Response to editor comments

Thank you for the comments. Below, we go through them and the corresponding changes in the manuscript. Line numbers in brackets "[]" indicate lines in the markup version of the document.

The current version still requires further improvements.

A further revision of the English language is crucial.

We revised several sentences. Various overly elaborate verbs such as "utilize" were replaced with simpler alternatives. [17-21, 23, 24-26, 46-47, 70, 75-79, 85, 96-107, 109-110, 116, 119, 122, 124-127, 130-131, 133, 135, 140, 153, 156, 163-165, 170-186, 188-189, 203, 209, 212, 220, 223, 226, 231-232, 238-242, 244-245, 272, 354-355, 365, 378, 381, 391-392, 395, 409-413]

All abbreviations that you need must be written out and/or explained.

We spelled out all remaining abbreviations [184,197-200,259-260,283,354,433]. We also added explanations of various technical terms. [197-200, 224-225, 240, 283, 354]

It is also necessary to check all punctuations, various commas are missing.

We added commas to [26, 46, 119, 347]. Simplifications were made to several sentences, reducing the need for commas [43-44, 67, 109-110, 116, 122, 126, 130]. We checked the punctuation of the text with the help of language tools in Word.

The paper should better highlight the functionalities and advantages of SMEARcore and explain them using the examples. Often we only find a description of the example, needs to be improved. (A good examples are the descriptions of Figure 6 and 7; the description of Figures 3, 4 and 5 needs to be improved to describe the advantage of SMEAR's functionality).

Added points about how the interactive and linked plots are an advantage. Also, the plots depend on data processed by SMEARcore. [307-308, 314-315, 321-323] Note that plots 3, 4, 5, do represent typical plots in this field, so the main difference any normal plotting is that they are automatically generated for the time periods in question. They are needed to show that we can detect the event based on a dashboard that has all three, but the plots themselves are not exceptional.

From my point of view, it is also crucial that the SMEARcore can also bring additional benefits in combination with a central network. This should be stated somewhere.

This is now mentioned in the abstract and the conclusions based on the other comments. [27, 409-410]

Revision of the Abstract (English and content) is required, examples see below.

The abstract was simplified, and the content additions below were added.

Content: e.g. SMAER can also be an asset in parallel to a central solution, see my comment in the Conclusions).

We added "be used in addition to them and to" to the last sentence of the abstract.

English: the following is no understandable for experts not experienced in this filed.

line 17: SMEARcore allows new SMEAR stations (Station for Measuring Earth Surface – Atmosphere Relations) to be built consistently with existing ones and to utilize pre-existing experience in data curation.

Suggestion: SMEARcore enables new SMEAR (Station for Measuring Earth Surface - Atmosphere Relations) stations to be integrated in a way that is consistent with existing stations and transfers existing data curation experience to the new station.

We changed the sentence as suggested.

Line 74-75: English revision

We changed the sentences from:

"The system indexes this ancillary data, so that it can be accessed for further analysis. This indexing enables the implementation of routine calculations, such as calibrations and visualizations, to be done automatically to aid operators to identify and solve problems with data collection."

To

"We store this metadata for further analysis. It allows us to automatically make calibrations and visualizations, which aid in identifying and solving problems with data collection."

Line 94-101: English revision

We simplified the paragraph from:

"Data management is not only about checking that consistent calculations have been made. As with any system, errors can occur, for example: computers crash, power gets interrupted, networks are throttled, reagents run out, somebody forgets to run a script or analyzer inlets foul. Some of these might cause problems for the measurements, some might just temporarily halt data transfers, but in any case, we need to know something unexpected happened. For this to happen in a timely manner, parts of the analysis must be automated and monitored. If we need to wait weeks or months for a responsible person to analyze the data and notice a problem, we cannot intervene when it matters most, and useful data is lost. Same goes for transferring data out of the measurement computers, monitoring the state of those computers and backups. For these reasons, routine operations should not be a manual process whenever it is possible to automate them."

To

"Data management is more than making consistent calculations. Any system can experience errors and a typical measurement station has many things that can go wrong. If errors are not detected, it is not possible to intervene in time and data is lost. It is critical to detect errors promptly and that requires automation of the detection. This monitoring should watch the data transfers, computing hardware, and calculations in addition to the measurements."

Line 160, capture figure 1

Include explanation of the black dot with circle, on the right of the figure in figure capture.

We explained the end point of the process in the figure caption.

Line 163-171: English revision

We changed the section from:

"Any data files usually need metadata to be interpreted correctly. This is information such as measurement units, calibrations, column names in the data files etc. In our case we also produce metadata about the data processing itself: what files were processed when, how much data was there, what ancillary data was used in the processing and where the files can be located.

One file format used to solve this problem in infrastructures is NASA-AMES used by ACTRIS. There the file metadata is stored within the file itself as extensive header lines. In our case this would lead to extensive duplication of the data in many cases, and it is not appropriate for the metadata about the processing itself. Thus, we store metadata mostly as database tables and link to other files as necessary. Limited sets of metadata can be exported with the files by workflows to produce other formats."

To

"Metadata is used to interpret data. Metadata consists of information such as measurement units, calibrations, and column names. We also create metadata about file processing, such as when files were processed, their sizes, where they were saved and what data files were combined to produce the resulting file.

There are several ways to store metadata. Some file formats such as NASA-AMES used by ACTRIS, or hierarchical data format (HDF) files store it in the data file itself. For us this would result in duplication of the metadata, and it is not well suited for file processing metadata. We store the metadata as database tables and link to files, as necessary. It is possible to export metadata from the databases to file formats when required by workflows."

Line 180: Figure description has to include all abbreviations used in the figure

We added expanded the abbreviations (https, S3, SSH, SCP, SQL, InfluxQL) in the caption and explained the terms.

Conclusions, line 382:

This makes it useful for measurements that are not controlled by centralized solutions

Add: or it can be used as a backup of the data owner in parallel with data transfer to centralized networks

We added the suggestion.

Conclusions, line 385.

Your current text:

We show in two case studies how continuous visualizations of the data and metadata, such as instrument diagnostics and datafile availability, can help quickly spot interesting phenomena and abnormal situations in the measurements.

Suggestion:

We demonstrated with two case studies how continuous visualizations of data and metadata, such as device diagnostics and data file availability, can help to quickly identify interesting phenomena and abnormal situations in measurements

We implemented the suggestion, with a small change: "help quickly" instead of "help to quickly".