Google Earth Engine Product Prototypes

Authors: Dominic Fawcett, Reik Leiterer, Holger Heisig, Hendrik Wulf, Tobias Kellenberger, Philip C. Joerg

Remote Sensing Laboratories
Geographisches Institut
Universität Zürich
Winterthurerstr. 190
8057 Zürich

www.npoc.ch
npoc@geo.uzh.ch
April 2017

Introduction

The Swiss National Point of Contact for Satellite Images (NPOC) offers information on access and processing of Earth observation (EO) data as well as support and consulting towards their use to Swiss stakeholders. The derivation of EO prototypes and the process documentation is one of the main services.

An increasingly large suite of Earth observing satellites provide the possibility to monitor all Earth spheres at multiple scales. High frequencies of observation also allow monitoring very dynamic processes such as crop production in agriculture. However, the sheer number of datasets needed to generate reliable results represents a limitation in terms of processing time and data management. This is where Google Earth Engine (GEE) comes in. It is a cloud computing platform which enables scientists to perform large scale data analysis using Google's infrastructure, free of charge. For this project we wanted to assess and demonstrate the capabilities of GEE for generating products of interest for Swiss stakeholders. This report demonstrates two of the prototypes developed: the "Switzerland Live" series of products as well as an agricultural field-use/vegetation development classification.

1 Switzerland Live

1.1 About the prototype

Earth observation satellites are capturing the current state of the Earth and Switzerland with different time intervals. From these images a vast number of products can be derived, be it the status of vegetation, the extent of snow-cover, the moisture content of soils etc. Optical data has the drawback of cloud-cover obscuring surface information, therefore many acquisitions can only partly be used for the mentioned products and to cover a certain region, many acquisitions must be composited together.

"Switzerland Live" is a product which reflects the current state of the whole of Switzerland as recorded by the latest Landsat 8 (LS8) observations. Due to frequent cloud cover, the product is a patchwork of most recent cloud-free observations. The oldest commonly date back 2 months for regions that are often cloudy and not as frequently captured.

We generate two versions of the "Switzerland live" product. One is a top of atmosphere (TOA) reflectance composite while the other represents surface reflectance (SR). They each have their advantages, though the SR composite is the one which can be used best for further derived products. Finally, the potential for a combined Landsat 8 and Sentinel 2 (S2) product is assessed by forming a latest cloud-free composite of TOA reflectances of both sensors. The methodology for generating the products differs slightly and is elaborated below.

1.2 LS8 TOA reflectance product

At the core of the "Switzerland Live" product is the masking of cloudy pixels. This represents a challenge. Although clouds show a rather unique spectral signature, mixed pixels with only very thin layers of cloud or fog exist, making the differentiation hard. Often it is also a question of what level of fog or haze is acceptable as there will always be a tradeoff of newest versus best information. For this product we decided to accept slightly suboptimal pixels in favour of the newest information, stressing the "live" aspect. TOA reflectance does not include an atmospheric correction and cloud-masking is therefore performed manually, using weighted scores of certain spectral combinations. This cloud-score function was originally written by Matt Hancher of the GEE engineering team. The key attributes for cloud identification are:

1) High reflectance over all visible and infrared bands

2) Low temperature in the thermal band

3) Not snow, therefore lower NDSI (Normalized Difference Snow Index)

```
var rescale = function(img, exp, thresholds) {
	return img.expression(exp, {img: img})
		.subtract(thresholds[0]).divide(thresholds[1] - thresholds[0]);
};

function landsatCloudScore(img) {
	// compute cloudyness indicators and take the minimum of them
	var score = ee.Image(1.0);
	// bright in the blue band
	score = score.min(rescale(img, 'img.blue', [0.1, 0.3]));
	// bright in all visible bands
	score = score.min(rescale(img, 'img.red + img.green + img.blue', [0.2, 0.8]));
	// bright in all infrared bands
	score = score.min(
	rescale(img, 'img.nir + img.swir1 + img.swir2', [0.3, 0.8]));
	// cool in temperature
	score = score.where(img.select(['temp']).mask(),score.min(rescale(img, 'img.temp', [300, 290])));
	// not snow
	var ndsi = img.normalizedDifference(['green', 'swir1']);
	score = score.min(rescale(ndsi, 'img', [0.8, 0.6])).multiply(100).byte();
	score = score.lt(cloudThresh).rename('cloudMask');
	img = img.updateMask(img.mask().and(score));
	return img.addBands(score);
}
```

(Comment: With clouds masked, the issue of cloud and topography related shadows still remains!)

