How Do Data Sources Affect the Accuracy of Data Fusion —— a comparison of MODIS products‘ performances in data fusion

Haiyin Ye, haiyin.ye@stud-mail.uni-wuerzburg.de
EAGLE Master Program innovation lab report

Introduction

Satellite datasets usually have an offset between temporal and spatial resolution. E.g. Landsat 8 has a high spatial resolution of 30m but a relatively long 16-day revisit cycle. After removing the images with high cloud coverage, there will be far less images of high quality in a given year (10 days in 2019, in this study). For MODIS, daily data is the trade off for a much lower spatial resoluaion at 250m or 500m. In order to merge the advantages of both, in another word, to improve data quality, data fusion is widely used between these 2 datasets. In this study, data fusion is conducted between Landsat 8 data and 5 different MODIS datasets separately and a comparison is made to find out which aspects of data sources might affect the fusion results in accuracy.

Study Area

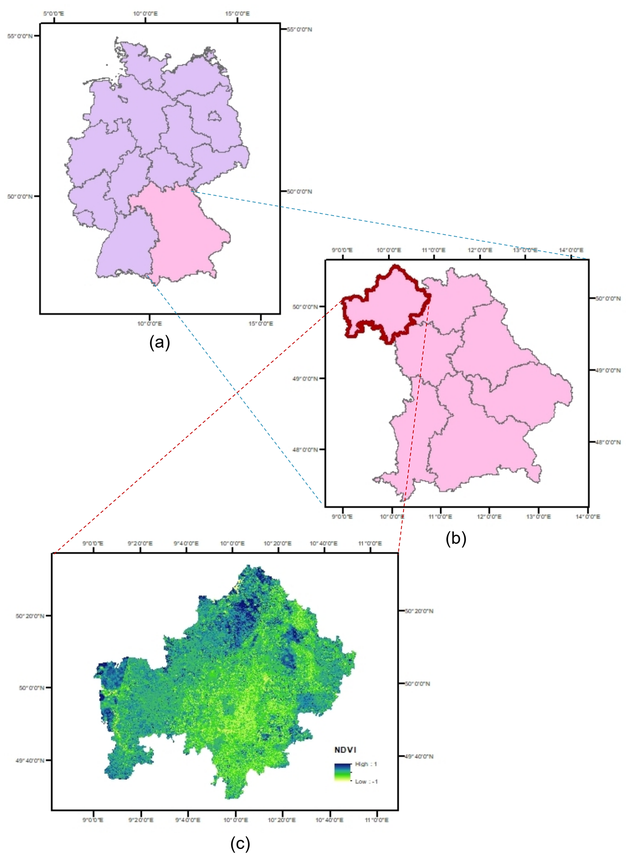

Figure 1. (a) Federal states in Germany; (b) Bavaria state; (c) NDVI map derived from Landsat 8 imagery on DOY 49, 2019, in Lower Frankonia, Bavaria, Germany.

Data

Year: 2019
High resolution imagery: Landsat 8, 30m res, on dates DOY49, DOY81, DOY113, DOY145, DOY177, DOY193, DOY209, DOY225, DOY241, DOY289.
Moderate resolution imagery: 4 MODIS products and 1 actual-acquisition-date-corrected [1] dataset, as listed in *Table 1*.
Landcover masks: Land cover classification in accuracy assessment is according to CLC(Corine Land Classification).

Table 1. MODIS Products selected.

Product	Topic	Temporal Resolution	Spatial Resolution
MCD43A4	Nadir BRDF	Daily	500m
MOD09GQ	Surface Reflectance	Daily	250m
MOD09Q1	Surface Reflectance	8-day	250m
MOD13Q1	Vegetation Index	16-day	250m
MOD13Q1_ADC	Acquisition Date Corrected	16-day	250m

