Task Definition

Satellite Images in Cloudy Weather

Exported on 11/01/2018

Table of Contents

1 Project attributes

Project name	Satellite Images in Cloudy Weather
Project case number	7699
Project task number	20
Project start date	01 Jan 2018
Project due date	31 Dec 2018
Project status	IN PROGRESS

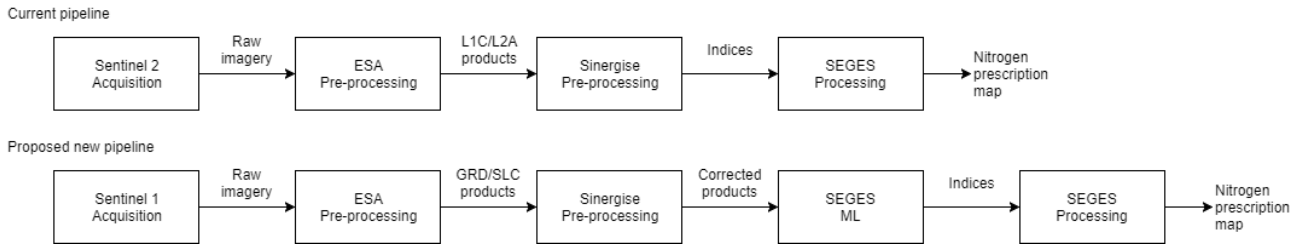
2 Stakeholders

Name	Organization
Jens Bligaard	SEGES
Rita Hørfarter	SEGES
Mikkel Mølgaard Mortensen	SEGES
Lars Pedersen	SEGES
Jacob Høxbroe Jeppesen	AU engineering
Jacob Dragsbæk	Landbrugsstyrelsen
Grega Milcinski	Sinergise
Rasmus Nyholm Jørgensen	AU engineering
Per Hejndorf	SEGES
Christian Schou Oxvig	SEGES
Unknown User (nicr)	SEGES
Peter Fogh	SEGES

3 Task description

In 2016 it became possible to analyze crop growth based on satellite images covering all of Denmark. The Sentinel 2 satellite acquires an image in the visible light spectrum of the entire country every 5'th day. Such images may be used to compute the green biomass of each field covering part of Denmark. The computed biomasses are displayed in the IT-systems CropSAT and CropManager which are extensively used by farmers and agricultural advisors. The only disadvantage is that Sentinel 2 cannot acquire useful images over areas with cloudy weather. A consequence of this is that only a small percentage of the images produced by Sentinel 2 in the spring of 2017 were useful for computing biomasses due to bad and cloudy weather.

In addition to the Sentinel 2 satellite, we also have access to radar images from the Sentinel 1 satellite. The quality of radar images is independent of cloudy weather. The goal of this project is to use machine learning to establish a connection between Sentinel 1 and Sentinel 2 images. It is expected that it is a possibility to find such a connection based on satellite images from at least two growing seasons. When a reliable connection has been found, the usefulness of CropSAT and CropManager no longer depends on the weather conditions.Essentially, this task is an investigation of the feasibility of replacing the current pipeline (see below) for computing nitrogen prescription maps with a new proposed pipeline (see below) based on Sentinel 1 imagery.