Additional, a layer containing the age of the dataset in reference to the current date is added. This is used for quality mosaicing and can be displayed per pixel as a pixel-age map. The executed steps are therefore adding a time and an NDSI band to the original collection, then performing the masking of cloud and shadow pixels and finally computing the quality mosaic based on pixel age.

```
var startDate = ee.Date('2016-01-01');
var endDate = ee.Date(Date.now());

var millitodays = 1.1574e-8;

var lasttime = ee.Number(endDate.millis())

var add_date = function(image) {
	var timestamp = ee.Image(lasttime).subtract(image.metadata('system:time_start')).multiply
	(millitodays).rename('daysold');
	var daysoldforQC = timestamp.multiply(-1).rename('daysoldforQC');
	return image.addBands(timestamp).addBands(daysoldforQC);
};

var collectionWithDates = collection.map(add_date).map(addNDSI);
var collectionMasked = collectionWithDates.map(maskL8);

var latestcloudfreepix= collectionMasked.qualityMosaic('daysoldforQC');
```

1.3 LS8 SR product

As the cfmask, a pixel quality information layer which is provided with the SR product by USGS, was found to be susceptible to labeling snow as cloud, a further method for differentiation was required. In literature, the NDSI is mentioned as ideal for just such applications[1]. The NDSI represents the normalized difference between green and SWIR bands which is maximized for snow targets. An NDSI band is therefore added to be utilized in the masking process.

```
// add NDSI band for better snow/cloud differentiation
var addNDSI = function(img){
	var NDSI = img.normalizedDifference(['B3','B6']).rename('NDSI');
	return img.addBands(NDSI);
}
```

In the masking function, clouds and shadows are masked. For both, the cfmask flags are used. Pixels which are flagged as cloud but have an NDSI above a threshold of 0.7 (as used here: http://landsat.usgs.gov/l8handbook_section4.php) are not masked as they are very likely snow pixels. Alternatively, cfmask cloud confidence metrics could also be used.

```
// masking function
var L8Cloudmask = function(image) {
	// select the Landsat8 QA band and mask all clouds determined by Fmask
	var cloudMask = image.select('cfmask').eq(4).not()
	// ignore cloudmask for likely snow pixels (NDSI greater than 0.7)
	var cloudMaskFin = image.select('NDSI').gt(0.7).or(cloudMask);
	// mask all shadows determined by Fmask
	var shadowMask = image.select('cfmask').eq(2).not()
	return cloudMaskFin.and(shadowMask);
};
```

[1] Hall, D. K., Riggs, G. A. & Salomonson, V. V. (1995). Development of methods for mapping global snow cover using moderate resolution imaging spectroradiometer data. *Remote sensing of Environment,* 54, 127-140.

1.4 Results

1.4.1 LS8 TOA reflectance product

Illustrated in Figure 1 is the result of the latest cloudfree pixel composite of LS8 TOA reflectance for 01.01.2017. Overall, only few patchy remnants of actual clouds are visible. However, the image is not entirely homogeneous over the whole of Switzerland. This is mainly due to varying visibility and thin fog layers over certain acquisitions. A slight blue tint can be attributed to aerosol scattering. As the TOA reflectance product does not include atmospheric correction which accounts for water vapor and aerosol content, this is to be expected. If surface reflectance data was used we'd expect a reduction of these effects.

Figure 1. Latest cloud-free LS8 TOA mosaic, snapshot from 01.01.2017.

The patterns in the resulting image of Figure 1 can be identified in the pixel-age map of Figure 2. Over parts of western and central Switzerland, older observation had to be used due to frequent cloud-cover. Another effect visible is the use of older data which seems to follow topography. On closer inspection these appear to be strongly illuminated snowy regions on slopes, which get identified as cloud and masked out. Due to the solar azimuth only varying slightly for LS8 acquisitions, certain regions will show topographic shadows which cannot be removed.

