

Publikationen der Deutschen
Gesellschaft für Photogrammetrie,
Fernerkundung und Geoinformation e.V.



Band 20

2011



Vorträge

**31. Wissenschaftlich-Technische
Jahrestagung der DGPF**

**13. – 15. April 2011
in Mainz**



Geodaten – eine Ressource des 21. Jahrhunderts

ISSN 0942-2870

Eckhardt Seyfert, Hrsg.

Spezifikation phänologischer Aufnahmezeitpunkte in Deutschland

MARKUS MÖLLER¹, JENS BIRGER², CORNELIA GLÄBER¹ & OLIVER BUCK³

Zusammenfassung: Der Beitrag beinhaltet die Beschreibung eines vereinfachten Temperatur-Pflanzen-Modells zur deutschlandweiten, automatisierten und kleinmaßstäbigen Interpolation von phänologischen Phasen zur Bewertung der Eignung der Aufnahmezeitpunkte von Satellitenfernerkundungsdaten für Landnutzungsklassifikationen. Der entwickelte Interpolationsalgorithmus basiert auf statistischen Beziehungen zwischen beobachteten phänologischen Ereignissen und Temperaturmessungen. Im Modell werden ein flächenhaft verfügbares SRTM-Höhenmodell mit klimatologischen und phänologischen Stationsdaten verknüpft, die im Rahmen des DWD-Sofortmelder-Beobachtungsnetzes erhoben und zeitnah zur Verfügung gestellt werden. Am Beispiel von zwei Testgebieten wird gezeigt, wie phasenspezifische Zeitfenster identifiziert und mit Aufnahmezeitpunkten von RapidEye-Satellitenbildern in Beziehung gesetzt werden können.

1 Einleitung

Die Satellitenbild-gestützte Differenzierung von Vegetationsklassen erfordert die Analyse multi-temporaler Satellitenbilddatensätze, um den saisonalen Entwicklungsverlauf von Pflanzen bzw. deren phänologische Stadien zu charakterisieren und von phänologischen Stadien anderer Vegetationsklassen unterscheiden zu können (VAN NIEL & MCVICAR, 2004). Die Repetitionsraten bisher verfügbarer optischer Fernerkundungsdaten sind unter den in Deutschland vorherrschenden meteorologischen Verhältnissen oft nicht ausreichend für die Bearbeitung vegetationsbezogener Fragestellungen. Neuere Satellitensysteme wie RapidEye sind aufgrund der hohen Repetitionsrate in der Lage, an die Vegetationsperiode angepasste Zeitreihendaten zu erzeugen (SANDAU, 2010). Daraus leitet sich erstmals die Möglichkeit ab, die Aufnahmetermine hinsichtlich ihrer zeitlichen Stellung in der phänologischen Vegetationsentwicklung einzuordnen, um den Klassifikationsprozess zu optimieren.

Der Beitrag beinhaltet schwerpunktmäßig die Beschreibung eines vereinfachten Temperatur-Pflanzen-Modells zur deutschlandweiten und automatisierten Interpolation ausgewählter phänologischer Stadien. Die Auswahl der phänologischen Stadien resultiert aus der Verfügbarkeit von DWD-Sofortmelderdaten. Die kleinmaßstäbigen Interpolationsergebnisse ermöglichen für beliebige Untersuchungsgebiete in Deutschland die Detektion von phänologischen Zeitfenstern, die zur phänologischen Einordnung von Aufnahmezeitpunkten verwendet werden können. Am Beispiel der phänologischen Phase *Gelbreife* der Ackerfrucht *Winterweizen* und zwei Untersuchungsgebieten wird gezeigt, wie phasenspezifische Zeitfenster identifiziert und mit RapidEye-Aufnahmezeitpunkten in Beziehung gesetzt werden können.

