Spatial Accuracy 2014, East Lansing, Michigan, July 8-11

Evaluation of estimated satellite images for filling data gaps in an intra-annual high spatial resolution time-series

Tobias Schmidt, Michael Förster, Birgit Kleinschmit

Technical University Berlin, Geoinformation in Environmental Planning Lab

Abstract

Interpolation approaches for filling temporal gaps of missing information due to clouds or haze are well established for MODIS or Landsat imagery. However, approaches for multi-temporal satellite data with higher spatial resolution are still missing. In this study three interpolation functions were used for filling gaps within a high spatial resolution RapidEye time-series. The first results indicate that a linear interpolation function can estimate adequate results for five different semi-natural grassland classes (overall RMSE = 0.077). However, the accuracy of the results is strongly related to the distribution of the observed data and the gap length between the observations.

Keywords: linear interpolation, polynomial interpolation, time-series, high spatial resolution, RapidEye

1. Introduction

Multi-temporal satellite observations are a valuable source for detecting annual vegetation dynamics. Systems such as NOAA AVHRR, MODIS (Jönsson & Eklundh, 2004) and Landsat have shown the possibility to utilize phenological information on medium to broad spatial scales. With the new generation of high temporal and spatial resolution satellite constellations, e.g. RapidEye or Sentinel-2 in 2014, phenological patterns can be combined with accurate land use classifications of very small vegetation patches (Schuster et al., 2011). Still, the revisiting time of RapidEye (approx. 14 days) and Sentinel-2 (approx. 5 days) will lead to unfavorable data gaps due to weather conditions. There are already approaches developed to mitigate gaps in MODIS or Landsat time-series (Brooks et al., 2012; Gao et al., 2006). However at present no utilization of these methods for high spatial resolution data (< 10m GSD) is available.

Therefore, the presented study aims to adapt the interpolation approaches, developed for MODIS and Landsat data, to interpolate high spatial resolution images.

2. Study area & Data

In this study a RapidEye (6.5m GSD, 5 spectral bands) (Sandau, 2009) time-series from 2010 with nine acquired images (see Figure 1) from a protected nature reserve in Germany was utilized. All scenes were ordered as Level 1B data, radiometrically calibrated without ortho-rectification. The scenes were geometrically corrected to sub-pixel accuracy (RMSE: ≤ 2) and atmospherically corrected with the ATCOR tool in the ERDAS Imagine software (Richter, 1996).

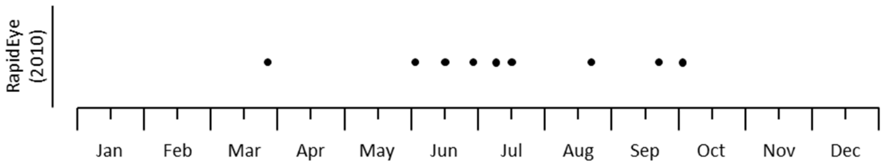


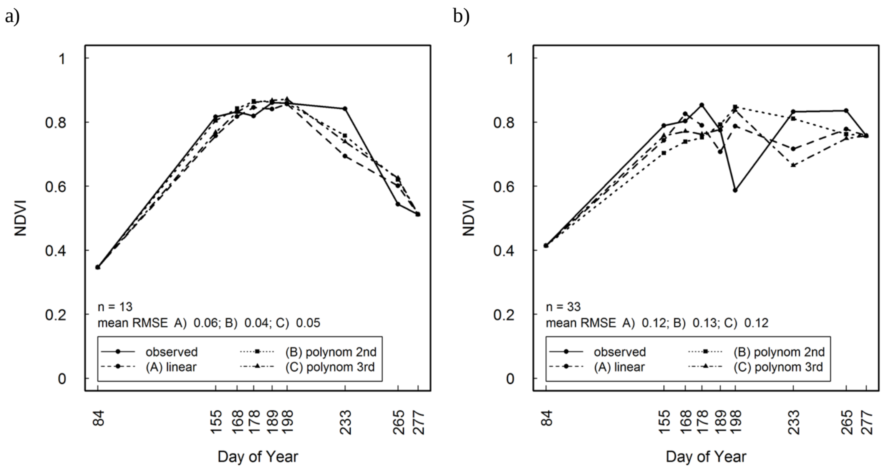
Figure 1: Temporal distribution of observed RapidEye imagery from 2010.

