# A dataset of spectral and biophysical measurements over Russian wheat fields

Allard de Wit[1*], Gerbert Roerink[1], Oleg Virchenko[2], Alexander Kleschenko[2], Sergey Bartalev[3], Igor Savin[4], Dmitry Plotnikov[3], Pierre Defourny[5], Raphael d'Andrimont[5]

[1] Wageningen Environmental Research, Wageningen, The Netherlands
[2] State Enterprise National Scientific & Research Institute on Agricultural Meteorology, Russian Federation
[3] Space Research Institute of Russian Academy of Sciences, Russian Federation
[4] Dokuchaev Soil Science Institute, Moscow, Russian Federation
[5] Université Catholique de Louvain, Louvain-la-Neuve, Belgium.
* e-mail: Allard.dewit@wur.nl

**Abstract:** From 2011 to 2013 the MOCCCASIN project (MOnitoring Crops in Continental Climates through Asimillation of Satellite INformation) was carried out financed by the European Commission 7th Framework Programme. During the project, two field campaigns (2011 and 2012) were carried out at two sites (Odoyev and Plavsk) in the Tula region of Russia. During these two campaigns, observations were made at selected winter-wheat fields consisting of phenological stage, biomass samples, hemispherical photographs, spectral properties of the canopy and the soil as well as ancillary information about the field. Meteorological observations from synoptic and agrometeorological stations were collected from the stations in and surrounding the Tula region. Finally, a large trajectory throughout the whole Tula region was surveyed in order to collect fields with different crop types.
**Keywords:** wheat, Russia, remote sensing, experimental data

**1 BACKGROUND AND ORIGINAL PURPOSE:** From 2011 to 2013 the MOCCCASIN project (MOnitoring Crops in Continental Climates through Asimillation of Satellite INformation) was carried out financed by the European Commission 7th Framework Programme. The overall objectives of the project were to improve the monitoring and prediction of yield for winter-wheat in Russia by combining several technologies: 1) satellite-based rapid with-in season mapping of winter-wheat fields, 2) deriving biophysical crop variables (Leaf Area Index, Fractional Cover and Fraction absorbed PAR) from low resolution satellite data (e.g. TERRA-MODIS) for the fields identified and 3) the simulation of wheat growth using the WOFOST crop simulation model with assimilation of satellite observations.

During the project, two field campaigns (2011 and 2012) were carried out at two sites (Odoyev and Plavsk) in the Tula region of Russia. During these two campaigns, observations were made at selected winter-wheat fields consisting of phenological stage, biomass samples, hemispherical photographs, spectral properties of the canopy and the soil as well as ancillary information about the field. Finally, a large trajectory throughout the whole Tula region was surveyed in order to collect fields with different crop types.

The field dataset collected had several purposes. First of all, the accuracy of the maps with the satellite identified winter-wheat fields was determined. Second, the algorithms for deriving crop biophysical variables from satellites need to be trained and validated using field observed data. Finally, the WOFOST crop simulation model needs to be calibrated for the specific Russian wheat varieties.

## 2 FIELD CAMPAIGN AND DATA COLLECTED

### 2.1 LOCATION AND NUMBER OF VISITS
The field campaigns focussed on two sites near the villages of Plavsk and Odoyev which are located 200 km south of Moscow. Due to the start of the project in January 2011, the 2011 field campaign started late with the first field visit being carried on 17 May 2011. The consequence was that the first part of the growing season is missing. A total of 11 field visits was carried out at roughly 10-day intervals with the last visit on 2011-07-22 for the Plavsk site and 2011-07-28 for the Odoyev site. Nevertheless, field visits were carried out alternating between the two sites and therefore 5 visits to Odoyev and 6 to Plavsk were carried out.

For the 2012 growing season the field campaign was started on time with the first field visits already before the winter on 14 October (Odoyev) and 21 October (Plavsk). If logistics and weather allowed, the

field visits were done on a weekly basis thereby alternating the visits to the two test sites. So in practice field observations for each site were made in fortnightly (two weeks) intervals.

## 2.2 DATA COLLECTED

Several types of data were collected during the field visits. First of all crop biomass samples were made using a procedure based on the The Guide on the Observation on Agro-meteorological Stations (Roshidromet, 2000 - in Russian). For delineating the sample area, a 50x50cm frame was used and one sample per field was taken. The sample location was determined visually based on the similarity of the selected place to the general plant state on the field. Furthermore, additional parameters were recorded including phenological phase, the presence of weeds, the average height of the plants and the plant density on a square meter.

