

Interactive comment on “Quantifying late-Holocene climate in the Ecuadorian Andes using a chironomid-based temperature inference model” by Frazer Matthews-Bird et al.

Anonymous Referee #2

Received and published: 22 February 2016

1. Does the paper address relevant scientific questions within the scope of CP? YES
2. Does the paper present novel concepts, ideas, tools, or data? YES
3. Are substantial conclusions reached? See comments. I don't think that the Conclusions are sufficiently supported by the data (fit-to temperature, representation of modern analogues in fossil samples, suitability of L Pindo wrt to the training set, possibly also data resolution, etc.)
4. Are the scientific methods and assumptions valid and clearly outlined? (See comments below)
5. Are the results sufficient to support the interpretations and conclusions? Partly

C1

6. Is the description of experiments and calculations sufficiently complete and precise to allow their reproduction by fellow scientists (traceability of results)? See comments below.
7. Do the authors give proper credit to related work and clearly indicate their own new/original contribution? Yes (see comments)
8. Does the title clearly reflect the contents of the paper? No
9. Does the abstract provide a concise and complete summary? YES
10. Is the overall presentation well structured and clear? YES (few minor comments)
11. Is the language fluent and precise? YES
12. Are mathematical formulae, symbols, abbreviations, and units correctly defined and used? Generally YES
13. Should any parts of the paper (text, formulae, figures, tables) be clarified, reduced, combined, or eliminated? NO
14. Are the number and quality of references appropriate? YES (see comments)
15. Is the amount and quality of supplementary material appropriate? I would include the chronology in the manuscript (NOT in SOM)

General remarks: This manuscript presents, to my knowledge, the first chironomid-based Transfer Function in tropical South America from >50 lakes in Peru and Ecuador, and provides a temperature reconstruction from fossil samples of Laguna Pindo (Ecuador, 1200 m) for the past 3000 yrs.

Establishing Transfer-Functions in this part of the world is very important and really novel (and much needed). Very little is known about (Holocene) TT changes because most of the available records are sensitive to precipitation. Indeed, the design of the TS has some problems (e.g. distribution of samples along the TT gradient) but, in

C2

the real world, there is often not much one can do about this. A further shortcoming is that nutrients were not measured for the training set (e.g. Lotter et al 1998) and I would have done some technical details in a different way (sample uppermost 3-4 cm sediments instead on only 0-1 cm; exclude Long/Lat in the TF or test what happens if excluded). However, this is a first start and deserves publication. It might be worth testing whether the TF could be enhanced and optimized by removing stepwise lakes with extreme properties (such as e.g. the very shallow lake with 10 cm water depth).

The temperature reconstruction from the sediments of Laguna Pindo (c. 3000 yrs) stands on much weaker foundations and requires careful (major) revision, further testing and likely adding more samples to test the robustness and reproducibility of the cold spells. The conclusions of the current version of this paper (mainly the cold spells and their temperature) are barely supported by the data and remain speculative. The main challenges with the reconstruction:

Amplitudes for (multi)decadal (10-20 yrs according to the sample resolution) mean annual temperatures on the order of 4°C within the last 500 yrs and ca 7°C for late Holocene TT changes seem unrealistically high compared with what is known from other parts of the world including the tropics (Marcott et al. 2013; PAGES 2k 2013). The finding reported by Polissar et al. 2013 (inferred from dELAs of two glaciers in Venezuela at 4600 and 5000 masl) is an exception, and has been explained with very special local conditions at high elevation sites. Yet a plausible physical explanation is missing for the very large amplitudes of the cold spells found in Laguna Pindo (this manuscript). TT variability in the tropics are a very important issue and, thus, require much better foundations and support by data (including replications). I rather suspect (which should be explored/discussed by the authors) that the large amplitudes are related to problems with the TF, with the fact that many taxa in the fossil downcore samples of L Pindo are not represented in the calibration data set (! Line 495; I think this is a real problem), with the fact that Laguna Pindo is not well represented in the TS lakes (Line497), other variables (such as precipitation) play a major role particularly

C3

in samples with poor fit-to-temperature (which is precisely the case in those samples with large TT amplitudes) and/or with some of the downcore samples (low numbers of hc, poor goodness-of-fit or poor modern analogues). These problems are honestly discussed in the text. It appears that L Pindo was not the best lake to perform a reconstruction (downcore analysis).

