The Impact of Multi-Sensor Data Assimilation on Plant Parameter Retrieval and Yield Estimation for Sugar Beet

Martina Hodrius a, Silke Migdall a, Heike Bach a, Tobias Hankb

a Vista Geowissenschaftliche Fernerkundung GmbH, Gabelsbergerstraße 51, 80333 München, Germany, Email: hodrius@vista-geo.de, migdall@vista-geo.de, bach@vista-geo.de
bLudwig-Maximilians-Universität München, Luisenstraße 37, 80333 München, Germany, Email: tobias.hank@lmu.de

KEY WORDS: yield modelling, crop growth model PROMET, sugar beet, multi-sensor data assimilation, SLC

ABSTRACT:

Yield Maps are a basic information source for site-specific farming. For sugar beet they are not available as in-situ measurements. This gap of information can be filled with Earth Observation (EO) data in combination with a plant growth model (PROMET) to improve farming and harvest management. The estimation of yield based on optical satellite imagery and crop growth modelling is more challenging for sugar beet than for other crop types since the plants’ roots are harvested. These are not directly visible from EO. In this study, the impact of multi-sensor data assimilation on the yield estimation for sugar beet is evaluated. Yield and plant growth are modelled with PROMET. This multi-physics, raster-based model calculates photosynthesis and crop growth based on the physiological processes in the plant, including the distribution of biomass into the different plant organs (roots, stem, leaves and fruit) at different phenological stages.

The crop variable used in the assimilation is the green (photosynthetically active) leaf area, which is derived as spatially heterogeneous input from optical satellite imagery with the radiative transfer model SLC (Soil-Leaf-Canopy). Leaf area index was retrieved from RapidEye, Landsat 8 OLI and Landsat 7 ETM+ data. It could be shown that the used methods are very suitable to derive plant parameters time-series with different sensors. The LAI retrievals from different sensors are quantitatively compared to each other. Results for sugar beet yield estimation are shown for a test-site in Southern Germany. The validation of the yield estimation for the years 2012 to 2014 shows that the approach reproduced the measured yield on field level with high accuracy. Finally, it is demonstrated through comparison of different spatial resolutions that small-scale in-field variety is modelled with adequate results at 20 m raster size, but the results could be improved by recalculating the assimilation at a finer spatial resolution of 5 m.

1. INTRODUCTION

Yield and biomass maps are basic information sources for smart farming. These maps can be used for the daily assessment of plant development and site-specific fertilization measures. Sugar beet yield is usually not mapped during harvesting [Schmittmann 2002], hence, spatially distributed yield information is hard to obtain. Earth Observation (EO) data fill this lack of information and support smart farming by delivering up-to-date information on plant growth independent from in-situ data [Migdall et al. 2013]. However, operational agricultural application requires reliable information for any date during the vegetation period, whatever the weather conditions are. Thus, the combination of EO data with crop growth modelling is necessary to monitor the vegetation development continuously over the growing season [Bach & Angermair 2013, Hank et al. 2015].

In this study, multi-sensor EO data from 2012 to 2014 is used to derive plant parameters for sugar beet using an inversion of the SLC (Soil-Leaf-Canopy) model [Verhoef & Bach 2003, 2007]. The resulting green LAI maps are then assimilated into the crop growth model PROMET to model the plant development at different phenological stages and estimate yield. This approach is well-established and validated for winter wheat as shown in several studies [Bach & Angermair 2013; Hank et al. 2015, Migdall et al. 2013]. The same approach is now applied for sugar beet. The estimation of sugar beet yield based on optical satellite imagery is more challenging than for other crop types since the plants’ roots are harvested, which are not directly visible from EO. However, PROMET calculates crop growth based on physiological processes, including the distribution of biomass into the different plant organs for each grid cell [Hank et al. 2015].

Three main issues are addressed by this paper: The first is to evaluate the impact of different sensor properties, e.g. of the spatial resolution, on plant parameter retrieval. Therefore, two data pairs have been selected, that are comparable in terms of acquisition time. The second question is, how accurate the yield estimation for sugar beet may become over 3 consecutive years using the multi-sensor approach (RapidEye, Landsat ETM+, OLI). This was done using field mean values provided by the farmer for validation. The third question is, how well small scale in-field heterogeneity is traced by the yield modelling approach. For this, yield was estimated at a 5 m and at a 20 m raster grid. The results were compared to evaluate the effect of spatial resolution on the results. As validation data, sampling points in two fields with a size of 4 m² were harvested manually to measure the spatial distribution of sugar beet yield.