- 1) FG Geofernerkundung, Institut für Geowissenschaften, Martin-Luther-Universität Halle-Wittenberg, Von-Seckendorff-Platz 4, 06120 Halle (Saale), E-mail: markus.moeller@geo.uni-halle.de
- 2) Geoinformationsservice Dr. Jens Birger, Hoppbergblick 12, 06118 Halle (Saale), E-mail: birger@geoinformationsservice.de
- 3) EFTAS Fernerkundung Technologietransfer GmbH, Oststraße 2-18, 48145 Münster, E-mail: oliver.buck@EFTAS.com

Die Untersuchungen sind eingebettet in das nationale Verbundprojekt DeCOVER 2 (<http://www.decover.info>), das als deutscher Beitrag zu europäischen und weltweiten Programmen (z.B. GMES, GEOSS) die Bereitstellung von operationell verfügbaren Grundlagendaten für Entscheidungsträger und Umweltbehörden zum Ziel hat. DeCOVER 2 entwickelt Methoden zur Aktualisierung und Erweiterung bestehender Landbedeckungsinformationen (BUCK, 2010).

2 Interpolation phänologischer Phasen

Viele phänologische Modelle basieren auf statistischen Beziehungen zwischen beobachteten phänologischen Ereignissen und Wetterbeobachtungen. Die Vegetationsentwicklung in Deutschland ist durch Vegetationsruhe im Winter und eine aktive Wachstumsphase im Frühjahr und Sommer gekennzeichnet, die in erster Linie durch die Temperatur gesteuert wird. Das Eintreten von phänologischen Phasen steht dabei im Zusammenhang mit akkumulierten Wärmephasen, die durch summierte Temperaturen ausgedrückt werden können. Weitere Einflussfaktoren sind die Anzahl der Frosttage sowie die Tageslänge (SCHABER & BADECK, 2003).

Die grundlegende Vorgehensweise zur deutschlandweiten Interpolation phänologischer Phasen geht aus Abbildung 1a hervor. Das vereinfachte Temperatur-Pflanzen-Modell basiert auf punkthaften täglichen Temperatur- (DWD_T) und phänologischen Sofortmelderdaten (DWD_P) des Deutschen Wetterdienstes, die mit flächenhaften SRTM-Höhendaten und Naturraumtypen verknüpft werden ($SRTM_{NTR}$) (Kap. 2.1). Der Algorithmus setzt sich aus folgenden Modellierungsschritten zusammen, die in Kapitel 2.2 näher erläutert werden:

- MODELL 1 generiert deutschlandweite tägliche Temperaturdaten ($SRTM_T$) durch die Anwendung eines Interpolationsalgorithmus.
- Nach der Übertragung der Temperaturwerte auf die phänologischen Sofortmelderstationen werden in MODELL 2 pro Station phasenspezifische kumulierte Temperatursummen ($DWD_{T(DOY,P)}$) berechnet.
- Im MODELL 3 werden – ebenfalls durch die Anwendung eines Interpolationsalgorithmus – phasenspezifische Flächendaten julianischer Tage ($SRTM_{DOY,P}$) auf Grundlage der Temperatursummen und des SRTM- bzw. Naturraumdatensatzes erzeugt.
- Bei der Qualitätskontrolle werden unter Verwendung von drei Qualitätsmaßen beobachtete und interpolierte Eintrittstermine gegenübergestellt.

2.1 Modelleingangsdaten

2.1.1 Phänologische und meteorologische Daten

Der Deutsche Wetterdienst (DWD^1) unterhält ein phänologisches Grund- sowie Sofortmeldernetz. Das Grundnetz besteht aus etwa 1300 Stationen, berücksichtigt 167 Phasen und erlaubt langfristige Analysen. Mit dem Sofortmeldernetz werden Informationen zu Entwicklungsstadien erhoben, die sich aus der Auswertung von etwa 500 Stationen ergeben. Der

¹ <http://www.dwd.de>

schnelle Zugriff auf die Ergebnisse phänologischer Beobachtungen ermöglicht die zeitnahe Ableitung aktueller Pflanzenentwicklungsstände. Pro Phase kann die Anzahl der Stationswerte zudem stark schwanken. Für das Jahr wurden durch den DWD deutschlandweite Informationen zu 34 Pflanzen und 101 phänologische Phasen bereitgestellt. Durch das DWD-Mess- und Beobachtungsnetz von etwa 500 Klimastationen wurden zudem tägliche Witterungsinformationen (Temperatur) bereitgestellt (Abb. 1b).

