HIGH-RESOLUTION IRRIGATION DETECTION WITH REGIONAL CROP MODELING AND DATA ASSIMILATION

A Data Management Plan created using DMPonline.be

Creator: Louise Busschaert

Affiliation: KU Leuven (KUL)

Funder: Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

Template: FWO DMP (Flemish Standard DMP)

Grant number / URL: 1158423N

ID: 193991

Start date: 01-11-2022

End date: 31-10-2026

Project abstract:
Irrigation is becoming an increasingly important component of the hydrological cycle in many regions across Europe and the globe. However, information on the location and amounts of water application is not easily obtained. Soil-vegetation models often poorly represent irrigation, but satellite signals should detect irrigation. Consequently, misfits between modeled soil moisture and vegetation simulations and observations could be used to locate irrigated areas. Recently, the plot-scale AquaCrop model was adapted to provide soil moisture and vegetation estimations on a regional scale, with the possibility to activate irrigation. In HIDRO-LAND, three methods will be explored to detect irrigation based on the fits and misfits between satellite observations and model simulations with and without irrigation. First, the irrigation detection can be based on deviations between AquaCrop and microwave-based satellite estimates of soil moisture and vegetation optical depth. Second, by using machine learning as an observation operator, backscatter can be simulated based on AquaCrop moisture and vegetation estimates. Analysing the correspondence between backscatter simulated with AquaCrop and observed with Sentinel-1 allows to mark irrigated areas. Finally, vegetation or water additions in a cycling Sentinel-1 backscatter data assimilation system can be indicative of where and when irrigation was applied, and at the same time improve simulations.

Last modified: 17-01-2023

HIGH-RESOLUTION IRRIGATION DETECTION WITH REGIONAL CROP MODELING AND DATA ASSIMILATION

FWO DMP (Flemish Standard DMP)

1. RESEARCH DATA SUMMARY

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

Dataset Name	Description	New or reused	Digital or Physical	Digital Data Type	Digital Data format	Digital data volume (MB/GB/TB)
1. Sentinel-1 satellite data	European Space Agency (ESA) Sentinel-1 backscatter satellite data	Reused	Digital	Observational	netcdf	< 10 TB
2. MERRA2 forcings	Daily global meteorlogical data from the Modern-Era Retrospective analysis for Reasearch and Applications (version 2)	Reused	Digital	Compiled/aggregated data	netcdf	< 50 TB
3. ERA5 forcings	Daily global meteorological data from the European Center for Medium-Range Weather Forecasts	Reused	Digital	Compiled/aggregated data	netcdf	< 50 TB
4. Ancillary data for land surface modelling	model parameters	Reused	Digital	Compiled/aggregated data	binary files, netcdf	< 10 TB
5. In situ data	in situ data (irrigation, rainfall) collected over different pilot sites for the ESA IRRIGATION+ project	Reused	Digital	Observational	netcdf	< 100 GB
6. Simulation output	irrigation estimations at local and regional scale	Generate new data	Digital	Simulation data	netcdf	< 1 TB

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

All reused data can be found via the following links. The numbers refer to the datasets in the table above.

1. https://step.esa.int/main/download/snap-download/
2. https://goldsmr4.gesdisc.eosdis.nasa.gov/data/MERRA2/
3. https://cds.climate.copernicus.eu/
4. https://lis.gsfc.nasa.gov/
5. https://osf.io/db5qz/ (access required)

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.

- No

Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.

- No

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.

- No

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/Data transfer agreements/ research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- No

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- No

2. DOCUMENTATION AND METADATA

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g., in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, Codebook.tsv etc. where this information is recorded).

All generated model output is produced in a standardized and documented way, either with metadata available inside the files or via user documentation. All reused data is accompanied by a source and information on how the data was downloaded and processed.

Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify (where appropriate per dataset or data type) which metadata standard will be used. If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reuse.

- Yes

The generated model output and metadata are standardized, as described in in the user guide of the system that is used (NASA Land Information System: https://lis.gsfc.nasa.gov/releases/lisf-public-release-741).

3. DATA STORAGE & BACK-UP DURING THE RESEARCH PROJECT

Where will the data be stored?

All data will be stored on the VSC HPC large volume storage (automatic backup).

How will the data be backed up?

All data will be stored on the VSC HPC large volume storage (automatic backup)

Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely.
If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.

- Yes

Collaborative grant on VSC HPC Tier1 (Storage4Climate 2.0) and Tier2 (/staging, /archive).

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

Access to the HPC is password and key protected.

What are the expected costs for data storage and backup during the research project? How will these costs be covered?

~12 000 euro for four years the costs are covered by the Tier1 grant and through related projects within our group (ESA projects: IRRIGATION+ and 4DMED, FWO projects: CONSOLIDATION, PEATBURN, KU Leuven project: C1)

4. DATA PRESERVATION AFTER THE END OF THE RESEARCH PROJECT

Which data will be retained for at least five years (or longer, in agreement with other retention policies that are applicable) after the end of the project? In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

Final generated data used in publications will be preserved and made open-source through Zenodo.
Re-used data will not necessarily be preserved, unless they remain used or are no longer supplied by the original data providers.

Where will these data be archived (stored and curated for the long-term)?

VSC HPC (/archive)

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?

The final output of this project will be limited compared to all storage needed during the project, and is estimated at ~1000 euro for 5 years. The costs will be covered by future Belspo, FWO, other projects, and the promotor's basic working fund.

5. DATA SHARING AND REUSE

Will the data (or part of the data) be made available for reuse after/during the project? In the comment section please explain per dataset or data type which data will be made available.

- Yes, in an Open Access repository
- Final irrigation estimations

If access is restricted, please specify who will be able to access the data and under what conditions.

NA

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)? Please explain in the comment section per dataset or data type where appropriate.

- No

Where will the data be made available? If already known, please provide a repository per dataset or data type.

All data will be made available via Zenodo (https://zenodo.org/).

When will the data be made available?

Upon publication of the research results

Which data usage licenses are you going to provide? If none, please explain why.

All data will be shared under a Creative Commons Attribution license (CC-BY 4.0). The data can be used, but users have to acknowledge the original data creators.

Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.

- Yes

All datasets related to publications will have a doi.

What are the expected costs for data sharing? How will these costs be covered?

None

6. RESPONSIBILITIES

Who will manage data documentation and metadata during the research project?

Louise Busschaert (PhD candidate)

Who will manage data storage and backup during the research project?

Professor Gabrielle De Lannoy, head of the research group, together with the HPC system administrators

Who will manage data preservation and sharing?

Louise Busschaert, Gabrielle De Lannoy, Michel Bechtold and the HPC system administrators

Who will update and implement this DMP?

Louise Busschaert