D2.3 First release of community software package with processing and quality control algorithms

Version 1.0

Working Group	WG2
Deliverable	D2.3
Date	05/07/2023
Version	1.0
Author	Christian Chwala
Reviewer	Aart Overeem

Description	This document describes the OPENSENSE software sandbox which is a collection of current open-source software packages for OS data processing and which allows running example code in a one-click online environment.
Keywords	Repository, software, methods

About OPENSENSE (COST Action CA20136). *OPENSENSE brings together scientists investigating different opportunistic sensors (e.g. microwave links, citizen science), experts from weather services, and end-users of rainfall products to build a worldwide reference opportunistic sensing community. The overarching goals of the COST are to overcome key barriers preventing data exchange and acceptance as hydrometeorological observations, define standards to allow for large-scale benchmarking of opportunistic sensing precipitation products and develop new methods for precipitation retrieval, coordinate integration of the opportunistic observations into traditional monitoring networks, and identify potential new sources of precipitation observations. Further details can be found here:*

Funded by the European Union

www.cost.eu
www.opensenseaction.eu

Table of contents

Glossary

WG	Working Group
MoU	Memorandum of Understanding
OS	Opportunistic Sensors
CML	Commercial Microwave Link
PWS	Personal Weather Station
SML	Satellite Microwave Link

1. Overview of WG2 tasks and synchronisation with D2.3

This document reports on the official OpenSense deliverable D2.3, which is part of the activities carried out by WG2 during the grant period two (GP2) of the OpenSense action. WG2 takes care of the homogenization of processing methods and the corresponding software tools. Details about WG2 activities and deliverables are reported in the OpenSense MoU [1] and listed in Table 1.

D2.3, the first release of community software packages with processing and quality control algorithms, is the first important step with respect to joint software development, building on D2.1 from GP1. The training school on OS processing methods (D2.2) held in GP2 also has a tight integration with the work on D2.3, because it builds on the joint work on processing software. During GP3 and GP4 development will continue on the basis of D2.3 resulting in the final OS processing software package (D2.5).

Table 1. A timetable of WG2 activities and deliverables.

	year 1				year 2				year 3				year 4			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
WG2 - Activities for:																
Software development				d21			d23									d25
OS processing methods									d24							
PhD&ECI Training School					d22											

D2.1 - Guidelines for contributing to Git repository
D2.2 - Content for training school on OS processing published at website
D2.3 - First release of community software package with processing and quality control algorithms
D2.4 - Report documenting benchmark algorithms
D2.5 - New release of OS processing package

2. First final version of OPENSENSE software sandbox

After a joint discussion during WG2 meetings in GP1, it was decided to start with a collection of the existing software tools and provide executable examples for each package using open datasets. This decision was motivated by the fact that the maturity of the existing software packages was very heterogeneous, not allowing to explore and understand the functionalities and limitations of the different packages without cumbersome installation processes, different for

different operating systems. We thus configured a github repository to serve as an environment in which all software packages can run. Using the free service of mybinder.com this environment, which can run both R and Python code in parallel, can be used to spin up a cloud compute instance using repo2docker in which users can interactively explore the functionality of each software package. Within GP2 we added several example notebooks that make use of selected software packages. One of these examples provides the full processing of CML data from the OpenMRG dataset, from raw data to CML-derived rainfall maps. In addition, we developed a prototype of the code for downloading existing open OS datasets and automatically transforming them to the OPENSENSE data format standards that WG1 has developed. Figure 1 shows the landing page of the OPENSENSE software sandbox on github at https://github.com/OpenSenseAction/OPENSENSE_sandbox.

Currently the following packages and additional functionalities are included:

- CML processing packages
  - pycomlink: A Python package with a collection of different published processing method, mainly developed by the Karlsruhe Institute of Technology (KIT)
  - PyNNcml: A Python package collection of mainly deep learning-based methods, developed at Tel Aviv University
  - RAINLINK: An R package for CML rainfall retrieval and mapping (with the nearby-link approach for wet-dry classification), developed at Wageningen University & Research and the Royal Netherlands Meteorological Institute (KNMI)
- PWS processing packages
  - PWSQC: An R package for quality control of PWS data, developed at Wageningen University & Research and the Royal Netherlands Meteorological Institute (KNMI)
  - pws-pyqc: A Python package for quality control and bias correction of PWS data, developed at the University of Stuttgart
- SML processing packages:
  - Currently, only a basic example for loading and plotting SML data is provided, developed by Czech Technical University Prague
- Custom additions
  - Prototype of the “opensense_data_downloader_and_transformer” using existing open datasets and the WG1 data format standards

