Final authors comments (AC) on egusphere-2025-4979

Title: The ECOSENSE forest: A distributed sensor and data management system for real-time monitoring of ecosystem processes and stresses
Author(s): Jasmin Tesch et al.
MS No.: egusphere-2025-4979
MS type: Research article

To improve clarity, we have reproduced the reviewers' original comments and provided our responses in blue. Revised text in the manuscript is presented in *italics*, while section headings are shown in **bold**.

Response to referee comment 1 (RC1)

The authors present the design and instrumentation of ECOSENSE forest, a large-scale environmental monitoring infrastructure that aims at studying forest ecosystem dynamics under climate change. The monitoring system incorporates more than 1100 sensors, generating many data to be processed. Such data may be used for archiving, thus enabling the forest state measurements evolution through time. The main contribution is the implementation of a practical framework for using distributed environmental sensor networks that consists of physical structures, dedicated power supply, communication networks and a data management system. This ECOSENSE forest shows the possibility of monitoring large-scale environments with reliability and resilience.

General comments

The authors are first thanked for their contribution. The manuscript is well written, clear, and comprehensive. The instrumentation presented is particularly impressive, both in terms of the number of deployed sensors and the overall system architecture.

Thank you for appreciating our work and for agreeing to serve as a reviewer. See our answers to your valuable comments below.

While the main contribution of the article is to present the ECOSENSE infrastructure as a blueprint for similar deployments, the conclusion would benefit from further discussion of the actual use of the collected data. In particular, providing examples of scientific analyses, preliminary results, or research questions already addressed (e.g. part 4) using the dataset would strengthen the paper and better illustrate the value of the proposed infrastructure.

We understand your point. As you rightly noted, this is not the focus of the present paper and paper writing and publication within the project are still ongoing. With that being said, several additional papers have been published or are at least submitted (and are available via open-access preprint server) since we first submitted this manuscript. We have gone through the manuscript and updated all references to preprints where the corresponding publications have since been published. In addition, we selected a few examples from recent manuscripts and added a new subchapter (4.1) briefly describing the data used and the main results. There, we have also included links to our project data repository (FreiData), as well as to our project website and FreiDok, where interested readers can find up-to-date information on publications by searching for "ECOSENSE".

The new paragraph reads:

"4.1. Data repository, ECOSENSE publications and examples for first results

Analysis and interpretation of the wide range of data collected is still ongoing. A current list of publications associated with ECOSENSE can be found on our website (https://uni-freiburg.de/ecosense/publications) and by searching for ECOSENSE in FreiDok, the institutional publication repository of the University of Freiburg (https://freidok.uni-freiburg.de). After publication, the data used in the respective paper will be made available via FreiData (https://freidata.uni-freiburg.de/communities/ecosense_sfb1537).

To illustrate the usage of the data collected at the ECOSENSE forest we provide a short summary of current publications and manuscripts: Lee et al. (2025) revealed pronounced spatial heterogeneity in forest floor VOC emissions, with strongly enhanced toluene emissions under waterlogged, low-redox conditions, demonstrating the importance of localized emission hotspots for ecosystem VOC budgets. Lee et al. (2026) further showed that tree stems are a highly variable source of VOCs across six major European tree species and, by scaling the measurements with the ECOSENSE forest inventory, quantified the contribution of stem emissions at the ecosystem level. Taking advantage of a variety of tree physiological measurements including our novel in situ leaf cuvette system, Dumberger et al. (2026) investigated inter- and intra-specific differences in leaf gas exchange. They found pronounced differences in seasonal dynamics of stomatal regulation and VOC emissions between the two investigated species and sun and shade leaves demonstrating the value of data at high temporal and spatial resolution. Making use of extensive soil respiration measurements, our detailed tree species inventory and data on soil physical properties, Brzozon et al. (2026) investigated drivers of soil respiration patterns. They found that temporal patterns were strongly affected by meteorological conditions, while tree species distribution explained some of the spatial heterogeneity. Stock et al. (2025) combined continuous measurements of active chlorophyll fluorescence with eddy flux and meteorological data to investigate the impact of cold spells on forest carbon dynamics. They found species-specific differences of photoinhibition in response to low air temperatures and highlight the value of chlorophyll fluorescence measurements for early stress detection in forest trees."

Furthermore, I have minor remarks and questions:

1. Data heterogeneity management

The monitoring system integrates a large number of sensors measuring different variables, likely with different acquisition frequencies, units, and possibly data dimensionality. The manuscript would benefit from a clearer explanation of how this data heterogeneity is handled.

