

Nachweis der Signifikanz einzelner Aufnahmezeitpunkte auf die Klassifikationsgenauigkeit am Beispiel einer RapidEye-Zeitreihe

Tobias Schmidt, Michael Förster, Birgit Kleinschmit

Technische Universität Berlin, Fachgebiet für Geoinformation in der Umweltplanung

Mit steigendem Bedarf Ergebnisse der verschiedenen Monitoringpflichten des Bundes und der Länder (u.a. NATURA 2000 oder WRRL) stieg in verschiedenen europäischen Ländern der Bedarf, diese Berichtspflichten durch Fernerkundungsdaten abzudecken. Davon zeugen verschiedene praxisnahe Forschungsprojekte mit dem Ziel einer Effektivierung der Forschungsaufgaben (z.B. CARE-X, Habit-Change, Habistat, SARA, SARA-EnMap, MS.Monina) (vgl. Gillespie et al. 2008). Im Rahmen des BMWi-Projektes CARE-X+ „Methoden für das sensorunabhängige phänologische-basierte Monitoring von Vegetation in Schutzgebieten“ wurde an operationellen Methoden gearbeitet um dem gesteigerten Bedarf Rechnung zu tragen. Gerade das Vegetationsmonitoring ist bisher nicht nur sehr plattformabhängig, sondern auch wegen der inter-annualen und intra-annualen hohen Variabilität des Untersuchungsgegenstandes stark an projektbezogene Rahmenbedingungen angelehnt.

Die hier präsentierte Studie hat das Ziel die Zeitpunktwahl von Satellitenaufnahmen für ein multi-temporales Vegetationsmonitoring zu unterstützen. Auf diese Weise soll sensorunabhängig überprüfbar sein, welche Vegetationseinheit zu welchem Zeitpunkt und mit welcher Sensoreinheit am besten erfassbar ist.

Aus dem Bereich der Hyperspektralanalyse sind verschiedene Ansätze bekannt, um aus der Vielzahl von *Features* jene auszuwählen, die eine optimale Klassifikation ermöglichen. Bei der sogenannten *Feature Subset Selection* stehen vor allem die Verbesserung der allgemeinen Klassifikationsgenauigkeit und die Beschleunigung und Effizienzsteigerung der Klassifikatoren im Vordergrund (Guyon & Elisseeff 2003). Diese Ansätze können auch auf die Bereiche des Vegetationsmonitorings übertragen werden. Hierbei steht allerdings nicht die Reduzierung spektraler *Features* aus Hyperspektraldaten im Vordergrund, sondern die Reduzierung und Bestimmung temporaler *Features* für eine adäquate Klassifikation untersuchter Vegetationseinheiten. Die Ergebnisse verschiedener *Feature Subset Selection* Methoden unterschieden sich allerdings erheblich. Es hat sich gezeigt, dass sowohl die Gewichtung als auch die Anzahl der ausgewählten *Features* für ein bestimmtes Klassifikationsniveau sehr unterschiedlich ausfallen können (vgl. Pal & Foody 2010). Des Weiteren tendieren die Ergebnisse zu einer Überanpassung (*overfitting*) des Klassifikators an die verwendeten Trainingsdaten (Loughrey & Cunningham, 2004). In der vorliegenden Arbeit wurde daher ein stärkerer empirischer Ansatz gewählt, um eine exaktere Aussage über den Einfluss der Aufnahmezeitpunkte auf das Klassifikationsergebnis zu erhalten. Hierbei wurden alle potentiell möglichen Kombinationen (*subsets*) von Aufnahmezeitpunkten, in Abhängigkeit von der Anzahl kombinierter Aufnahmen, mit einer *Support Vector Machine* (SVM) klassifiziert und jene Kombinationen gesucht, die das beste Klassifikationsergebnis (*accuracy*) für die verwendeten Daten liefern. Auf diese Weise können, abhängig von den verwendeten Daten, exakte Aussagen über geeignete Aufnahmezeitpunkte getroffen werden.