2. DATA AND TEST SITE

Plant parameter retrieval and yield estimation for sugar beet have been conducted for a test site in Southern Germany. The test site is located near Straubing next to the river Danube, at the centre of the so called Gäuboden, a region in Lower Bavaria, which covers one of the largest loess regions in Southern Germany.
From 2012 to 2014 the development of the Leaf-Area-Index (LAI) during the vegetation period was monitored for several sugar beet fields. After assimilation into the crop growth model, yield has been estimated. A multi-sensor approach has been used for the crop monitoring, since the whole vegetation period should be covered with satellite data to ensure an accurate yield model result.

The freely available data of the Landsat missions is well-suited for plant parameter retrieval. The spatial resolution of 30 m is sufficient for many applications in agriculture and the spectral range from Visible to SWIR (Short Wave Infrared), the latter of which is sensitive to plant water content, is ideal for agricultural analyses. Besides Landsat, also RapidEye data is used. The spectral configuration containing the red edge band and the spatial resolution of 5 meters make the sensors very suitable for vegetation monitoring, but the missing SWIR reduces the spectral information content. With the constellation of five satellites and pointing capabilities RapidEye allows to cover the region of interest every day. Accordingly, data can be acquired as soon as there is no cloud cover. Since the presented yield estimation methods are used in an operational mode, a multi-sensor approach is preferred for a reasonable compromise between data availability, data quality and economic interests (mix of cost-free and commercial data). For the test site, the availability of data has been varying in the different years. For 2012, only Landsat 7 ETM+ images have been used. Since 2013, also Landsat 8 OLI data is available. For 2014 the input data used are from Landsat 7 ETM+, Landsat 8 OLI and RapidEye. Table 1 shows a list of the used satellite data. For the analysis of these data, atmospheric correction was carried out to obtain surface spectral reflectance. Based on the calibrated data, the photosynthetically active leaf area is retrieved by using the satellite data has to be pre-processed extensively: it has to be atmospherically corrected and spectrally corrected. For this, the radiative transfer model MODTRAN, which was developed by the US Airforce, 2015 is used. This physically-based model calculates the path of the light from the sun through the atmosphere, the interactions with the surface and the path back through the atmosphere to the sensor. A model inversion delivers the surface spectral reflectance (MODTRAN interrogation technique [Verhoef & Bach 2007]). This way, the reflectance of each pixel can be calculated. Reflectance is a property of the surface that is independent of the atmospheric condition and allows for comparison between different images and hence for the analysis of time series.

The data processing and LAI retrieval was done for all three years at 20 m resolution. In 2014, for which the sampling points are available, it was also carried out at 5 m, which is the resolution of the RapidEye Level 3A Product.

For studying the impact of multi-sensoral data on yield estimation it would be desirable to have two complete coverages of each sensor type following the whole growing cycle. This, of course, is not realistic (see Table 1). RapidEye and Landsat vary not only in their spectral configuration but also in spatial resolution. In order to mimic these differences...
and to allow for a similar temporal LAI monitoring [Hank et al. 2013], an alternative approach was chosen. The “Landsat 20m” based assimilation uses spatially degraded RapidEye scenes from 22 May and 19 July to fill temporal gaps at the beginning and peak of the LAI development. The “RapidEye 5m” based assimilation is complemented with Landsat derived LAI values during crop maturity (7 Sep and 9 Oct). Table 2 lists the data used for the respective yield estimations.

Table 2: Satellite data used for 5m and 20m yield estimation; light green indicates that these dates have been used for inter-sensor comparison (RE= RapidEye, OLI = Landsat 8 Operational Land Imager, AZA= Acquisition Zenith Angle, GSD = Ground Sampling Distance of original satellite image, DOY=Day of Year)