You are right that with the amount of data collected, instruments and people involved there are inherently some differences in the collected frequencies, units and data types across measurements. We thought about this already during/before field site setup and provided project guidelines for data structure, e.g. we assigned sensor IDs and metadata is provided by the researchers. We briefly elaborated on this in the submitted manuscript in section 3 “From measurements to data”.

Going over this section again, we realized that we did not clearly state the efforts that went into data management at the beginning of the project. We will add this information in the revised document. The extended paragraph now reads:

“3.3. Data standardization, metadata and quality control

As mentioned above, we aim to ensure the FAIR principles for all data collected. Essential components of FAIR data are the systematic documentation and preservation of metadata associated with all spatial and temporal datasets (Wilkinson et al., 2016). To foster standardized and consistent research data management across all sub-project, the data manager organized a series of workshops covering: (1) an introduction to RDM structure, including data collection and transmission; (2) file data upload, including input data parsing, metadata, QA/QC, and time series storage in Aquarius; (3) technical data management, covering FAIR principles, documentation tools and formats, and both commercial and open-access software (e.g., GitLab); (4) geographic information systems, centered on a GIS tutorial introducing users to the ECOSENSE QGIS project; and (5) Aquarius, focusing on the web-based GUI for finding, filtering, visualizing, and downloading data. Acquisition frequencies are set by the respective researchers according to their specific research questions, and in some cases are further constrained by sensor characteristics (e.g., heating phase of sap flow sensors). Rather than homogenizing frequencies across all datasets, we instead rely on Aquarius' processing capabilities to aggregate raw data as needed, for instance, high-frequency precipitation data (one-minute resolution) are aggregated into hourly and daily sums. Since measurement frequency requirements vary across modelling approaches and applications, users are expected to further process the data to fit their specific needs. Measurement units are standardized through a parameter and unit list distributed to the consortium at the onset of the project. While this approach worked well for programmable data loggers, other cases required calculation of time series within the database to convert sensor outputs into standardized units (e.g., analog signals from mV, or hPa to kPa).

A key component of the established framework is the assignment of unique sensor IDs that are linked with meta data. Each sensor and its measured time series is accompanied by key contextual information, including spatial coordinates, instrument identifier, data ownership, measured parameter, unit of measurement, sampling interval, observation period and time reference.”

In particular:
- How are different acquisition frequencies managed or synchronized?

The original acquisition frequencies are determined by the respective researchers and depend on what they consider necessary for their specific research question. In addition, some measurements have an inherent limitation in acquisition frequency, e.g., sapflow can only be measured every 15 minutes due to the heat pulse produced by the sensor. Therefore we decided not to homogenize frequencies across all data collected.

However, our database Aquarius provides options for further processing, including aggregation of collected raw data. For example, for high-frequency precipitation data (with a data point every minute) we calculated both hourly and daily sums. As the requirements for the measurement frequency can be quite different for different modelling approaches or other applications, users of our datasets will have to further process the data to fit their specific needs. This aspect is implemented in the revised paragraph above.

- How are measurement units standardized across sensors?

Good question. In the beginning of the project we provided a parameter and unit list to the consortium to align measurements across sensors. For programmable data loggers that worked well. In other cases we created calculated time series in our database to convert analog sensors from mV to the agreed unit, or converted e.g., hPa to kPa. We added this remark in the extended paragraph.

- Are there procedures for harmonizing data structures before storage or distribution?

We shortly explained the harmonization in the submitted manuscript (L 372-374): *"The server functions as a central hub for initial data handling (7), i.e., parsing and preprocessing data with Python scripts into a unified format to ensure consistency across data sources and reduce data volume."*

We realize that this might not be sufficient information for the readers of this paper. More specifically, we load the raw data into the python script, format it to fit the SQL database structure and then upload it to the database with an API call directly from the formatting script. We keep the raw data files as backups. To make that process clearer in the manuscript, we adjusted the paragraph above and it now reads as following in revised document:
"The server functions as a central hub for initial data handling (7), i.e., parsing and preprocessing data with Python scripts into a unified format to ensure consistency across data sources and reduce data volume. Once formatted to fit the database structure, the processed data are uploaded directly from the Python script via a dedicated API call, and transmitted to the database within the university network over a VPN-secured, satellite-based internet connection."

We hope that our specific workflow is now more understandable for the reader. Thank you for this comment!