3. Methods

Three interpolation functions (linear and polynomial of 2nd and 3rd degree) of the NDVI of the images were used to estimate data for temporal gaps (Colditz et al., 2008). In a leave-one-out process, each image from the observed image layerstack, with the exception of the first (day of year 84) and the last image (day of year 277) was omitted. For samples of five different semi-natural grassland classes within the study area, each simulated gap was filled with estimated data from the interpolation functions. The resulting estimates were validated by a regression analysis (estimated vs. observed).

4. Results

The accuracy of the estimated NDVI observations of the three interpolation approaches vary according to the semi-natural grassland classes and the omitted days between a mean RMSE = 0.04 and 0.13. In Figure 2 temporal mean NDVI profiles of the investigated classes are shown. For all classes an observed and three estimated NDVI curves (linear, polynomial 2nd and 3rd degree) of the omitted days are depicted. It can be recognized that the accuracy of the estimations is lower when the observed NDVI profile have a multi-modal shape. In case of mono-modal profiles, the accuracy ascends. Therefore, the estimated profiles for *Molinion caeruleae*, *Caricion elatae* and *Phragmitetum australis* matches mainly to the observed NDVI profile. On contrary, the match of the estimated profiles of *Calthion palustris* and *Arrhenatheretum elatioris* deviate to a larger extend from the observed profile.



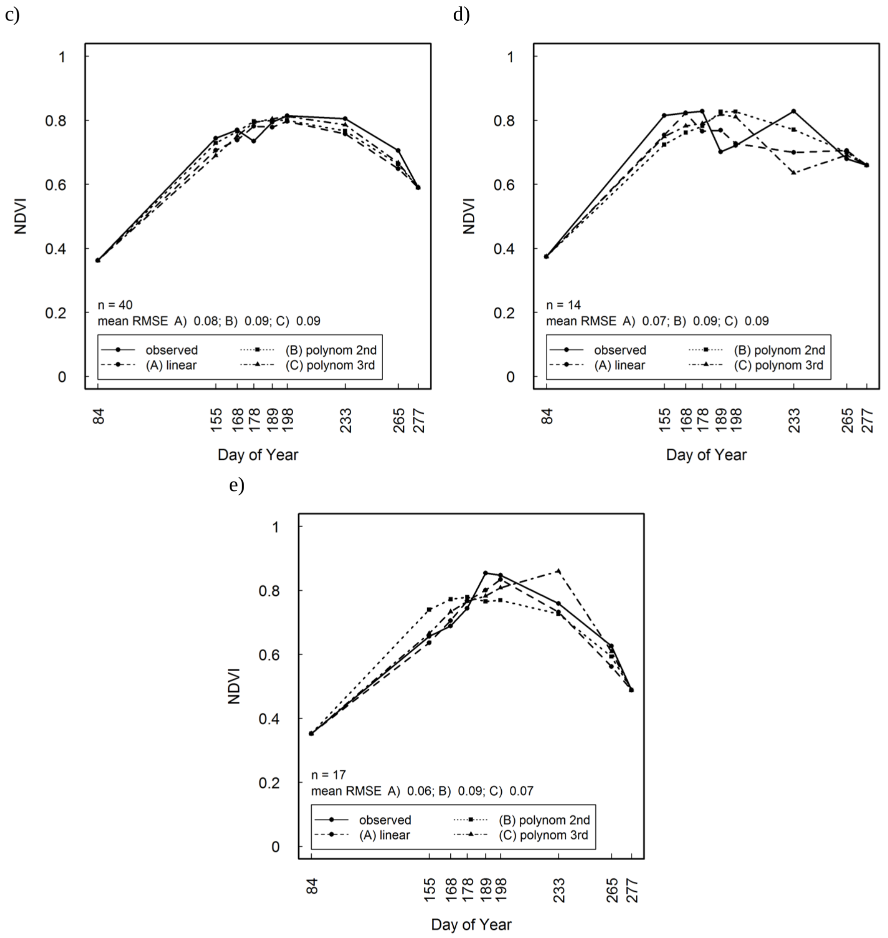


Figure 2: Observed and estimated temporal mean NDVI profiles of the three interpolation approaches for the grassland species a) *Molinion caeruleae*, b) *Calthion palustris*, c) *Caricion elatae*, d) *Arrhenatheretum elatioris*, e) *Phragmitetum australis.*