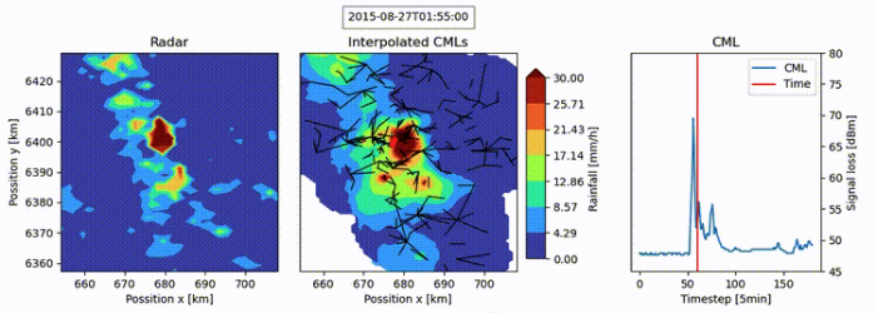
- Notebook that does one full processing of the OpenMRG CML dataset to produce the animation shown in the README of the repository

README.md

OpenSense

launch binder 👈 Click here to run the examples online. Please retry if the build process does not finish after some minutes.

One of the examples will walk you through a full CML processing from raw data provided by the OpenMRG dataset to a rainfall map animation like this one:



OPENSENSE software sandbox

The OPENSENSE software sandbox is a collection of software packages for processing data from opportunistic rainfall sensors, developed within the COST Action OPENSENSE.

The purpose of this repo is to showcase existing codebases and existing open datasets in one reproducible environment that can be run online via mybinder.

Not all notebooks are in a final state, but the OpenMRG usecase and data exploration notebooks are nicely prepared to give new users a quick overview of code and data that is available.

Run code online

This repository is inteded to be run online via the "launch binder" button above. Note that it can take some minutes to spin up the online environment. In some cases it might not start at all due to limited resources of the binder service. In that case, just try again some minutes later.

Fig 1.: The README file of the OPENSENSE sandbox with the animation of the CML map that is derived in one of the working example notebooks. The button to launch the online environment to interactively run all examples is on the top right.

In addition to the work on the sandbox, several new CML processing methods have been implemented in the package pycomlink during GP2. The main addition was the nearby-link approach for wet-dry classification from RAINLINK and the associated outlier filter to remove malfunctioning links, for which the translation to Python and optimised implementation was carried out during the STSM of Maximilian Graf (KIT) at Delft University of Technology (TU Delft). As a result, the core of the most commonly used CML processing methods is now available in pycomlink and WG2 will promote its usage as the common CML processing package. For SML and PWS processing, planning for joint software packages is underway.

3. Next steps

Based on the insights from making all packages executable in the online environment and based on the prototype that was developed for transforming existing open datasets according to the WG1 data format conventions, we started with a joint software package to combine central aspects of OS data processing. This package, for which a proof-of-concept exists, will provide the following functionalities:

- plot data on a map using webmap tiles, including a simple way to plot lines with a colormap
- find neighbouring or intersecting sensor data (e.g., to collocate CML observations with reference point observations from rain gauges or path-average observations from gridded radar rainfall products)
- do standard validation of results including a set of metrics and thresholds to facilitate a fair and uniform comparison of algorithms
- use xarray.Dataset as default data model throughout all functions allowing to carry metadata through all processing steps
- enforce the usage of the OPENSENSE data format conventions and provide a function to check whether input and output data comply with these conventions

The implementation will be done during joint WG1-WG2-WG3 meetings in GP3. The package will facilitate intercomparison studies of the performance of retrieval and merging methods and it will help to preprocess the datasets e.g. by providing a unified way of identifying collocated CMLs and

rain gauges. It will also help to ensure a common data format throughout the processing chain. Hence, this package will provide the foundation on which individual processing packages, or packages for merging sensor data, can be based on.