Methodology

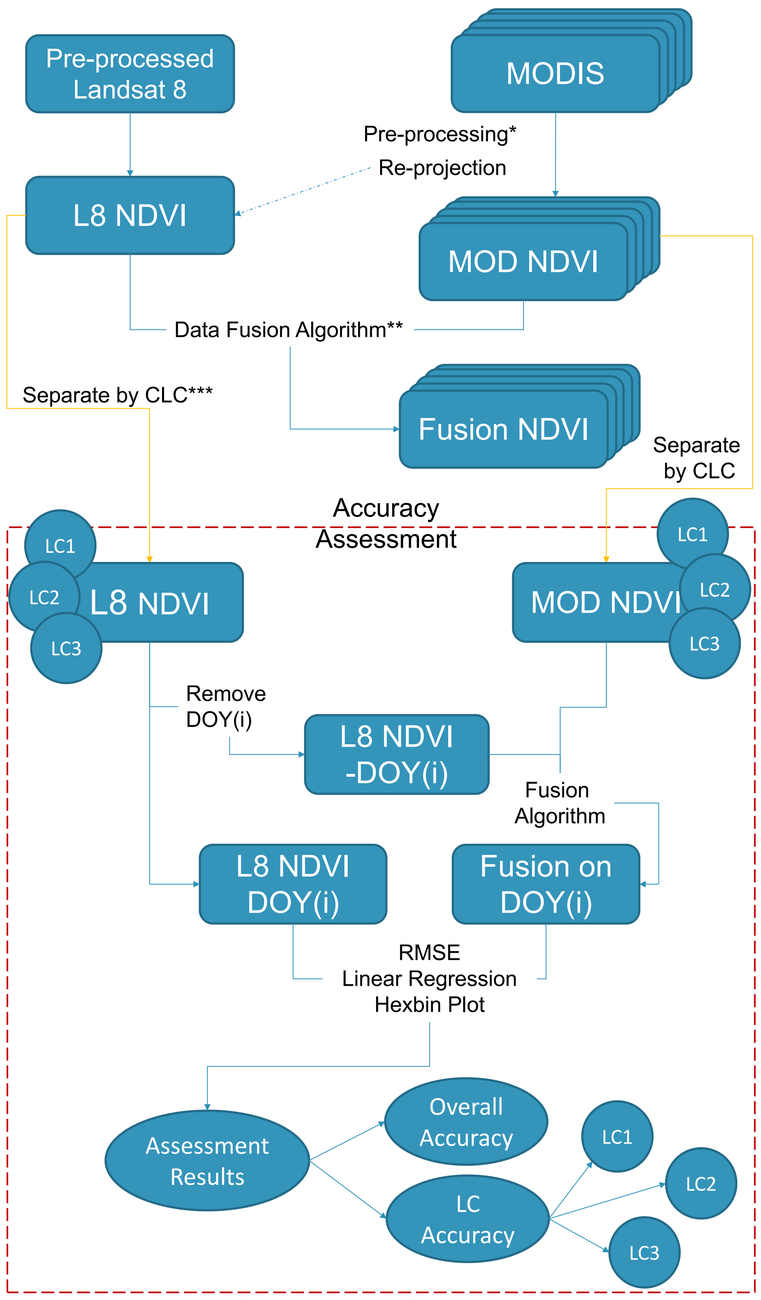

Chart 1. Work flow chart.

*Pre-processing includes cloud masking, NDVI calculation & data interpolation to fill in the gaps of cloud mask
**STARFM[2] is the data fusion algorithm applied in this study
***Datasets are separated according to Corine Land Classification [3] , LC1: aritificial surface, LC2: agriculture area, LC3: forest & seminatural surface.

Results

The overall performances of the 5 products are listed in *Table 2*. The acquisition-date-corrected MOD13Q1_ADC dataset performs the best in general. Hexbin plot comparison in Figure 2 demonstrates a significantly higher aggregation of pixels around 1:1 line in MOD13Q1_ADC result than in MOD09GQ result.

Table 2. Overall accuracy assessment.

Product	RMSE	R^2
MCD43A4	0.120	0.606
MOD09GQ	0.135	0.546
MOD09Q1	0.124	0.581
MOD13Q1	0.124	0.562
MOD13Q1_ADC	0.115	0.633

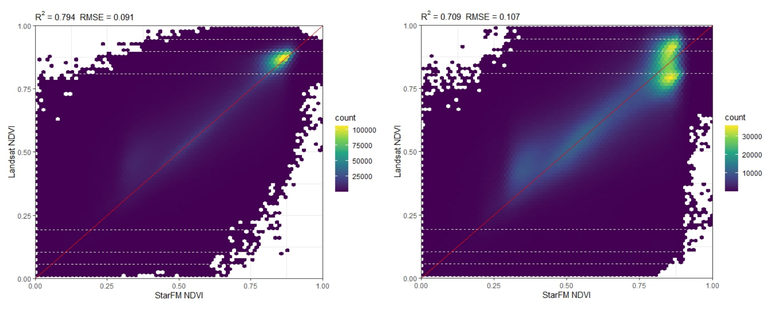

Figure 2. Hexbin comparison plots showing the accuracy of STARFM product on DOY193, with input moderate imagery (a) MOD13Q1_ADC; (b) MOD09GQ.