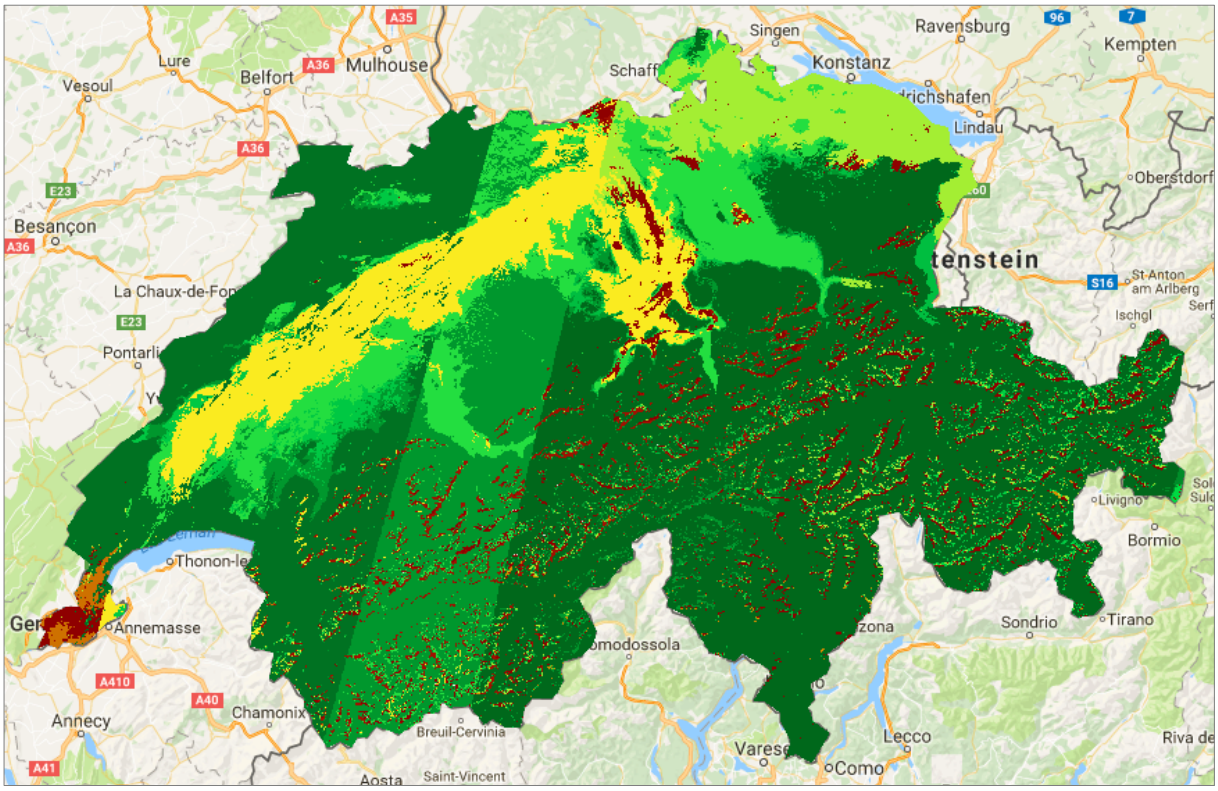

Figure 2. Age of latest cloud-free LS8 TOA pixel in days, snapshot from 01.01.2017.

The stripe clearly visible in Figure 2 is an effect of lack of overlap in LS8 acquisitions resulting in fewer datasets available for this region. The number of times a pixel is covered by an LS8 observation per year is illustrated in Figure 3.

