

## ***Interactive comment on “The Global Terrestrial Network for Permafrost Database: metadata statistics and prospective analysis on future permafrost temperature and active layer depth monitoring site distribution” by B. K. Biskaborn et al.***

### **Anonymous Referee #1**

Received and published: 18 June 2015

#### **A. General comment**

The main objective of this paper is to present the new GTN-P database. The mission of this database is the long-term and world wide collection of permafrost-related data: permafrost temperatures (TSP) and active layer thickness (ALT and CALM).

The paper has the following structure: after an introduction on the steps which leaded

C131

to the present shape of the database, the authors face the main topics of the paper that are (i) the description of the database in terms of principles and structure, (ii) the origin and quality of metadata and data currently stored and (iii) a collection of analysis on monitoring site's metadata. Then section 5 with conclusions.

In general the paper has a good structure and no substantial changes are required: in my opinion, some paragraphs can be moved, merged or shortened for making the reading more easy (see section comments).

The section 4 is interesting but seems to have the claim of driving the future localization of new permafrost monitoring site. In my opinion, the spatial analysis conducted has a strong arctic bias and are not relevant in mountain ranges or complex topographies. These analysis should be presented as examples of the high-potential of the analysis on big (meta)dataset and not as guidelines to identify new monitoring sites. Also because (meta)data actually stored in the database seems of dubious quality thus still not ready for analysis.

Thus follow my main objection. The “forced” inclusion of e.g. sites with coordinate accuracy of 4 decimals is counterproductive and gave me the impression that a lot of monitoring sites of bad quality (incomplete metadata (374 sites?), bad accuracy (?), big data gaps (?), ...) are actually included in the database (section 3). A global database must collect and maintain only key dataset of high-quality ready for the analysis, no garbage or incomplete data should be included. In this sense your site selection criteria or “minimum quality requirements” are neither explicit nor clear. This lack of quality-standards is not attractive neither for data-provider nor for modelers and data users. Quality section must be deeply reworked with new objectives.

In conclusion, the paper is relevant for the permafrost researchers community because it provides a first overview on a tool which will, hopefully, lead to the publication of big homogeneous datasets fundamental for including permafrost in global models and analyze permafrost evolution at bigger scales. After the suggested major revisions (see

C132

following sections) I think the paper is ready for publication.

Nevertheless, looking at the “Aims and scope” of ESSD, I’m not sure that the datasets actually included in the database match the minimum-standards of this journal. I would suggest a more permafrost-oriented journal but, in this regard, I refer to the editor judgment.

## B. Section comments

Title: Is too long, I suggest something simple like: The new database of the Global Terrestrial Network for Permafrost (GTN-P).

Abstract: Rework the abstract in accordance to the revised version of the paper. A database cannot be an Early Warning System by definition, moreover the words in quotation marks are not suited to scientific articles.

### 1. Introduction

3 paragraphs has no sense, merge all in one.

Paragraph 1.1 is too long. The definition of permafrost, the relation permafrost-climate change and the impacts of permafrost degradation are not relevant for this paper, one general sentence as introduction with 4 or 5 key reference is enough.

Paragraph 1.2 – more than a state of the art, this is an overview on permafrost data collection initiatives around the world. This is boring, but I know, necessary. I suggest to invert the structure: starting from the past (norperm, permis, ...) then the near-present (ipa snapshot) and finally the perspectives (gtn-p).

### 2. Description of the data management system

Paragraph 2.1: I would change the title as: Database design and principles. Site selection criteria or “minimum quality requirements” must be clearly stated here. I would keep here some concepts of paragraph 2.4 (data policy): pangaea, creative commons, downloaded data,...

C133

Paragraph 2.2 and 2.3 and partially the 2.4 can be merged all in one. I would change the title as: Data collection, browsing and publication. Actually the structure and order of the sentences is confusing. I suggest to keep the structure of the proposed title, clearly stating: collection (frequency, principles, modality), browsing (instruments, modality, netCDF...), publication (frequency, principles, references criteria for both published in pangaea (by system admin) or manually downloaded, not clear at the moment).

### 3. Data quality

Also refer to general comments.

Not principles here, just analysis and results on the quality of the data (is not clear if some TSP or CALM data are already available) and metadata actually contained in the database. In other words provide a picture of the completeness (under differing point of view) and characteristics of the data stored in the database. The objective must be the purging of garbage from the database: reduce the numbers for improving the quality. A real challenge. All the relevant quality criteria actually stated in these paragraphs must be synthesized and moved in paragraph 2.1 (see specific comments).

### 4. GTN-P metadata statistics

Also refer to general comments.

Except for paragraph 4.2 all the analysis are spatial-analysis, thus the title of the section is not appropriate. You can change in many way, some examples: Spatial analysis on GTN-P dataset, Analytical potential of the GTN-P dataset, .... or whatever. Such analysis must be conducted excluding bad data (see general comments and specific comment P292,r5-10) and must be treated just as examples of the potential of the data contained in the database (see also general comments).

### 5. Conclusions I would rename this section as Conclusion and perspectives.

Rework this section in accordance to the revised version of the paper. Includes in

C134

this section the future perspective (also financial) and expected development for the database.

