving day and night temperatures and photoperiod. Int. J. Biometeorol. 12:191-223.

Rogers, J. C. 1997. North Atlantic storm track variability and its association to the North Atlantic oscillation and climate variability of Northern Europe. Journal of Climate 10:1635-1647.

Roumet, C., and J. Roy. 1996. Prediction of the growth response to elevated CO_2: A search for physiological criteria in closely related grass species. New Phytologist 134:615-621.

Rousseeuw P.J. 1984. Least Median of Squares Regression. J. Amer. Statist. Assoc. 79(388):871-880.

Rousseeuw P. J., A.M. Leroy. 1987. Robust Regression and Outlier Detection. New York, John Wiley & Sons. 329p.

Roy, D. B., and T. H. Sparks. 2000. Phenology of British butterflies and climate change. Global Change Biology 6:407-416.

Ruimy, A., P. G. Jarvis, D.D Baldochi, and B. Saugier. 1995. CO_2 fluxes over plant canopies and solar radiation: a review. Advances in Ecological Research 21:1-68.

Sachs, L. 1997. Angewandte Statistik. Springer, Berlin. 877p.

Samish, R. M. 1954. Dormancy in woody plants. Annu. Rev. Plant Physiol. 5:183-204.

Sarvas, R. 1972. Investigations on the annual cycle of development of forest trees. Active period. Communicationes Instituti Forestalis Fenniae 76:5-110.

Sarvas, R. 1974. Investigations on the annual cycle of development of forest trees. Autumn dormancy and winter dormancy. Communicationes Instituti Forestalis Fenniae 84:5-101.

Scamoni, A. 1938. Über Eintritt und Verlauf der männlichen Kiefernblüte. Z. f. F. u. Jagdw. 70:289-315.

Schaber, J. 1997. FARMSIM: A dynamic model for the simulation of yields, nutrient cycling and resource flows on Philippine small-scale farming systems. Master Thesis. University of Osnabrück, Osnabrück. 114p.

Schaber, J., F.-W. Badeck, and P. Lasch. 1999. Ein Modell der Sukzessionsdynamik europäischer Wälder - Forest Ecosystems in a changing Environment (4C). Pages 212-217 in D. R. Pelz, O. Rau, and J. Saborowski, editors. Deutscher Verband forstlicher Versuchsanstalten - Sektion forstliche Biometrie und Informatik. 11. Jahrestagung und Internationale biometrische Gesellschaft. Deutsche Region. Arbeitsgruppe Ökologie, Herbstkolloquium, Freiburg. p. 212-217.

Schimel, D. S., T. G. F. Kittle, A. K. Knapp, T. R. Seastedt, W. J. Parton, and V. B. Brown. 1991. Physiological interactions along resource gradients in a tallgrass prairie. Ecology 72:672-684.

Schimel, D. S., J. I. House, K. A. Hibbard, P. Bousquet, P. Ciais, P. Peylin, B. H. Braswell, M. J. Apps, D. Baker, A. Bondeau, J. Canadell, G. Churkina, W. Cramer, A. S. Denning, C. B. Field, P. Friedlingstein, C. Goodale, M. Heimann, R. A. Houghton, J. M. Melillo, B. Moore, D. Murdiyarso, I. Noble, S. W. Pacala, I. C. Prentice, M. R. Raupach, P. J. Rayner, R. J. Scholes, W. L. Steffen, and C. Wirth. 2001. Recent patterns and mechanisms of carbon exchange by terrestrial ecosystems. Nature 414:169-172.

Schnelle, F. .1950. Hundert Jahre phänologische Beobachtungen im Rhein-Main-Gebiet 1841-1939, 1867-1947. Meteorologische Rundschau 3 (7/8):150-156.

Schnelle, F. 1952. Über die Abhängigkeit der Laubverfärbung von der Temperatur. Ber. d. D Wetterd. i. d. US-Zone Nr. 38:227-228.

Schnelle F. and F. Witterstein 1952. Beiträge zur Phänologie Deutschlands II. Tabellen phänologischer Einzelwerte von etwas 500 Stationen der Jahre 1936 bis 1944. Berichte des Deutschen Wetterdienstes in der US-Zone Nr. 41. Deutscher Wetterdienst. Bad Kissingen. 120p.

Schnelle, F., and E. Volkert. 1957. Vorschläge zur Errichtung 'Internationaler Phänologischer Gärten' als Stationen eines Grundnetztes für internationale phänologischen Beoachtungen. Meteorologische Rundschau 10:130-133.

Schnelle F. and F. Witterstein 1964. Beiträge zur Phänologie Deutschlands IV. Tabellen phänologischer Einzelwerte von etwas 500 Stationen der Jahre 1922 bis 1935. Berichte des Deutschen Wetterdienstes Nr. 95. Deutscher Wetterdienst. Offenbach a. M. 123p.

Schnelle F. 1955. Pflanzenphänologie. Leipzig, Geest & Portig. 299p.

Schnelle, F., and E. Volkert. 1964. Internationale Phänologische Gärten. Stationen eines Grundnetzes für international phänologische Beoachtungen. Agricultural Meteorology 1:22-29.

Schönwiese, C.-D., and J. Rapp. 1997. Climate Trend Atlas of Europe: Based on observations 1891 - 1990. Kluwer Academic Publishers, Dortrecht. 228p.

Schrepfer, H. 1922. Blüte- und Erntezeiten des Winterroggens in Deutschland. Arb. d. Deutsch. Landwirtsch. Ges. 321.

Schwartz, M. D., and Karl. T. R. 1990. Spring phenology: nature's experiment to detect the effect of green-up on surface maximum temperatures. Monthly Weather Rev. 118:883-890.

Schwartz, M. D. 1992. Phenology and springtime surface layer change. Monthly Weather Rev 120:2570-2578.

Schwartz, M. D. 1997. Spring Index Models: An Approach to Connecting Satellite and Surface Phenology. Pages 23-38 in In Lieth, H; Schwartz, M. D. (ed): Phenology of Seasonal Climates, Backhuys Leiden.

Schwartz, M. D., and B. E. Reiter. 2000. Changes in North American spring. International Journal of Climatology 20:929-932.

Searle S.R. 1971. Linear Models. New York, Wiley. 532p.

Searle S.R. 1987. Linear Models for Unbalanced Data. New York, Wiley. 536p.

Siegmund, A. 1999. Das Klima der Baar. Mannheimer geographische Arbeiten 51. Geographisches Institut der Universität Mannheim. Mannheim. 294p.

Smith H., and N. P. Kefford. 1964. The chemical regulation of the dormancy phases of bud development. Amer. J. Bot. 51:1002-1012.

Sneyers R. 1990. On the statistical analysis of series of observations. Technical Note No. 143. Geneva, Switzerland, World Meteorological Organization. 192p.

Sokal R. R., and F. J. Rohlf. 1995. Biometry, 3rd Edition. Freemann & Co., New York. 887p.

Sparks, T. H., and P. D. Carey. 1995. The Responses of Species to Climate over two Centuries - an Analysis of the Marsham Phenological Record, 1736-1947. Journal of Ecology 83:321-329.

Sparks, T. H. 1999. Phenology amd changing pattern of bird migration in Britain. Int. J. Biometeorol. 42:134-138.

Sparks T. H., E. P. Jeffree, and C. E. Jeffree. 2000. An examination of the relationship between flowering times and temperature at the national scale using long-term phenological records from the UK. International Journal of Biometeorology 44:82-87.

Spieker, H., K. Mielikäinen, M. Köhl, and J. P. Skovsgaard. 1996. Growth trends in European forests. Springer, Berlin. 273p.

Suckow, F., F.-W. Badeck, P. Lasch, and J. Schaber. 2001. Nutzung von Level-II-Beobachtungen für Test und Anwendungen des Sukzessionsmodells FORESEE. Beiträge für Forstwirtschaft und Landschaftsökologie 35:84-87.

Suzuki, S. 1951. Agricultural Meteorology. 267p.

Sykes, M. T., I. C. Prentice, and W. Cramer. 1996. A bioclimatic model for the potential distributions of North European tree species under present and future climates. Journal of Biogeography 23:203-233.

Thompson R. 1969. Iterative estimation of variance components for non-orthogonal data. Biometrics 25:767-773.

*Toepfer H. 1882-1910. Phänologische Beoachtungen in Thüringen 1881 bis 1909. Mitt. d. Botan. Vereins Irmischia zu Sonderhausen and Mitt. d. Vereins f. Erdk. z. Halle.

Thornley, J. H. M., and M. G. R. Cannell. 2001. Soil carbon storage response to temperature: an hypothesis. Annals of Botany 87:591-598.

unknown author. 1781-1792. Ephemerides Societatis Meteorologicae Palatinae.

Vanninen, P., and A. Mäkela. 2000. Needle and stem wood production in Scots pine *(Pinus sylvestris*) trees of different age, size and competitive status. Tree Physiology 20:527-533.

Vassella A. 1997. Phänologische Beobachtungen des Bernischen Forstdienstes von 1869-1882: Witterungseinflüsse und Vergleiche mit heutigen Beobachtungen. Phänologie von Waldbäumen: Historische und aktuelle Beobachtungen. Bern, Bundesamt für Umwelt, Wald und Landschaft. 73:9-75.

Vegis A. 1964. Dormancy in higher plants. Annu. Rev. of Plant Physiol. 15:185-224.

Vegis A. 1973. Dependence of growth processes on temperature. In H. Precht, J. Christophersen, H. Hensel and W. Larcher, editors. Temperature and life. Springer-Verlag, Berlin. 145-169 .

Walker, B. H. 1991. Ecological consequences of atmospheric and climate change. Climatic Change 18:301-316.

Walker, B. H. 1994. Landscape to regional-scale responses of terrestrial ecosystems to global change. Ambio 23:67-73.

Wang, J. Y. 1960. A critique of the heat unit approach to plant response studies. Ecology 41:785-790.

Wareing, P. F. 1953. Growth studies in woody species. V Photoperiodism in dormant buds of *Fagus sylvatica* L. Physiol. Plant. 6:692-706.

Wareing, P. F. 1956. Photoperiodism in woody plants. Ann. Rev. Plant Physiol. 7:191-214.

Wareing, P. F., and P. F. Saunders. 1971. Hormones and Dormancy. Annu. Rev. Plant Physiol. 22:261-288.

Wigley, T. M. L. 1987. Radiative contribution of different trace gases to the Greenhouse Effect. Climate Monitor 16:14-28.

White M.A., S.W. Running, P.E. Thornton. 1999. The impact of growing season length variability on carbon assimilation and evapotranspiration over 88 years in the eastern US deciduous forest. Int. J. Biometeorol. 42:139-145.

Wilson, K. B., D. D. Baldocchi, and P. J. Hanson. 2000. Spatial and seasonal variability of photosynthetic parameters and their relationship to leaf nitrogen in a deciduous forest. Tree Physiology 20:565-578.

Worrall, J., and F. Mergen. 1967. Environmental and genetic control of dormancy in *Picea abies*. Physiologia Planetarum 20:733-745.

Worrall, J. 1983. Temperature-bud burst relationships in *Picea amabilis* and subalpine fir provenance tests replicated at different elevations. Silvae Genet. 32:203-209.

Worrall, J. 1993. Temperature effects on bud burst and leaf fall in Subalpine larch. Journal of Sustainable Forestry 1:1-18.

*Ziegler J. 1891a. Pflanzenphänologische Beoachtungen zu Frankfurt am Main. Frankfurt.

*Ziegler J. 1891b. Vegetationszeiten zu Frankfurt a. M. (Pflanzenphänologischer Kalender). Frankfurt.

*Ziegler J. 1904. Thermische Vegetationskonstanten. Frankfurt.

Zotz, G., M. Franke, and M. Woitke. 2000. Leaf phenology and seasonal carbon gain in the invasive plant, *Bunias orientalis* L. Plant Biology 2:653-658.

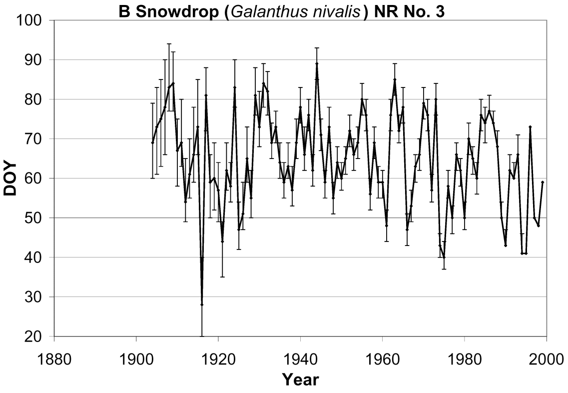
Appendix A1 Phenological time series from 1880 until 1999 in the Natural Regions No. 3, 11 and 23 in Germany

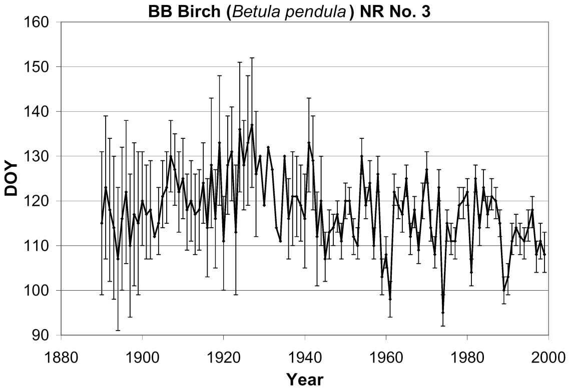
In the following the phases are coded as follows

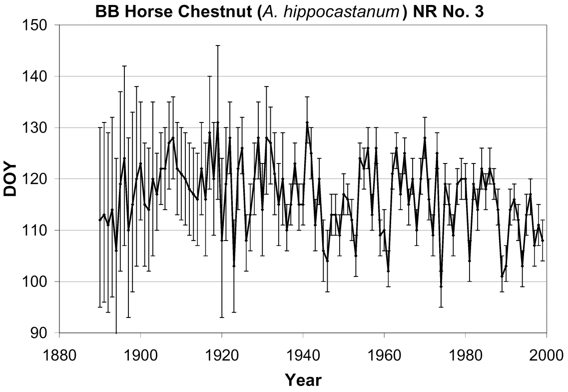
- B Blossoming
- BB Bud Burst
- LC Leaf Coloring

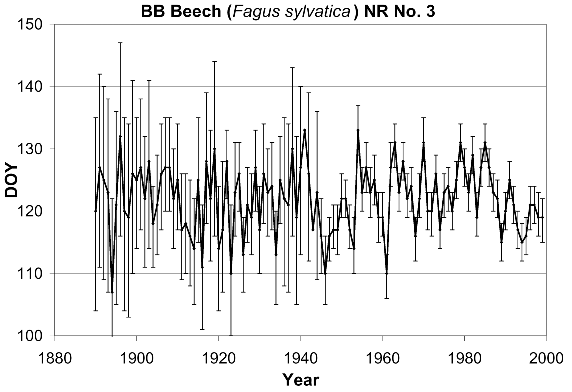
Error bars indicate 95% confidence intervals of the estimated combined time series.