4 Tasks prioritization

4.1 Main tasks for the data science team

- **Data collection**:
  - The data science team must collect field polygons for at least 200 000 fields that have been stable (i.e. they have not changed their spatial form) for at least two consecutive growing seasons (2016, 2017).
  - The data science team must collect all available non-cloud-occluded hyperspectral images (all 12 raw L2A bands) from Sentinel 2 detailing all of the fields covered by the collected field polygons for the two growing seasons (2016, 2017). Each hyperspectral image (including all of its band) must be resampled to a 10m x 10m grid fixed for each field.
  - The data science team must collect Ground Range Detected (GRD) imagery from Sentinel 1 acquired at the nearest point in time and at same fields as the Sentinel 2 data. The GRD imagery must be resampled to fit the fixed grid used for Sentinel 2 images.
  - The data science team must compute and collect biomass index maps (i.e. NDVI, MSAVI2, and NDRE) for each non-occluded Sentinel 2 image.
- **Analyses**:
  - For each of the biomass index maps (NDVI, MSAVI2, and NDRE), the data science team must train one or more machine learning algorithm(s) to predict the biomass index map (on a 10m x 10m grid across each field) based solely on the corresponding Sentinel 1 GRD imagery.
- **Evaluation:**
  - Based on the analysis results, the data science team must evaluate the validity and performance of the proposed algorithm(s) in order to assess the feasibility of estimating the biomass index maps solely from the Sentinel 1 GRD imagery.
- **Report**:
  - Based on the data collection, analysis results, and evaluation, the data science team must write an IMRAD report covering the results.

4.2 Prioritized additional tasks (if time permits)

1. The data science team may include other data than satellite images (e.g. crop type or harvest date).
2. The data science team may include Sentinel 2 times series data as input to the prediction.
3. The data science team may train one or more machine learning algorithm(s) to predict aggregated values (mean, std dev, min/max) of a field, since these are used in inter-field comparison in CropManager.
4. The data science team may experiment with resampling of input or output to different grid resolutions with the goal of improving performance.

4.3 Not doing

- The data science team does not combine the satellite data with weather data and assess if it improves the results of the machine learning algorithms.
- The data science team does not include Sentinel 1 SLC imagery in the analysis.
- The data science team does not train a machine learning algorithm for predicting the prescription maps directly.

4.4 Risks and blockers

- **The required Sentinel 1 data is currently unavailable in the necessary preprocessed format (corrected products) and for the required timespan (only available from May 1st, 2017 and onward).**
  - **The required corrected GRD products are currently only available via the experimental EOCloud platform. Sinergise expect to make these products available via SentinelHub at some point.**
  - **Alternatively, AU is expected to release the source code for Fieldbabel "soon", which may potentially be used as an alternative provider of corrected GRD data (instead of SentinelHub).**
- **If neither SentinelHub or Fieldbabel provide the necessary corrected Sentinel 1 data and the EOCloud platform is too unstable, we must use ESA's archive.**

4.5 Time estimation

As seen in the table below, we expect that we will use all of the allocated hours to solve the task defined in this document.

Time	Hours
Allocated time	857
Scoping time (already spent)	230
Estimated time needed for project management (meetings, task refinements, etc.) (to be spent)	130
Estimated time needed for solving the task (to be spent)	500
Remaining time	0

4.6 Description of end product

This is a proof-of-concept of a machine learning system to relate Sentinel-1 GRD imagery to biomass index maps. If the system turns out to produce useful results, it is expected that a final version of the system is to be used in CropSAT and CropManager.

4.7 Current state-of-the-art

4.7.1 Current usage of SentinelHub within SEGES

SentinelHub is currently used in connection with the nitrogen prescription maps shown in CropManager. The DMDB database [NdviHub] is also used as part of this.

The most important details in this current use of SentinelHub are:

- An NDVI map is computed based on the L2A data product.
- A custom cloud mask is used.
- Only fields that have a zero percentage cloud coverage is used for computing prescription maps.

- When requesting data from SentinelHub, 50km x 50km tiles at 10m x10m resolution are used.

4.7.2 Definitions of biomass/vegetation indices

There exist many different vegetation indices (see https://www.indexdatabase.de/ for an overview). In this project, we focus on three indices, defined as follows:

- Normalized Difference Vegetation Index (NDVI)[1]: $\frac{(NIR - R)}{(NIR + R)}$
- Modified Soil Adjusted Vegetation Index 2 (MSAVI2)[2]: $\frac{(2 \cdot NIR + 1 - \sqrt{{(2 \cdot NIR + 1)}^2 - 8 \cdot (NIR - R)}}{2}$
- Normalized Difference Red Edge Index (NDRE)[3]: $\frac{(NIR – RE)}{(NIR + RE)}$

where

- Near Infra-red (NIR) is band 8 of Sentinel 2,
- Red (R) is band 4 of Sentinel 2, and
- Red Edge (RE) is band 5 of Sentinel 2.