In my view (also your statement in Line 531) the TT reconstruction is, thus, rather qualitative than quantitative, it is not known what the (mixed) TT signal actually is. This might, in turn, explain the unrealistically high TT amplitudes of the L Pindo reconstruction.

For a publication in CP I would expect that a few additional samples should be analyzed to assess whether the prominent results (cold spells) are robust and can be reproduced or whether these single samples (e.g. at 20 cm sediment depth) could be artifacts, outliers or coincidence. This information is most relevant for the quality of the paper and the implications. Two examples:

(i) L335: The 'sudden drop at 1600 cal BP' is inferred from just one (1) sample which has a substantial error (Fig. 7), poor goodness-of-fit (Fig 9), no good modern analogue (Fig. 9) and very low hc concentrations (Fig 9); the number of hc is not known. Is this really robust and significant?

(ii) L340: the short minimum around AD 1850: only one single sample at 20 cm; the error is substantial (Fig. 7), goodness-of-fit poor (Fig 9), no good modern analogue (Fig. 9) and very low hc concentrations (Fig 9); again, the number of hc is not known. This requires additional support by data from adjacent sediment samples or parallel cores. In general, counts (number of hc) should be given in all figures (in addition to hc concentrations). This is also important for the assessment of other parameters (e.g. L361-363) and the overall quality/robustness of the results.

Moreover, it is very important (and in my view conditional) that the raw data of the Training Set and TF (environmental variables and chironomids) as well as the raw data

C4

of the chironomid stratigraphy of L Pindo are made available in digital form (as SOM to the publication).

In summary: The Transfer Function is important, although not perfect, but deserves publication. The L Pindo reconstruction has major deficits. I see two options: (i) The TF is optimized, undergoes additional testing, the quality of the reconstruction is substantially improved (robustness of the TT amplitudes, robustness of the cold anomalies, etc.) and/or (ii) the profile of the reconstruction is lowered; given the pertinent deficits the reconstruction is qualitative and NOT quantitative, not overstating the results and conclusions. I keep regarding this as publishable. All data (Transfer Function and reconstruction) should be made digitally available in a data repository. While this is standard in many scientific communities this does not yet seem to be the case with chironomids.

Specific Comments:

The Introduction could be shortened (quite lengthy).

Chapter 3: I would not make too many sub-chapters (only one paragraph in 3.4 and 3.5)

The sampling design for the downcore analysis should be described in detail (continuous, discrete sampling, regular intervals, stratigraphically...?) What is the percentage of sediment that is actually covered in the analysis? (e.g. 1 cm slice every 10 cm sediment makes 10% coverage and 90 % is not covered; this has serious implications regarding the robustness of the reconstruction).

I would include the Suppl Fig (Chronology) in the manuscript.

Title: reconcile. It is quantitative indeed, but how robust and how good are the numbers? => Qualitative

L77: Shulmeister

C5

L93: ... preceding Glacial and Late-Glacial period ... (if you refer to 25-11.7 kBP; 20-25 kPB is not Late Glacial)

L95 ff: Make also reference to Marcott et al 2013. This is the most comprehensive dataset.

L99-100: Growing evidence from the tropics? I'm not sure about this. In fact it is still very controversial whether cold events (depending on the time scale) were globally, hemispherically or regionally synchronous (Wanner et al. 2011 QSR, Neukom et al 2014. NatCC;PAGES 2k 2013). The PAGES 2k Consortium 2013 has shown that, with a few exceptions (with solar-volcanic downturns) multi-decadal long cold phases were not coherent across the globe. Maybe rephrase sentence.

L108-109. References not appropriate (these are not climatology papers). Make reference to Garreaud et al 2009 or Stefan Hastenrath 1991 Climate Dynamics of the Tropics or similar.

L150. ... gradient of 24°C (not 25°C; from 0.8 to 25°C)

L151: How reliable are WTs in a 10 cm deep water body? It should be assessed how sensitive the TF is with/without such lakes. In such water bodies the difference between MAT and WT is typically very large (in particular Tmax). I guess that the TF stats could be improved.