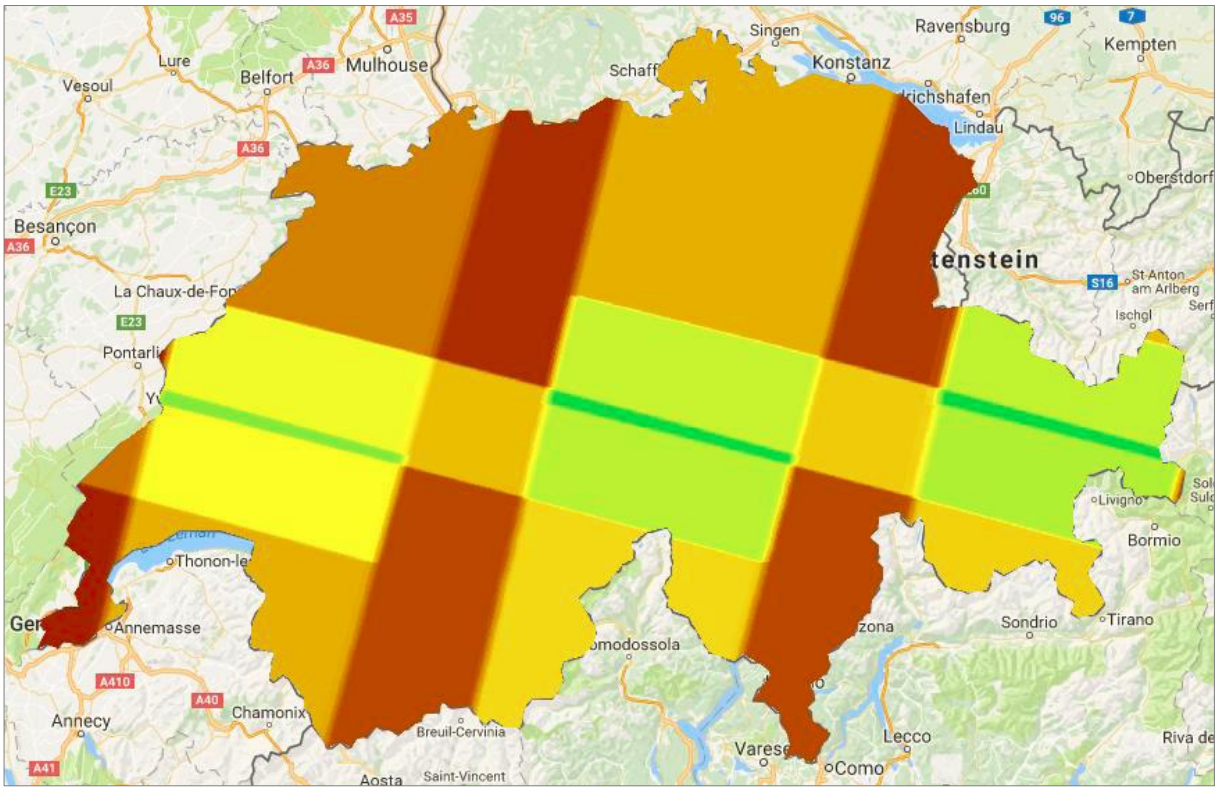

Figure 3. Number of times a pixel is covered by LS8 datasets in a year.

1.4.2 LS8 SR product

Visible in Figure 4 is the product based on LS8 SR data. It appears more homogenous than the TOA product, with some patches of remaining fog. Using cfmask, more data is masked out as cloud in general. This is visible when comparing the pixel-age maps (Figure 2 and Figure. 5). This product can provide a good basis for further processing, as long as the more extreme differences in time of acquisition per pixel are considered. For time-dependent products, observations which date back too far could be masked out. The NDSI threshold used in combination with the cfmask classification appears to do a better job identifying snow pixels. Many of the topography effects seen in Figure 2 are absent here.

Figure 4. Latest cloud-free LS8 SR mosaic, snapshot from 01.01.2017.

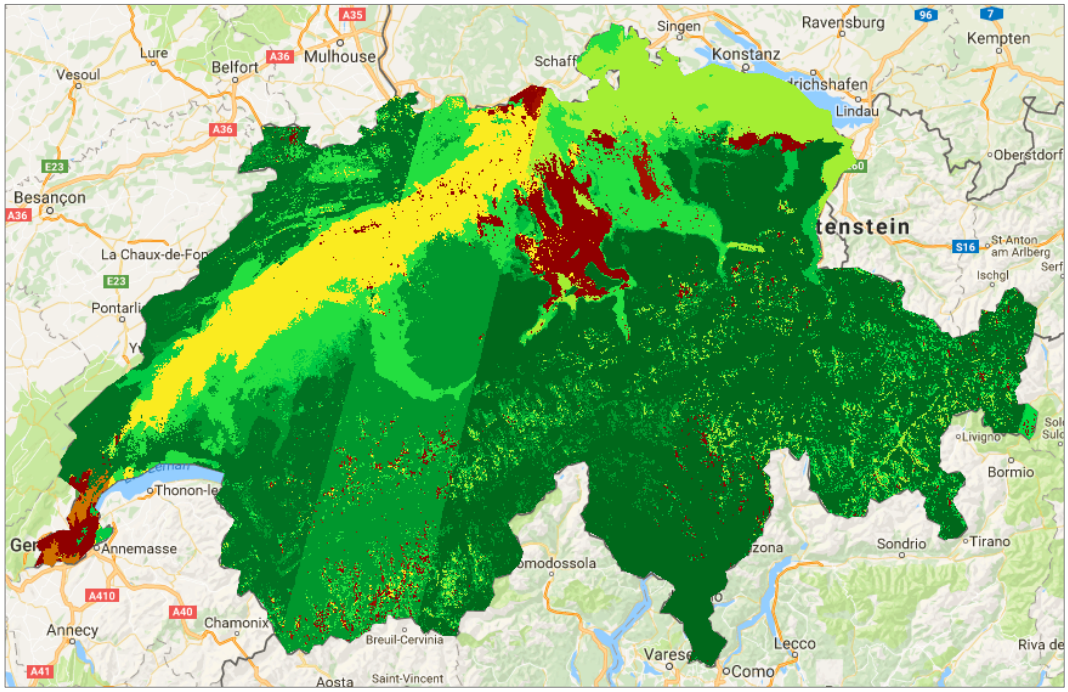

Figure 5. Age of latest cloud-free LS8 SR pixel in days, snapshot from 01.01.2017.

1.4.3 LS8 + S2 TOA reflectance product

Visually the TOA reflectance composite of both sensors differs only little from the LS8 composite at the large scale (Figure 6). It does appear slightly more homogenous where the increased number of observations thanks to S2 allowed capturing cloud-free conditions. The difference is clearer when comparing the pixel age maps (Figure 7). The large strip of older data due to lacking overlap is filled in with S2 observations as well as some smaller regions which were cloud covered in recent LS8 observations but not at the time of S2 overpass. Zoomed in, the patchwork of different resolutions becomes visible (LS8's 30m vs S2's 10m pixels), illustrating the benefit of using S2 data where possible. In LS8 and S2 compositing there are often methods employed which make use of S2 resolution to upscale LS8 data. This cannot be done here as we are interested in the most recent pixels only. Due to this, a resampling to LS8 resolution might be the best option for a homogenous product.