Figure 3 shows the average RMSE from the linear regression analiysis for a) each class and b) for each omitted day. *Calthion palustris* obtained the highest mean RMSE (up to 0.128) and standard deviation (up to 0.80) of all classes. This can be reasoned by a mowing event on day 198 for some samples of this class (see Figure 4). Thus, day 198 shows the highest standard deviation (up to 0.112) of all omitted days. The highest mean RMSE occurs at day 233 (up to 0.147), since this date has the largest temporal distance to its neighbor dates and therefore a high uncertainty for prediction. The temporal distance of day 155 to its previous date (84) is also large, but the uncertainties were strongly mitigated by the short distance to its following date (168) (RMSE up to 0.089).

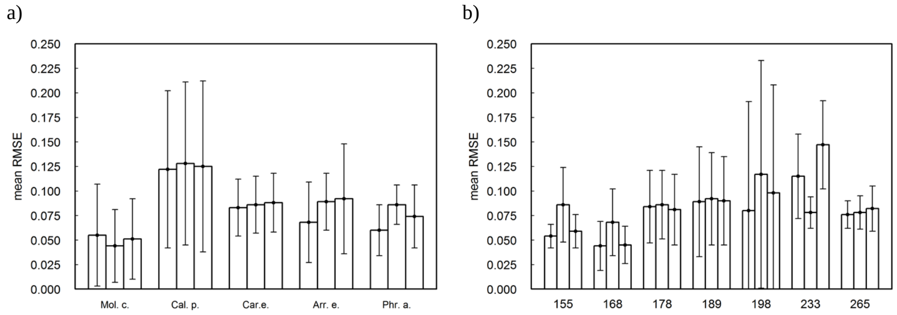


Figure 3: Mean RMSE of estimated NDVI profiles of a) the semi-natural grassland classes and b) the omitted days for the three interpolation approaches (left bars = linear, central = polynomial 2^{nd} degree, right = polynomial 3^{rd} degree). Error bars indicating the 1st standard deviation.

Figure 4 shows and exemplary linear regression for *Calthion palustris* on days 198 and 233 for the used interpolation approaches. Day 198 (above) indicates an anthropogenic measure (mowing event) for some areas of this class and therefore partly a strong overestimation for all three approaches. Day 233 (below) indicates the regrowth of biomass on these areas but an underestimation when linear or 3^{rd} degree polynomial function were used. The 2^{nd} degree polynomial function seems to be less sensitive against this specific local minima in the observed profile. Consequently the RMSE of the 2^{nd} polynomial is higher for day 198 (0.313) but lower for day 233 (0.077).