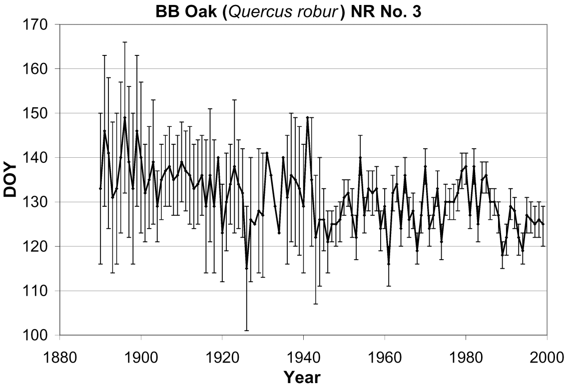
A1.1 Combined phenological time series in NR No. 3

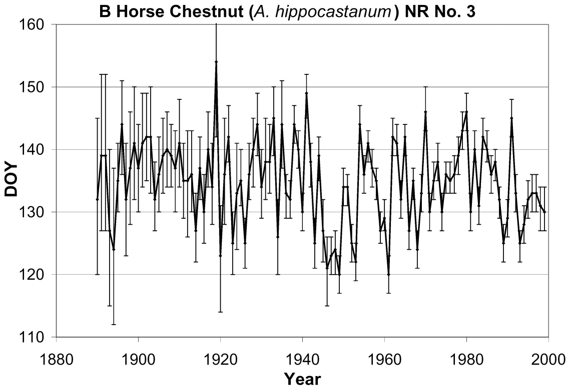


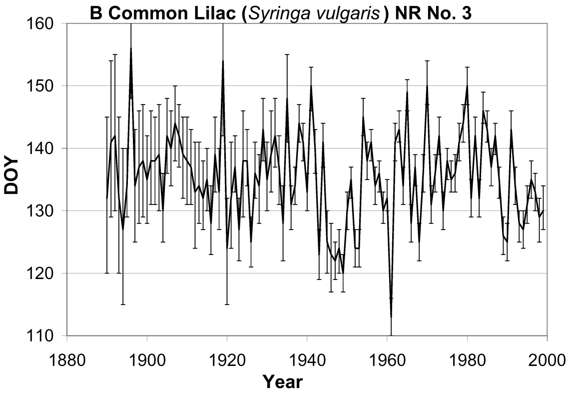


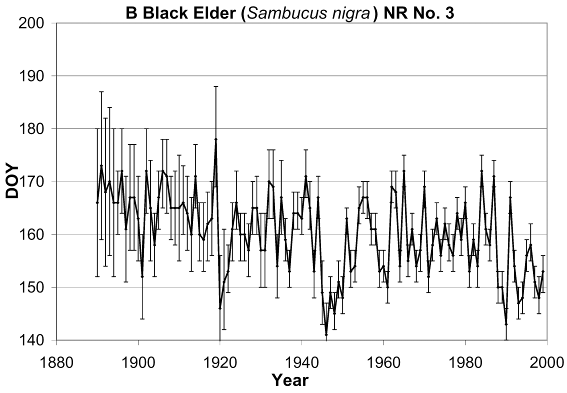


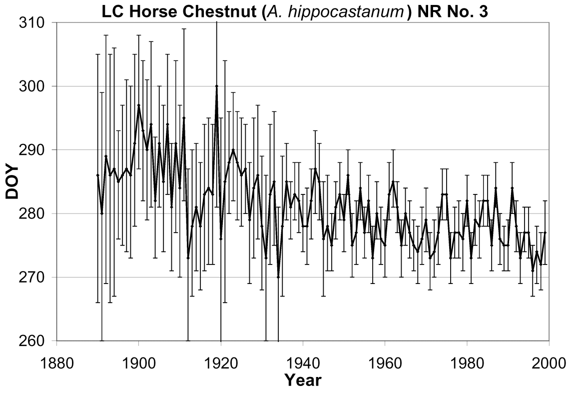


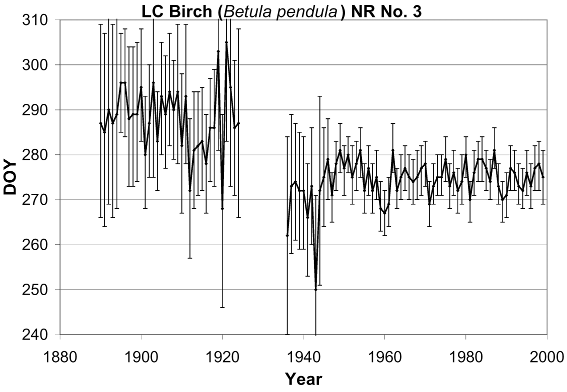


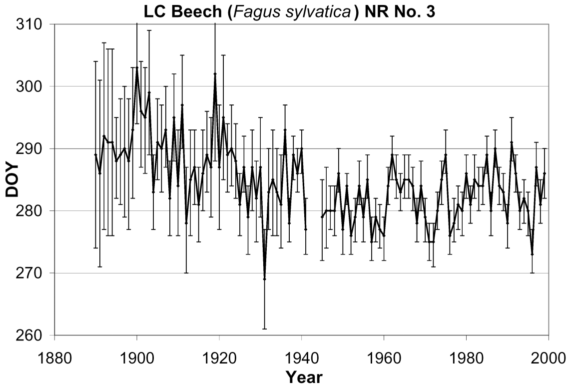


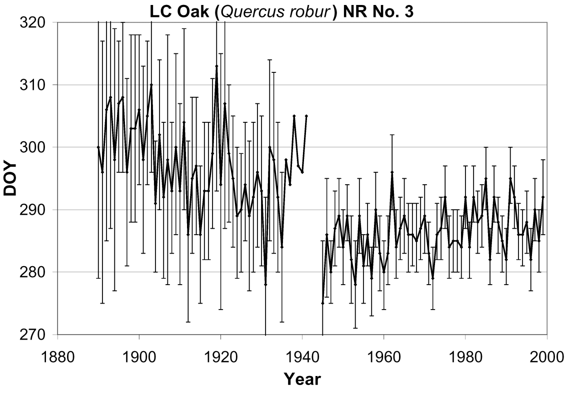

B Common Lilac (*Syringa vulgaris*) NR No. 3


B Black Elder (*Sambucus nigra*) NR No. 3

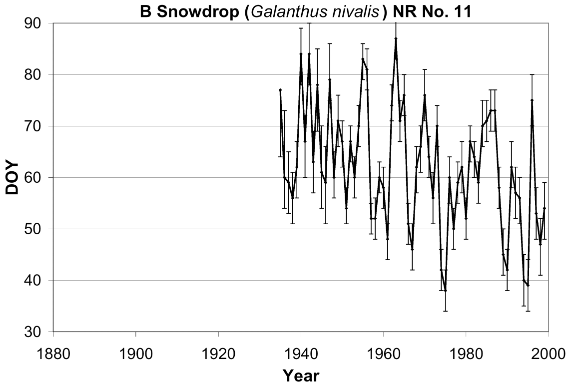

LC Horse Chestnut (*A. hippocastanum*) NR No. 3

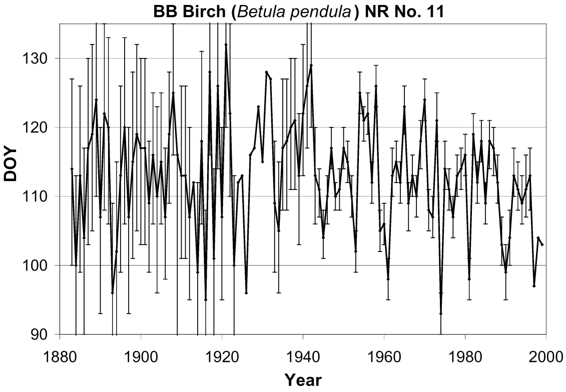

LC Birch (*Betula pendula*) NR No. 3

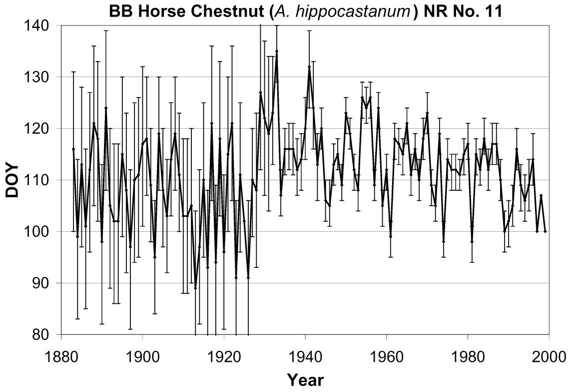

LC Beech (*Fagus sylvatica*) NR No. 3

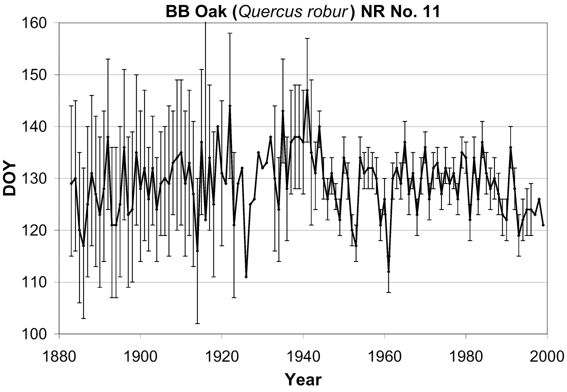

LC Oak (*Quercus robur*) NR No. 3

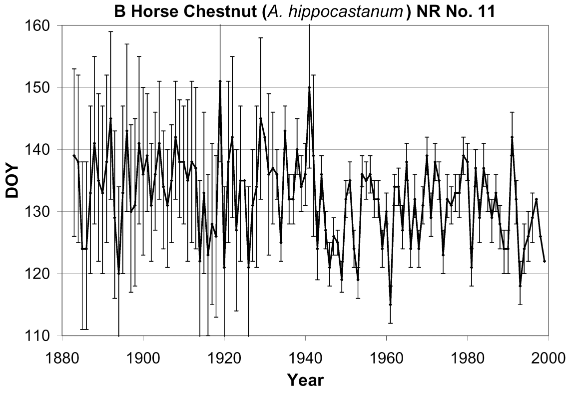
A1.2 Combined phenological time series in NR No. 11

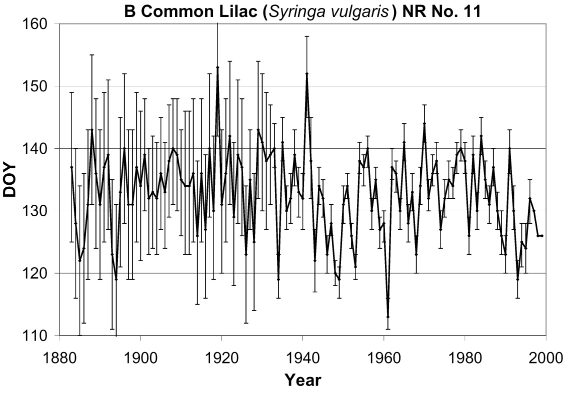

B Snowdrop (*Galanthus nivalis*) NR No. 11

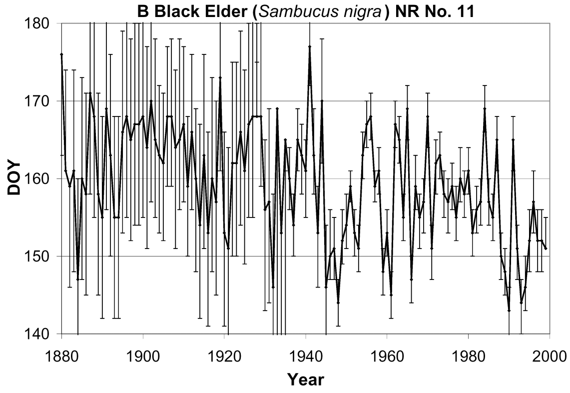

BB Birch (*Betula pendula*) NR No. 11

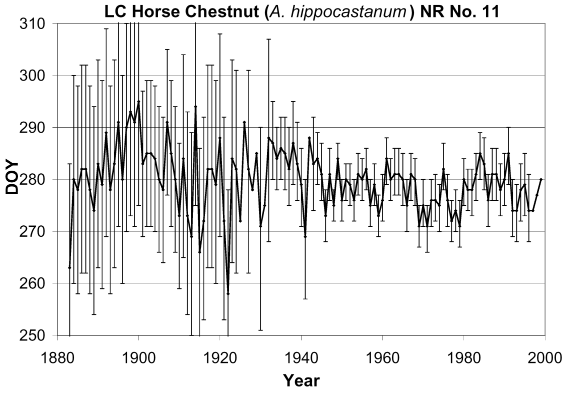

BB Horse Chestnut (*A. hippocastanum*) NR No. 11

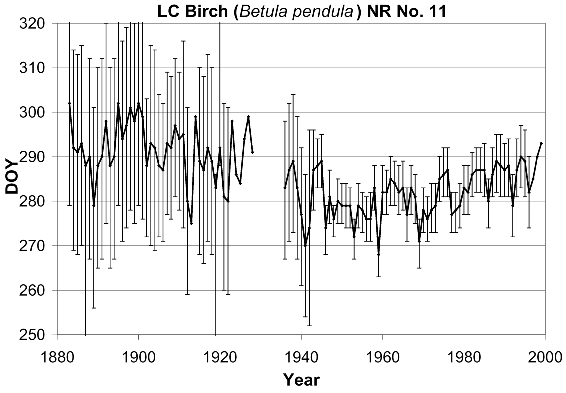

BB Oak (*Quercus robur*) NR No. 11

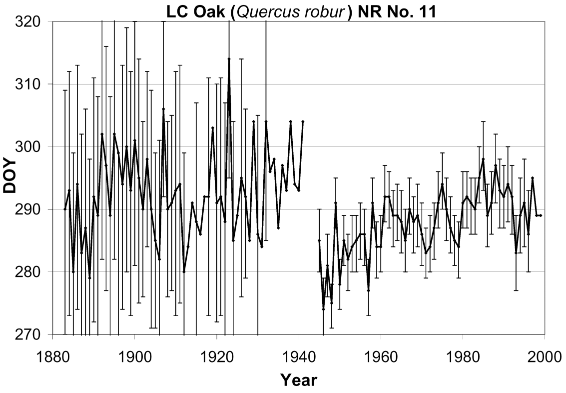

B Horse Chestnut (*A. hippocastanum*) NR No. 11


B Common Lilac (*Syringa vulgaris*) NR No. 11

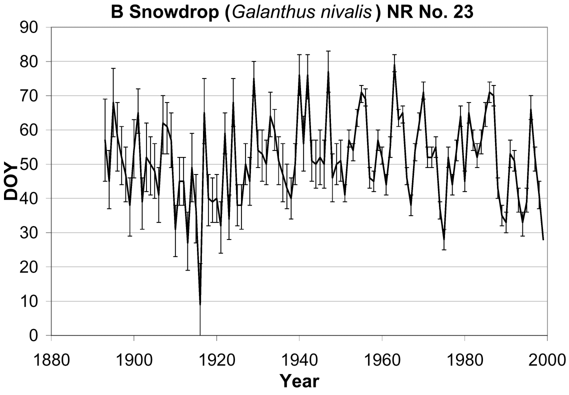

B Black Elder (*Sambucus nigra*) NR No. 11

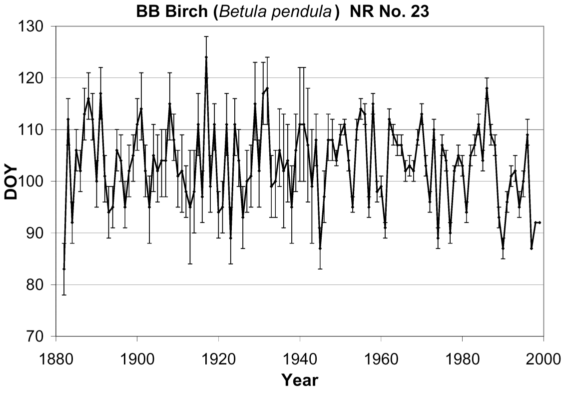

LC Horse Chestnut (*A. hippocastanum*) NR No. 11

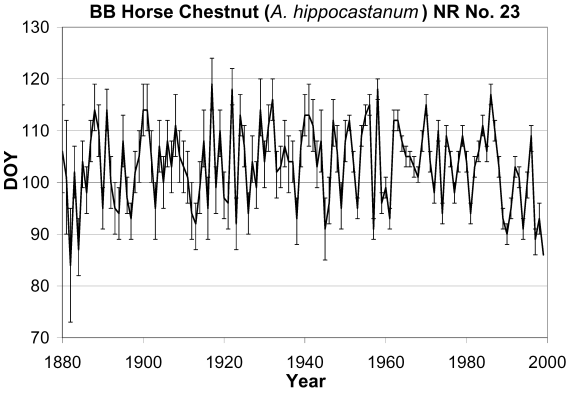

LC Birch (*Betula pendula*) NR No. 11
DOY
Year
320
310
300
290
280
270
260
250
1880
1900
1920
1940
1960
1980
2000