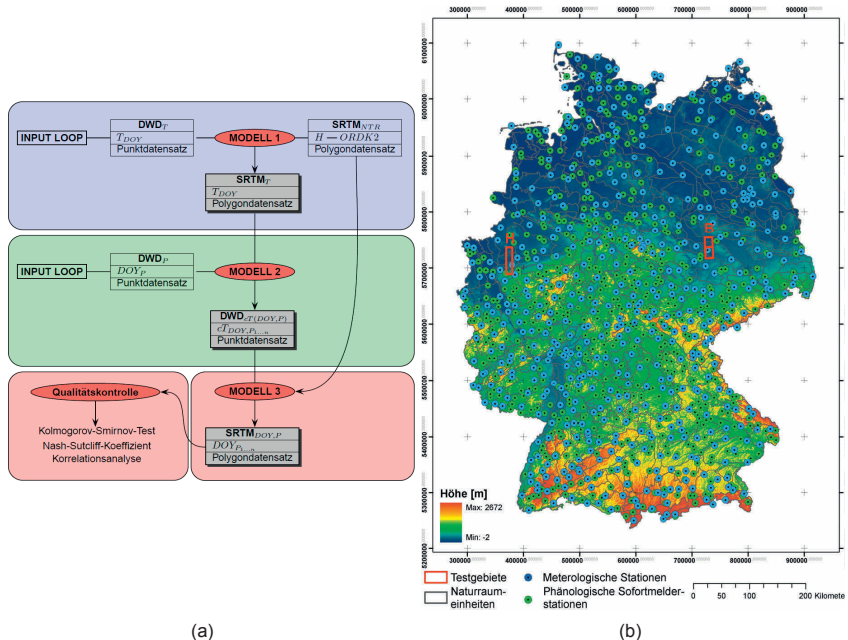


Abb. 1: Fließschema zur Interpolation phänologischer Phasen (a) sowie Modelleingangsdaten und Lage der Testgebiete Herne (H) und Bitterfeld/Dessau (B) (b)

2.1.2 Naturraumeinheiten

Das phänologische Meldersystem zielt auf die Kennzeichnung der wichtigsten Phasen innerhalb der 85 deutschen Naturraumtypen ab. Als Naturraumtypen werden Gebiete betrachtet, die einen ähnlichen Landschaftscharakter aufweisen. Die Ähnlichkeit wird u.a. bestimmt durch die Höhenlage, das Mittel der Lufttemperatur, die jährlichen Niederschlagssummen und die Eintrittstermine der phänologischen Phasen (MEYNEN et al., 1962). In Abbildung 1b sind die Naturräume für das Gebiet von Deutschland dargestellt. In den Interpolationsprozess ging die mittlere Gliederungstiefe „ORDK2“ ein.

2.1.3 SRTM

Mit dem SRTM-Datensatz steht ein nahezu weltweit verfügbares Höhenmodell zur Verfügung,

das im Zuge einer amerikanisch-europäischen Shuttle-Radar-Mission generiert wurde². Das digitale Höhenmodell (DHM) hat eine geometrische Auflösung etwa 90 x 90 m. Die horizontale bzw. vertikale Genauigkeit beträgt etwa 20 m bzw. 16 m. Die Datensätze können über verschiedene Server frei bezogen werden³. Zu beachten ist, dass das DHM die Oberfläche samt Vegetation, Gebäude etc. repräsentiert, woraus ein starkes Rauschen der Oberfläche resultiert. Um die Rauscheffekte zu minimieren, ist das DHM deshalb innerhalb der Softwareumgebung RSAGA⁴ einer Filterung unterzogen worden (LEE, 1980). Weiterhin wurde das DHM einer Segmentierung innerhalb der Softwareumgebung eCognition 7.0⁵ unterzogen. Dabei werden Rasterzellen entsprechend ihrer statistischen Ähnlichkeit und räumlichen Nachbarschaft solange zusammengefasst, bis durch den Nutzer zu definierende Abbruchkriterien erfüllt sind. Dazu gehören bereits existierende Grenzen (hier: Naturraumeinheiten), die in den Segmentierungsprozess mit einbezogen werden können. Das entscheidende Abbruchkriterium ist die statistische Heterogenität der entstehenden Segmente, mit der der Aggregierungsgrad der Segmente und damit indirekt deren Anzahl beeinflusst werden kann (MÖLLER et al., 2008). Auf diese Weise sind etwa 68 000 000 Rasterzellen zu etwa 280 000 Segmenten zusammengefasst worden. Die Segmente können als Relief-Naturraum-Einheiten betrachtet werden.

