Authors' reply to reviewer comments GC-2019-22

Weather and Climate Science in the Digital Era, Martine G. de Vos et al.

Reviewer Comment 1

General:

The paper is an interesting read on an important and timely topic. The length is appropriate, and the authors represent a broad spectrum of subjects within Geosciences. However, the paper could improve significantly if the text would be more specific and if more specific examples would be provided for the claims in the text (see list of points below but there are many more places in the text).

We agree with the reviewer that we could be more specific on how we have collected our observations and how these support the claims in the text. We will add a dedicated 'Methods' section (see below for the suggested text). Besides, throughout the paper we will rephrase text to be more specific on our observations and how these support our story.

" **Methods :** The focus of the conference session was on data and compute intensive approaches that are applied in weather and climate science. The session comprised 10 oral abstract presentations,
one keynote talk, and 6 short poster pitches. The 16 participants were either presenters or involved in the organization of the session, and represented domain science, as well as computer and data sciences.
The first part of the session was dedicated to the presentations.The second part was interactive. In three groups of each 5 or 6 persons the participants discussed
the "challenges and opportunities regarding open weather and climate science" and noted their findings on a flipchart. The findings of each group were presented and discussed in a following plenary session. Observations and insights from the plenary discussion were documented.
The observations in this paper are based on both the insights from the studies presented in the session, and the notes made during the interactive part of the session. The majority of the participants from the session also contributed to this paper. As such this paper represents a shared view of the participants, i.e., a group of experts in weather and climate science, on the digital and open science developments in their field."

I think that it would make the paper much more credible if the authors would provide a list of action points to improve the situation at the end of the paper. However, I leave this for the authors to decide.
We thank the reviewer for this great suggestion. At the end of the paper, we will provide a list of action points or conclusions that are described in the different sections of the paper.

The English should be improved.
We agree with the reviewer and get a native speaker to edit the manuscript

Specific comments:

• *ll.6-9: How is this shown?*
We will rephrase the paragraph to clarify its meaning and add concrete examples that illustrate the importance of shared data and software:
“The majority of studies (roughly 80 %) presented in the conference session depended in some way or another on shared data and software. For example, many studies included open datasets from disparate sources to improve accuracy of forecasts on the local scale, or to extend analyses beyond the domain of weather and climate. Furthermore, shared software is a prerequisite for the studies that presented systems like a model coupling framework or a digital collaboration platform. Although these studies showed that sharing code and data is important, the consensus among the participants was that this is not sufficient to achieve open weather and climate science and that there are important issues to address.”

• *ll. 10-14: What is special about the origin, scalability and legal barriers?*
For instance, many data sources come from private industry who may see a competitive advantage to maintaining privacy. But those data may prove useful to the weather community for improving initial conditions of forecast models. Such corundums may be solved by signing nondisclosure agreements and allow weather service to act as trusted agents who use the data for the public good without disclosing their details.

We will include this explanation in the abstract and in the corresponding sections.

• *ll. 10-14: Why does the complexity limit collaboration? Can you give examples?*
We will elaborate the text in the abstract and the corresponding section. Please see the last comment on software platforms for the text suggestion.

• *l. 14: Why is there a need for new roles?*
Data management and programming have become an integral part of current research practice, and these activities require specific digital skills. It is therefore important to acknowledge and

define roles, responsibilities and mandates concerning data stewardship and research software engineering.
The aforementioned trusted agents can also be considered a new role
We will include this explanation in the corresponding sections.

• *l. 36: Was this really both short and long wave? If you refer to the 90s, you should also cite the original papers by Chevallier et al.*
We can confirm that neural nets have been used for both short and long wave radiation. We will rephrase the sentence and add the corresponding references.

• *l. 56: Lagging behind whom? Can you give an example?*
The reviewer rightly points out that it is not clear who, or which field we compare to. In fact, open sharing of data, software and vocabularies is only true common practice in a few fields such as astronomy and genomics. Most scientific fields, including weather and climate science, can be considered lagging behind. Furthermore, the actual point was to show that these weather and climate science are mature in terms of applying digital technologies, while the implementation of open science methodologies is less advanced

We will rephrase the corresponding paragraph accordingly.