Zum Nachweis der Effekte des Aufnahmezeitpunktes von Satellitendaten auf das Klassifikationsergebnis wurde eine Signifikanzanalyse für die Klassifikationsunterschiede durchgeführt. Hierfür wurde der Monte-Carlo Permutationstest verwendet (vgl. Foody 2004, McKenzie 1996). Dadurch war es möglich die Kombinationen mit den höchsten Klassifikationsgenauigkeiten mit einer Referenzkombination auf signifikante Unterschiede zu testen. Als Referenzkombination wurde jene verwendet, die am dichtesten am Mittelwert der Klassifikationsgenauigkeiten aller Kombinationen lag.

Die vorliegende Methodik wurde anhand einer RapidEye-Zeitreihe bestehend aus 24 Aufnahmen entwickelt und getestet. Die Aufnahmen stammten aus den Jahren 2009 bis 2011 und sind flächendeckend für die Döberitzer Heide, einem geschützten ehemaligen Truppenübungsplatz westlich von Berlin, vorhanden. Untersuchungsgegenstand waren naturnahe Offenlandbiotoptypen innerhalb dieser Heidelandschaft. Hierzu wurden *in situ* Daten aus verschiedenen Bereichen des Areals verwendet. Insgesamt standen 853 Messpunkte als Pixel (*samples*) für die Klassifikationen zur Verfügung. Diese Daten wurden für die Klassifikationen in je eine Trainings- und Validierungsmenge aufgeteilt. Für erste Ergebnisse wurden für jeden Kombinationsumfang jeweils bis zu 1000 zufällige Kombinationen generiert und diese nach obenstehender Vorgehensweise klassifiziert.

Für eine optimale Klassifikation der Vegetationseinheiten wurden die RapidEye-Aufnahmen geometrisch und atmosphärisch korrigiert. Für eine gesteigerte Klassifikationsgenauigkeit wurden Vegetationsindizes für die Klassifikationen verwendet. Hierfür wurde der NDVI (*normalized difference vegetation index*) (Rotes Band zum NIR Band) sowie dessen Modifikationen NDVI RedEdge (Rotes Band zum RedEdge Band) und NDVI RedEdge-NIR (RedEdge Band zum NIR Band) berechnet. Der NDVI stellt in der Fernerkundung den gängigen Vegetationsindex für die Erkennung von Biomasse dar (Studer et al. 2007). Die Modifikationen wurden ausgewählt, da das von den RapidEye-Sensoren abgedeckte RedEdge-Spektrum sehr sensibel auf Chlorophyll innerhalb der Biomasse reagiert und somit eine bessere Klassifikationsgenauigkeit von Offenlandbiotoptypen zu erreichen ist (Schuster et al. 2011).

Die ersten Ergebnisse zeigen, dass sich Mittelwert und Maximalwert der 1000 zufälligen Kombinationen merklich unterscheiden (Abbildung 1). Besonders bei noch geringer Anzahl von Aufnahmen (*features*) sind die Unterschiede groß. Erst mit zunehmendem Umfang verringert sich der Unterschied, bleibt jedoch fast immer über 3%-Punkte. Der durchgeführte Signifikanztest verdeutlicht eine generelle signifikante Verbesserung der Klassifikationsgenauigkeit durch die gezielte Auswahl von wichtigen Aufnahmezeitpunkten. In Abhängigkeit von der Anzahl der Aufnahmen ist zudem eine Sättigung der Klassifikationsgenauigkeit mit steigendem Umfang zu erkennen. Diese tritt bei ungefähr 7-10 Aufnahmen ein.

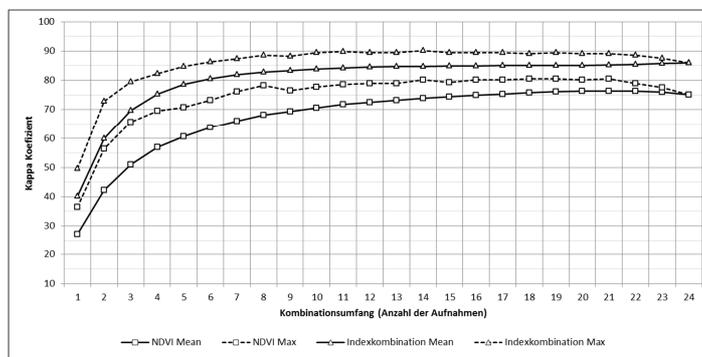


Abbildung 1: Mittel- und Maximalwerte Kappa Koeffizient bei jeweils 1000 zufälligen Kombinationen für den NDVI und die Indexkombination.