2.2 Modellierung

Der Algorithmus liegt als ein innerhalb der Programmumgebung **R**⁶ lauffähiges Skript vor⁷. Kernbestandteile sind Import und Exportfunktionen, ein Interpolationsverfahren sowie eine Summenfunktion, die alle in Schleifenstrukturen implementiert worden sind. Für jedes Interpolationsergebnis sind statistischen Qualitätsmaße abgeleitet worden.

Der direkte Import der phänologischen und klimatologischen Sofortmelderdatensätze ist mit den Funktionen *read.fwf* und *read.xls* realisiert worden, die jeweils mit der *for*-Schleifenfunktion⁸ und horizontalen Verknüpfungsfunktion Funktion *rbind*⁹ des *base*-Paketes gekoppelt wurden. Die Funktion *read.fwf*¹⁰ des R-Paketes *utils* ermöglicht das automatische Einlesen von Tabellen mit festen Breiten, wobei exakte Spaltenbreiten definiert werden können. So ist aus den zweispaltigen Eingangsdatensätzen der klimatologischen Sofortmelderdaten ein fünfspaltiger Gesamtdatensatz generiert worden. Mit der Funktion *read.xls*¹¹ des R-Paketes *xlsReadWrite* sind die phänologischen Sofortmelderdatensätze importiert worden.

Mit der R-Funktion *randomForest*¹² kam ein von BREIMAN (2001) entwickeltes regressions- und ensemblebasiertes Entscheidungsbaumverfahren zur Anwendung, mit dem sowohl kontinuierliche Attribute (z.B. Höhe, Temperatur) als auch thematische Klassen (z.B.

² <http://www.dlr.de/srtm>

³ z.B. <http://srtm.csi.cgiar.org>

⁴ <http://cran.r-project.org/web/packages/RSAGA/index.html>

⁵ <http://www.ecognition.com/> (Modul *multi-resolution segmentation*)

⁶ <http://www.r-project.org>

⁷ http://mars.geographie.uni-halle.de/geovlexcms/downloads/pdfdocs/poster_temporal-windows.pdf

⁸ <http://stat.ethz.ch/R-manual/R-devel/library/base/html/Control.html>

⁹ <http://stat.ethz.ch/R-manual/R-devel/library/base/html/cbind.html>

¹⁰ <http://stat.ethz.ch/R-manual/R-patched/library/utils/html/read.fwf.html>

¹¹ <http://cran.r-project.org/web/packages/xlsReadWrite/index.html> und <http://www.swissr.org/>

¹² <http://cran.r-project.org/web/packages/randomForest/>

Naturraumtypen) verarbeitet werden können. Entscheidungsbaumverfahren unterteilen den Merkmalsraum von erklärenden und Zielvariablen solange, bis die entstandenen Teilmengen die größten (hier: Regressions-)Zusammenhänge zeigen. Der *randomForest*-Algorithmus generiert eine Vielzahl bzw. ein Ensemble von unkorrelierten Entscheidungsbäumen. Alle Entscheidungsbäume unterscheiden sich in der Variablenauswahl und Stichprobenziehung.

Jede stationsspezifische phänologische Phase ist durch ein definiertes Eintrittsdatum gekennzeichnet, das das Ende der Temperatursummierung markiert. Die Berechnung beginnt am Tag 0 und berücksichtigt nur positive Temperaturwerte. Die Funktion *rowSums*¹³ des *base*-Paketes erlaubt die reihenweise Summierung von Datensätzen, wobei die jeweilige Anfangs- und Endspalte definiert werden können. Durch die Kopplung mit der *for*-Schleifenfunktion konnte das Summierungsende dynamisch bestimmt werden.