• *End of section 1: It would be good to give a hint about the structure of the following. The reader does not know what to expect from the rest of the paper.*
We adopt the advice of the reviewer and will clarify the structure of the rest of the paper at the end of the introduction section

• *l. 71-74: Which or at least how many countries? How many Funders and Research institutes? Can you give examples?*
We will try to be more specific and add examples and references to this paragraph.
“Europe and the United States have made efforts to adapt legal frameworks and implement policy initiatives greater openness in scientific research (OECD, 2015; National Science Foundation, 2018). Several countries provide digital infrastructure based on rich metadata for the resources in the research environment, that support their optimal re-use (Mons, 2017). Examples include the European Open Science Cloud in Europe (Directorate-General,2018) , NIH Data Commons projects in the United States, AARnet in Australia (AARNet, 2018) and the African Data Intensive Research Cloud in South Africa (Simmonds,2016). Funders and research institute have announced policies encouraging, mandating, or specifically financing open research practices (Mckiernan et al.,2016; Wilkinson et al, 2016). Examples include the National Science Foundation (NSF) in the United States (National Science Board,2011), CERN in Switzerland (CERN, 2014), the Netherlands Organization for Scientific Research (NWO) (NWO, 2019) and the United Nations Educational, Scientific and Cultural Organization (UNESCO) (Unesco, 2013).”

Unfortunately, we are not able to provide quantitative information.

Mons, B., Neylon, C., Velterop, J., Dumontier, M., Da Silva Santos, L. O. B., & Wilkinson, M. D. (2017). Cloudy, increasingly FAIR; Revisiting the FAIR Data guiding principles for the European Open Science Cloud. Information Services and Use, 37(1), 49–56. https://doi.org/10.3233/ISU-170824

Directorate-General for Research and Innovation. (2018). Prompting an EOSC in practice. Final report and recommendations of the Commission 2nd High Level Expert Group on the European Open Science Cloud (EOSC). https://doi.org/10.2777/112658

AARNet. (2018). ANNUAL REPORT / 2018 DATA CONNECTOR FOR THE FUTURE. Chatswood, Australia.

CERN-OPEN-2014-049. (2014). Open Access Policy for CERN Physics Publication.

R. Simmonds, Taylor, R., Horrell, J., Fanaroff, B., Sithole, H., Rensburg, S. J. van, & Al., E. (2016). The African data intensive research cloud. IST - Africa Week Conference.

National Science Board. (2011). Digital Research Data Sharing and Management.

NWO executive board. (2019). Connecting Science and Society - NWO strategy 2019-2022.

UNESCO Executive board. (2013). Open Access Policy concerning UNESCO publications.

• *l.82: As you elaborate later on, OpenIFS is not Open Source as it has a (free) license.*
The reviewer is right. We will rephrase the sentence.

• *End of section 2: You could also mention Reanalysis data here.*
We adopt the reviewers suggestion and will add the following text:
“Already since the 1990s the international meteorological and climate research communities started sharing data. Examples of data sharing with common file and metadata formats are reanalysis data, starting with NCEP/NCAR reanalysis and ECMWFs ERA reanalysis data products (e.g. Dee et al 2011, Kalnay et al. 1996) and coupled model intercomparison projects (Taylor et al. 2012).”

• *l.108: “clearly enrich their research” Can you give an example how?*
Examples include the various reanalysis datasets published by the ECMWF and NOAA/NCAR that are made freely available to the community and application of the open models, such as WRF.
We will add these examples to the section.

• *l. 114: What is “CF”?*
CF conventions provide guidelines for the use of metadata in the netCDF file. We will rephrase the paragraph and include the meaning and use of CF

• *l. 124: What do you mean by “performance scalability”. Software tools that allow to evaluate data at scale on supercomputers? How is data interoperable?*

• *l. 132: Which tools? Can you name them?*
We will rephrase the corresponding sentences to clarify the challenges of producing FAIR weather and climate model data:
“Regarding open and interoperable weather and climate model data, i.e. data and metadata that are formatted according to community standards (CF, CMIP, WMO), we consider performance scalability as the foremost technological challenge. Whereas high-resolution weather and climate data is predominantly produced on large clusters using many compute nodes, subsequent data processing and analysis is often still confined to a single CPU, and hence does not scale easily with, e.g., increased model resolution. Producing FAIR model data via traditional post-processing pipelines is quickly becoming unfeasible for high-resolution climate model data due to the sheer volume and complexity of the model output as noted above.”