### C. Specific Comments

P282,r3-24. Too long and not relevant for the paper (also refer to section comments). Reduce to 1 (max 2) sentence(s) with some key references. You can replace your introduction with a thought on why it's important today aggregate and share scientific data (for inspiring, [http://link.springer.com/chapter/10.1007/978-94-007-2351-1\\_17](http://link.springer.com/chapter/10.1007/978-94-007-2351-1_17)).

P282,r.22. ...northern and high-mountain infrastructure... (add Bommer et al, Permafrost and Periglac. Process. 21: 97–104 (2010))

P282,from r.27. to the end of paragraph. Remove quotation marks. Too long and complex. Simply say that: the monitoring of ECV for permafrost is demanded to the GTN-P (one reference for the history and context of GTN-P) which has activated 2 specific monitoring programs (i) the TSP for permafrost temperature and (ii) the CALM for the active layer.

P283,r9. Merge all paragraph in one (see section comments)

P283,r23. Remove thermal. ...USA are archived...

P283,r26. NSF explain acronym

P284,r5. ...collection of permafrost...

P284,r11-12. ...with the capability to interlink the permafrost community and scientists working in other fields of research like e.g. climate modelers, biologists or engineers.

P284,r17. Remove the concept of early warning system (see also section comments)

P284,r27 to end of paragraph. Review the objective of the paper, also refer to general comments.

P285,r3. Refer to section comment for rebuilding this section.

C135

P285,r16. Replace submitter with provider (everywhere). What do you mean with data product?

P285,r21. What is the sense of storing not-accessible data?

P286,r5-9. Explain better. Add CSV acronym.

P286,r13. Replace "input" with provide

P286,r27. Move "(comma separated values)" to P286,r.8.

P287,r10-13. Not clear, explain better. This seems a big step forward in borehole data homogenization. I would empathize this explaining in more depth the processing steps. A database cannot be a black box.

P287, r14-16. This can be moved into Conclusions and perspective section. What do you mean with "flags"? Why a "regular grid" is needed? (for global model I guess), explain better.

P287,r19. Sentence not clear, what do you mean with "mediates", why the database needs a mediator?

P287,r21-27. Remove, not necessary here to explain what pangaea is.

P288,r1-13. Move this block in the new paragraph 2.1 (are principles)

P288,r14. Please refer to section comments for deeply rebuilding this section. Some possible ideas: what is the completeness of sites' metadata (partially faced in your paragraph 3.3, starts from there); how many sites have also TSP and/or CALM dataseries; how many sites have no dataseries; what is the continuity in time of the available dataseries, what periods, what overlap between periods; what are the most frequent sensors depth (your paragraph 4.2, good); ... You can try also to regionalize the analysis for e.g. arctic, antarctic, alps, himalaya, ... Suggestions: express always the results as percentage when possible not as number of sites (es P291,r.14); remove bad sites from the analysis (es P292,r5-10)

C136

P288,r15. This block is partly a repetition of the introduction (your paragraph 1.2). Avoid this repetition and say “who are the data providers” in the new section 2.1 or 2.2.

P289,r12. What does these dataset fit into a TSP and CALM database? Not clear.

P289,r20-27. Repetition, already said in section 2.1

P290,r9-13. I'm not confident about that, refer also to general comments. Whatever, in my opinion, this cannot be considered a data quality control criteria.

P290,r27 to P291,r3. Not clear what do you mean with this analysis.

P292,r1. Please refer to both general and section comments for reviewing this section.

P292,r2. Interesting analysis. In my opinion has sense only in flat, gentle terrains not in complex morphologies. I would make this analysis only in the Arctic (or better to a smaller area as example) providing 1 or 2 zoom (for figures 5 and 6, see also figures comments) in interesting zones with differing characteristics of the Voronoi cells.

P292,r5-10. Remove this data from the analysis.

P292,r25-26. Move this sentence to P292,r19.

P293,r26. Move this in section 3 (see comment P288,r14).

P294,r13. See comment P292,r2.

P294,r14-19. Repetition.

P295,r11. I'm skeptical about the “monitoring bias” you mention in the slope orientation analysis conducted only on sites north 60°N. This bias exists only if global models approximate the terrain geometry to flat surfaces otherwise the correction of the expected potential solar radiation by slope orientation is quite elementary and surely (or probably) included into a global model. In my opinion a metadata analysis on slope orientation would be more interesting in complex topographies (low-latitude mountains) which present more relevant slope angles. Just a suggestion.

C137

P296,r24. Very good, simple and focused.

P297,r5-7. Not clear, explain better.

P297,r5. Please refer to section comments for rebuilding this section.

D. Figures and tables comments

Figure 1. In my opinion this figure is not relevant, thus not really necessary. Up to you the decision. Caption: replace “from” with collected by

Figure 2. This figure could be very useful for having an overview of the metadata asked by the database, instead just provide the structure. Moreover to show the methods (+index(),+view(),etc...) in each box has no-sense. I suggest to replace in the box below each object the associated metadata e.g. Under the box Site put (id, lat, lon, etc...). Modify the caption in accordance.

Figure 3. In my opinion this figure is not relevant, thus not really necessary. Up to you the decision.

Figure 5. Figure too small. I suggest a zoom-in to a significant area.

Figure 6. Figure too small. I suggest a zoom-in to a significant area.

Figure 9. Figure too small. I suggest a zoom-in to a significant area.

---

Interactive comment on Earth Syst. Sci. Data Discuss., 8, 279, 2015.

C138