Die Qualitätsmaße Korrelationskoeffizient r (Pearson), NSK und D wurden mit den Funktionen *cor.test*¹⁴ und *ks.test*¹⁵ sowie *nashS*¹⁶ berechnet. Der Korrelationskoeffizient r ergibt sich aus dem Verhältnis der Kovarianz mit dem Produkt der Standardabweichung zweier Variablen. Der nicht-parametrische *Kolmogorov Smirnov*-Test gestattet den Vergleich von zwei Verteilungen unabhängig von statistischen Voraussetzungen (z.B. Normalverteilung) und der Anzahl. Basierend auf empirischen Summenfunktionen (Empirical Cumulative Distribution Function = ECDF) wird mit dem Test untersucht, ob zwei Verteilungen gleich sind (Null-Hypothese) oder signifikante Unterschiede aufweisen. Der Grad der Differenz wird durch die größte Distanz D zwischen den ECDF-Graphen zweier Variablen ausgedrückt (DAVIS, 2002). Mit dem *Nash-Sutcliff*-Koeffizienten NSK können Abweichungen simulierter Werte vom Mittelwert der Referenzwerte berechnet werden. Je näher der NSK -Wert zum Optimalwert von 1 tendiert, desto geringer sind die zu erwartenden Abweichungen (NASH & SUTCLIFF, 1970).

3 Ergebnisse

Die Ergebnisse werden am Beispiel der phänologischen Phase *Gelbreife* der Ackerfrucht *Winterweizen* für das Jahr 2010 erläutert. Die Phase startet in Deutschland am 25. Juni (Julianischer Tag 176) und endet am 4. August (Julianischer Tag 214). Das Interpolationsergebnis spiegelt typische meteorologische Bedingungen in Deutschland wider (Abb. 2a). In der westlichen Mitte Deutschlands sind die ersten Eintrittstermine zu verzeichnen. Die Phase setzt sich in nord-östlicher Richtung fort. Späte Phaseneintrittstermine sind im Nordwesten und Süden konzentriert.

¹³ <http://stat.ethz.ch/R-manual/R-patched/library/base/html/colSums.html>

¹⁴ <http://finzi.psych.upenn.edu/R/library/stats/html/cor.test.html>

¹⁵ <http://finzi.psych.upenn.edu/R/library/stats/html/ks.test.html>

¹⁶ <http://finzi.psych.upenn.edu/R/library/tiger/html/nashS.html>

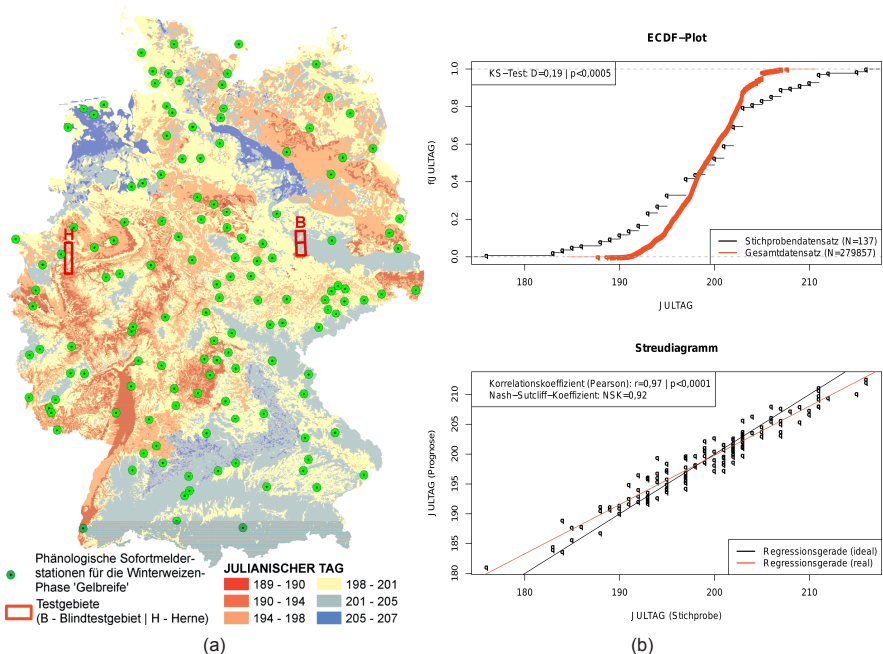


Abb. 2: Deutschland-weite interpolierte Eintrittstermine der phänologischen Phase *Gelbreife* für die Ackerfrucht *Winterweizen* (a) und zugehörige Qualitätsmaße (b)