Note that MSAVI2 is derived from SAVI and MSAVI which is defined as follows:

- Soil Adjusted Vegetation Index (SAVI)[4]: $\frac{(NIR - R)(1 + k)}{(NIR + R + k)}$ where $k=0$ means high vegetation cover $k=1$ means low vegetation cover. Sentinel hub[5] use k=0.5
- Modified Soil Adjusted Vegetation Index (MSAVI)[6]: $\frac{(NIR - R)(1 + k)}{(NIR + R + k)}$ where $k = 1 - \frac{2 \cdot s \cdot (NIR - R) \cdot (NIR - s \cdot RED)}{(NIR + R)}$

4.7.3 Other applications of Sentinel 1 GRD data

AgriFish and DHI Gras have done a preliminary study [1] on the use of Sentinel 1 + 2 data to identify mowed meadows and plowed catch crops. Their findings were that NDVI maps based on images from Sentinel 2 gave the highest number of correctly identified mowed meadows and plowed catch crops. They also concluded that it was possible to use Sentinel 1 SAR backscatter values (GRD imagery) to identify mowed meadows and plowed catch crops, though this method gave less certain results than the NDVI based methods. The best results, though, were obtained by using a combination of NDVI values and SAR backscatter values.

FieldSense is currently using Sentinel-2 data to identify inter-field crop stress problems [2]. It is based on various machine learning algorithms for detection of crop stress which includes a feedback mechanism for feedback from the farmer. Also includes assistance on nitrogen application, treatments, and optimal harvesting times.

4.7.4 Other works on the subject of SAR to optical image transcoding

Several works already propose solutions to the SAR (Sentinel 1) to optical (Sentinel 2) image conversion. They are all based on the idea of using (conditional) generative adversarial nets (GANs) [Goodfellow2014]/[Mirza2014] for learning a loss function for image to image mapping as popularized by the pix2pix[7] library detailed in [Isola2017] in combination with a convolutional neural network (CNN) generator based on the U-Net architecture [RonneBerger2015]. These works (ordered by relevance) are:

1 http://www.hiphen-plant.com/plant-phenotyping/products/normalized-difference-vegetation-index-ndvi_46.html
2 http://wiki.landscapetoolbox.org/doku.php/remote_sensing_methods:modified_soil-adjusted_vegetation_index
3 http://www.hiphen-plant.com/plant-phenotyping/normalized-difference-red-edge-index-ndre_51.html
4 http://wiki.landscapetoolbox.org/doku.php/remote_sensing_methods:soil-adjusted_vegetation_index
5 https://www.sentinel-hub.com/develop/documentation/eo_products/Sentinel2EOproducts
6 http://wiki.landscapetoolbox.org/doku.php/remote_sensing_methods:modified_soil-adjusted_vegetation_index
7 https://github.com/phillipi/pix2pix

1. [Bermudez2018]: An attempt at training a GAN with a U-Net based generator for Sentinel 1 to Landsat 8 optical images (all bands) transcoding. This work applies real image pair patches of crop fields and evaluates the performance visually and based on the a crop classification using a random forest classifier.
2. [Grohnfeldt2018]: An attempt at training a GAN with a U-Net based generator for inferring a non-cloud-occluded Sentinel 2 image (all bands) from a cloud occluded Sentinel 2 image (the 10 medium and high resolution bands) and a Sentinel 1 image. The experimental study is based on real image pairs but with artificially created clouds. Performance is evaluated based on RMSE and SAM of the true Sentinel 2 image to the inferred one.
3. [Ley2018]: An attempt at training a GAN with a U-Net based generator for Sentinel 1 (SLC products) to Sentinel 2 maps that are used as a proxy in Sentinel 1 based land classification. The performance is evaluated using the land classification accuracy.
4. [Wang2018]: An attempt at training a GAN with a CNN based generator for SAR despeckling and colorization.
5. [Schmitt2018]: A description of a large set of matching Sentinel 1 and Sentinel 2 image patches with motivating remote sensing deep learning applications. The work described in [Grohnfeldt2018] is based on these image patches.