Figure 6. Latest cloud-free LS8 + S2 TOA reflectance mosaic, snapshot from 01.01.2017.

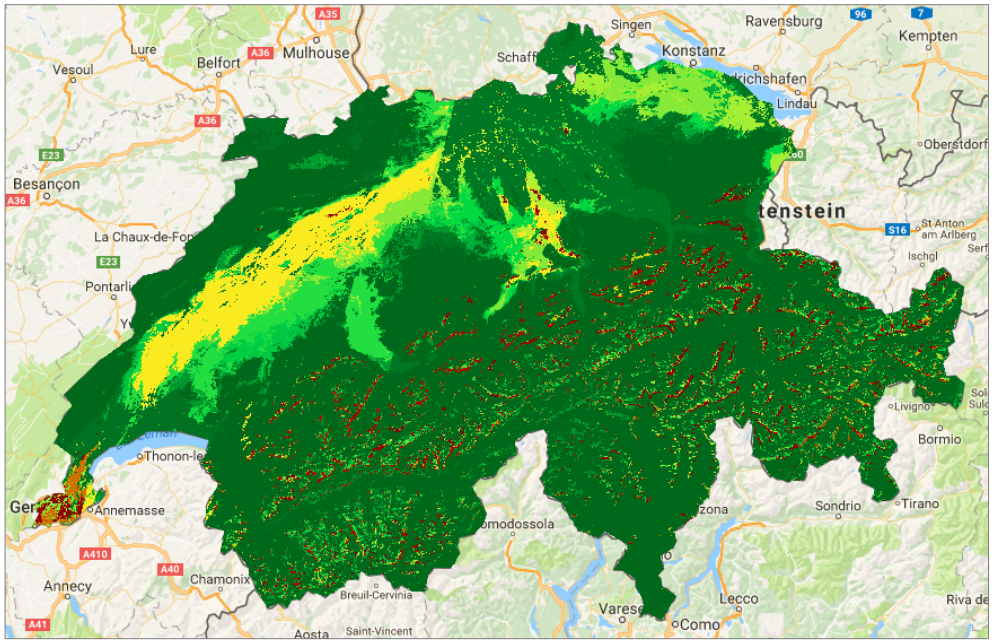

Figure 7. Age of latest cloud-free LS8 + S2 TOA reflectance pixel in days, snapshot from 01.01.2017.

1.5 Further development of the prototype

There are several ways in which the prototypes presented here can be further developed and improved, making them even more useful for applications. There is potential for improving the “Switzerland Live” products by addressing cloud-masking in a more sophisticated way. The cloud score function for TOA data could be further optimized for the Swiss situation, especially improving the snow-cloud differentiation. The cloud threshold would ideally be adaptive and not a fixed number as used here, depending on the situation.

Recently Sentinel 2b was launched which will start to produce useful data after its commissioning phase. We can expect this to further improve the actuality of the "Switzerland Live" TOA reflectance product. If the processing of S2 TOA reflectances to surface reflectances is performed on a large scale (ideally by ESA) in the future, the SR composite will also benefit from the higher temporal coverage and resolution offered by S2a and S2b.

1.6 Links to code

LS8 TOA latest cloud-free pixel:

https://code.earthengine.google.com/b231940afbb7105062ca1edc02c36773

LS8+S2 latest cloud-free pixel:

https://code.earthengine.google.com/42c8c3ddcfd2a02b83d1fbd6a9f82fd7

LS8 SR latest cloud-free pixel:

https://code.earthengine.google.com/56e81f1aed36d3091d42e1c20c3f651d

LS8 and S2 observation frequencies:

https://code.earthengine.google.com/88b9ff736d453c8fe11f4ce64575ac52

2 Swiss agriculture types

2.1 About the prototype

Frequent satellite observations of Switzerland not only allow the generation of very current products but also the monitoring of change over time. Time-series are often used to reveal changes in extent of land cover types. However, typical changes over a year also reveal useful information on the usage of areas, as is the case for agriculture. For this prototype, we seek to use NDVI time-series to firstly, identify regions which are likely used for agricultural purposes and secondly, to classify their use into easily identifiable categories. The categories we decided on are all related to vegetation growing seasons, the patterns of which can be retrieved from the time-series. These categories only allow limited interpretation but illustrate what differentiations between field usage are possible.