2. Data format specification

The manuscript refers to a "standardised format" (L.132) and mentions file access (L.380), but the actual data format used for storage and distribution is not clearly specified, if I am not mistaken. Providing details about the file format (e.g., HDF5, CSV, Zarr, etc.) and the general structure of the datasets would improve reproducibility and usability.

All time series data are stored in a central MS SQL database. The format is .mdf, .ndf and .ldf. The structure of the database is specified by the software provider "Aquarius". It includes the fields: Label, Parameter, Units, Interpolation, UTC Offset, Computation, Instrument, Data Owner and a Unique ID. We did not include this information as this is a given software architecture. The interaction with the database is done via the UI and APIs, not with the database directly. Data can be selected and retrieved from the database UI as CSV files. The standardized format refers to the pre-determined dataset structure, not the file format. We clarified this in section 3 in the revised manuscript:

"Our choice of a relational database compared to alternatives such HDF5 or ZARR, as the primary data storage system reflects the structure of our time series data, which consist predominantly of tabular sensor time series rather than large multidimensional gridded datasets. However, similar to standards such as the HDF5 group recommendations, we follow the general principle of self-describing, co-located metadata by storing core contextual metadata (as listed above) directly alongside the observations."

3. Integration of external data sources

External data sources are briefly mentioned, such as LiDAR scans (L.281). Is there other external data sources integrated into the datasets?

We assume that with "external" you mean everything except time series data that is collected and fed into our database (Aquarius) automatically. See our answers to your following questions for details.

If so:

- How are these external datasets incorporated?

The data architecture is two-fold, an SQL database for time series data and a fileshare for spatial data and "unstructured" data like CAD files, calibration algorithms and code. Depending on the datasets, they are incorporated in one of those. All data that can be considered a time series, also very short ones, are uploaded to the SQL database.

- How are they synchronized with in-situ measurements?

We are not completely sure what you mean with synchronization in this context. Spatial or campaign data like LiDAR scans are matched to in-situ data with their respective timestamp. As an example, researchers use LiDAR scans to draw conclusions on observed precipitation patterns.

- Are they stored within the same data structure?

Reading your comments, we are under the impression our different data storage platforms and processing pathways might not be described sufficiently. We tried to depict this in Figure 6 but try to describe it in more detail here:

There are two different types of platforms: 1) the SQL database "Aquarius" for all timeline data, 2) a share folder for all other data, e.g. geospatial data. Regardless of the data storage location, data will be made available via FreiData following publication.

4. Security considerations

Section 2.5 discusses security measures against potential network attacks.

- Is there any protections against local access threats, from Wi-Fi access points, or LoRaWAN gateways (L.245)?

The Wi-Fi access points are centrally managed, allowing continuous monitoring of connected clients. The network is secured using WPA2 encryption, which significantly increases the difficulty of unauthorized access. The LoRaWAN network is secured with AES-128 encryption. The entire forest network is protected by a Fortinet firewall managed by the RZ (Computing Center) of the University of Freiburg, ensuring that intrusion into the university network is not possible.

We updated the manuscript to incorporate the mentioned security considerations and it now reads:

"The gateway+, which establishes the connection via VPN tunnel, is protected by a Fortinet firewall managed by the RZ (Computing Center) of the University of Freiburg hosts a firewall. To restrict and track access from the outside, the University of Freiburg hosts a MAC filter, meaning only computers/servers with pre-registered MAC addresses can connect to the ECOSENSE forest directly. Another possibility for remote access is through the university's VPN client, which requires a login with a personalised user ID. To further enhance network security, the Wi-Fi access points are centrally managed, allowing continuous monitoring of connected clients. The network is secured using WPA2 encryption, which significantly increases the difficulty of unauthorized access. The LoRaWAN network is secured with AES-128 encryption. This allows us to prevent attacks against the ECOSENSE forest network from both outside and inside"

5. Architecture terminology (Figures 4 and 6)

Figure 4 provides a useful overview of the system architecture, and the consistent color and naming scheme with Figure 6 helps readability.

Thanks for appreciating the thoughts that went into this :)

However, some terminology may be imprecise, it is a difficult task to formalize accurately the overall process.

- The term "2. Detection" may not be fully appropriate. In many contexts, detection refers to extracting information from raw or processed data, whereas this stage appears to correspond to data measurement. The term "Measurement" may therefore be more accurate.

Agreed. We will change “2. Detection” to “2. Measurement” in Figure 4 and 6.

- Some labels are not immediately clear. For instance, examples "Selection, Formats" under "7. Processing" could be clarified. At least those single words are not obvious.

We understand the criticism, and therefore exchanged the examples from “Selection, Format” to “Restructuring, Mapping”.