Der Vergleich der Beobachtungs- und Simulationswerte basiert nicht auf unabhängigen Testdaten, da die phasenspezifischen Grundgesamtheiten der Sofortmeldungen gering sind. Allerdings wird eine Stichprobenunabhängigkeit durch die Verwendung des randomForest-Interpolationsverfahren gewährleistet, da der „beste“ Entscheidungsbaum auf einer zufälligen Stichprobenauswahl beruht. Aus Abbildung 2b wird deutlich, dass zwischen beobachteten und interpolierten Phaseintrittsterminen signifikante Beziehungen bestehen. So sind sowohl der Korrelationskoeffizient r mit 0,97 als auch der Nash-Sutcliffe-Koeffizient NSK mit 0,92 als hoch einzuschätzen. Die Abweichung zwischen idealer ($r = 1$) und realer Korrelation zeigen die zugehörigen linearen Regressionsfunktionen. Je größer die „Schere“ zwischen beiden Gerade ist, desto mehr weichen Beobachtung und Prognose von einander ab. Der maximale Abstand zwischen den Summenfunktionen des Beobachter- und interpolierten Gesamtdatensatzes wird durch die Distanz D ausgedrückt. Trotz der geringen Stichprobenanzahl von 137 sind die D -Werte für die Phase *Gelbreife* mit 0,19 als gering einzuschätzen. Aus der Visualisierung der Summenkurven ist zudem ersichtlich, welche Tage von den Prognoseunsicherheiten besonders betroffen sind. So ist im Bereich der julianischen Tage 190 bis 195 sowie um 205 von größeren Prognoseunsicherheiten auszugehen.

Abbildung 3 veranschaulicht, wie für beliebige Testgebiete in Deutschland Aufnahmetermine von Satellitenbildern (hier: RapidEye) mit phänologischen Phasen verknüpft werden können. Innerhalb der beiden Testgebiete Herne (H) und Bitterfeld/Dessau (B; Abb. 2a) sind Ackerschlag-bezogene Kartierungen durchgeführt worden. Die Kartierungsergebnisse für die Feldfrucht *Winterweizen* sind mit den Interpolationsergebnissen der beobachteten Phasen sowie den NDVI-Werten der jeweiligen RapidEye-Satellitenbilder verschnitten worden. Die einzelnen Phasen sind über den Jahresverlauf der Julianischen Tage (y-Achse) schraffiert dargestellt, die Phase *Gelbreife* ist rot hervorhoben. Zusätzlich werden die Phasen durch die Mediane als Zahl bzw. horizontale Linie charakterisiert. Danach tritt die *Gelbreife* im Testgebiet Herne eher ein als im Testgebiet Bitterfeld/Dessau, was auch durch die Mediane 195 und 201 ausgedrückt wird. Für beide Testgebiete steht jeweils eine RapidEye-Satellitenbildszene zur Verfügung, die allerdings eher am Anfang der Phase aufgenommen worden ist. Die zugehörigen NDVI-Werte der Kartierungsflächen sind durch Boxplots dargestellt. Die NDVI-Mediane repräsentieren einen typischen jahreszeitlichen Verlauf. Die Werte steigen mit Beginn des Jahres, um während des Ährenschiebens ihren Höhepunkt zu erreichen. Die NDVI-Werte nehmen dann mit zunehmender Reife ab und zeigen nach der Ernte ein Minimum. Nach der Bestellung und dem Auflaufen im Herbst kommt es zu einer Werteerhöhung.