Hexbin plots compare pixels in STARFM and Landsat NDVI one by one and state all results by counting and coloring with density (yellow as high density, dark blue as low density). The more pixels around the 1:1 red line, the higher the accurarcy is.

Input datasets are separated into 3 land cover classes according to CLC (see *Chart 1*). Results of different land cover are shown in *Table 3, Table 4, Table 5,* separately. MOD13Q1_ADC remains in the first place in accuracy for all land cover.

Table 3. Artificial surface accuracy assessment.

Product	RMSE	R^2
MCD43A4	0.116	0.539
MOD09GQ	0.127	0.494
MOD09Q1	0.119	0.514
MOD13Q1	0.113	0.509
MOD13Q1_ADC	0.113	0.555

Table 4. Agricultural area accuracy assessment.

Product	RMSE	R^2
MCD43A4	0.134	0.475
MOD09GQ	0.132	0.468
MOD09Q1	0.130	0.450
MOD13Q1	0.141	0.441
MOD13Q1_ADC	0.128	0.502

Table 5. Forest & seminatural area accuracy assessment.

Product	RMSE	R^2
MCD43A4	0.094	0.273
MOD09GQ	0.106	0.213
MOD09Q1	0.109	0.212
MOD13Q1	0.122	0.145
MOD13Q1_ADC	0.089	0.325

Discussion

According to the results of this study, the following aspects of data sources might affect the accuracy of data fusion results:

1. Actual acquisition date correction
Accuracy assessment results of MOD13Q1 & MOD13Q1_ADC show that **actual acquisition date correction might increase the data fusion accuracy**. But whether it always helps, remains to be further tested with data in other years or with other datasets.

2. Spatial & temperal resolution
Among the 4 regular MODIS products, with which only data pre-processing (see *Chart 1*) is applied, the MODIS product **MCD43A4 with higher temporal resolution but lower spatial resolution performs relatively better than others**.
On the other hand, MOD09GQ performs the worst in this study. Significant large gaps in this dataset might be the main contribution to this result. This might have affected the results in two ways:
(1) MOD09GQ doesn't provide a valid QA data layer for cloud masking, so it was cloud-masked by the QA layer in MOD09GA. This might result in an over-masked output dataset.
(2) The gaps in datasets are fiiled in by interpolation in this study. Further studies on the quality of interpolation might be needed to discuss how this might have affected the final data fusion results.

3. Landcover types
All datasets perform the best in forest and seminatural areas and the worst in artificial surface areas. The tested datasets are NDVI data layers derived from Landsat and MODIS products, so they are better at depicting and differentiating vegetations. The idea of landcover separation is also based on that. Possible further studies could lie in a further separation of vegetated areas into second or third CLC levels.

Conclusion

In conclusion, data quality (spatial and temporal resolution, cloud cover rate) as well as certain treatment (e.g.actual acquisition date correction) applied to the source datasets, might affect the accuracy of data fusion results. The topic of the datasets (e.g. Vegetation Index) also influence the results in different land cover areas.

References

[1] S. Testa, E. C. B. Mondino, and C. Pedroli, "Correcting MODIS 16-day composite NDVI time- series with actual acquisition dates Correcting MODIS 16-day composite NDVI time-series," *Eur. J. Remote Sens.*, vol. 47, pp. 285–305, 2014.
[2]F. Gao, J. Masek, M. Schwaller, and F. Hall, "On the Blending of the Landsat and MODIS Surface Reflectance : Predicting Daily Landsat Surface Reflectance," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 8, pp. 2207–2218, 2006.
[3] Corine Land Cover classification data [2018]. Retrieved from Copernicus Land Monitoring Service [accessed on Nov.23, 2020], processed by ESA.
[4] M. S. Dhillon, T. Dahms, C. Kuebert-flock, E. Borg, C. Conrad, and T. Ullmann, "Modelling Crop Biomass from Synthetic Remote Sensing Time Series : Example for the DEMMIN Test Site , Germany," *Remote Sens.*, vol. 12, no. 1819, 2020.