- The association of "Broadcast" with VPN is somewhat confusing (see also L.376). A VPN generally provides a secure point-to-point communication channel rather than a broadcast mechanism. Conceptually, this functionality might fit better under "10. Access."

Thank you for pointing this out. We understand that this is not the most appropriate term and changed “Broadcast” to “Routing” in Figure 4 and 6.

Specific comments

- Figure 3 should be placed after 2.5 because the network infrastructure (right part) is presented in 2.5 and illustrated in Figure 3.

Ok. We changed the Figure placement. It now appears after Section 2.5.

- Figure 4 "8. Broadcast" is not readable (white font color on light gray background)

Thank you for this comment. We changed “Routing” (previously “Broadcast”) to black font in Figure 4 and 6.

- L.373: "NodeRed" instead of "NoteRed" (I think)

We realized it is spelled Node-RED. We changed it in the revised manuscript.

- L.234: "prevent" instead of "protect" would be more accurate.

We changed this sentence. It now reads: “*This allows us to prevent attacks against the ECOSENSE forest network from both outside and inside.”*

- L.396: "ECOSENSE" instead of "EOCSENSE"

Thank you for pointing that out, we changed it in the revised manuscript.

End of response to RC1

Response to referee comment 2 (RC2)

This paper addresses the design and the implementation of large-scale instrumentation system called ECOSENSE forest. This large-scale environmental monitoring infrastructure aims at studying forest ecosystem dynamics under climate change. The monitoring system incorporates more than 1100 sensors, generating many data to be processed. Such data may be used for post analysis and also for new modeling up to virtual reality developments.

The main contribution is the implementation of a practical framework for using distributed environmental sensor networks that consists of physical structures, dedicated power supply, communication networks and a data management system.

General comments

The manuscript is well written, clear, and comprehensive. The instrumentation presented is diverse, from on the shelf sensors up to in-house developed sensors. The scale of the development in a natural environment is impressive.

Thank you for appreciating our efforts and work invested. We are also grateful for your comments provided, see our detailed answers below.

Anyway, though the ECOSENSE blueprint is clearly described, some aspects require further explanations.

Detailed remarks and questions:

Large scale instrumentation system are implemented and used in different industrial domains or for transport infrastructures monitoring. Their deployment are still gathered with a maintenance strategy. Could you precise the ECOSENSE strategy implemented for ECOSENSE sensors and the whole system maintenance ?

The maintenance strategy highly depends on the deployed instruments and is managed by the responsible researchers (that usually have most experience with the respective sensors). Additionally, regular maintenance and checks of every sensor installed would exceed the capacities and capabilities of a few people responsible for the field site management. For example, the tipping buckets needed regular cleaning and recalibration and sapflow sensors needed yearly reinstallation, for most instruments and sensors maintenance has been conducted on demand, e.g., when sensors stopped sending reliable data.

However, as central management, we conducted regular checks to verify that the field site infrastructure remained intact, i.e., confirming stable power supply, network access, and proper data transfer and storage. Whenever a researcher identified an issue with a sensor, data logger, or data transmission, we typically worked together to find a solution.

In field sensors are ageing and could drift with time. Could you precise how you manage for instance periodic calibration or autonomous in-situ solutions to control sensor deviation? Due to the large amount of sensors deployed, it could be interesting to make this description for a macro, a meso and a micro scale for a typical sensor.

The sensors regularly retrieve the network time from the fieldsite server to prevent a time drift. The periodic calibration is managed by the respective researchers and not centrally organized. Sensor deviation might only become visible after a measurement period and will then get corrected for influencing factors, e.g., the needle distance from the sapflow sensors. As this procedure is highly heterogeneous and lies within the responsibility of each researcher, we believe that a general description cannot be compiled.

How metadata and data management in the ECOSENSE systems takes into account or was influenced by OGC standards, HDF5 group recommendations, ZARR, etc.? Could you precise these aspects in your paper ?

We thank the reviewer for raising this point and have expanded the manuscript to clarify our metadata management approach. Our raw data are mainly tabular sensor time series, for which a relational database is more suitable than hierarchical array formats such as HDF5 or ZARR, which target large multidimensional gridded data rather than structured time series. Nonetheless, our schema follows the HDF Group's underlying principle of self-describing, co-located metadata: core contextual information (coordinates, instrument identifier, measured parameter, unit, sampling interval, quality flags) is stored directly alongside the observations. As the database architecture cannot accommodate all metadata required for full standards compliance, remaining metadata are appended externally as XML/JSON schema files at the point of publication, following the DataCite standard, in consultation with the Central Data Facility (CDF) of the University of Freiburg and NFDI4Earth. Regarding OGC standards, sensor locations are visualized in an internal QGIS project using standard coordinate reference systems and geometries compatible with OGC Simple Feature Access. This layer serves internal site management only and is not linked to the observational time series data, which reside separately in our relational database.