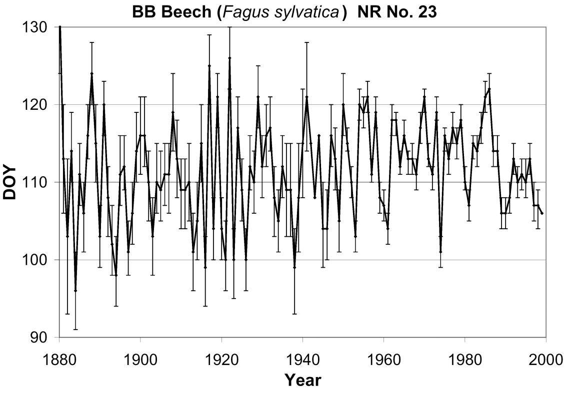

LC Oak (*Quercus robur*) NR No. 11
DOY
Year
320
310
300
290
280
270
1880
1900
1920
1940
1960
1980
2000

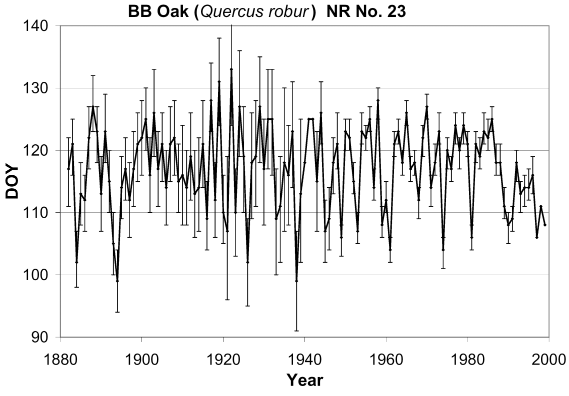
A1.3 Combined phenological time series in NR No. 23

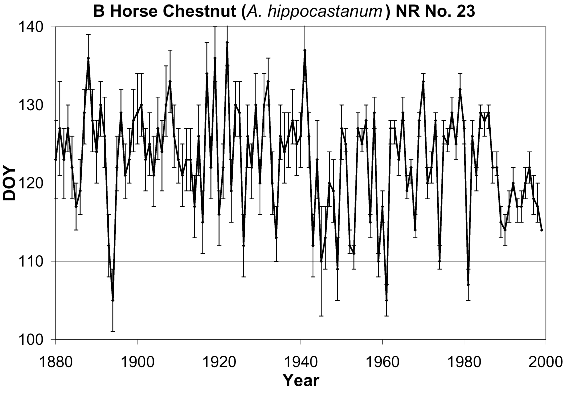


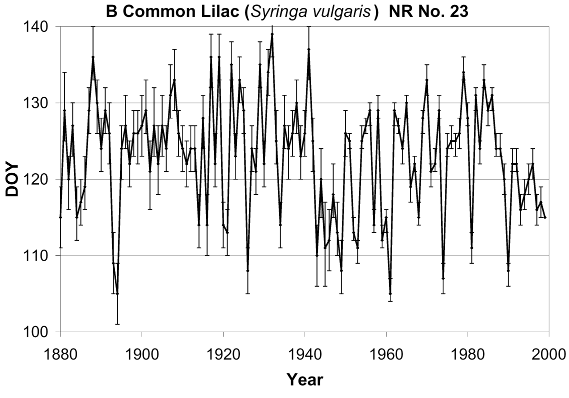


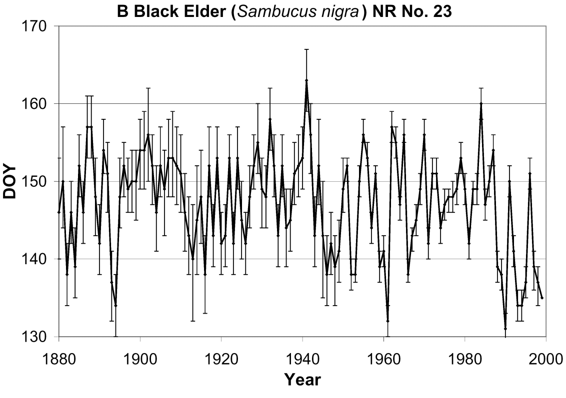


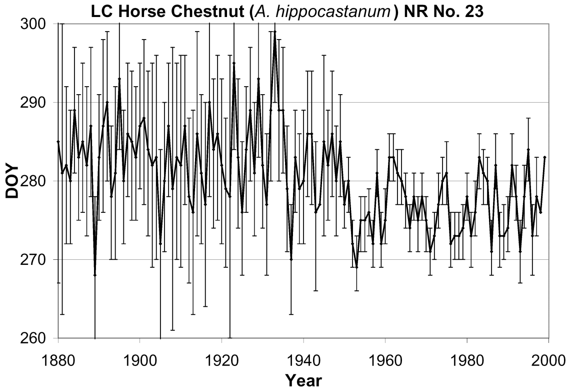


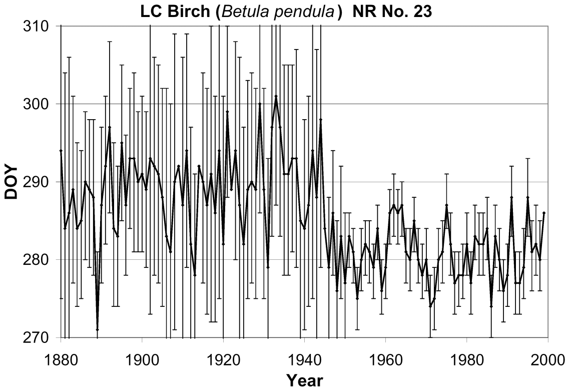


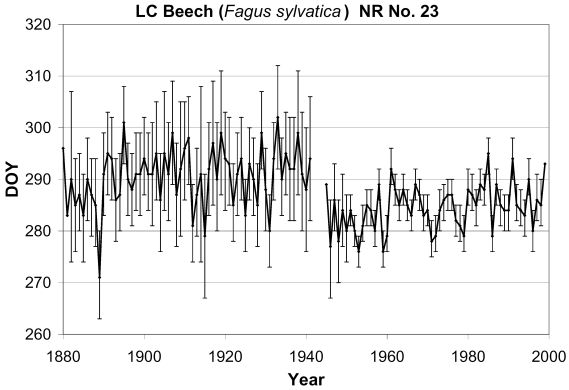


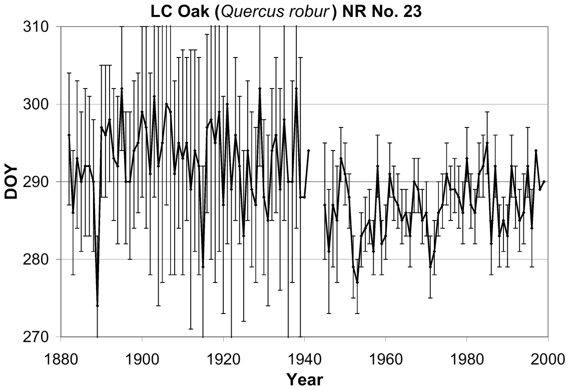





LC Horse Chestnut (*A. hippocastanum*) NR No. 23


LC Birch (*Betula pendula*) NR No. 23

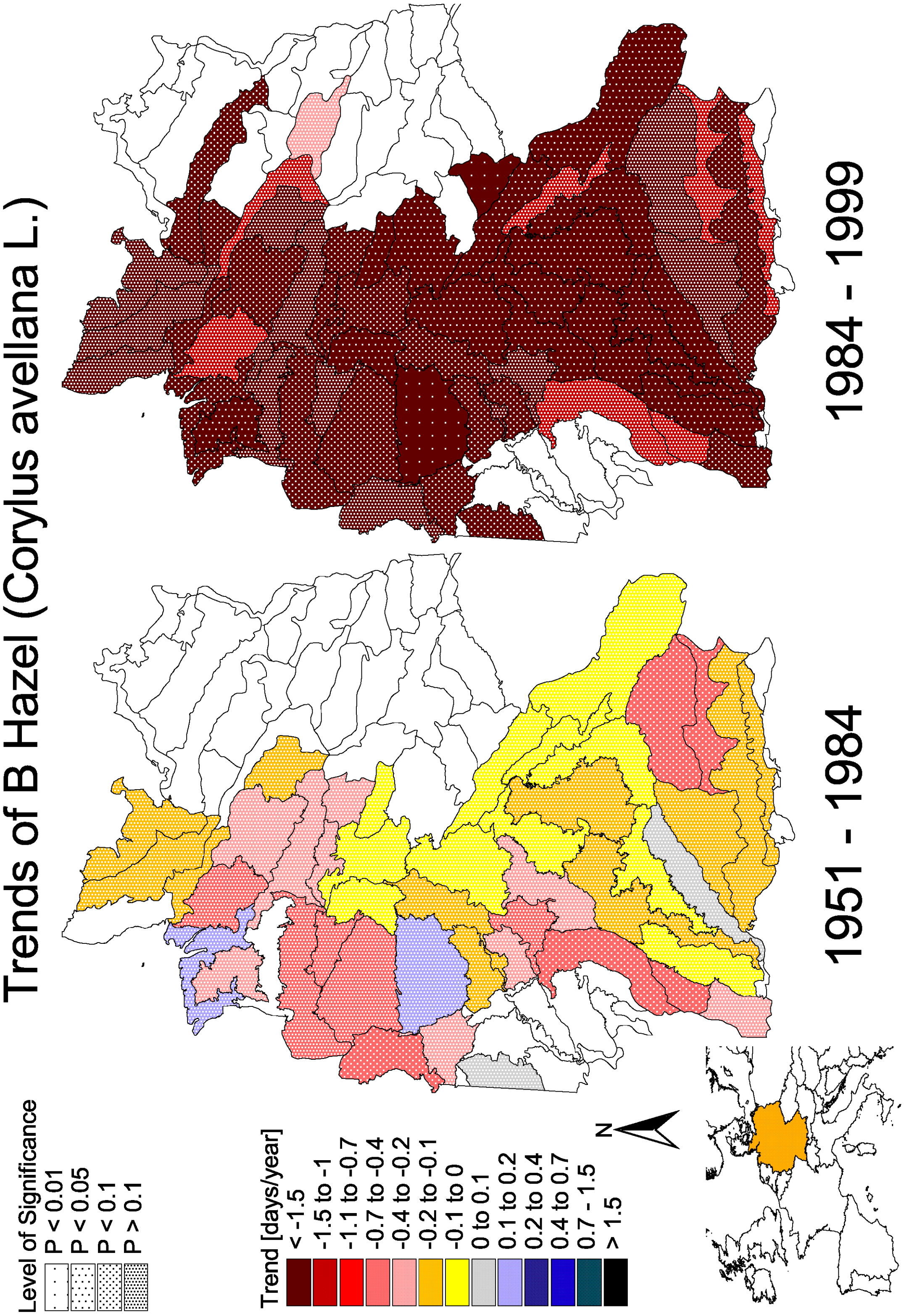

LC Beech (*Fagus sylvatica*) NR No. 23

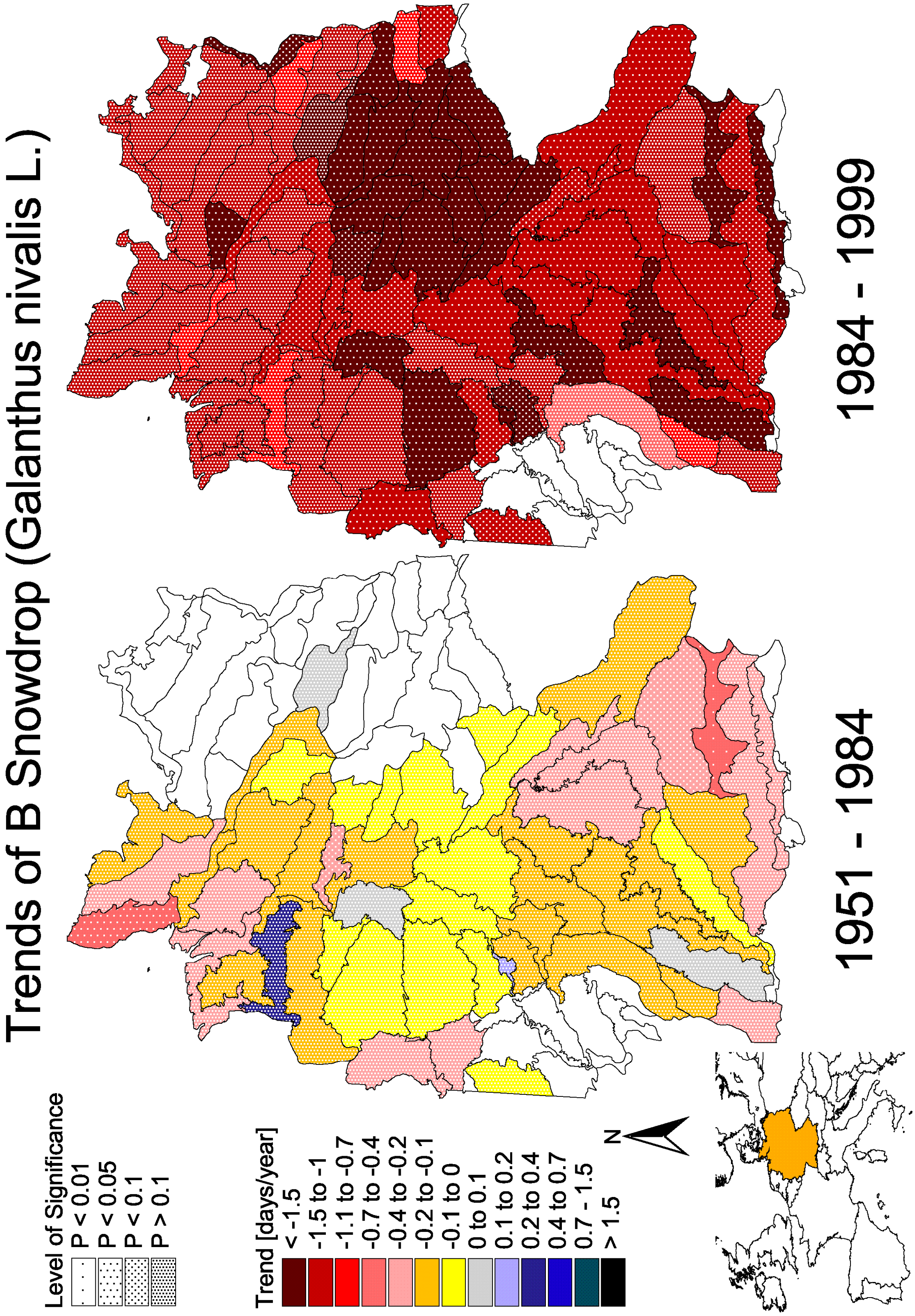

LC Oak (*Quercus robur*) NR No. 23

Appendix A2 Trends maps of phenological phases for Natural Regions in Germany

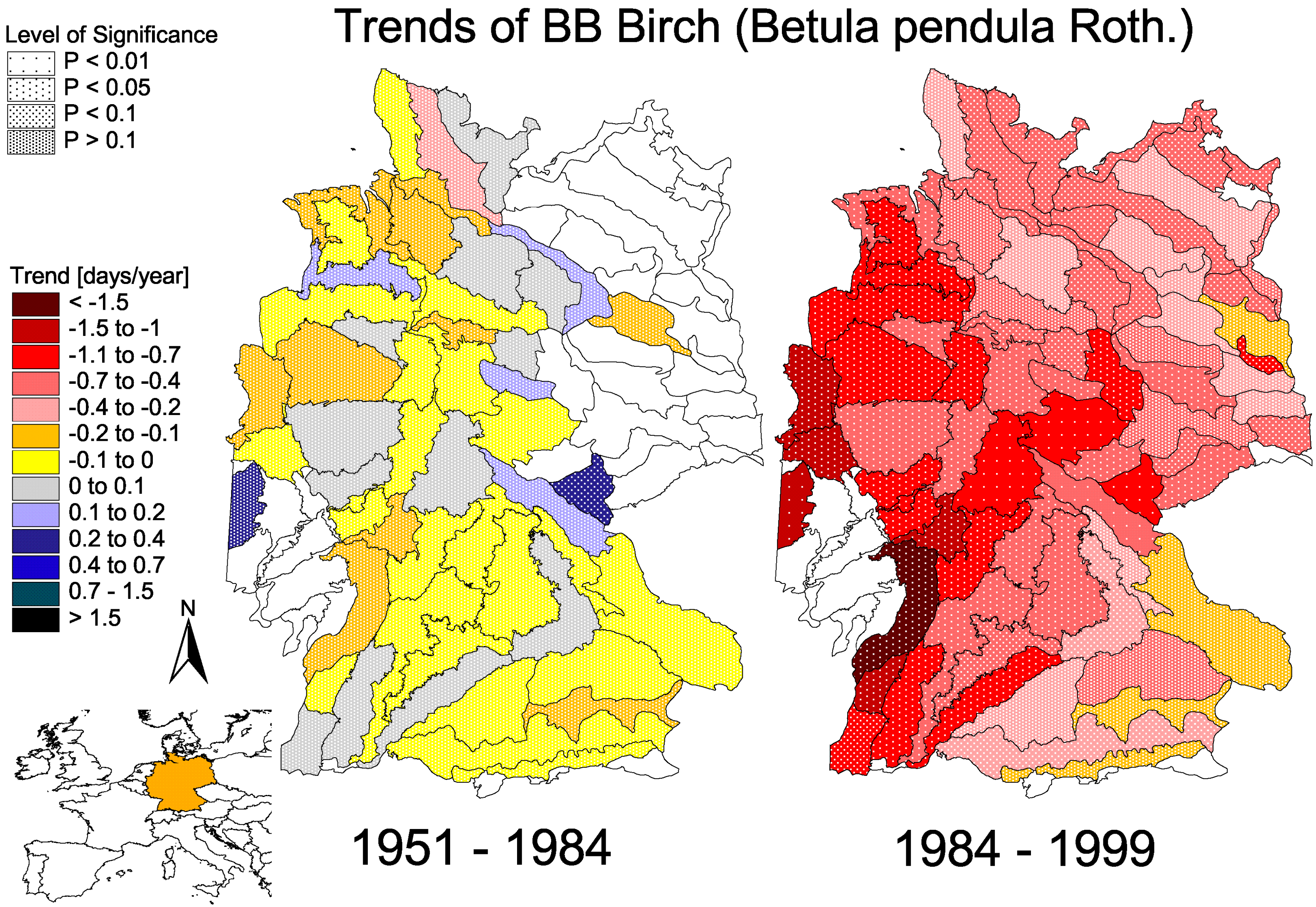
A2.1 Trends maps for early spring phases

