Clim. Past Discuss., 10, C1224–C1230, 2014
www.clim-past-discuss.net/10/C1224/2014/

Climate of the Past Discussions

Interactive comment on "CREST: Climate REconstruction SofTware" *by* M. Chevalier et al.

M. Chevalier et al.

chevalier.manuel@gmail.com

Received and published: 1 August 2014

R#2: My apologies for the delayed review of this paper, while I bear some of the responsibility, I would like to make clear that I have no problem reviewing a paper on methods, but a paper that purports to discuss software, as this paper does, must include the software and a sample dataset if it is going to make claims about the software.

Authors: We apologise for any confusion, and have changed the title and the few instances where the software was referred to accordingly. The intent of this paper is to describe the methods we have developed and included in the CREST software package, rather than the specific functioning or user-friendliness of the software itself. A sample dataset will be included with the software when it is distributed.

R#2: The paper itself is interesting. The increasing development of paleoecological records from regions that are species rich does pose a quantitative challenge for pollen-

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

based climate reconstruction, and it is good to see that the authors are developing methods to assist in this endeavor. I believe that this method provides a way forward for climate reconstruction, but I believe that several issues with this paper need to be addressed. I think the introductory discussion is well framed, however I don't know of many who believe that WA-PLS is one of the best methods. In recent paleolimnological work by Telford and others it was my impression that WA (using monotone spline de-shrinking) performs better than most methods, and does not suffer from the kinds of spatial autocorrelation issues that other methods appear to have (see discussions around Guiot and de Vernal, 2011). Given the methodological similarity between WA and pft type models it is interesting that WA is not mentioned. Birks et al. (2010) provide an excellent summary of these methods and ought to be cited. I am also surprised that, given recent papers by Salonen et al. (2012) using BRT, machine learning techniques are not mentioned.

Authors: We agree, and will extend this part of the introduction.

R#2: In some cases pollen production can vary across species within a pollen taxon. In this model each taxon is assumed to have the same weight based on presence, however it is not difficult to think of examples where presence does not scale (for example) to biomass within a pollen taxon, and where pollen production varies across a taxon (Cupressaceae offers a good example for both situations), although I have no appropriate South African examples. This seems like a limitation that should be addressed.

Authors: As elsewhere, this is indeed a factor that needs to be kept in mind, and we will highlight this more clearly in the text. That said, comparisons of reconstructions from fossil pollen sequences with independent records (as yet unpublished) indicate that this is not usually a significant limitation. But of course each sequence needs to be evaluated carefully for such factors, and decisions regarding what taxa will be included will depend on the authors' expertise and experience.

CPD
10, C1224–C1230, 2014

R#2: The paper posits that this model should theoretically perform better than MAT or strictly pollen-based methods (as opposed to inverse methods). To back this up I would have liked to see these results compared to MAT predictions. Do they actually perform better? How do these error estimates compare to the simpler WA or MAT reconstructions?

Authors: Apologies if we have given the impression that this model should perform better than other techniques. We certainly haven't claimed this in the text. The goal in the development of this method is to provide a tool that can be used in areas where existing methods based on modern pollen surface samples are not possible/reliable. Pollen preservation in surface samples is very poor in drylands, where soils are often poorly or undeveloped, and other archives (such as ephemeral wetlands, pans, etc.) are not representative of the broader vegetation. We do not claim that ours is the 'best' method, only that it provides a functional and reliable alternative in some regions where MAT, etc. can't be employed.

R#2: I am also curious why the authors reconstruct every climate variable they have. Birks et al. discuss why this is inappropriate, and Telford and Birks (2011) has proposed methods for determining which climate variables may be appropriate for reconstruction, and these are also discussed in Juggins and Birks (2012). It seems to me, for example, that it is simply inappropriate to assume that mean diurnal range is a significant ecological indicator, and as such it shouldn't be considered. Eliminating some of these variables would improve the readability of figures (for example Figure 9 which is almost unreadable) in which all variables are presented.

Authors: Absolutely correct, and this is why we reconstructed such a wide range of variables; to highlight the limitations of the method. As expected, variables describing climate phenomena that directly impact the plant growth/reproduction/survival can be more reliably reconstructed. But we felt that this had to be explicitly demonstrated and described. For Fig9, the landscape page format of the online version of CPD does make it unreadable. However, in usual portrait A4 format, the figure is dense, but

CPD
10, C1224–C1230, 2014

readable.

R#2: Table 4 adds further support. If the R2 is only 4% why reconstruct the variable? Table 4 needs to be clear about what each of the terms means in the table caption.

Authors: As described in the results section, we have tried to explain the repartition of our errors by different variables, namely the number of species, the type of vegetation and the 'availability' of the climate value in the climate space. This table of R2 just indicates the percentage of variance explained by each model for each variable. It should be noted that a small R2 does not indicate how accurately the variable has been reconstructed, just that the factor selected does not describe well where or why the errors exist. Based on this table, we find that the factor that best explains the observed variance is the availability of the climate value in the climatic space. We will expand the caption to explain this more clearly and make suitable changes in the text.

R#2: As it is, I believe that this software presents a methodological advance in many regards, and potentially provides the field with a method that is robust, however I think that the implementation of the software, and the accompanying manual are far from user friendly as purported in the paper. I believe that I am somewhat technologically proficient, but I find the software to be far from straightforward.