L154: ... uppermost 1-2 cm ... representing 5-20 years... Well, it was done like this and is usually done like this. But this implies that the sample for the Training Set depicts in one lake interannual/subdecadal variability (which may be very different from climatology!) and in another sample it is rather climatology (20-30 yrs). I suspect that this adds substantial errors to the TF. Suggestion: if such large TT gradients are used (24°C) use the uppermost 3-4 cm of sediment to make sure that 20-30 yrs (climatology) are represented. The TT trends during 30 yrs are relatively small and similar in all lakes of the training set.

C6

L171: Fourteen 14C samples? Fig SOM shows six of them. Where are the others?
Pls change and make it consistent with L 324 ff.

L188: Were nutrients (N and P) not measured? This might be a problem (Lotter et al 1998 J Paleolimnology)

L272 and 275: avoid references in the results section. This reads like 'Discussion'

L280 ... optimum... (?)

L295: in general, hc counts should be given in all Figures and Tables. L295 ff is rather Discussion than Results. Move this paragraph.

L300: Yes, this is critical (number of hc). It should be assessed whether the number of hc has an effect on the calibration statistics, in particular the residuals. (see also L303, I am not sure if this is the only criterion according to which the TF could be optimized)

L302: Table 1 does not show these details (which are important), Table 1 shows the summary only. The details (hc) should be given (in the SOM)

L327: The sampling design must be clarified (in the Methods section). You took 30 samples spread over 420 cm. How did you take the samples? 1 cm slice every 10-15 cm? Stratigraphically (according to which criteria?) or continuously (complete sediment section)?

L353: .. only seven samples? According to Fig 9 and the vertical dashed line there are many more.

L485: I don't think that anything is known about the precip/temperature relationship during the Late Holocene.

L 495: I think this is a substantial problem.

L531: according to this statement I would conclude that the temperature reconstruction of Laguna Pindo is qualitative at best.

C7

L539: maybe also refer to Kanner et al (speleothems) and Ledru et al (N Ecuador)

L548: Jones & Mann 2004 is not the best (has been criticized; S-Hemisphere is very poor). Suggestion: PAGES 2k 2013.

L552/553: I don't think this is true. There's a large body of literature pointing out the role of volcanoes, or a combination of S+V ... rephrase sentence.

L555: No, I don't think this is true (cool from 400 yr BP onwards). The sample at 250 yr BP is still among the warmest of the entire record, almost as warm as today (!). There is only 1 sample (at 1850 AD) that shows cool conditions, and it is very questionable how robust that is (see your comment and my comments above)

L569ff: It has been repeatedly demonstrated that the Andean ice cores (stable isotopes) record precipitation and not temperature (as claimed by Thompson et al).

L572: TT drop of 3-4°C during the LIA. Yes, this value has been reported for two Venezuelan glaciers (at 4600 and 5000 masl, mainly inferred from a drop in ELA by 300-500 m; Polissar et al. 2006). I doubt that similar (special high-elevation) conditions apply for L Pindo, given the limitations of the reconstruction (see above). This value seems extraordinarily high to me. Alternatively an explanation should be provided showing that such large TT amplitudes are physically plausible at local scales.

L598. Yes, the potential is shown (with the TF). But the reconstruction has major problems and severe limitations (see above). I would say: qualitative at best.

L605: ...). Special...

L624: reference listed twice

L634: Dryas-Holocene

L667: check carefully

L668: ...Science 289,

C8

L680: Vol missing

L702: Lemke

L810: ... Science 234, ...

L814: Ref listed twice (also L819)

L818: Holocene

L841 Woodward, C.

L855: LOI: specify 550 or 950; ditto L858, Table 1 and Table 2, L883

Table 1: Data set should be made available in full detail

Table 3: add units (where appropriate), also Caption Fig 5

Fig 2: pH

Fig 3 (all Figs where appropriate, Fig 6, Fig 9): numbers of hc should be shown. It would be interesting to see the 'unusual lakes' (e.g. those with water depth of 10 cm).

Interactive comment on Clim. Past Discuss., doi:10.5194/cp-2015-186, 2016.