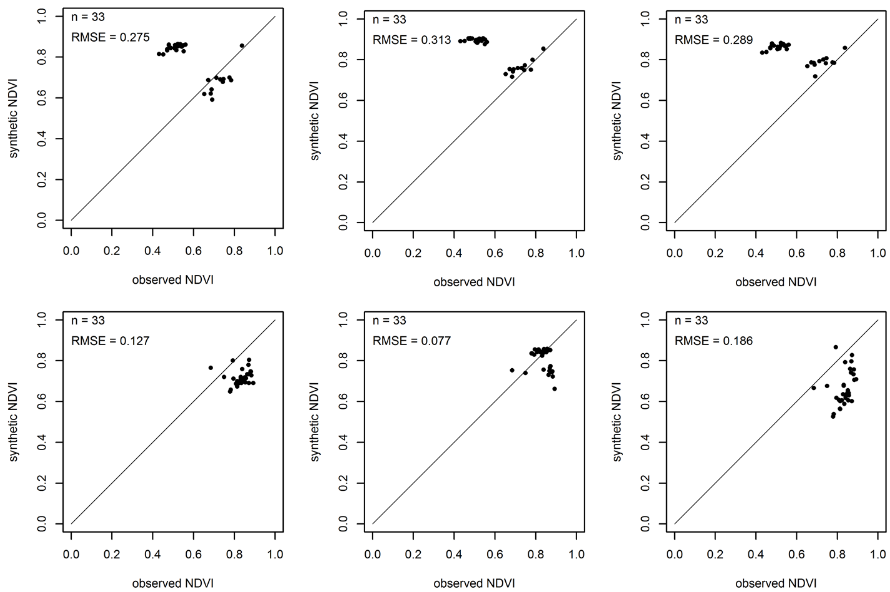


Figure 4: Exemplary linear regressions of observed and synthetic data for the *Calthion palustris* class (left: linear, central: polynomial 2^{nd} degree, right: polynomial 3^{rd} degree) for two days (above: day 198, below: day 233).

5. Discussion

The comparison of the observed and estimated temporal profiles for the five semi-natural grassland classes indicates that all three interpolation functions can be sufficient for filling missing data gaps. However, if the gap is too large these functions do not capture the natural phenological development or anthropogenically-induced changes in the investigated land cover classes. Especially in the case of one or more small local minima within an annual NDVI (multi-modal) profile, biased values were returned. However, the linear interpolation function seems to be most suitable for filling data gaps for these specific semi-natural grassland classes, since this method achieved the lowest overall RMSE (0.077) in this study. Both polynomial functions performing similar but yielding a slightly higher overall RMSE (2^{nd} degree = 0.087 and 3^{rd} degree = 0.086). An additional in this study realized utilization of higher-order polynomial functions achieved obviously lower accuracies.

6. Conclusion

The first results of this study indicate that it is possible to fill gaps of missing information on a class by class basis in a high spatial resolution time-series with the utilized interpolation functions. However, distinct limitations of the methods restrict the application for some cases. If the gap length between two observations is too long, the uncertainty in the estimated NDVI increases. This problem can be theoretically solved with a higher density of observed images, which is practically limited due to the sensor's repetition rate and the average weather conditions in this case for mid-Europe. Another way to minimize this problem is the application of other estimator (e.g. Fourier regression) or filtering algorithms (e.g. Savitzky-Golay filter (1964)) to acknowledge the heterogenic spectral behavior (Brooks et al., 2012).

References

Brooks, E. B., Thomas, V. A., Wynne, R. H., & Coulston, J. W. (2012). "Fitting the Multitemporal Curve: A Fourier Series Approach to the Missing Data Problem in Remote Sensing Analysis". *IEEE Transactions on Geoscience and Remote Sensing*, 50(9), 3340–3353.

Colditz, R. R., Conrad, C., Wehrmann, T., Schmidt, M., & Dech, S. (2008). TiSeG: "A Flexible Software Tool for Time-Series Generation of MODIS Data Utilizing the Quality Assessment Science Data Set". *IEEE Transactions on Geoscience and Remote Sensi*ng, 46(10), 3296–3308.

Gao, F., Masek, J., Schwaller, M., & Hall, F. (2006). „On the blending of the Landsat and MODIS surface reflectance: predicting daily Landsat surface reflectance". *IEEE Transactions on Geoscience and Remote Sensing*, 44(8), 2207–2218.

Jönsson, P., & Eklundh, L. (2004). "TIMESAT—a program for analyzing time-series of satellite sensor data". *Computers & Geosciences*, 30(8), 833–845.

Richter, R. (1996). "Atmospheric correction of satellite data with haze removal including a haze/clear transition region". *Computers & Geosciences*, 22(6), 675–681.

Sandau, R. (2009). „Status and trends of small satellite missions for Earth observation". *Acta Astronautica*, 66(1-2), 1–12.

Savitzky, A., & Golay, M. J. E. (1964). "Smoothing and differentiation of data by simplified least squares procedures". *Analytical Chemistry*, 36(8), 1627–1639.

Schuster, C., Ali, I., Lohmann, P., Frick, A., Förster, M., & Kleinschmit, B. (2011). „Towards Detecting Swath Events in TerraSAR-X Time Series to Establish NATURA 2000 Grassland Habitat Swath Management as Monitoring Parameter". *Remote Sensing*, 3(7), 1308–1322.