<table>
<thead>
<tr>
<th>Date in 2014</th>
<th>DOY</th>
<th>Sensor</th>
<th>AZA[°]</th>
<th>GSD [m]</th>
<th>“Land sat 20 m”</th>
<th>“RapidEye 5 m”</th>
</tr>
</thead>
<tbody>
<tr>
<td>22 May</td>
<td>142</td>
<td>RE</td>
<td>2.95</td>
<td>6.5</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>10 June</td>
<td>161</td>
<td>OLI</td>
<td>Nadir</td>
<td>30</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>11 June</td>
<td>162</td>
<td>RE</td>
<td>9.75</td>
<td>6.5</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>19 June</td>
<td>170</td>
<td>OLI</td>
<td>Nadir</td>
<td>30</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>19 July</td>
<td>200</td>
<td>RE</td>
<td>6.42</td>
<td>6.5</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>2 Aug</td>
<td>214</td>
<td>RE</td>
<td>0.27</td>
<td>6.5</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>6 Aug</td>
<td>218</td>
<td>OLI</td>
<td>Nadir</td>
<td>30</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>7 Sept.</td>
<td>250</td>
<td>OLI</td>
<td>Nadir</td>
<td>30</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>9 Oct.</td>
<td>250</td>
<td>OLI</td>
<td>Nadir</td>
<td>30</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

Figure 2 on the right-hand side lists the required input parameters. On the left-hand side it shows the four radiation fluxes considered in the SLC model. The parametrization of the SLC model takes the Spectral-Response-Function of the sensors into account. Therefore, it can be adapted to any sensor according to the sensor’s spectral and geometrical properties [Migdall et al. 2009]. The sun-observer geometry is recorded during the image acquisition and therefore is known. The soil reflectance and its variation with moisture are described by a soil BRDF (bi-directional reflectance distribution function) sub-model, based on the soil model by Haípe, 1981. Some of the leaf and canopy parameters are assumed to be constant for one crop type or within one specific phenological stage and are either obtained from literature or were determined using hyperspectral and in-situ data. The remaining parameters, which are highly variable (e.g. LAI, chlorophyll), can be retrieved by model inversion using the RMS error between the simulated and the measured spectra as criterion for the best fit [Migdall et al. 2009].

3.3 The Crop Growth Model PROMET

While optical remote sensing data can retrieve accurate information on the developed leaf area, it cannot see the absolute biomass or its distribution into the different plant compartments. It can definitely not directly observe the root biomass, which in case of sugar beets makes up the actual yield.

Therefore, the green leaf area serves as spatially distributed input for crop growth and yield modelling with PROMET [Mauser & Bach 2009, Hank et al. 2015]. This multi-physics, raster-based model calculates crop growth based on the physiological processes in the plant, including the distribution of biomass into the different plant organs (roots, stem, leaves and fruit) at different phenological stages. The model calculates the plant growth in hourly time-steps for the whole growing period, using background data such as a Digital Terrain Model and Soil Maps as well as up-to-date meteorological data. The model generates in an ensemble mode different scenarios for varying soil conditions [Hank et al. 2015].

Green LAI maps, retrieved with the SLC model, are used to find the scenario that fits the current growth conditions best. For this, the LAI maps are assimilated into the plant growth simulation as raster data sets. Small-scale soil variations due to e.g. different water holding capacity, which cannot be included in the more generalized background data, will thus be considered in the modelling [Migdall et al. 2009]. Since the unknown spatial heterogeneity of soil conditions is considered a major cause for in-field variations of plant development, the assimilation of remotely sensed data into the model improves the model outcomes significantly [Hank et al. 2015]. The concept of assimilating multi-sensoral EO data into PROMET is shown in Figure 3.
4. RESULTS

Based on the pre-processed data, the inversion of the SLC model was applied to retrieve the green LAI. Then, the green LAI was assimilated into the crop growth model at 5m and 20m resolution to estimate yield. The following section shows the results.

4.1 Impact of different Sensors on Plant Parameter Retrieval

Figure 4 shows the results for the green LAI retrieval for all available scenes in 2014 over the whole sugar beet growing period. Two different effects can be observed in this figure. For one, the LAI time series derived from different sensors is consistent. There is no large deviation due to sensor characteristics, only small differences of about 0.15 LAI occur at Day of Year 218. Secondly, true temporal coverage of the crop growing period is only possible through the combination of multi-sensor data. From the four RapidEye acquisitions, just two are at about the same time as the Landsat data. For these data pairs, quantitative comparisons of atmospherically corrected spectral signatures and retrieved LAI values are given.

Figure 5 shows the comparison of the average reflectance measured for sugar beet with Landsat 8 OLI and RapidEye.