2.2 Agricultural cycle product

2.2.1 Generation of agriculture mask

Taking the mean of the NDVI standard deviation over summer months proved an effective method for separating agricultural fields from other landcover types such as built-up areas but also pasture and forest. The variability over areas used for agriculture appears to be particularly high as many popular crop types are harvested during this time. The result is a mask for agricultural fields which in general clearly follows field boundaries.

```
//calculate summer stdev per year
var stdvyearlysummer= ee.List([]);


for (var i=0; i<4; i++){
        var currentsummervals = collectionMasked2.filter(ee.Filter.calendarRange({start:2013+i,field:'year'}));
        var currentstdv = currentsummervals.select('NDVI').reduce(ee.Reducer.stdDev());
        stdvyearlysummer = stdvyearlysummer.add(currentstdv);
}
var stdvyearlysummercoll = ee.ImageCollection(stdvyearlysummer);
// mean over summer stdevs
var stdvsummermeans = stdvyearlysummercoll.mean();
// empirical threshold, should exclude forest and pasture
var stdvsummermeansagriculture = stdvsummermeans.updateMask(stdvsummermeans.gt(0.13));
// generate mask for further use
var summeragriculturemask = stdvsummermeans.gt(0.13);
// export for further use
print(summeragriculturemask)
Export.image.toAsset({image:ee.Image(summeragriculturemask),description:'agrimask',scale:30,region:swis
sBoundsNew,maxPixels: 200000000})
```

2.2.2 Harmonic fitting, local maxima identification and classification

Through strict masking of cloud, snow and shadow there are many gaps which emerge in the NDVI time-series. One method of interpolating these and describing the overall NDVI development is to fit a harmonic curve through the NDVI points available. The number of harmonics to be used can be specified. This allows a closer description of NDVI variability but increases processing time and the risk of bad fits where we have data gaps. The harmonic fitting is essentially a linear regression of the independent (harmonic components, time and constant) versus the dependent variable (NDVI). This procedure of fitting in the GEE was presented by Nick Clinton (Earth Engine developer relations) during the GEE 2016 User's Summit where details on the method can be found[2].
In a next step, local maxima in the fitted curve are identified. These are the peaks of the growing seasons. In order to eliminate insignificant peaks and fitting artifacts, only peaks higher than the mean NDVI of the pixel for this year are flagged as such. The peaks per pixel are then summed up and the time of the highest peak is retrieved. Based on this, the pixels are assigned to four distinct classes. The classes are defined as follows, along with illustrated examples of typical time series:

Class 1: Fields showing only one significant growing season

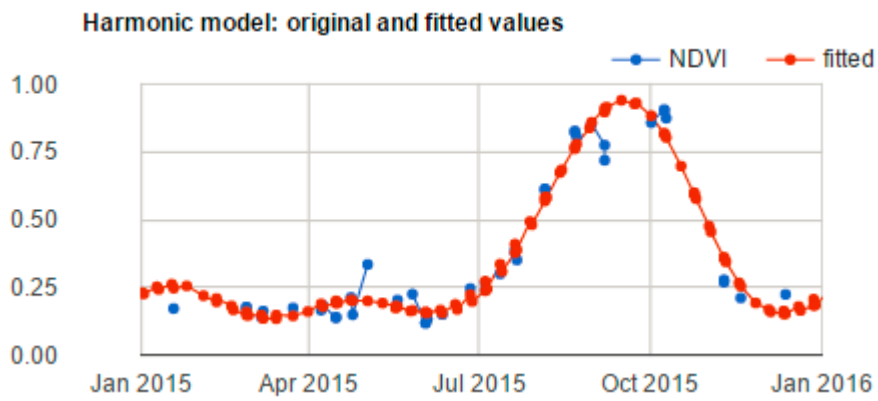

Figure 8. Example of «class 1» - pixel.

Class 2.1: Fields showing two growing seasons with the dominant season occurring before August 1st.

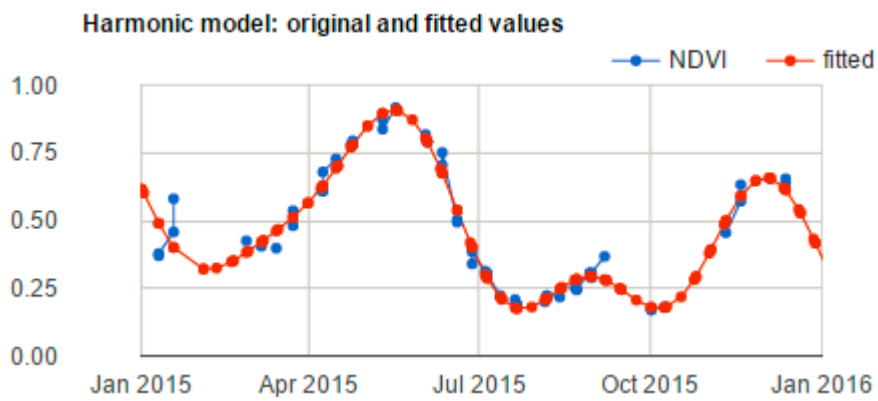

Figure 9. Example of «class 2.1» - pixel.