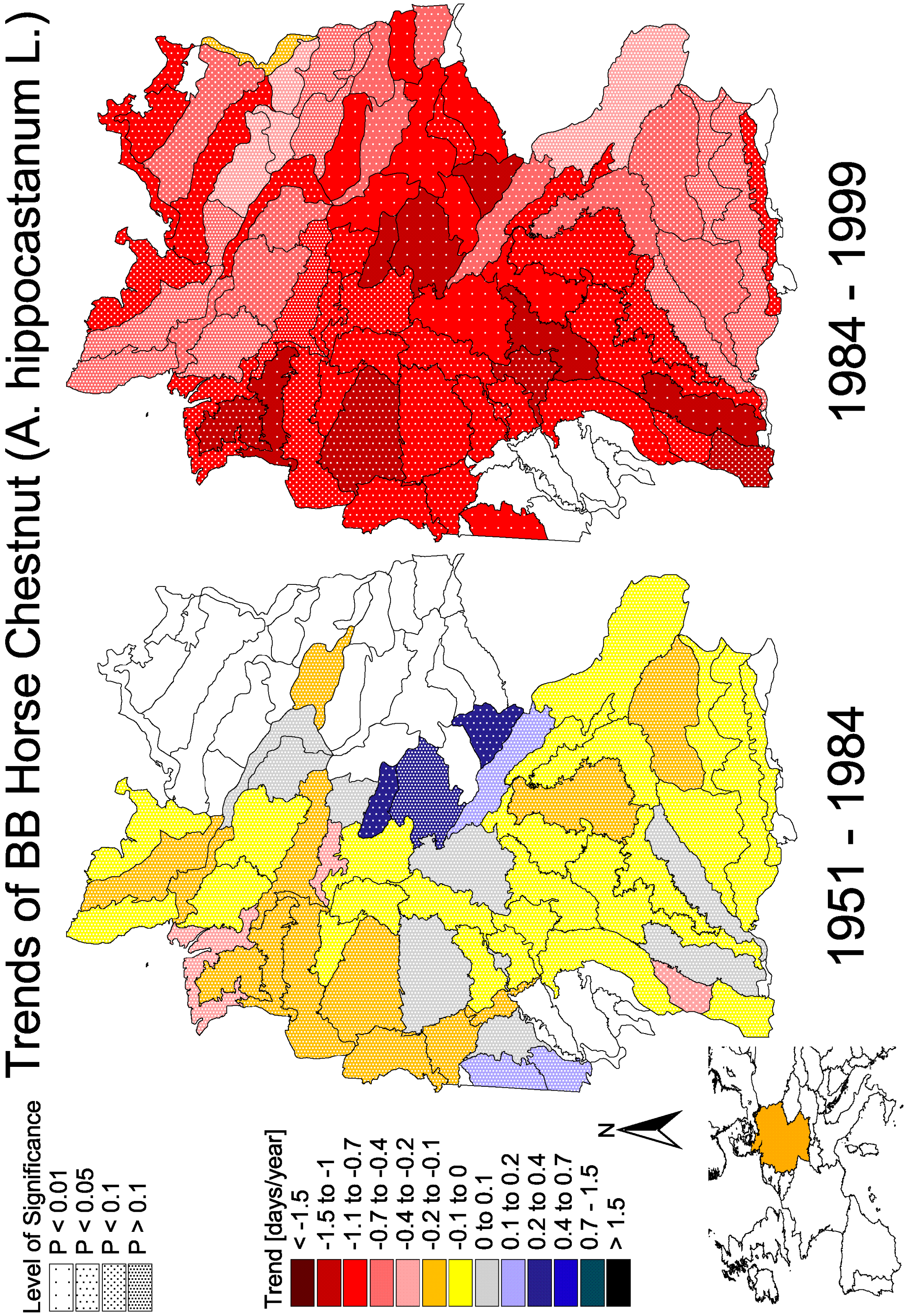
B blossoming, BB bud burst, ordered by average date of appearance




Trends of B Snowdrop (Galanthus nivalis L.)
1951 - 1984
1984 - 1999
Level of Significance
P < 0.01
P < 0.05
P < 0.1
P > 0.1
Trend [days/year]
< -1.5
-1.5 to -1
-1.1 to -0.7
-0.7 to -0.4
-0.4 to -0.2
-0.2 to -0.1
-0.1 to 0
0 to 0.1
0.1 to 0.2
0.2 to 0.4
0.4 to 0.7
0.7 - 1.5
> 1.5
N

Trends of BB Birch (Betula pendula Roth.)

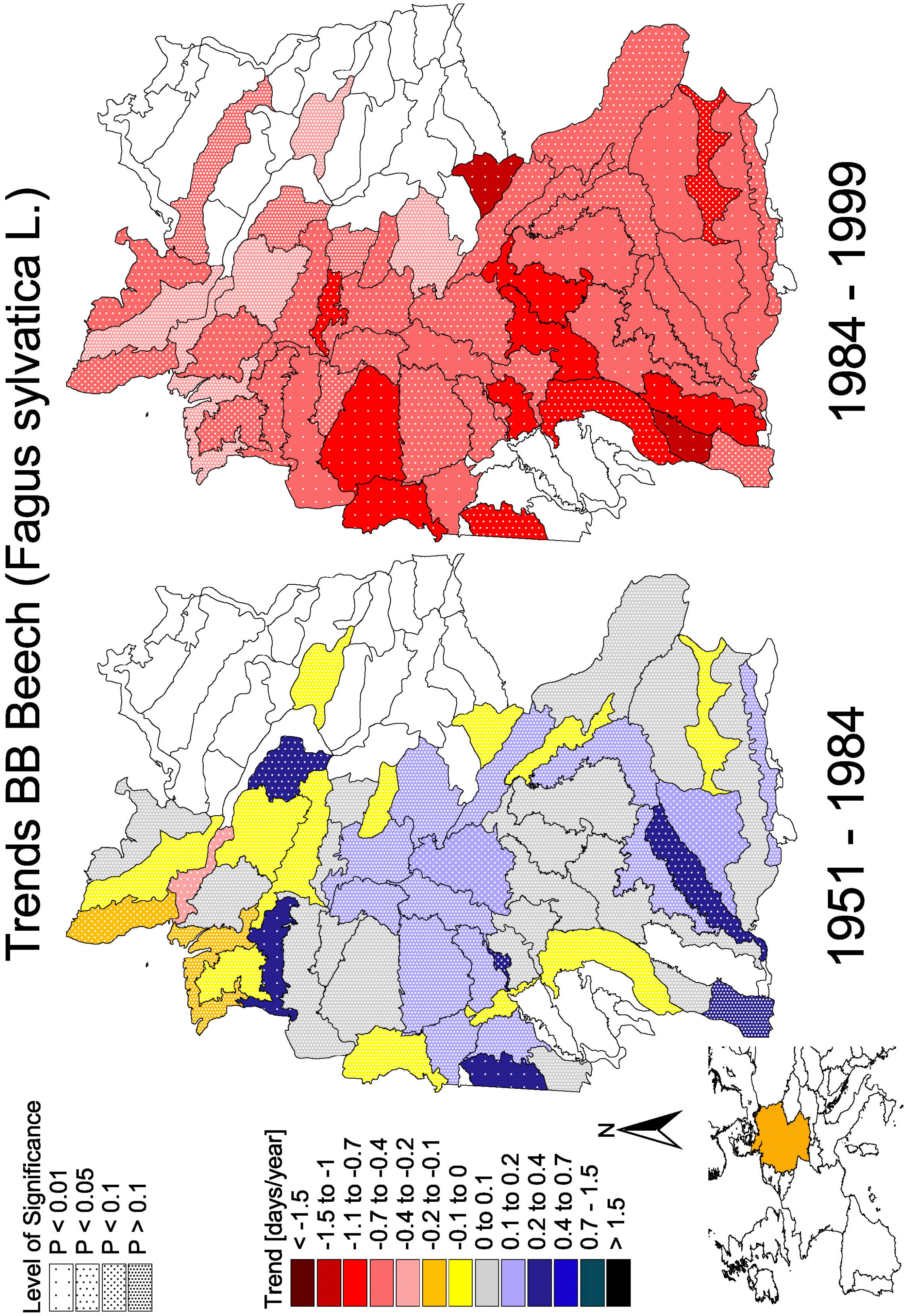


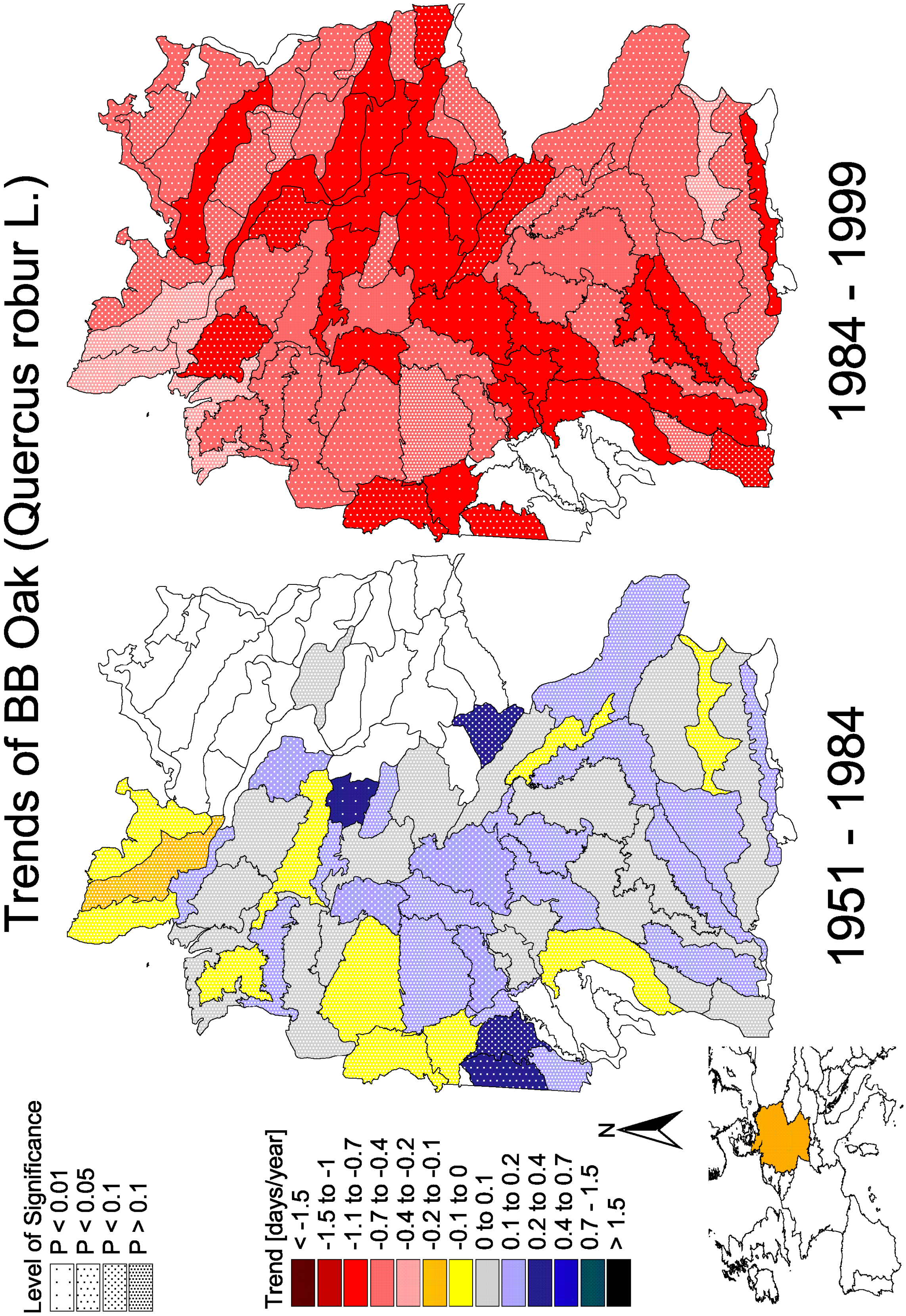


Trends of BB Horse Chestnut (A. hippocastanum L.)