While the visible bands of Landsat and RapidEye correspond to a large extent in their spectral response function (SRF), the NIR bands differ substantially [Blackbridge 2012, NASA 2015]. This can also be recognized in the shift of central wavelength of the NIR band in Figure 5. The challenge of the multi-sensor approach is to take these different spectral configurations for the plant parameter retrieval into account. The parametrization of the SLC model uses the SRF of the sensors [Migdall et al. 2009]. Thus, the influence of the spectral configurations should not lead to differences in the retrieved plant parameter. Whether those requirements are fulfilled is analysed by comparing the LAI derived from different sensors acquired during the same timeframe as described above.

Figure 6 shows the LAI retrieved with RapidEye compared to the LAI retrieved with Landsat 8 OLI for each 20m pixel. The discrete steps visible in this scatter plot are caused by the applied Look-Up-Table inversion. The steps of the tables can be recognized and the non-linear stepping is visible (smaller steps for lower LAI values where higher accuracies are targeted). The absolute values of LAI retrieval show a high congruency, as the gain with 0.97 is very close to 1. The RMSE between the two sensor retrievals amounts to 0.6 m²/m². The scattering of the values increases with increasing LAI values. This is caused by a saturation effect that occurs at very high LAI values and makes a distinction between LAI 5 and 6 much more difficult than between 2 and 3 [Bach et al. 2012].
Table 3 compares the averages of the retrieved LAI. The disagreement of the LAI is 0.4 and 0.1 for the two analysed dates. There is no systematic offset observable, as the deviation occurs in both directions.

<table>
<thead>
<tr>
<th>DOY</th>
<th>Plant Parameter</th>
<th>OLI</th>
<th>RE</th>
</tr>
</thead>
<tbody>
<tr>
<td>161/162</td>
<td>Green LAI</td>
<td>2.4</td>
<td>2.8</td>
</tr>
<tr>
<td>214/218</td>
<td>Green LAI</td>
<td>4.8</td>
<td>4.7</td>
</tr>
</tbody>
</table>

Table 3: Plant Parameter retrieved from OLI and RE

4.2 Multi-Sensor-based Yield Estimation for Sugar Beet 2012 – 2014

Sugar beet plant parameter retrieval was done for three consecutive years. The next question to answer is how accurate this multi-sensor approach is. In the years 2012 and 2013, only Landsat ETM+/OLI data was available, whereas in 2014 all three sensors have been used.

Figure 7 shows the mean LAI development of all three observed years. Sugar beet development has differed significantly in those years. The peaks of the LAI development are in 2012 and 2014 much earlier and higher in absolute value than in 2013. Furthermore, the increase of the LAI was slower in 2013. These differences are mainly caused by the weather conditions. In 2013, seeding took place almost one month later then in 2014 due to snow cover, rain and wet soil conditions. Extreme weather conditions were dominant during the whole year. A phase with heavy rainfall in June was followed by drought in July. Altogether, the growing season was four weeks shorter in 2013 than in 2014. In contrast, the conditions in 2014 were ideal. The seeding took place very early and the weather conditions were optimal for sugar beet growing during the whole season.

The varying LAI development leads to a different amount of accumulated biomass in the roots and thus to yield differences. This results in a very high modelled yield in 2014, where the LAI is constantly higher than in the other years. In contrast, the modelled yield in 2013 is very low which is expected due to the late start and slow LAI increase.

![Figure 7: Development of LAI in the years 2012 to 2014](image)

The varying LAI development leads to a different amount of accumulated biomass in the roots and thus to yield differences. This results in a very high modelled yield in 2014, where the LAI is constantly higher than in the other years. In contrast, the modelled yield in 2013 is very low which is expected due to the late start and slow LAI increase.

These modelled results are validated with available in-situ data as shown in figure 8. The different colours represent the three years. The linear regression, which is very close to the 1-to-1-line, proofs that the approach is very suitable for sugar beet yield estimation. Both, the absolute values of yield and variations on field level are well reproduced by the model. Not only the variation between different years but also the spatial variations between fields in each year are well represented. A gain value very close to one shows that there is no offset in the modelling results, neither in very low nor in very high yield ranges. This is also indicated by the low RMSE of 4.4 t/ha, which is only 4.5% of the yield mean over all 3 years.

![Figure 8: Yield validation on field level](image)

Concluding, it could be shown that with the multi-sensor approach it is possible to retrieve plant parameters and to model yield with a high accuracy on field level.