Authors: We apologise if we have implied that the software is "user-friendly". We have not claimed this in the manuscript. CREST is scientific software and thus has some degree of expectable complexity. We have left many choices open to engage users to think about their data, but more importantly to allow for a higher degree of freedom. In comparison with the set of equations presented in a paper, we do believe that CREST will facilitate the reconstruction process for most users.

R#2: The reliance on SQL databases surprises me somewhat. In most cases climatic data comes from flat text files (or ASCII grids), rasters or (more recently) netCDF files. I know of few datasets that come specifically bundled as SQL files. This adds complexity to the methods here since it would require a user to also get data into an SQL format.

CPD
10, C1224–C1230, 2014

Authors: Databases seemed to be a logical choice to us, as they are the most efficient format for dealing with what may be vast datasets when hundreds of requests must be performed. Grids can also be easily included in a database (nb: freeware exists to convert flat files (such as csv files) into sqlite3 database file, so it is really easy to create a database from that type of files).

R#2: My feeling is that this paper should be published as a "Methods only" paper, excluding the software. The software should then be published separately in a disciplinary journal (Computers and Geosciences for example) with the example from the manual presented directly. Authors: This was intended to be a methods paper, and any necessary changes will be made in the text to indicate this. The associated software will be available via an external link for interested parties.

R#2: In this case the Manual should be improved with my suggestions above and the technical discussion in CotP should be revised somewhat to (1) indicate limitations of the current method, and to provide comparison to other standard methods (with fewer climate variables), although it's not clear to me whether the extent of data in South Africa would support this.

Authors: Indeed, suitable data for a comparison with other methods is not available/obtainable from much of southern Africa, which was the rationale for the development of the method presented.

R#2: Some notes on the manual: 1. The link to download the python software gives a 404 Page not found error. 2. The instructions for installing the easy_install are not clear. The link takes me to a file called ez_setup.py, not easy_setup.py. There is no clear executable file, unless the authors are talking about ez_setup.py. 3. The instructions for installing pyRserve do not seem to install pyRserve. 4. I found the manual very frustrating to follow. While the example is straightforward to some degree, the structure of the manual means that figures do not follow from the users' experience. For example, the user needs to scroll down two pages (and past two figures) after

CPD
10, C1224–C1230, 2014

loading the SQLite3 file before they can tell whether they've done things right. 5. This structure means that user steps get buried in the manual. 6. Figure references are used throughout the text as [8.1], [8.2] where the numbers refer to numbered boxes in the figures, but the figure captions do not refer to these numbered boxes. The figure caption should identify each of these numbered boxes and providea small stand-alone explanation for each box, especially since the h yperlink for the references shows only the caption and not the figure. 7. The step at Figure 11 is not straightforward. It would make it significantly easier for the user if they could either copy & paste, or build the request from drop down options.

Authors: We appreciate and will consider these detailed comments on the manual and execution of the software, but as this is a methodological paper, these will only be include in the distributable software and manual. Please keep in mind that we cannot be responsible for the third-party providers. Web pages will change, and not all elements are equally easy to install. Generally, however, the necessary information and advice can be found through searches and online forums.

R#2: 8. The program crashed after I performed the second re-construction: Traceback (most recent call last): File "C:\Users\Late Reviewer\Dropbox\Review\CPD\CREST\CREST\ResultFrame.py", line 505, in OnSaveButton s=self.make_SWM_effect([self.Robj.distrib_list[x] for x in idx],[self.Robj.sp_poids[x] for x in idx],[self.Robj.prob[tt][x][j] for x in idx]) File "C:\Users\Delinquent Reviewer\Dropbox\Review\CPD\CREST\CREST\ResultFrame.py", line 666, in make_SWM_effect s+=distrib_list[pol][0].upper()+"\t"+distrib_list[pol][1][0]+"\t%.5f\t%.5f\n"%(sp_poids[pol][0] TypeError: cannot concatenate 'str' and 'int' objects 9. The manual could use revision for language throughout.

Authors: That bug was caused by the example dataset we quickly developed for this review. It has been corrected now. Thanks for the feedback.

References

CPD
10, C1224–C1230, 2014

Birks, H. B., Heiri, O., Seppä, H., & Bjune, A. E. (2010). Strengths and weaknesses of quantitative climate reconstructions based on late-Quaternary biological proxies. Open Ecology Journal, 3, 68-110.

Guiot, J., and A. de Vernal, 2011a: Is spatial autocorrelation introducing biases in the apparent accuracy of paleoclimatic reconstructions? Quaternary Science Reviews, 30 (15-16), 1965-1972.

Guiot, J., and A. de Vernal, 2011b: QSR Correspondence "Is spatial autocorrelation introducing biases in the apparent accuracy of palaeoclimatic reconstructions?" Reply to Telford and Birks. Quaternary Science Reviews, 30: 3214-3216.

Juggins, S., & Birks, H. J. B. (2012). Quantitative environmental reconstructions from biological data. In Tracking environmental change using lake sediments (pp. 431-494). Springer Netherlands.

Telford, R. J., & Birks, H. J. B. (2011). A novel method for assessing the statistical significance of quantitative reconstructions inferred from biotic assemblages. Quaternary Science Reviews, 30(9), 1272-1278.

Salonen, J. S., Seppä, H., Luoto, M., Bjune, A. E., & Birks, H. J. B. (2012). A North European pollen–climate calibration set: analysing the climatic responses of a biological proxy using novel regression tree methods. Quaternary Science Reviews, 45, 95-110.

Interactive comment on Clim. Past Discuss., 10, 625, 2014.