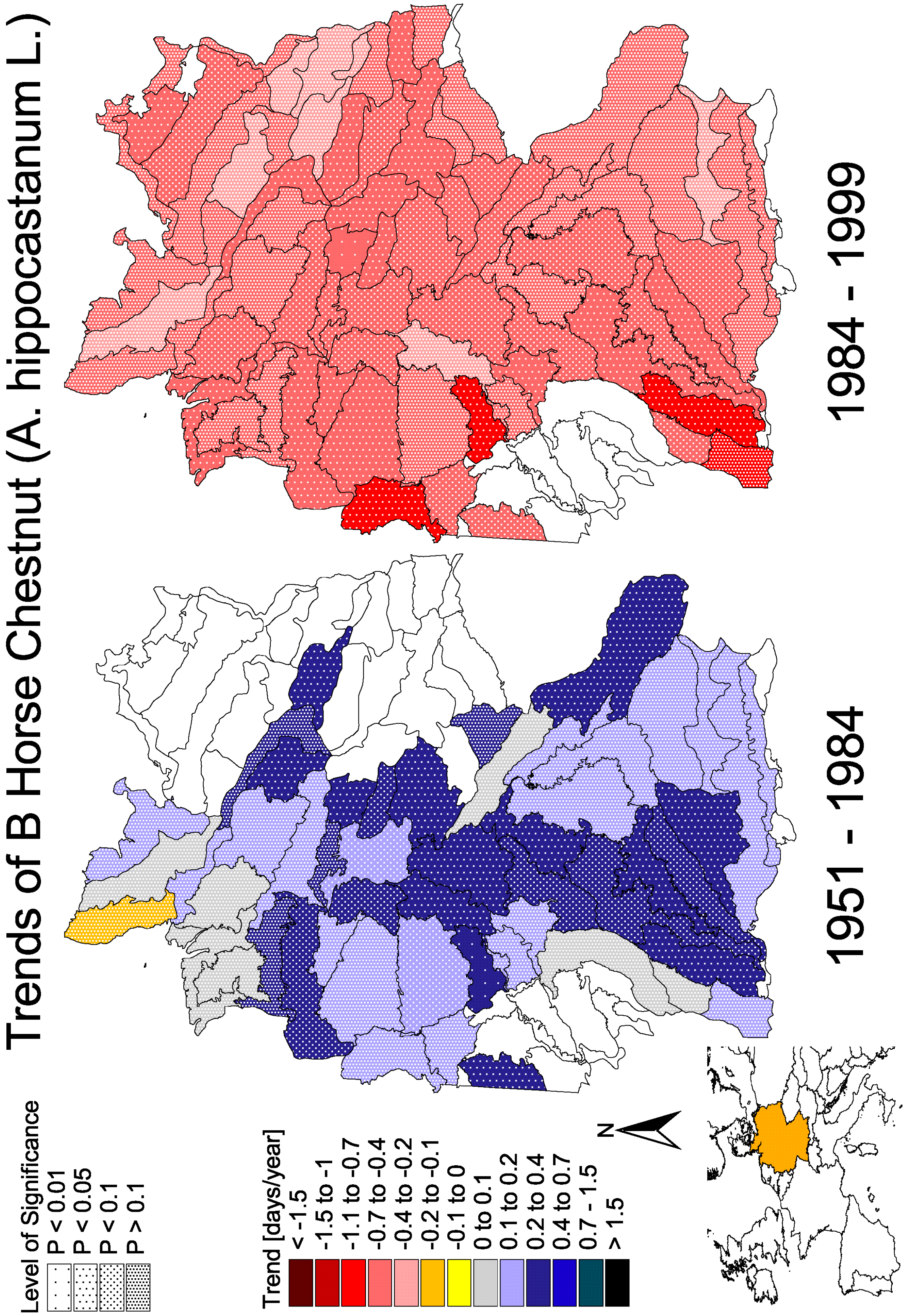
A2.2 Trend maps for spring and early summer phases

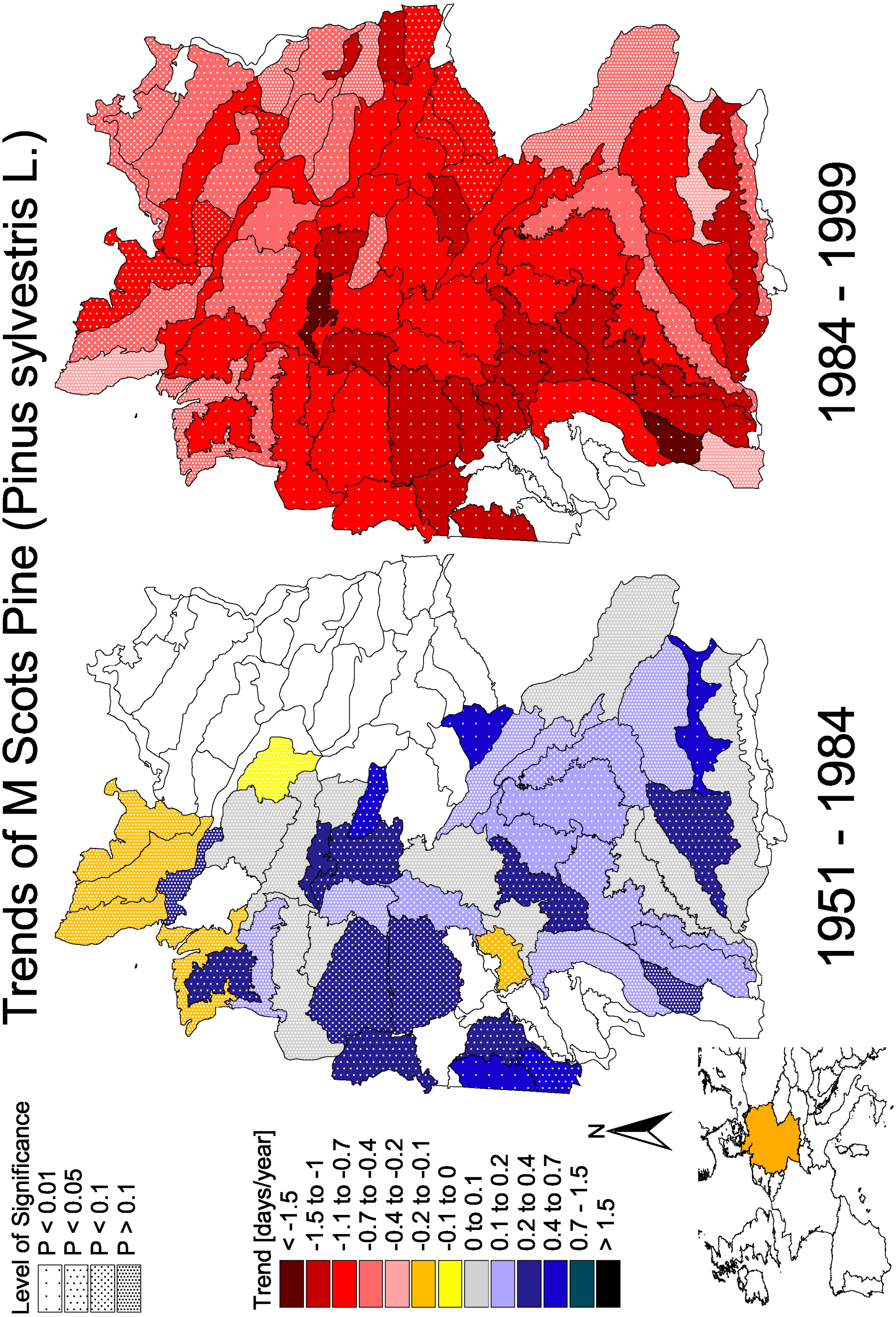
BB bud burst, B blossoming, ordered by average date of appearance





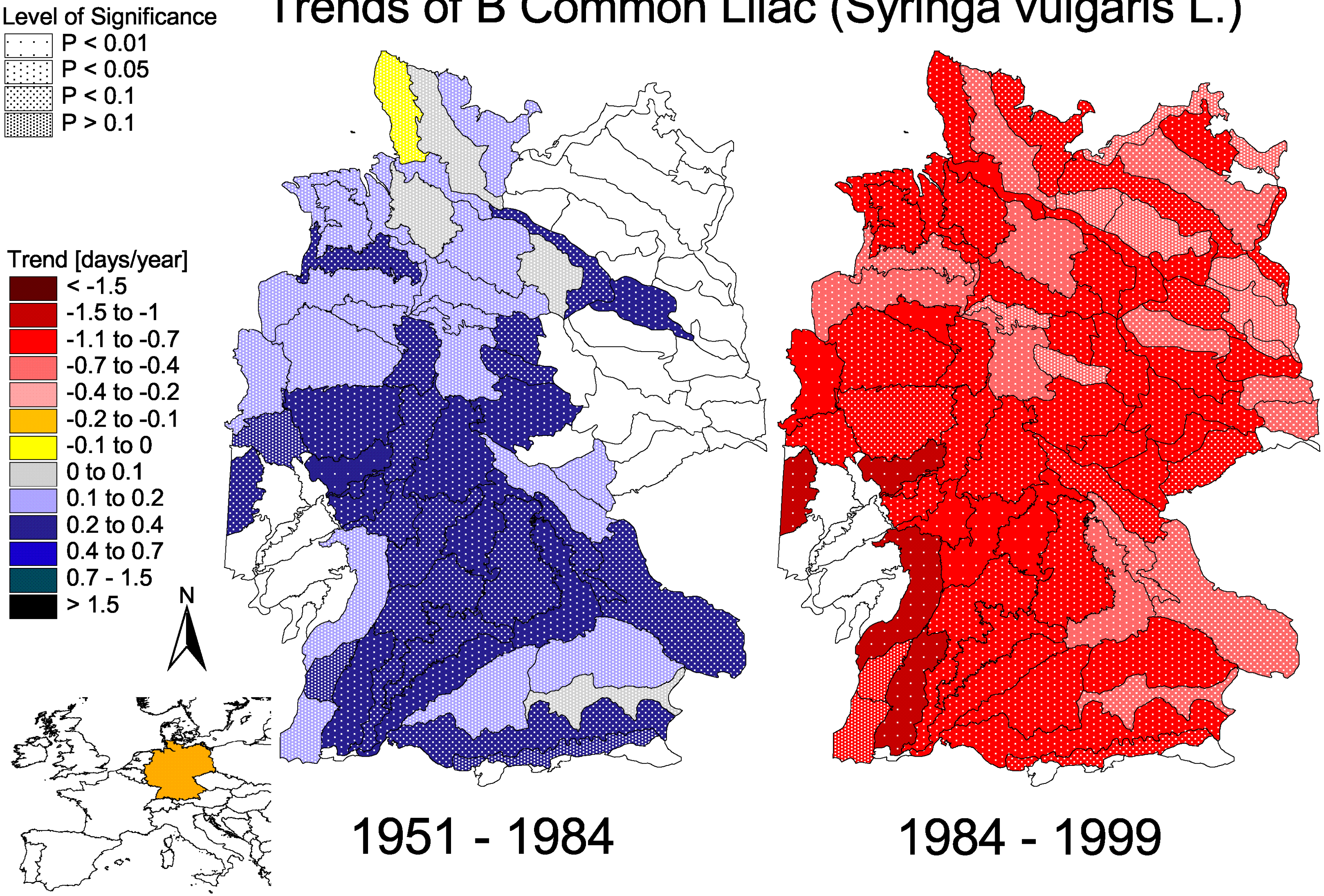
Trends of BB Oak (Quercus robur L.)

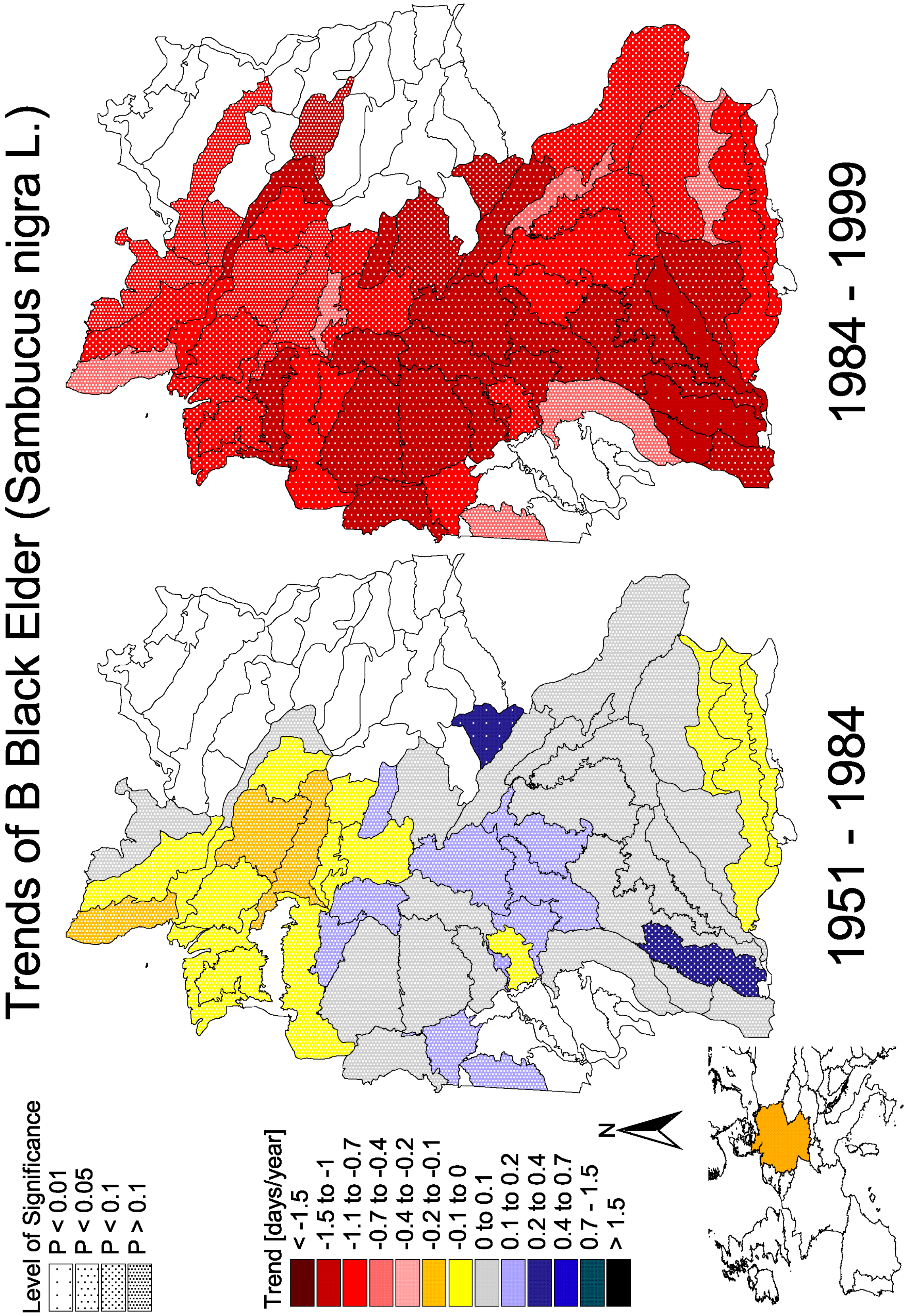

Trends of B Horse Chestnut (A. hippocastanum L.)
1984 - 1999
1951 - 1984
Level of Significance
P < 0.01
P < 0.05
P < 0.1
P > 0.1
Trend [days/year]
< -1.5
-1.5 to -1
-1.1 to -0.7
-0.7 to -0.4
-0.4 to -0.2
-0.2 to -0.1
-0.1 to 0
0 to 0.1
0.1 to 0.2
0.2 to 0.4
0.4 to 0.7
0.7 - 1.5
> 1.5
N



Trends of M Scots Pine (Pinus sylvestris L.)

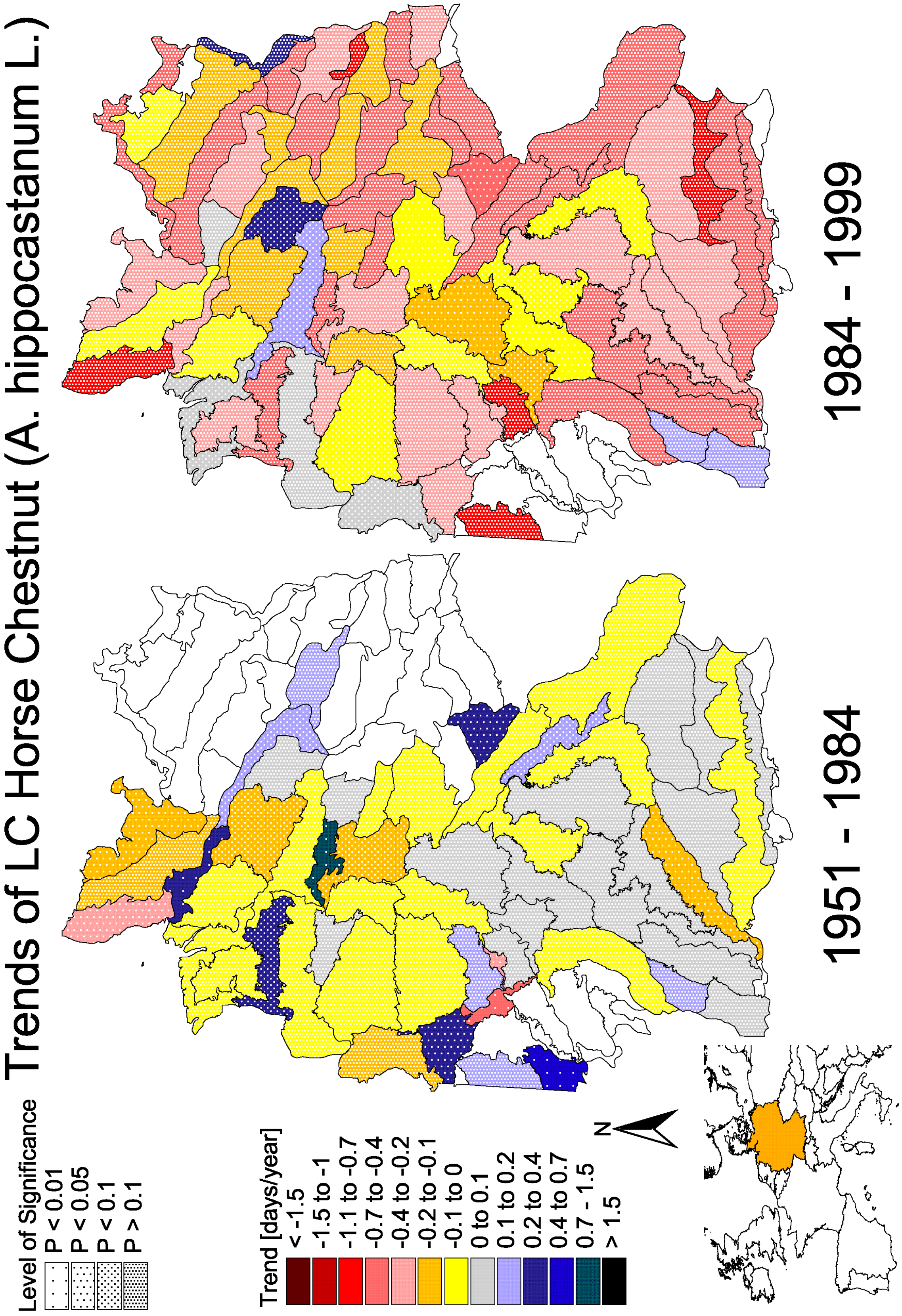
Trends of B Common Lilac (Syringa vulgaris L.)