4.7.5 General background on remote sensing and deep learning

A review the use of data fusion, i.e. combining data from different sensors, in remote sensing is given in [Schmitt2016]. A review of deep learning ideas and applications in remote sensing is given in [Zhu2017]. A lightweight introduction to the fundamentals of remote sensing is given in [3].

4.7.6 Reference list

For the named references, see the references database.

Number	Title	Author	Link
1	Sentinel 1 og 2 lige på kornet: Monitering af landbrugsaktiviteter - græsslåning og detektering af efterafgrøder med henblik på landbrugskontrol	Lotte Nyborg and Sanne Eskesen	6_sentinel-1-2_agricultural_monitoring_sannelotte.pdf Sorry, the widget is not supported in this export. But you can reach it using the following URL: http://youtube.com/watch?v=hF6xSRyb-F4

Number	Title	Author	Link
2	Satellitovervågning af afgrøder - Opdag problemer i dine afgrøder i tide og forøg din indtjening	John Smedegaard	https://site.fieldsenseapp.com/ Sorry, the widget is not supported in this export. But you can reach it using the following URL: https://www.youtube.com/watch?v=Ymi9Md19G5A
3	Fundamentals of Remote Sensing	The Canada Centre for Mapping and Earth Observation	https://www.nrcan.gc.ca/node/9309 

4.8 Current related SEGES projects

Project name	Relation to this project
CropSAT	Uses Sentinel 2 images to build nitrogen application maps from biomass estimation (currently NVDI but is being replaced by MSAVI2)
CropManager	Uses Sentinel 2 images to build nitrogen application maps from biomass estimation (currently NVDI but is being replaced by MSAVI2)
Perceptive Sentinel	A collaboration with Sinegise on building algorithms for processing of satellite data for use in agricultural applications.
Big Data in Crop Growing	Also based on a raster model of field related data.
Future Cropping	Various precision agriculture elements, e.g. the yield map cleaning and the soil water capacity study.

5 Tentative time schedule

This project is part of the time schedule of the Data Science team.

6 Data description

Data from the Sentinel satellites is freely available in a partially pre-processed form from ESA's Copernicus Open Access Hub (https://scihub.copernicus.eu/). However, most of this data need further pre-processing (e.g. cloud removal, orthorectification, resampling, etc.) to be useful in the processing pipelines depicted above. Sinergise provides the SentinelHub platform that offers easy access to Sentinel data for which such pre-processing has been done. Specifically, as of February 2018, SentinelHub offers:

- Data from the Sentinel 2 satellites in the necessary formats.
- An early access to partially pre-processed data from the Sentinel 1 satellites. This is still too incomplete to be useful in the proposed new processing pipeline depicted above.

6.1 Data sources

Source name	Source format	Access	Storage and usage restrictions	Additional information
SentinelHub Sentinel 2	Lossless: png, tiff, raw / Lossy: jpeg 8-bit, 16-bit ints or 32-bit floats https://www.sentinel-hub.com/develop/documentation/api/output-formats	sentinel-hub.com OGC API: https://www.sentinel-hub.com/develop/documentation/api/ogc_api, https://www.sentinel-hub.com/develop/capabilities/configuration_utility	Load limitations: • do not query too large area; 5000x5000px is the limit • when using statistical API, do it for up to 3 months at once	Available products: • Color composites (combinations of color bands) • Indices (e.g. NDVI) • Raw bands (B1, B2, B3, B4, B5, B6, B7, B8, B8a, B9, B10, B11, B12) • User scriptable custom products (including e.g. band combinations and custom cloud removal) https://www.sentinel-hub.com/develop/documentation/eo_products/Sentinel2EOproducts

Source name	Source format	Access	Storage and usage restrictions	Additional information
			• when using statistical API over larger areas, do not query with full resolution - it's statistics anyway, so using e.g. 100m resolution will provide comparable results with much faster performance	

Source name	Source format	Access	Storage and usage restrictions	Additional information
			• if you plan to run several processes in parallel, e.g. more than 5 concurrent requests, let Sinergise know in advance, so that they can estimate potential infrastructure requirements.	

