

Interactive comment on “Ideas and perspectives: enhancing the impact of the FLUXNET network of eddy covariance sites” by Dario Papale

Dario Papale

darpap@unitus.it

Received and published: 30 July 2020

Also in this case let me thanks Jason Beringer for the review and comments that are useful and appreciated. Here below my comments and answers on the specific points. My comments and answers are in blue italic

In addition to the demand for real time data that you have outlined, there is also an emerging area of ecological forecasting, for example, The EFI-NEON Research Coordination Network that is an NSF project to create a community of practice that builds capacity for ecological forecasting using NEON. There is also potential demand from data assimilation of flux tower data into short range weather forecast (e.g. <https://doi.org/10.1175/MWR-D-19-0370.1>). Also real-time agricultural monitoring

C1

should be possible. Finally, data can also be used to generate regional real time evapotranspiration estimates using a fusion of flux tower, remote sensing and modelling all of which have potential for use by the public.

Thanks for the additional suggestions on the uses; I will improve this section because I agree that a larger set of potential user communities would work in the direction of a stronger and more central FLUXNET

With respect to the proposed re-organisation from a single large database into sub collections. On the one hand this helps make the network more sustainable by delegating work and responsibility to continental clusters so that not all the work is being done by a small group. On the other hand this will require continental clusters to be functional, accessible and have open data sharing policies. I would be concerned that many clusters may not have capacity to do this or there may be a difference in data sharing between groups and individuals, such that individual sites that may want to contribute are unable because of the inability of the cluster to participate for technical, personnel or other reasons.

This is a good point and helps me to understand that my text was not clear enough (and so I will work on it). The Continental clusters are a way to become more interoperable, with the specific meaning that the competences, capacity and resources to do the same type of activity are shared by different groups. This would give to the system the needed robustness thanks to redundancy.

However Jason Beringer is right saying that the “Continental clusters” should be able to do this and this is not always the case, in particular in a starting phase. Here there are few possible solutions/clarifications that I would better clarify in the new version:

- Continental is probably a wrong term. A cluster can be also just a single county if they have the capability or can be a joint initiative across continents (Asia+Australia just to make an example). In other words it is a scalable concept: the main point is to have number a groups able to produce the FLUXNET agreed products and care about their access.

C2

- We can also have, for a period, an additional generic "FLUXNET basket" managed by a group that has resources/capacity, where single sites or clusters not ready can be processed and shared. But the aim must be to move in the direction that these groups become first able to use shared resources to do the processing (e.g. a cluster hosted in Europe to process themselves the data) and then to organize a Continental cluster that can then be a catalyst for regional networks development.

- On the data policy: it is not needed that the continental cluster adopts the common open policy in general, only that is able to share under this policy data that the PIs agree to share and document them with the needed metadata to be FAIR.

Following on from this I can see that there could be many, many sites (even some that were not part of FLUXNET2015) that want to contribute to FLUXNET but they are unable too because they don't have a functional continental cluster. It will be crucial to make sure there is a mechanism for them to contribute. The ONEFLUX processing code is not designed for this purpose. So I wonder if there could be an online processing tool that registered site users could upload their site data at what ever intervals they are able to do so?

This is also important, true and somehow connected (and partially answered) above. In a "Research Infrastructure" view, the online tool does not solve the problem, rather make it more complex because someone then has to ensure that the processed data are according to the standard, add the metadata and share them. All this with high risk of losing the traceability.

In my view the solution is to 1) stimulate and support the creation of 3-4 Continental Clusters that can offer the service to all the sites, also independently respect to the physical inclusion in the geographic continent; 2) prepare and maintain for the moment a generic cluster to host sites or small groups where a Continental cluster is not ready. It is a medium term process, having intermediate steps is needed.

C3

I'm not entirely sure for the rationale and need to move away from a large centralised data base to Continental cluster collections? There has been a lot of effort gone into making the database and I'm not sure if it is broken in some way or has reached its capacity technically. It seems to add another layer of complexity to have Continental cluster collections and a shuttle they queries each of them. It then relies to the ability of the clusters to maintain the data in real time in these clusters. Why not just have a continually updated big database where data is added at any time (daily or annually as it becomes available). You then rely on sites and clusters to push the data continually rather than calling and hassling for sites to submit data (current model). Users can query the data anytime too. We could build a set of tools that allow data to be accessed and queried in more sophisticated ways.

In theory, a large database could also work. But it would require a constant and substantial funding to do the call for data submission, quality check and processing. Even if the centralized database is only for distribution, it would still require funding and support (by all the networks, with all the complications due to trans-continent funding) for the curation and maintenance. The database, if we want to have PID, citations counts, good metadata etc. is something substantially different respect to a "simple" repository.

However, the main reasons why I think that the centralised database is not the robust solution in the medium term are: 1) it doesn't help to build the distributed competences needed for the robustness and 2) as said also in the last comment, regional and continental clusters want also their cake and being more connected to FLUXNET (from collection to production and distribution) would help.

The submission of data to a central database does not work automatically, believe me, it requires a lot of efforts as demonstrated by the activities in LaThuile and FLUXNET2015. The submission to "your network" is much easier, thanks to the more easy communication and membership spirit. That said, as also answered above, in a first phase it is possible to have a "mix model", but I'm sure that we have 3 groups globally ready for the test.

C4

I like the idea of a persistent identifier (PID) or something similar. I would envisage that initially the processing would follow the FLUXNET2015 (i.e. ONEFLUX) pipeline and the PID would reflect that. But processing methods do evolve and a Fluxnet steering group could endorse any changes to the pipeline and periodically the PID would change to say FLUXNET2025 for example. It is probably important to think about changes in processing and reprocessing the whole database. This may well happen in the future if we have a new pipeline and you would want to apply the new pipeline retrospectively across all the data I presume? This would be manageable under a single large database but may be difficult under Continental cluster collections.

The PID identifies the object, the metadata reports the processing. These two together make the data transparent. It is clear that evolution in the processing will happen (hopefully and fortunately) and these will be reflected in the metadata and generate new objects with new PIDs.

I agree, in case of new processing a complete reprocessing is needed (like it was between e.g. LaThuile and FLUXNET2015) but it is exactly in cases like this that the Continental clusters are crucial: it is requested an interaction with the PI to get the last version of the data, corrections if needed and then do the processing. A shared workload is the only option with a network growing like FLUXNET. Processing and reprocessing are not only machine time after pushing a button, this could be centralized. There are a lot of interactions needed and this is the key aspect where the Continental clusters are crucial.

As Papale acknowledges, it will be important full flexibility of each single network to decide what to share and when in FLUXNET and the possibility to distribute different formats and versions through their Data portals. I would envisage that in OzFlux, one could have rapid processing of near real time fluxes using the FLUXNET pipeline on a daily basis and this would have PID that differentiated these data streams as 'beta'

C5

datastreams or something. These data can be used for applications that require real time data but the data comes with the caveats of not being quality controlled by site investigators. The data should be fit for the purpose that it is being applied. As Ray Leuning would tell us, "know thy site" and investigators would still produce the finalised data sets with human skill and site knowledge. So we can have our cake and eat it too. So I think there are potential pros and cons and maybe the best way is to get a working group together and flesh out an operational model for the future. I am very keen and you can count us in.

Yes, this is exactly the spirit and basis for the system: make each network able to do what they want and at the same time be part (practically, so really feeling this) of FLUXNET. Linking networks to save resources if they want (the continental clusters), gaining visibility and making all the PIs also more engaged in the global networking.

Interactive comment on Biogeosciences Discuss., <https://doi.org/10.5194/bg-2020-211>, 2020.

C6