[2] https://docs.google.com/presentation/d/1J1rUtf-bkfaJwYJY-tU17kzKI4U8FnF7Q2_VWqWdaak/edit#slide=id.g494020760_291

Class 2.2: Fields showing two growing seasons with the dominant season occurring after August 1st.

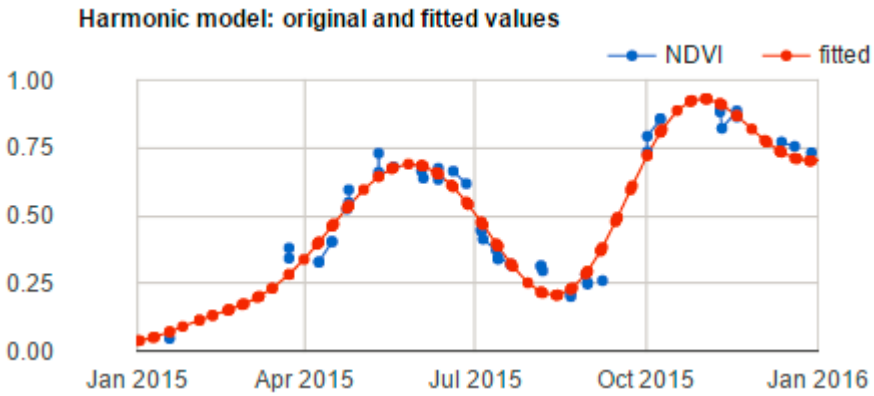

Figure 10. Example of «class 2.2» - pixel.

Class 3: Fields showing three or more peaks. Here the signal is complex and may include corner cases which would be attributed to «class 2». Typically these fields show vegetation activity over a longer time period.

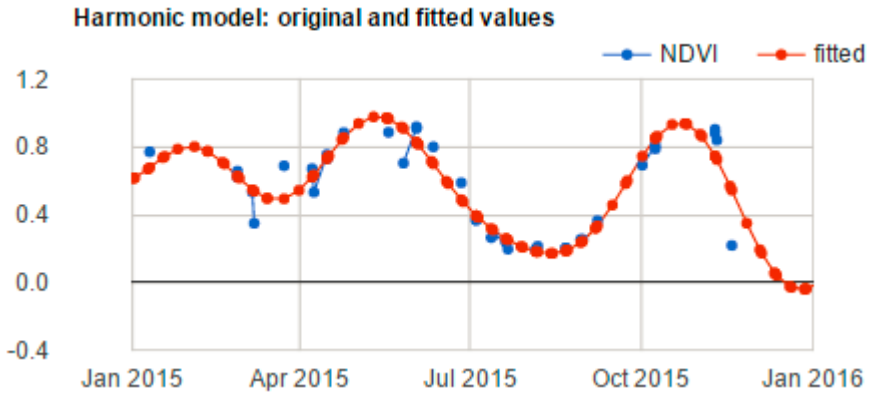

Figure 11. Example of «class 3» - pixel.

2.3 Results

The resulting map is presented in Figure 12, for a study area containing relatively large agricultural fields. As there was no validation data readily available, the product was only inspected visually. One indicator of classification quality and meaningfulness is the spatial contiguity of the classes. This is not given in all regions and appears to vary with the amount of data available. Data gaps can lead to large errors in the fitted NDVI curve and growing seasons which are artifacts of the fitting method. For our study area we observe mostly coherent fields despite pixel-wise classification, showing that our classes do allow a clear differentiation between field usages. The nature of these usages is however not easily identified and some advanced knowledge on agriculture is necessary to correctly interpret these results. For example, if a field is classified as «class 2.1» or «class 2.2» it does not mean that two growth and harvest cycles have occurred. Instead, many fields in Switzerland are planted with winter wheat, where last year's batch is harvested in spring or summer and the new batch is planted in autumn and shows a first germination, visible as a second peak. A possible second peak can also stem from the planting of crops for nitrogen fixation, e.g. clover.