Second, downward images were made of the wheat canopy using digital SLR camera's equipped with hemispherical lenses. For the field campaigns, who cameras were used which were both GPS enabled ensuring that geographic coordinates were stored in the images itself. The location of the observation within each plot was approximately the same for the whole vegetation period. Images were taken started from one corner of the field plot and moving to opposite corner.

Third, spectral measurements were made with a HandHeld2 portable spectrometer with a wavelength range of 325-1075 nm with a spectral Resolution <3.0 nm at 700 nm. The measurements were conducted with the arm pointed horizontally towards the sun with the spectrometer pointed downwards. The spectrometer was calibrated using the spectralon white reference before starting measurements for each field and more often if conditions are changing fast (strong changes in solar radiation due to clouds). The height of the spectrometer was near 1.5 m from the crop canopy.

A number of measurements were made on one point when the crop canopy is heterogeneous. Usually a bare soil spectrum was taken at each field as well, but which was not marked explicitly. However, usually the bare soil spectra can be readily identified by the vast difference between the bare soil and the vegetation spectra. Spectra were taken started from one corner of the field plot and moving to opposite corner

Finally, for validating the winter-wheat masks, it was necessary to have a dataset with fields with known crop type within the 2012 cropping season. For this purpose, personnel drove a transect by car through the Tula region, thereby observing the crop types along the roads and making pictures of the fields in order to geolocate the fields using the camera GPS. Afterward these observations were connected to field boundaries yielding a trajectory of fields with known crop types through the Tula region. In total 158 fields were visited and mapped.

## 2.3 SUPPORTING AND ANCILLARY DATA

Besides the data collected by the project team itself, several ancillary data layers are available. Vector datasets with field boundaries of all fields in the Odoyev and Plavsk site are available and field boundaries have been uniquely labelled in order to connect observations to specific fields in a relational database. Labelling was done separately for the 2011 and 2012 campaigns as the crop types change from year to year.

Next, meteorological and agro-meteorological observations were collected from 20 weather stations in and surrounding the Tula oblast. The observations from these stations cover the period 2006 to 2012. The synoptic observations were taken at time intervals of 12, 8, 6, 3, or 1 hour depending on the type of station and include temperature, precipitation, cloudiness, wind speed and sunshine duration. Synoptic observations have been processed in order to obtain daily values.

Agrometeorological observations were carried out at 10 day intervals and the reports include agronomic and meteorological observations that are relevant for the crops grown in a the region such as soil temperature, information on winter conditions (snow depth, frost kill events), crop management and damage as well as crop details (number of tillers, grain mass).

**3 PROCESSING OF DATA**

The biomass samples were processed at the Russian Agrometeorological institute and consisted of the following steps:
- Measuring the weight of the fresh biomass (including weeds) with an accuracy below 0.1 gram;
- Drying of the wet biomass in the desiccator  until all water was removed from the plant material (at least 20 hours at 105 degrees Celsius);
- Measuring the weight of dry plants;
- Separating the grains from the other plant materials (only after Anthesis);
- Measuring the dry weight of grains.

The results are stored in a relational database together with the other elements that were recorded during the field visits. **Table 1** shows the table structure that we as used to store the results.

| Column name | Description | Units |
|---|---|---|
| LocalFieldId | Local field Identifier | |
| Crop | Description of crop type | |
| Site | Test site, e.g. Plavsk or Odoyev | |
| Field_ID | Site field Identifier | |
| tField_ID | Tula field Identifier | |
| dDate | Day of measurement | |
| FrameNumber | Frame number of the measurement (e.g. multiple samples in one field) | |
| PhenologicPhase | Description of the phenologic phase | |
| WeedsLevel | Description of weed infection level | |
| Height | Height of the crop | Cm |
| Density | Number of plants per square meter | |
| WaterContentBiomas | % of water of total biomass, calculated as (wet-dry)/dry * 100 | % |
| DryBiomassPlot | weight of total dry biomass in the plot | $g/m^2$ |
| DryBiomassField | weight of total dry biomass on the field | 100 kg/ha |
| waterContentGrainmass | % of water of grains, calculated as wet/dry mass | % |
| DryGrainmassPlot | weight of total dry grain mass in the plot | $g/m^2$ |
| DryGrainmassField | weight of total dry grain mass on the field | 100 kg/ha |

**Table 1.** Structure and contents of the results from the field sampling of crop biomass

Canopy characteristics (LAI, FPAR and FCover) of the wheat fields were estimated from the images obtained with digital SLR cameras with hemispherical lenses. For this purpose we processed the images with the CanEye software (Weiss et al. 2004), available from https://www4.paca.inra.fr/can-eye. For the processing we used CanEye version 6.39 (date of executable 21/Sep/2012).