• *l. 144: Which Journals? Can you name them?*
We will add some examples to the text:
“Data journals, like Geoscience data journal (Royal Meteorological Society), Scientific Data (Springer Nature) and Earth System Data (Copernicus Publications), are a partial remedy, as these provide open access platforms where scientific data can be peer-reviewed and formally published.”

• *l. 150: Can you outline some of the examples in more detail?*
We will elaborate the examples and rephrase the paragraph as follows:
“The conference session provided excellent examples of tools and approaches that were developed and made openly available to the research community. For example, approaches to reduce the computational or post processing costs of existing simulation models (Stringer et al., 2018; Behrens et al., 2018; van den Oord et al., 2018, Jansson et al., 2018) and approaches to integrate data sets from different sources (van Haren et al., 2018; Schultz et al., 2018). Several of the studies in the session presented an approach for which open data and software is a prerequisite, for example because these comprise a model coupling framework or a digital collaboration platform (Pelupessy et al., 2018; Ramamurthy, 2018; Hut et al., 2018; Bendoukha, 2018).”

• . *“The studies show that use of machine learning methods has added value because models are built with data beyond standard meteorological data. For example, local conditions related to the natural and built environment that cannot be captured easily in simulation models can be taken into account through trained models.” I do not understand this. Can you rephrase?*
This paragraph is about the use of data beyond the standard meteorological datasets. We will rephrase the paragraph to clarify this.

• *l. 177: Can you name examples for hardware and software platforms. And can you define what you mean by “platform” in this context?*
These platforms refer to digital platforms that use cloud technologies to create a virtual research

environment where scientific end-users can store, analyze and share their data. In the conference session several of these platforms were presented. An example of a current platform is the Open geospatial Consortium. We will rephrase the paragraph to clarify this.

• *“data such as that of the environment and citizen science sources.” I do not know which data sets you are referring to here.*
This sentence is referring to the data sets described in the section on open data, i.e., social media posts and observations from amateur weather stations. We will rephrase the sentence to make this clear

• *“The increase in accuracy and skill of forecasts at local scales are shown, improved consistency of data products and improved efficiency and skill of simulations, often crossing different disciplines.” Again, I do not understand this. Do you mean “show” instead of “are shown,”?*
The reviewer is right, it should have been “show”. We will rephrase the sentence accordingly.

• *l. 194: Which issues?*
This term refers to the issues described in the next paragraphs in the same section. The reviewer rightly points out that this should be clear from the text. We will rephrase the text in the section correspondingly.

• *“Technologically, the promise of using modern digital technologies is not always met due to the complexity of software platforms.” I do not understand this.*
The cloud appears to be a potential avenue, as it enables individual researchers to gain access to high computing resources, vast amounts of storage and a suite of software tools. In our session, several digital platforms were presented, that use cloud technologies to create a virtual research environment where scientific end-users can store, analyze and share their data. The participants also observed, however, that current platforms, like the Open Geospatial Consortium and JRC Earth Observation Data and Processing Platform, do not seem to increase the extent of scientific collaboration, especially across disciplines. This may be partly due to the fact that these platforms each have implemented their own set of standards for both data formats and interfaces to access these data. Since scientists are required to invest time and effort in working with a specific platform, the heterogeneity poses hurdles to their collaboration with researchers on another platform.

We will rephrase the paragraph to clarify this:

Minor points:

• *l. 9: Rephrase: “that here are”*

We will rephrase the sentence

• *l. 32: Rephrase “since ensured”*
We will rephrase the sentence

• *l.45: Rephrase: “use of using”*
We will rephrase the sentence