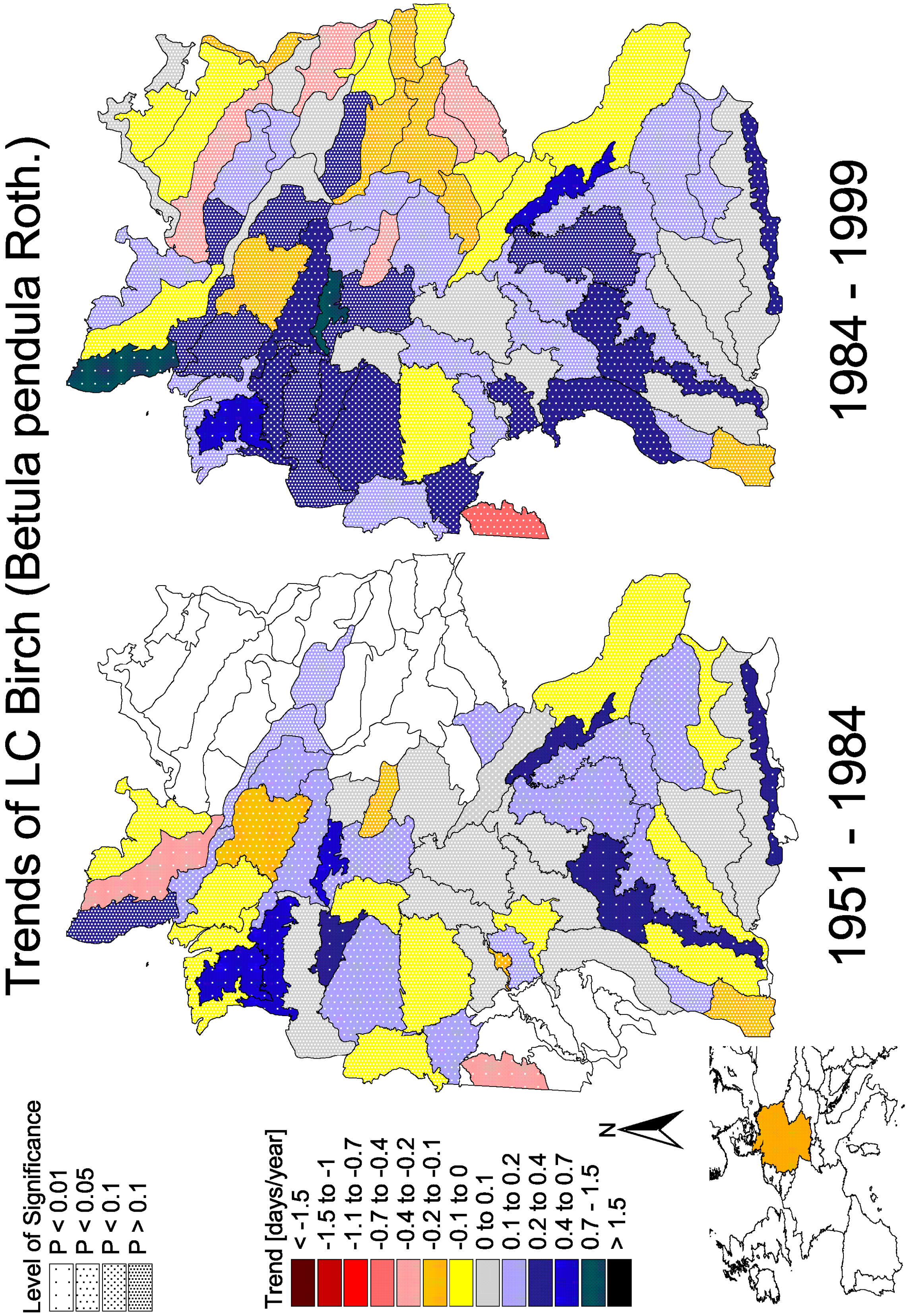



Trends of B Black Elder (Sambucus nigra L.)
1984 - 1999
1951 - 1984
Level of Significance
P < 0.01
P < 0.05
P < 0.1
P > 0.1
Trend [days/year]
< -1.5
-1.5 to -1
-1.1 to -0.7
-0.7 to -0.4
-0.4 to -0.2
-0.2 to -0.1
-0.1 to 0
0 to 0.1
0.1 to 0.2
0.2 to 0.4
0.4 to 0.7
0.7 - 1.5
> 1.5
N

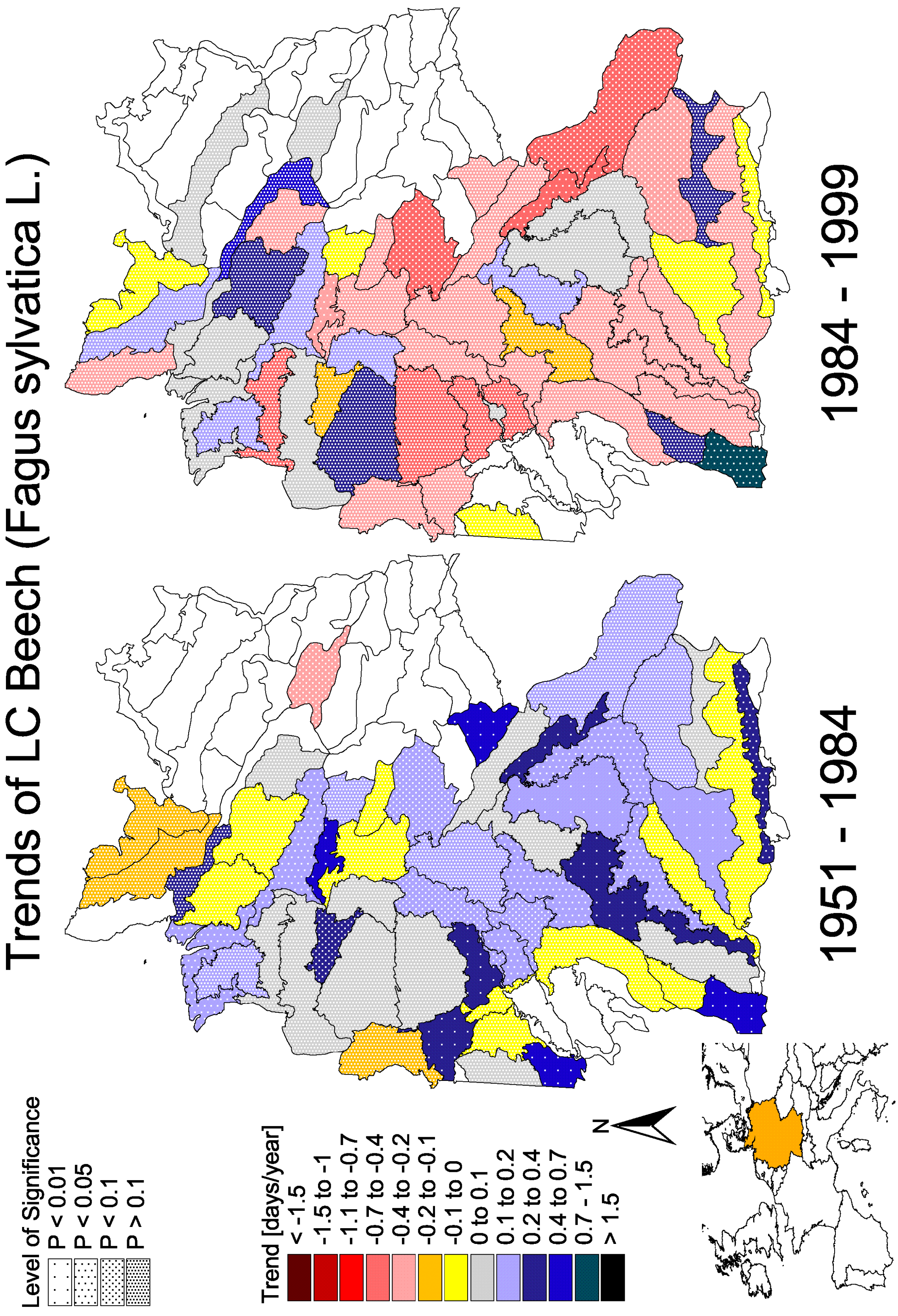
A2.3 Trend maps for autumn phases

LC leaf coloring, ordered by average date of appearance

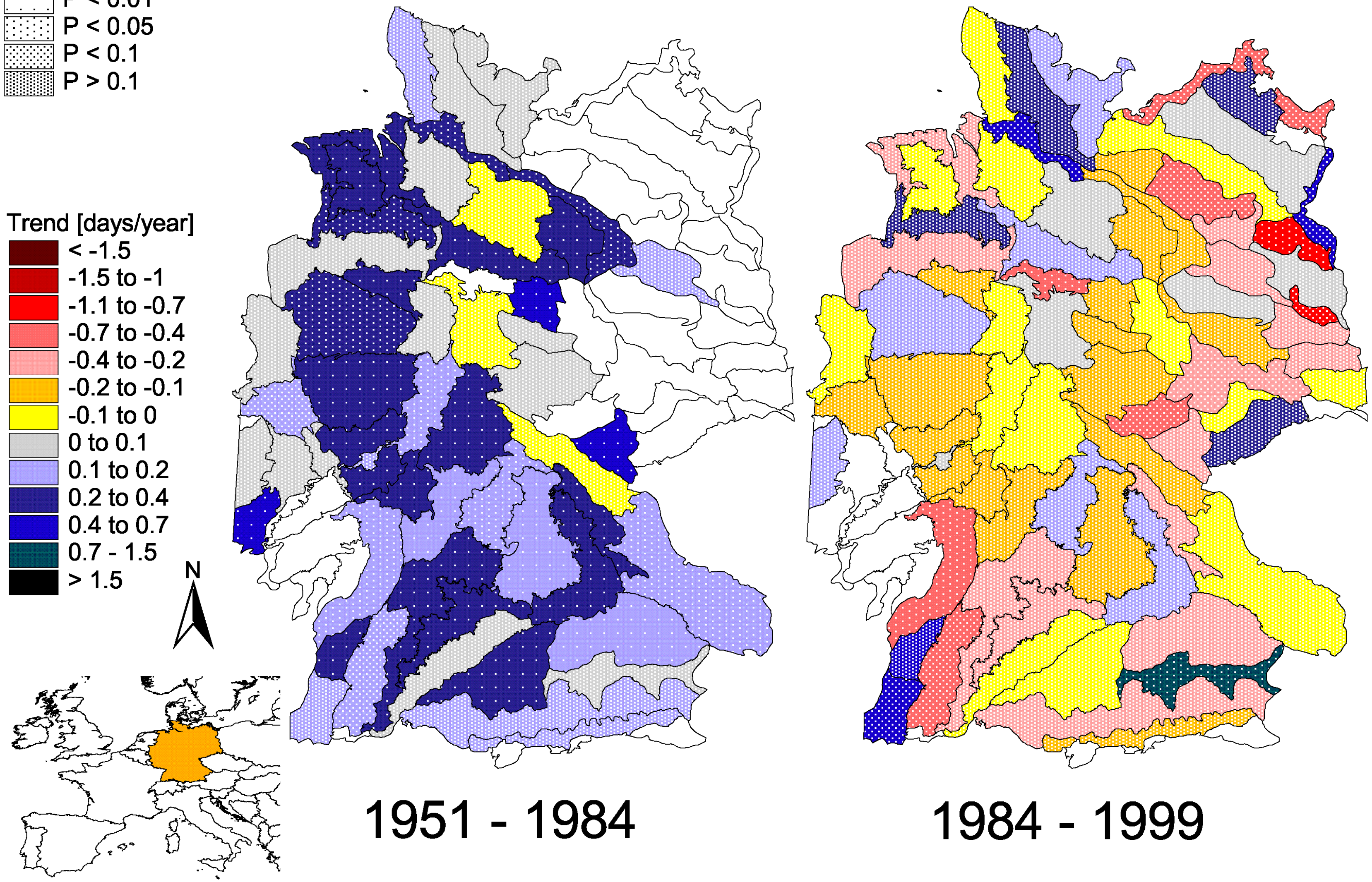




Trends of LC Birch (Betula pendula Roth.)

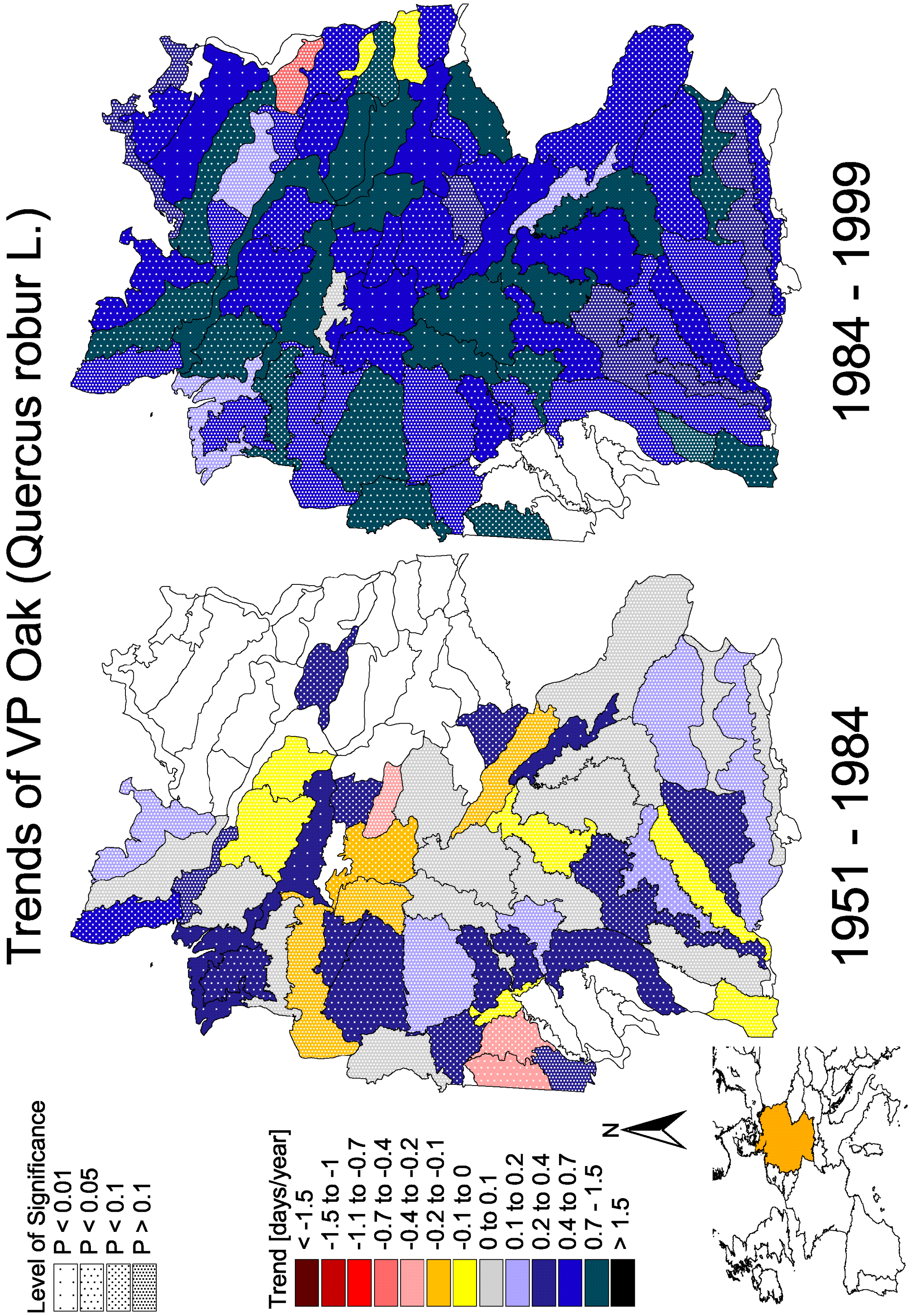

Trends of LC Beech (Fagus sylvatica L.)
1984 - 1999
1951 - 1984
Level of Significance
P < 0.01
P < 0.05
P < 0.1
P > 0.1
Trend [days/year]
< -1.5
-1.5 to -1
-1.1 to -0.7
-0.7 to -0.4
-0.4 to -0.2
-0.2 to -0.1
-0.1 to 0
0 to 0.1
0.1 to 0.2
0.2 to 0.4
0.4 to 0.7
0.7 - 1.5
> 1.5
N