Source name	Source format	Access	Storage and usage restrictions	Additional information
SentinelHub Sentinel 1	Lossless: png, tiff, raw / Lossy: jpeg 8-bit, 16-bit ints or 32-bit floats https://www.sentinel-hub.com/develop/documentation/api/output-formats	Currently available via either EOCloud (https://eocloud.cloudferro.com/knowledgebase.php?action=displayarticle&id=10), Amazon AWS (http://sentinel-pds.s3-website.eu-central-1.amazonaws.com/), or SentinelHub (https://apps.sentinel-hub.com/configurator). Sinergise does not plan to support SLC products. Further information: https://www.sentinel-hub.com/develop/documentation/eo_products/Sentinel1EOproducts	AWS Freely available within EU. EOCloud data can only be accessed from within EOCloud VMs (at some price). SentinelHub freely available for SEGES until 2020. Same limitations as for Sentinel 2 data from SentinelHub.	AWS restrictions: • Rolling window data one year back, starting 1st of May 2017. • Only GRD products. • Only IW and EW beam modes. • Only VV+VH, HH + HV, VV, HH polarizations. • Only medium and high sensing resolutions. • Only Gamma0 backscatter coefficient. • Images are not orthorectified. • Not all combinations of the above are necessarily available. EOCloud restrictions: • Only GRD products. • Only IW and EW beam modes. • Only VV+VH polarizations for IW. • Only HH+VH and HH polarizations for EW.

Source name	Source format	Access	Storage and usage restrictions	Additional information
				• Only High sensing resolution (10m/px) for IW. • Only Medium resolution (40m/px) for EW. • Only Gamma0 backscatter coefficient. SentinelHub restrictions: • The same as the AWS restrictions since the SentinelHub offs the GRD data from AWS (https://medium.com/sentinel-hub/sentinel-1-data-available-on-aws-and-through-sentinel-hub-ecc51c910db6)
AU Fieldbabel	TBD	https://fieldbabel.eng.au.dk/ or the underlying source code, once it has been publicly released.		The intent is to provide both Single Look Complex (SLC), and Ground Range Detected (GRD) SAR data.

Source name	Source format	Access	Storage and usage restrictions	Additional information
Google Earth Engine	Custom API	Currently in open testing: https://developers.google.com/earth-engine/	Full Sentinel 1+2+3 archive	https://developers.google.com/earth-engine/datasets/catalog/sentinel/ Detailed in [Gorelick2017].

6.2 Data attribute information SentinalHub

Note that when fetching data, we must use the file format MimeType.TIFF_d32f, because it returns the original data.

Data source	Configuration dashboard	Configuration Name	Layer ID	Source	Additional information

All Sentinel 2 bands from Sentinel Hub	https://apps.sentinel-hub.com/configurator/#/	SIICW instance	S2-L2A	Sentinel 2 L2A	**Layer processing script**:
					`// Returns all bands of Sentinel 2 L2A`
					`// as definere here: https://www.sentinel-hub.com/develop/documentation/eo_products/Sentinel2EOproducts`
					`return [B01, B02, B03, B04, B05, B06, B07, B08, B8A, B09, B11, B12];`

All Sentinel 1 "bands" from EOCloud/ Sentinel Hub	http://apps.eocloud.sentinel-hub.com/wms-configurator/	SIICW instance	S1_GRD_IW	Sentinel 1 IW GRD	**Layer processing script**: `return [VV, VH];` GRD IW for VV, VH polarizations.

6.3 Additional information

This project is part of the PAF project: "Øget konkurrencekraft i landbruget gennem brug af kunstig intelligens." - Ansøgning til promilleafgiftsfonden for landbrug 2018.pdf