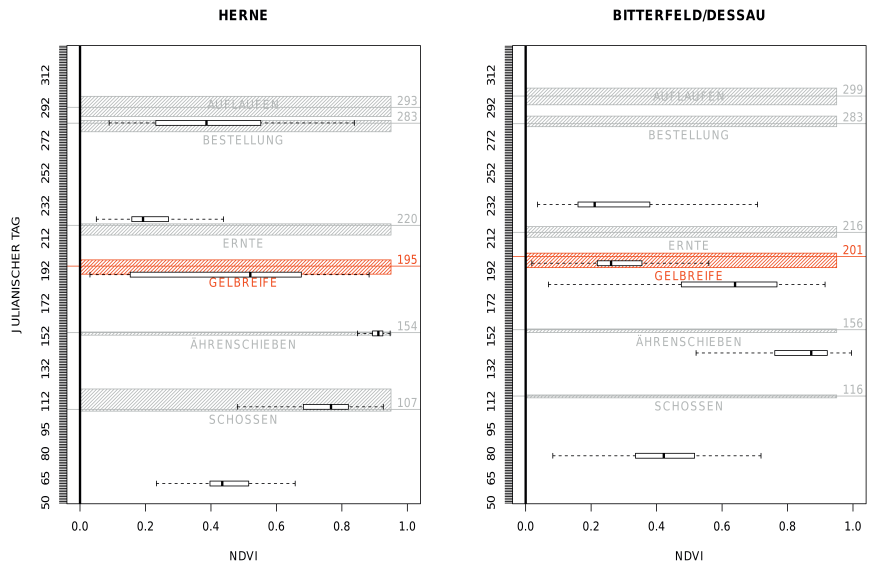


Abb. 3: Interpolierte Eintrittstermine phänologischer Phasen der Ackerfrucht *Winterweizen* für Trainingsdatensätze der Testgebiete Herne (links) und Bitterfeld/Dessau (rechts)

4 Zusammenfassung und Ausblick

Das nationale Verbundprojektes DeCOVER 2 entwickelt Methoden zur räumlichen bzw. thematischen Aktualisierung und Erweiterung von bestehenden Landbedeckungs- bzw. Landnutzungsinformationen. Als Datengrundlagen dienen Fernerkundungsinformationen (z.B. RapidEye), die durch eine hohe räumliche und zeitliche Verfügbarkeit gekennzeichnet sind. Die vorgestellte Studie zeigt, wie aus phänologischen und meteorologischen Punktdatensätzen des DWD-Beobachternetzes deutschlandweite Karten der Eintrittstermine phänologischer Stadien interpoliert werden können. Der Algorithmus erlaubt eine Bereitstellung der Ergebnisse unmittelbar nach Erhalt der Sofortmelderdaten. Die Interpolationsergebnisse erlauben die Einordnung der RapidEye-Aufnahmetermine in den phänologischen Entwicklungsverlauf von Vegetationstypen, um den DeCOVER-Klassifikationsprozess zu optimieren. Der entwickelte Algorithmus ist als offenes System konzipiert. Das heißt, dass einzelne alternative Interpolationsverfahren und/oder zusätzliche Eingangsdaten (z.B. Niederschlag, MODIS) in den Programmablauf integriert werden können. Schließlich ist geplant, die Interpolationsergebnisse mit Daten des DWD-Beobachtungsnetzes einer unabhängigeren Prüfung zu unterziehen.

5 Literaturverzeichnis

- BUCK, O., 2010: DeCOVER 2 – The German GMES extension to support land cover data systems: Status and outlook. Proceedings 'ESA Living Planet Symposium', 28 June – 2 July 2010 (ESA SP-686, December 2010), Bergen, Norway
- BREIMAN, L., 2001: Random forests. *Machine Learning*, **45**, S. 5-32.
- DAVIS, J., 2002: Statistics and data analysis in geology. John Wiley & Sons, New York.
- LEE, J.-S., 1980: Digital image enhancement and noise filtering by use of local statistics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-2** (2), 165 - 168.
- MEYNEN, E., SCHMITHÜSEN, J., GELLERT, J., NEEF, E., MÜLLERMINY, H. & SCHULTZE, J.H., 1962: Handbuch der naturräumlichen Gliederung Deutschlands, **1**, Bad Godesberg
- MÖLLER, M., VOLK, M., FRIEDRICH, K. & LYMBURNER, L., 2008: Placing soil genesis and transport processes into a landscape context: A multi-scale terrain analysis approach. *Journal of Plant Nutrition and Soil Science*, **171**, S. 419 – 430.
- NASH, J. & SUTCLIFF, J., 1970: River flow forecasting through conceptual models – Part I: A discussion of principles. *Journal of Hydrology*, **10** (3), 282-290.
- SANDAU, R., 2010: Status and trends of small satellite missions for earth observation. *Acta Astronautica*, **66** (1-2), 1 - 12.
- SCHABER J. & BADECK, F.-W., 2003: Physiology-based phenology models for forest tree species in Germany. *International Journal of Biometeorology*, **47**, S. 193-201.
- VAN NIEL, T. & MCVICAR, T., 2004: Determining temporal windows for crop discrimination with remote sensing: a case study in south-eastern Australia. *Computers and Electronics in Agriculture*, **45** (1-3), S. 91 - 108.