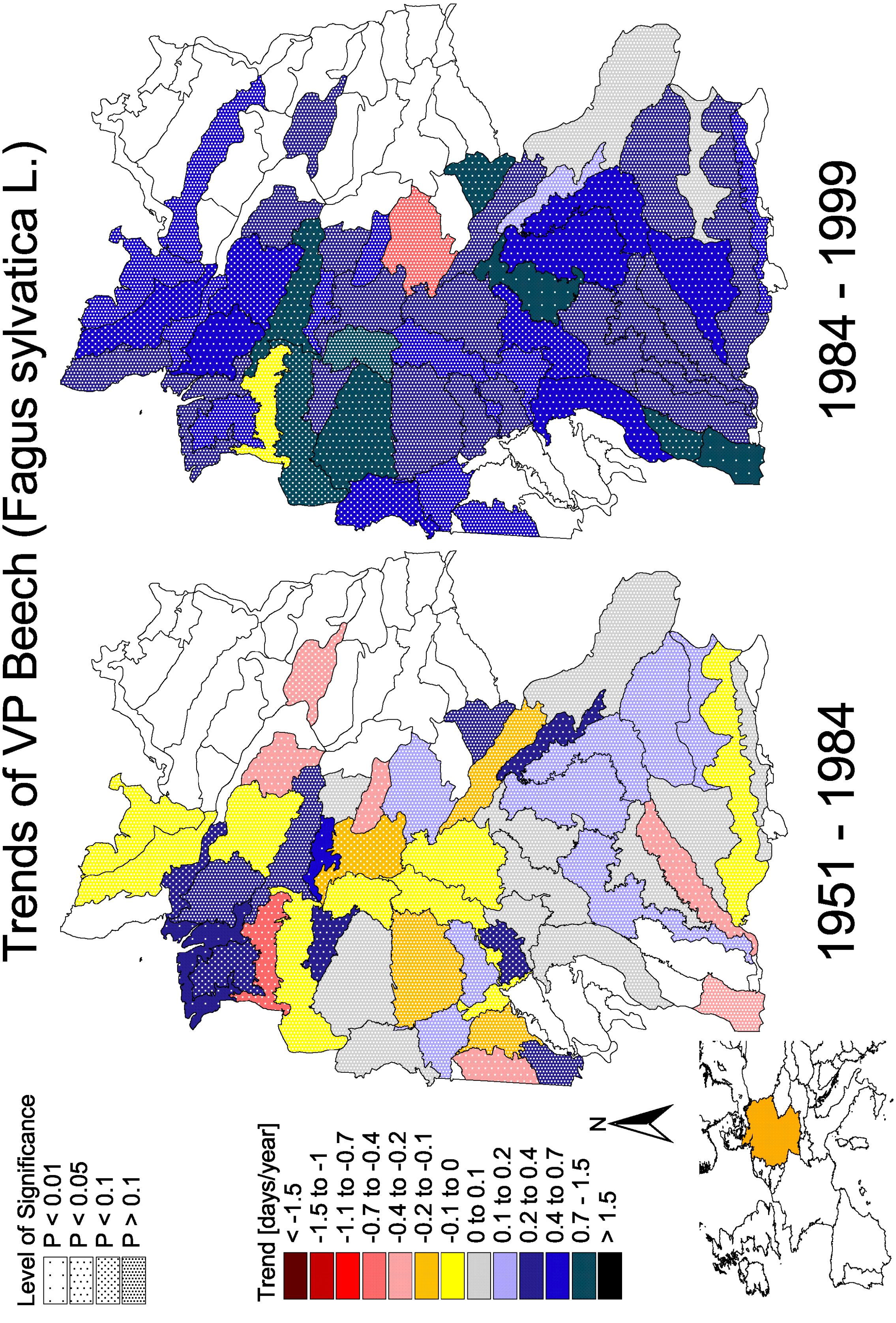
Trends of LC Oak (Quercus robur L.)

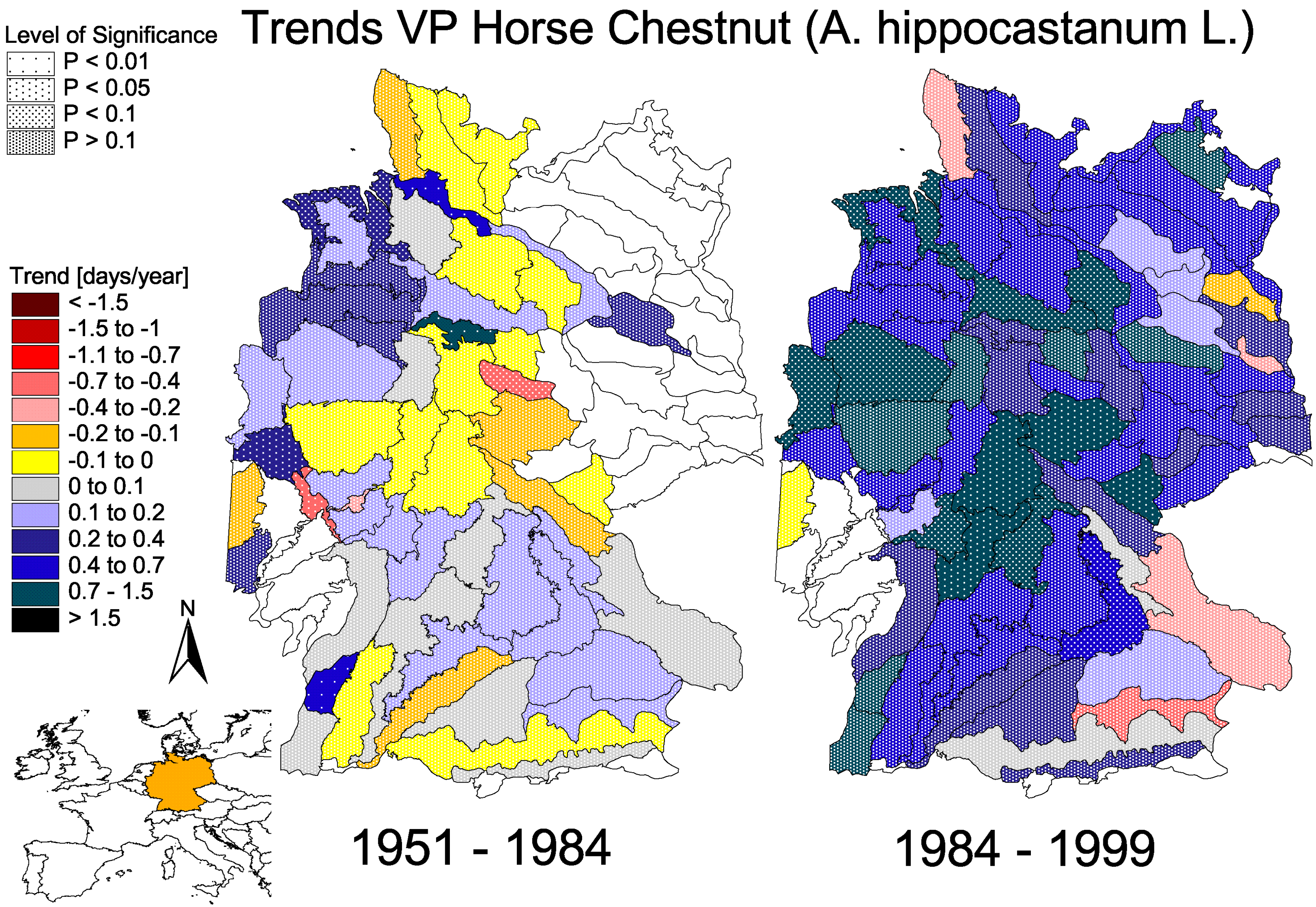


A2.4 Trend maps for the length of the vegetation period

VP vegetation period, ordered by the average length of the VP



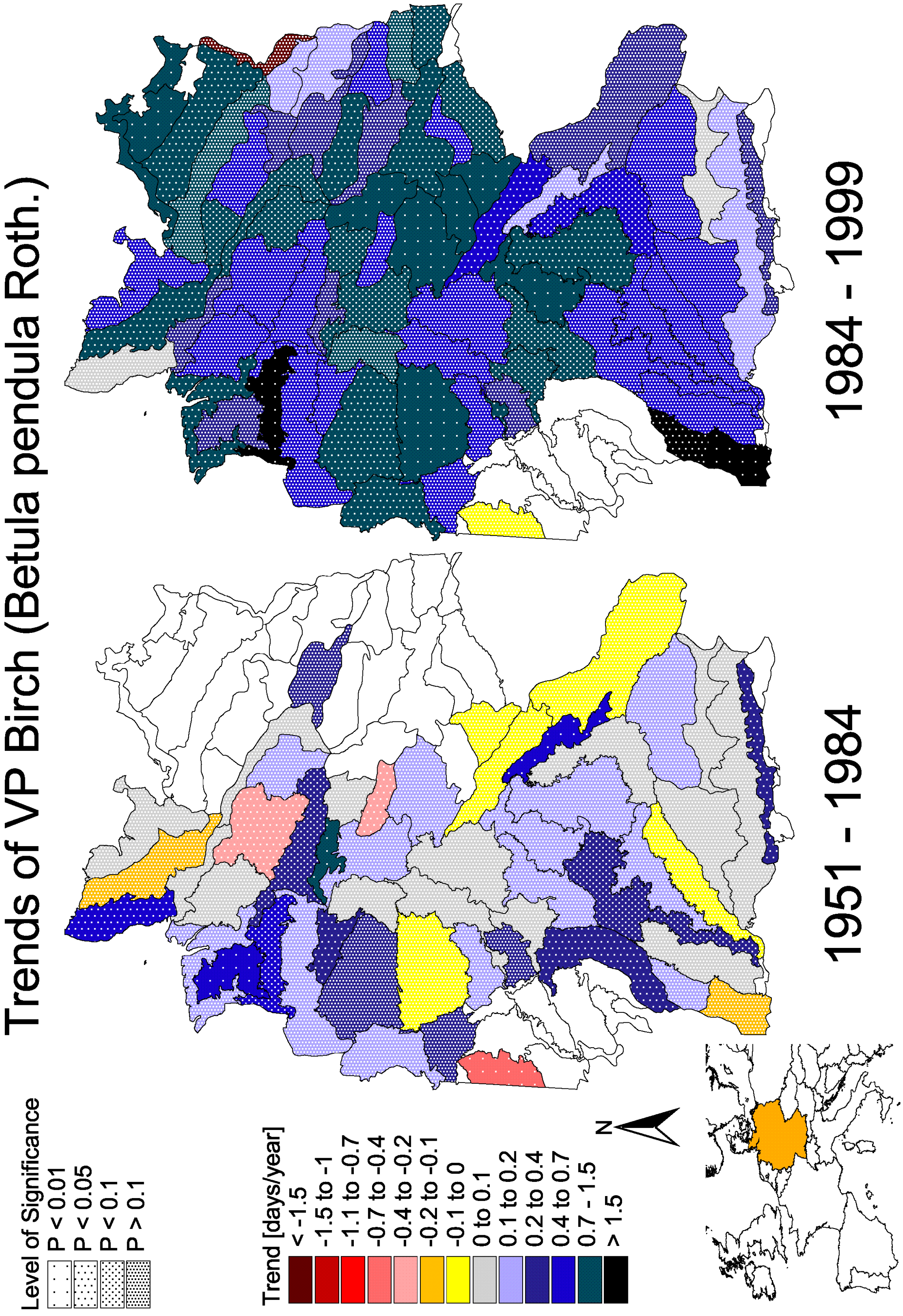

Trends of VP Beech (Fagus sylvatica L.)
1984 - 1999
1951 - 1984
Level of Significance
P < 0.01
P < 0.05
P < 0.1
P > 0.1
Trend [days/year]
< -1.5
-1.5 to -1
-1.1 to -0.7
-0.7 to -0.4
-0.4 to -0.2
-0.2 to -0.1
-0.1 to 0
0 to 0.1
0.1 to 0.2
0.2 to 0.4
0.4 to 0.7
0.7 - 1.5
> 1.5
N



Trends VP Horse Chestnut (A. hippocastanum L.)