The revised paragraph in section 3 now reads:

“A key component of the established framework is the assignment of unique sensor IDs that are linked with meta data. Each sensor and its measured time series is accompanied by key contextual information, including spatial coordinates, instrument identifier, data ownership, measured parameter, unit of measurement, sampling interval, observation period and time reference. Due to predetermined and unalterable database architecture, it is not possible to integrate all metadata into the active datasets using a standardized schema. Therefore, upon completion of each time series metadata are appended in the form of XML or JSON schema files, in accordance with an appropriate metadata standard, i.e., “DataCite” (Brase, 2009). The specific metadata standard is selected in consultation with the Central Data Facility (CDF) of the University of Freiburg and NFDI4Earth, with the goal of ensuring long-term interoperability and broad usability. Our choice of a relational database compared to alternatives such as HDF5 or ZARR, as the primary data storage system reflects the structure of our time series data, which consist predominantly of tabular sensor time series rather than large multidimensional gridded datasets. However, similar to standards such as the HDF5 group recommendations, we follow the general principle of self-describing, co-located metadata by storing core contextual metadata (as listed above) directly alongside the observations.”

For the reading of the paper, o,e will appreciate to have a separate list of your in-house sensor collection . Could you add such a list and make a schematic view of their location into the ECOSENSE system ?

A list of deployed sensors is provided in the Supplement (Table S2). All in-house sensors are marked as “developed by ECOSENSE” in column “Manufacturer”. We indeed maintain a central GIS project where all sensors are marked on top of a georeferenced orthoimagery scan. However we believe that this image does not add value to the manuscript as details about type and measurements cannot be shown in this zoomed-out snapshot. See screenshots of this below.

And from a close-up:

Could you give an overview of the volumetric amount of data daily or monthly produced (including additional periodic field campaigns)?

We added some average numbers to the manuscript to give an impression of the volumetric magnitude of data. The entire, revised paragraph now reads:

"2.6 Environmental monitoring setup

We established a comprehensive monitoring system in the ECOSENSE forest across multiple scales, i.e., from leaf to stand level. More than 670 commercial and 430 self-built sensors monitor over 90 distinct parameters, fluxes, or processes generating upwards of 4,500 time series that capture soil, tree and atmospheric processes with high spatial and temporal resolution. The volume of incoming data varies considerably throughout the year: on average, we receive 1 GB per month, while high-throughput instruments such as leaf angle cameras, phenocams, terrestrial LiDAR, isotope and VOC analyzers, and (relaxed) eddy covariance systems generate substantially larger raw data volumes. The field site server handles over 66 TB of traffic per month, of which 240 GB is transferred to the university, as most raw output is preprocessed on-site and only a small fraction is retained long-term. Campaign data, particularly the monthly LiDAR flights, can produce many gigabytes per flight, but only a much smaller processed version is retained. This processing is carried out outside our server infrastructure."

Finally, as mentioned in the paper such instrumentation system pave the way to the development of new models up to virtual reality developments. Could you precise geolocalisation and time requirements accuracies required for such research works?

We thank the reviewer for this helpful comment. This information was indeed missing from the manuscript. We have now added a clarification in Section 4.4, specifying that georeferencing in the current ECOSENSE setup primarily ensures consistent relative co-registration of sensors, trees, infrastructure and repeated 3D observations. We further clarify that temporal synchronisation is tailored to the relevant process dynamics, typically within seconds to minutes, enabling the coupling of rapidly interacting variables such as irradiance, chlorophyll fluorescence and leaf temperature.

We added this information in the revised manuscript:

"In the current setup, the main purpose of georeferencing is to ensure consistent relative co-registration of sensors, trees, infrastructure and repeated 3D observations within the ECOSENSE forest. This enables sensor observations to be linked to individual trees, canopy positions and plots, and provides the spatial basis for tree- and plot-level modelling as well as digital twin and VR applications. Temporal synchronisation is tailored to the relevant process dynamics, typically within seconds to minutes, enabling the precise coupling of rapidly interacting processes such as changes in irradiance and resulting responses in chlorophyll fluorescence or leaf temperature."

End of response to RC2