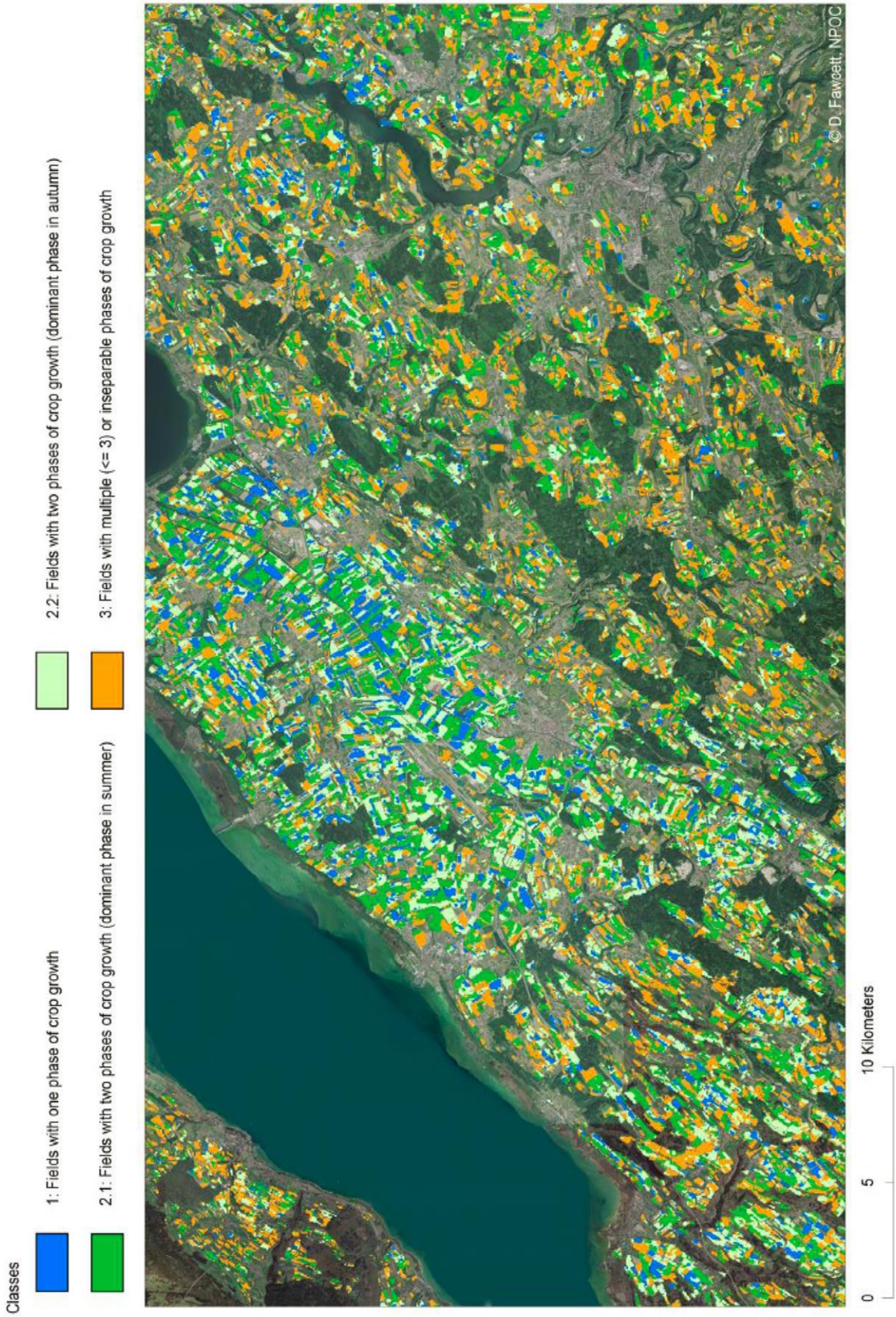

Figure 12. Classification result for a region of interest close to lake Neuchâtel.

2.4 Further development of the prototype

For this prototype we made use of only LS7 and LS8 data. Gap filling utilizing MODIS NDVI was discussed but ultimately was not employed due to the high resolution needed to resolve individual agricultural fields as well as differences in derived NDVI between instruments. S2 TOA reflectance based NDVI is not ideal due to the large expected atmospheric effects which could influence the time series. An interesting opportunity however lies in the possible use of Sentinel 1 data as radar backscatter intensity change over time may be a good indicator of field usage (growth and harvest should be clearly visible in intensity time series) and, importantly, not reliant on cloud-free observations.

The classes used in this example are not very intuitive. For better understanding and communication of results, the connection between time-series signature and typical crop planting cycles in Switzerland has to be identified. Once the dominating types of field usage have been defined, it is advised to use a more sophisticated classification technique by providing training data and independent validation in a next step. A number of common machine learning based classification methods are implemented in GEE for easy use, given adequate training and validation data are available. These types of methods could for example allow a further differentiation between crop types but possibly identify much more nuanced differences in field usage.

2.5 Links to code

Agriculture mask via NDVI variability:

https://code.earthengine.google.com/d1dc66e597dbdd1389af237a81ae4445