CanEye requires a classification of the image into green and brown image elements which was done based on visual interpretation of the hemispherical images according to the CanEye manual. This classification of images is fairly time-consuming and due to time limitations, not all images have been processed with CanEye. The following images have been processed:
- For the 2011 campaign, all images have been processed into estimates of green area index, FAPAR and fraction coverage;
- For the 2012 campaign, images were processed for a selection of the fields. These fields were selected because they has a good temporal coverage in terms of camera images and crop biomass samples. Moreover,  the final yield figures showed that these fields had normal to above normal yields.

Regarding the spectral measurements, the most important step in processing of the spectra was the conversion of the raw spectra to absolute reflectance. This step was carried out by calibrating the reflectance of the spectralon reference panel that was supplied with the spectrometer, against a panel with known absolute reflectance. This calibration was done in the laboratory before the instrument was brought to Russia in April 2011 as well as when the instrument was returned in December 2011.

After comparing the calibration measurements it was found that the reference spectralon panel that was brought to the field had become dirty during the field visits resulting in a considerable lower reflectance
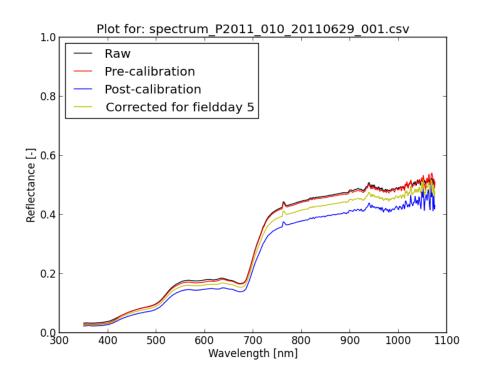
that was used as a reference value during the field visits. This implies that the measurements would gradually start to overestimate the true reflectance during the course of the field campaign.

We tried to correct for this problem by assuming that the reference panel gradually got dirtier with each field visit and we could correct by linear interpolation between the pre-calibration constants (before the panel went to Russia) and the post-calibration constants (after the panel was returned). The CSV files containing the spectrometer measurements therefore contain 5 columns:

1. Wavelength: The wavelength at which the measurement took place;
2. Raw_reflectance: the raw reflectance directly coming from the instrument;
3. Pre_reflectance: the raw reflectance correct for the whiteness of the reference panel using the correction factors determined during the pre-calibration;
4. Post_reflectance: the raw reflectance correct for the whiteness of the reference panel using the correction factors determined during the post-calibration;
5. Corrected_reflectance: The reflected calculated by interpolating between the pre_reflectance and the post_reflectance based on the number of field visits.

Figure 1 gives an example of the reflectance measurements from the spectrometer and the corrections that were applied to it.

**Figure 1.** Example of a spectrum with corrections applied.



For the measurements starting in April 2012, a different solution was used: the spectralon reference panel was stored in a container that could be opened to do the calibration. After the calibration, the container could be closed to avoid that the panel was becoming dirty. Therefore the influence of the degrading of the reference panel is not (or to a very limited extent) applicable on the spectral measurements collected in 2012.

## 4 STRUCTURE OF THE MOCCCASIN DATASET

The total size of the MOCCCASIN data is around 18.5 Gb which is mainly due to the size of the hemispherical photographs. The data has been ordered in a folder structure as is visualized in figure 2. The top folder contains a PDF document with a more detailed description of the field campaigns and characteristics of the data.

Below the top folder there are two folders related to the 2011 and 2012 campaigns, a 'misc' folder with miscellaneous data and finally a folder with SQLite databases contain (agro)meteorological data as well as a PDF document with a detailed description of the meteorological data.