Unterschiede sind auch zwischen den drei Vegetationsindizes sowie bei deren Kombination zu erkennen. Sowohl beim Mittelwert, als auch beim Maximalwert der Klassifikationsgenauigkeit hebt sich die Kombination der drei Vegetationsindizes merklich von den einzelnen ab. Der NDVI RedEdge zeigt innerhalb der Vegetationsindizes die höchste Erkennbarkeit der Offenlandbiotoptypen, während der NDVI die geringste Genauigkeit aufweist.

Bei der Gewichtung der Aufnahmen an den besten 100 Kombinationen können klare Tendenzen für bestimmte Aufnahmezeitpunkte erkannt werden (Abbildung 2). Besonders die Aufnahmen aus dem Hochsommer (Tage 188, 198, 208) scheinen eine wichtige Rolle bei der Erkennung von Offenlandbiotoptypen zu spielen. Hingegen werden die Aufnahmen aus dem Frühjahr teilweise als nicht relevant für die Klassifikation bewertet. Hierbei ist jedoch zu berücksichtigen, dass die häufig gezogenen Aufnahmen nicht zwangsläufig auch in Kombination das beste Klassifikationsergebnis aufweisen.

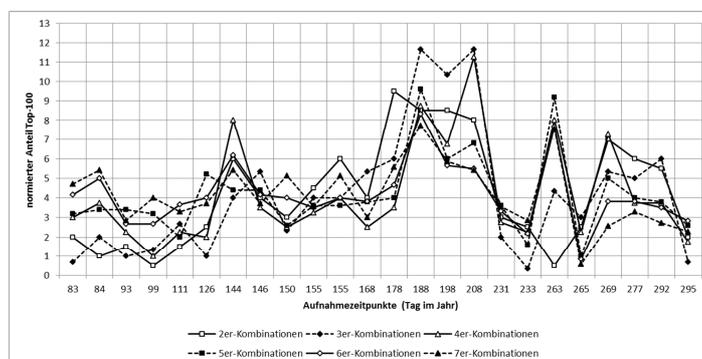


Abbildung 2: Anteile der Aufnahmen an den besten 100 Kombinationen für die Kombination der Vegetationsindizes. Da die Wahrscheinlichkeit einer Aufnahme in eine Kombination gezogen zu werden mit steigendem Kombinationsumfang ansteigt, müssen diese für einen Vergleich Normiert werden. Hierzu wurden alle gemessenen Anteile der Aufnahmen durch den Kombinationsumfang dividiert.

Literatur

Foody, G. (2004). Thematic Map Comparison: Evaluating the Statistical Significance of Differences in Classification Accuracy. *Photogrammetric Engineering & Remote Sensing*, Vol. 70, No. 5, S. 627-633.

Gillespie, T. W., Foody, G. M., Rocchini, D., Giorgi, A. P., & Saatchi, S. (2008). Measuring and modeling biodiversity from space. *Progress in Physical Geography*, Vol. 32, No. 2, S. 203-221.

Guyon, I., Elisseeff, A. (2003). An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research*, Vol. 3, S. 1157-1182.

Loughrey, J., Cunningham, P., (2004). Overfitting in Wrapper-Based Feature Subset Selection: The Harder You Try the Worse it Gets. In: *Research and Development in Intelligent Systems XXI*. Bramer, M., Coenen, F., Allen, T. London, Springer-Verlag, 2004, S. 33-43.

McKenzie, D. P. (1996). Comparing correlated kappas by resampling: is one level of agreement significantly different from another? *Journal of Psychiatric Research*, Vol. 30, No. 6, S. 483-492.

Pal, M., & Foody, G. M. (2010). Feature Selection for Classification of Hyperspectral Data by SVM. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 48, No. 5, S. 2297-2307.

Studer, S., Stöckli, R., Appenzeller, C., Vidale, P. L. (2007), A comparative study of satellite and ground-based phenology. *International Journal of Biometeorology*, Vol. 51, S. 405-415.

Schuster, C., Förster, M., & Kleinschmit, B. (2011). Testing the red edge channel for improving land-use classification based on high-resolution multi-spectral satellite data. *International Journal of Remote Sensing*, S. 1-19.