1951 - 1984

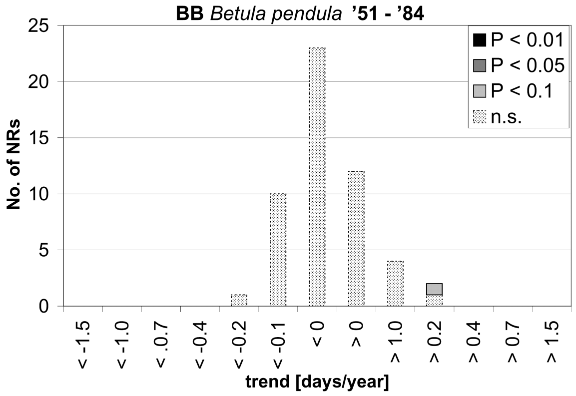
1984 - 1999

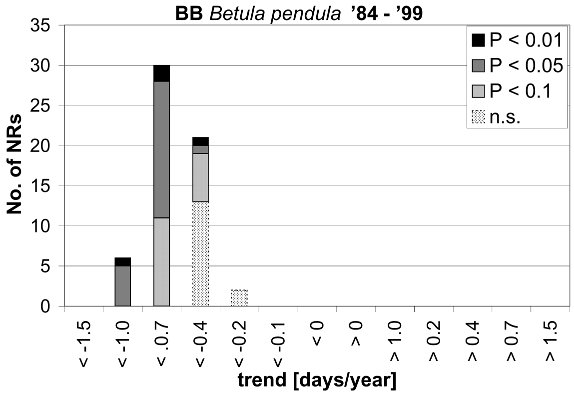

Trends of VP Birch (Betula pendula Roth.)
1984 - 1999
1951 - 1984
Level of Significance
P < 0.01
P < 0.05
P < 0.1
P > 0.1
Trend [days/year]
< -1.5
-1.5 to -1
-1.1 to -0.7
-0.7 to -0.4
-0.4 to -0.2
-0.2 to -0.1
-0.1 to 0
0 to 0.1
0.1 to 0.2
0.2 to 0.4
0.4 to 0.7
0.7 - 1.5
> 1.5
N

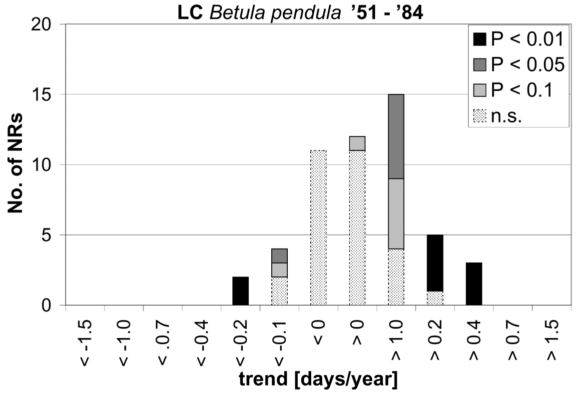
Appendix A3 Histograms of phenological trends of BB, LV and VP in Germany

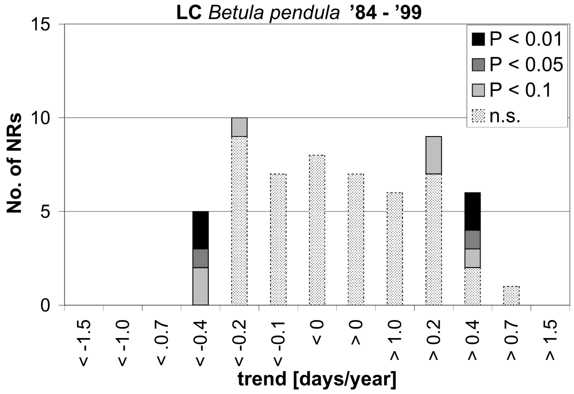
BB bud burst, LC leaf coloring, VP vegetation period

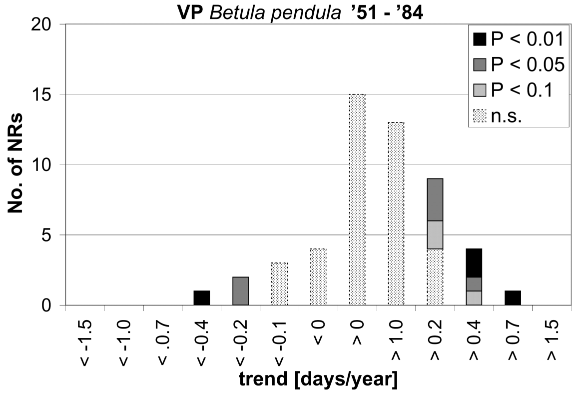
A3.1 Histograms of trends of BB, LV and VP of birch (*Betula pendula*) in Germany

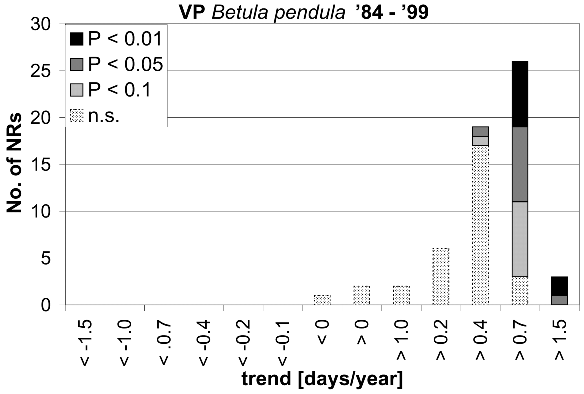

BB *Betula pendula* **'51 - '84**


BB *Betula pendula* **'84 - '99**

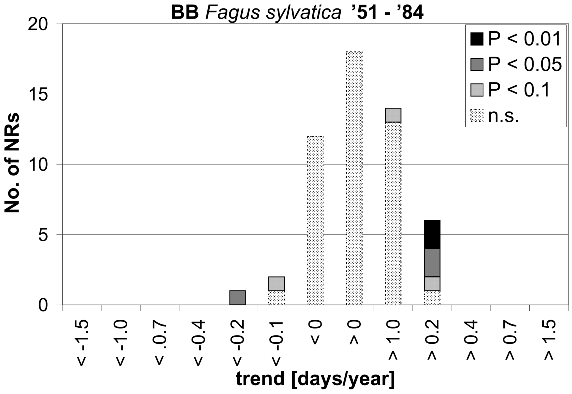

LC *Betula pendula* **'51 - '84**

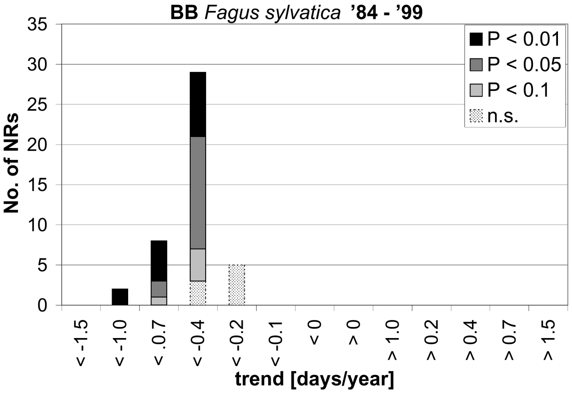

LC *Betula pendula* **'84 - '99**

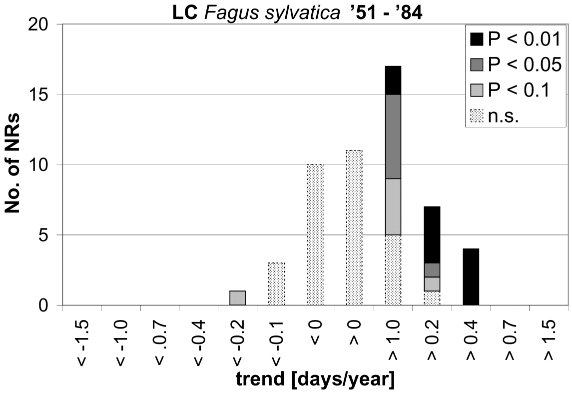

VP *Betula pendula* **'51 - '84**

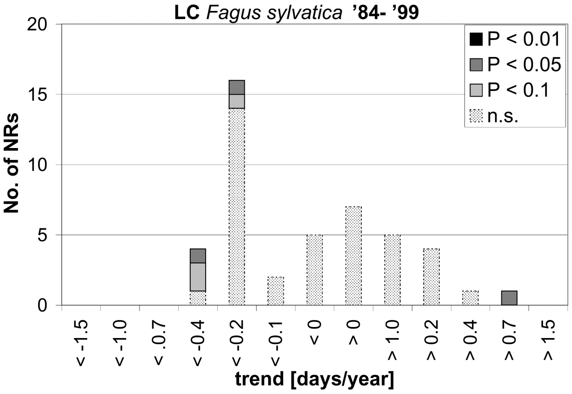

VP *Betula pendula* **'84 - '99**

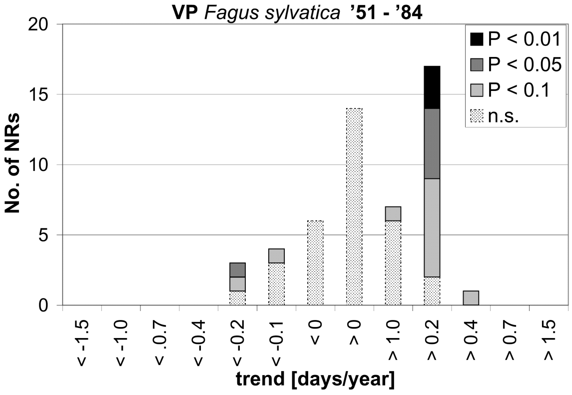
A3.2 Histograms of trends of BB, LV and VP of beech (*Fagus sylvatica*) in Germany

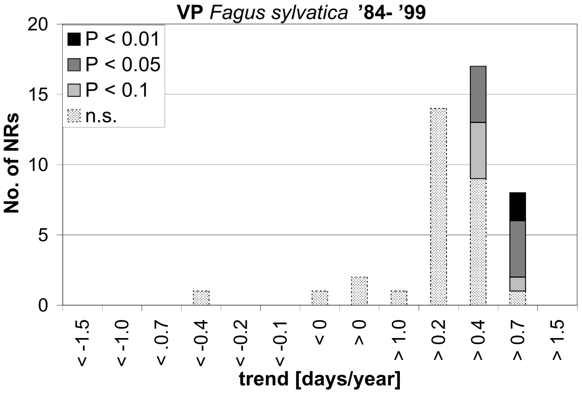












Appendix A4 Model descriptions

A4.1 The sequential model (SM)

The sequential model (SM) (Hänninen 1994, Kramer 1994) considers rest and quiescence as two distinct phases that follow sequentially in time. Transition from rest to quiesence does not take place unless a critical state of chilling C_{ch} is attained. Similarly, bud burst only takes place after a critical state of forcing C_{frc} is reached. The current state of chilling S_{ch} is determined by the discrete integral of the rate of chilling R_{ch}. Similarly, the current state of forcing S_{frc} is determined by the discrete integral of the rate of forcing R_{frc}. The model is driven by daily mean temperature T.

$R_{ch}(T)$: Rate of chilling (parameters see also Fig. 5-1 and Eqn. 5-2).

$$R_{ch}(T) = \begin{cases} \dfrac{T - T_{min}}{T_{opt} - T_{min}} & \text{for } T_{min} \le T \le T_{opt} \\ \dfrac{T_{max} - T}{T_{max} - T_{opt}} & \text{for } T_{opt} \le T \le T_{max} \\ 0 & \text{else} \end{cases}$$

$R_{frc}(T)$: Rate of forcing.

$$R_{frc}(T) = \begin{cases} K\dfrac{1}{1 + e^{b(T + c)}} & \text{for } T > 0 \\ 0 & \text{for } T \le 0 \end{cases}$$

S_{ch}: State of chilling

$$S_{ch} = \sum_{t_1}^{t} R_{ch}(T$$

S_{frc}: State of forcing

$$S_{frc} = \sum_{t_1}^{t} R_{frc}(T$$

$$K = \begin{cases} 0 & S_c < C_{ch} \\ 1 & S_c \ge C_{ch} \end{cases}$$

Bud burst takes place when $S_{frc} > C_{frc}$.

The start time t_1 is set to the 1st November of the previous year.

The model has seven parameters including T_{min}, T_{opt}, T_{max}, b, c, C_{ch}, C_{frc}.

A4.2 The Cannel and Smith model (CSM)

The Cannel and Smith model (CSM) empirically models the observation that increased chilling in winter, decreases the required temperature sum for bud burst in spring. It was developed by Cannel and Smith (1983) and modified by Menzel (1997).

Chilling C is modelled as a simple counter of ’chill days’, i.e. days whose mean temperature falls below a certain threshold T_c, starting on t_1 = 1. Nov.

$$C = \sum_{t_1}^{t} 1 \qquad \text{for} \quad T \leq T_c$$

Chilling reduces the required temperature sum for bud burst T_{crit} in an logarithmic manner.

$$T_{crit} = a + b\ln(C)$$

Temperature sum T_s is formed by integrating daily mean temperatures T above a certain threshold T_b starting on t_2 = 1. Feb.

$$T_s = \sum_{t_2}^{t} (T - T_b) \qquad \text{for} \quad T > T_b$$

Bud burst occurs on the day when $T_s > T_{crit}$.

Including T_c, a, b and T_b the model has four parameters.

A4.3 The linear temperature sum model (TSM)

The temperature sum model (TSM) (Wang 1960, Robertson 1968, Kramer 1994, Menzel 1997) simply integrates daily mean temperatures T above a certain threshold T_b starting from a defined date t_1 up to a fixed critical value T_{crit}. The starting date t_1 was assumed to vary between the 1st November and the 31 of March, i.e.between 1 and 131.

$$T_s = \sum_{t_1}^{t} (T - T_b) \qquad \text{for} \quad T > T_b$$

Bud burst occurs on the day when $T_s > T_{crit}$.

Including T_b, T_{crit} and t_1 the model has three parameters.

Appendix A5 Parameters of the model fits for SM, CSM and TSM

A5.1 Parameters for the model fits of the SM

Table 5-1. Parameter values for the SM. For a description of the parameters see ”A4.1 The sequential model (SM)” on page 141. min, max: allowed ranges for the parameters during model fits. min SSR: Parameters for the fit that performed best in terms of SSR among ten different runs. mean: Average parameters after ten optimization runs. stderr: Standard error of the average parameters after ten optimzation runs.

		C_{Ch}	C_{fr}	T_{min}	T_{opt}	T_{max}	a	b
allowed range	min	10	0	-30	-10	0	-1	-50
	max	500	150	30	20	100	0	30
A. hippo-castanum	min SSR	112.77	0.55	-29.74	19.56	95.08	-0.94	-1.4
	mean	113.61	2.11	-29.72	19.2	84.58	-0.38	-0.8
	stderr	0.45	1.73	0.02	0.19	6.04	0.09	0.56
B. pendula	min SSR	115.09	1.31	-29.8	18.09	67.23	-0.15	6.28
	mean	114.35	42.57	-29.57	18.41	59.58	-0.15	-7.1
	stderr	1.09	14.41	0.16	0.44	6.81	0.08	8.16
F. sylvatica	min SSR	128.38	40.75	-29.88	0.88	99.94	-0.88	11.18
	mean	126.02	42.54	-29.3	6.2	87.83	-0.55	6.05
	stderr	0.97	10.91	0.37	2.65	5.63	0.13	6.2
Q. robur	min SSR	124.19	59.31	-30.0	19,87	90.31	0.0	-3.69
	mean	127.31	28.61	-29.81	18.4	83.25	-0.19	5.23
	stderr	0.85	15.0	0.07	0.37	3.1	0.08	2.8
L. decidua	min SSR	113.25	0.41	-27.67	17.0	71.86	-0.8	-1.89
	mean	109.03	5.47	-27.81	15.89	57.21	-0.77	1.51
	stderr	1.47	4.99	0.16	2.33	7.82	0.06	2.77

A5.2 Parameters for the model fits of the CSM

Table 5-2. Parameter values for the CSM. For a description of the parameters see ”A4.2 The Cannel and Smith model (CSM)” on page 142. min, max: allowed ranges for the parameters during model fits. min SSR: Parameters for the fit that performed best in terms of SSR among ten different runs. mean: Average parameters after ten optimization runs. stderr: Standard error of the average parameters after ten optimzation runs.

		T_c	T_b	***a***	***b***
allowed range	min	-5	-5	-1000	0
	max	20	20	-0	3000
A. hippo-castanum	min SSR	19.74	3.52	2989.58	-550.12
	mean	19.32	3.69	2948.26	-543.46
	stderr	0.12	0.12	31.43	5.22
B. pendula	min SSR	18.42	4.19	3075.99	-573.67
	mean	15.8	3.83	2694.38	-498.02
	stderr	1.67	0.42	247.02	48.16
F. sylvatica	min SSR	11.03	6.5	3039.75	-574.49
	mean	10.26	5.39	2810.75	-519.51
	stderr	0.75	1.16	127-14	31.2
Q. robur	min SSR	18.31	5.58	3451.18	-631.5
	mean	17.65	5.55	3008.55	-546.96
	stderr	0.85	0.04	137.41	26.06
L. decidua	min SSR	20.0	-1.15	2956.39	-503.1
	mean	17.61	-1.12	2857.41	-485.08
	stderr	1.47	0.37	221.6	46.97

A5.3 Parameters for the model fits of the TSM

Table 5-3. Parameter values for the TSM. For a description of the parameters see ”A4.3 The linear temperature sum model (TSM)” on page 142. min, max: allowed ranges for the parameters during model fits. min SSR: Parameters for the fit that performed best in terms of SSR among ten different runs. mean: Average parameters after ten optimization runs. stderr: Standard error of the average parameters after ten optimzation runs.

		T_b	T_{crit}	t_1
allowed range	min	-10	0	1
	max	10	4000	131
A. hippo-castanum	min SSR	-0.59	316.81	121
	mean	-3.69	527.64	113
	stderr	0.8	55.4	2.11
B. pendula	min SSR	-6.07	672.9	108
	mean	-4.77	586.02	111.3
	stderr	0.77	52.82	2.29
F. sylvatica	min SSR	-6.98	664.88	131
	mean	-6.97	664.07	131
	stderr	0.16	7.87	0.0
Q. robur	min SSR	0.49	372.06	131
	mean	0.5	372.06	131
	stderr	0.03	2.0	0.0
L. decidua	min SSR	-4.53	561.39	101
	mean	-4.59	564.8	101
	stderr	0.25	18.12	0.58