Within the folders related to the field campaigns (2011 and 2012) several subfolders can be found:
- The folder 'CanEyeResults' contains Microsoft Excel files storing the estimates of canopy properties derived from the hemispherical images with CanEye;
- The folder 'CropCutting' contains an SQLite database with the results from the biomass sampling (table 1). The folder 'Fields' contains ESRI shapefiles and KMZ (compressed Keyhole Markup Language) with the outline of all fields in the Odoyev and Plavsk test sites. Each polygon has its unique FieldID that is used in excel sheets and SQLite tables to reference a particular field;
- The folder 'Fisheye' contains the hemispherical images organized by site, FieldID and date. All hemispherical images are stored as ordinary JPG files with unique file names and containing EXIF tags with details including local time, exposure and camera type but also GPS coordinates and GPS time in UTC. Additionally, thumbnails of each file are available starting with a 'T' in the file name;
- The folder 'spectrometer' contains the spectral measurements as CSV files and PNG quicklooks. These files are also organized by site, FieldID and date;
- Finally, only the 2012 campaign folder contains a subfolder 'transectValidationFields' that contains an ESRI shapefile with field polygons with associated crop type. A PDF map of the shape is also available in that folder.
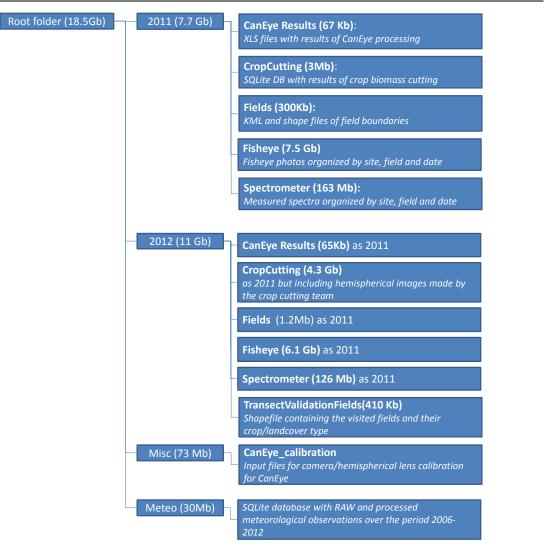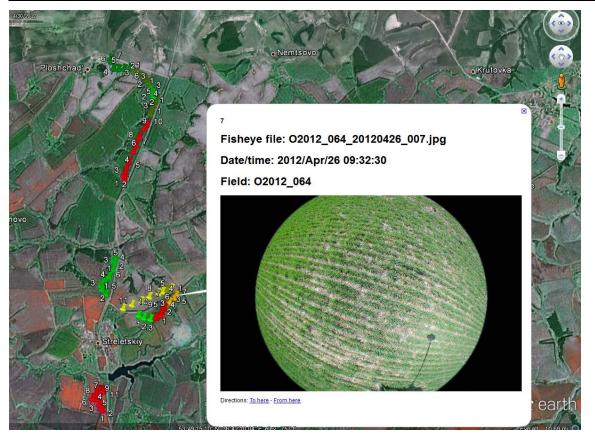
---

**Figure 2.** *Structure of the MOCCCASIN dataset.*



Because of the size of the dataset, it can be difficult to get an overview and to find a particular set of images or spectra for a given field or date. To improve accessibility, each subfolder within the spectrometer and fisheye folders also contains a KML file that can be loaded in Google Earth which allows to visualize the locations, dates (using the date slider) and thumbnails of the spectra and

hemispherical images made (Figure 3). For viewing the KML files it is needed to run the `start_webserver.bat` script which starts a local webserver serving the image files needed by Google Earth. It is assumed that a python3 interpreter is installed and its location is include in the system PATH.

**Figure 3.** Screenshot of the visualization of the field visits in Google Earth. The thumbnail can be displayed by selecting a particular pin on the map.



## 5 ACKNOWLEDGEMENTS

## REFERENCES

Weiss, M., Baret, F., Smith, G.J. and Jonckheere, I., 2004. Methods for in situ leaf area index measurement, part II: from gap fraction to leaf area index: retrieval methods and sampling strategies. *Agric. For. Meteorol., 121: 17-53*.

Roshidromet, 2000. Guide to hydro-meteorological stations and posts. Edition 11. Agro-meteorological observation on stations and posts. Part 1. Basic observation. Book 1., Management directive.  RD 52.33.217-99. Moscow, 2000, 348 p.  (in Russian).