4.3 Validation of Modelled In-Field Heterogeneity

The validation of the multi-sensor yield modelling should not be limited to field averages, but also consider the accuracy of the modelling of the in-field heterogeneity of yield. Sampling points within the sugar beet fields were harvested by hand to assess the spatial distribution of the yield. This was done for two fields in 2014 and the collected data was used for validation of the modelled yield maps (see Figure 1). There is usually some loss of yield during harvesting with a root-crop harvester. This loss is calculated as 7% by comparing the mean of the sampling points with effective yield of the fields. Accordingly, the yield samples were multiplied with the factor 0.93 for comparison with the model results.

The LAI retrieval and yield modelling was performed at 5 m (“RapidEye like”) and at 20 m (“Landsat like”) resolution. The average yield of the “Landsat-like 20m” and “RapidEye-like 5m” yield results vary only slightly (see Table 4). This supports the conclusion that the multi-sensoral approach is reliable and produces comparable results.

![Figure 7: Development of LAI in the years 2012 to 2014](image)
Table 4: Results of the yield estimation for the “Landsat like 20m” and the “RapidEye like 5m” estimation

On the other hand, this approach also allows to compare the effect of sensor resolution in terms of spatial accuracy of yield. Figure 9 shows the yield maps in 20 m and 5 m resolution for one field in comparison with the sampling points for both fields. While the same overall structures can be seen in both resolutions, the additional detail in the 5m version is visible. In the 5 m yield map even row structure becomes visible.

Figure 10 shows the result of the validation for 20 m and Figure 11 for 5 m. The higher resolution could improve the coefficient of determination ($R^2$), which means that the spatial variance is better reproduced. But also it could be shown that with only the Landsat data small scale variability can be modelled with adequate results.

5. CONCLUSIONS

For precision farming applications, the management unit is not the whole field. Management is rather conducted on smaller units depending on the spatial distribution of site-characteristics and the working width of the machinery (e.g. 24 m or 36 m). With sugar beet, site-specific applications are not common yet (in opposition to e.g. wheat), because site-specific information about sugar beet growth and especially yield is hard to come by, as the main biomass is under ground and there is no technology for site-specific harvesting of sugar beets available on the market yet. Therefore, information derived from EO data and crop growth modelling is a new and exciting spatial data source for new site-specific sugar beet applications.

Yield estimation based on SLC and PROMET was successfully conducted for sugar beet during 3 consecutive years. It could be shown that the accuracy of the yield estimation is very high on field level. Additionally, the small-scale in-field variety is modelled with adequate results at 20 m raster size, but even
better results are achieved at a 5 m raster. The comparison of the LAI retrieval based on RapidEye and OLI shows that the SLC model and the data assimilation concept in PROMET is very suitable for the multi-sensor approach, since it is physically based and SLC takes the individual spectral configurations of the different sensors into account.

The demonstration that satellite data of variable spectral and spatial characteristics can be successfully used in crop yield estimation is of special importance, since using only one sensor often does not allow monitoring the LAI development very well. Thus, the multi-sensor approach can improve the accuracy of the yield estimation by increasing the number of assimilated LAI maps.

The presented methods can be used in an operational mode to support site-specific farming for sugar beet. The up-to date plant monitoring based on satellite imagery can be used for the daily assessment of the sugar beet crop. Thus, the occurrence of plant diseases, pests and other challenges can be detected early and the necessary measures can be conducted. Some phenological stages are very important for the vegetation development, e.g. the phase of row closure. This information can be provided spatially distributed. Additionally, the crop growth model delivers information on the root development, which is not observable from above. Using this information, the harvesting can be optimised in terms of logistics and time planning [Angermair & Bach 2015].

ACKNOWLEDGEMENTS

This work was supported by ESA within the IAP Demonstration Project TalkingFields, funded under contract number 4000100879/10/NL/US (ARTEMIS 20 Programme) and the EU FP7 project MELODIES (Grant Agreement 603525) coordinated by the University of Reading. Thank you to our TalkingFields partners Wolfgang Angermair and Stefan Burgstaller from PC Agrar GmbH for organising the field trial. Thanks to the Poschner-Bray’sche Güterverwaltung in Irlbach, namely Jürgen Schwarzensteiner, for the cooperation and providing validation data for our work. Thank you also to the Südzucker AG for supporting the field trial and harvesting of sampling points.

REFERENCES


