

## ***Interactive comment on “***

# **A nonlinear multi-proxy model based on manifold learning to reconstruct water temperature from high resolution trace element profiles in biogenic carbonates” by M. Bauwens et al.**

**M. Bauwens et al.**

msbauwen@vub.ac.be

Received and published: 14 October 2010

### RESPONCE ON GENERAL COMMENTS:

- Can you explain why many studies find linear relationships while other fail to find such linear systems?

C457

A big problem in sclerochronology remains shell growth rate, which is mostly not known. As recognized by the reviewer, unless shell growth is monitored intensively, growth rate remains an unknown and the relation between proxy and external condition will depend largely on the assumptions made about shell growth rate. In such cases forcing the growth axis to yield a linear relationship between proxy and environmental condition is often the most logical step.

- Can you compare qualitatively the manifold approach with the linear approach. What are the advantages and disadvantages of the manifold approach. Where will it fail?

I recognize it might seem odd that the WDMR model is described without first mentioning the potential of linear multiple regressions. Therefore, I have added a small paragraph on linear multiple regression in which I answer your question:

#### “3.1. Linear multiple regression

Linear multiple regression is the most commonly used multivariate model to describe the linear relationship between two or more explanatory variables (here proxies) and a response variable (here temperature). This is done by fitting a linear equation to observed data. An equation similar to the equations in equation 1 describes how temperature co-varies with the proxies. A limited number of parameters  $\alpha_1, \alpha_2, \dots, \alpha_n$  define the slope of the regression line and a coefficient  $C$  defines the offset. The main advantages of linear multiple regressions are that appropriate toolboxes are available on all statistical software packages and that models have a limited number of parameters and model outputs which renders interpretation easier. A large disadvantage, however, is that linear models are not able to fit nonlinear relationships which are likely to occur in biogenic archives.”

- Why did you reconstruct only temperature and not for example salinity and temperature from the proxy records?

Both modeling methods (WDMR and linear multiple regression) are multiple-input,

C458

single-output methods. So, besides the assumption that the studied proxies do contain some salinity information, the methods could indeed be used to compute salinity models. Unfortunately I could not test this since for most of the proxy records contemporary salinity information is not available for all data samples.

- Elaborate on how the trace elements were gathered or give the appropriate reference.

In the section 'raw data' and 'data pre-processing' I do refer to Vander Putten et al. (2000) and Gillikin et al. (2006a), Gillikin et al. (2006b). These are the papers in which the original proxy data have been published and described. I reformulated the paragraph to indicate to the interested readers where to find detailed information about the used data:

"The reader interested in more details about these data sets is referred to the papers by Vander Putten et al. (2000) and Gillikin et al., (2006a; 2006b)."

- Personally, I do not know statistical studies which are clearly able to explain physical or biological effects from a black box model. For that reason, I think that the interpretation in terms of kinetic effects is rather speculative. On the other hand, I am aware that most readers will probably be most interested in this interpretation, rather than in the statistics.

I agree, black box models do not directly explain physical or biological effects, but I do believe that the fact that the models are working properly does have a physical or biological basis. Therefore I think the results deserve to be discussed in this direction, although I agree that the discussion is speculative. However, as mentioned in the answer to comment#1 of reviewer#1, I realized that the statistical novelty of this study was not clearly stated; therefore I slightly adapted the introduction and the abstract. I trust it now is clear to the reader that the reasons to use nonlinear multi-proxy models are purely statistical. Nevertheless, the results obtained by the statistical black-box models may still contribute to improving the mechanistic understanding of the processes

C459

involved.

- Systematic errors: I am wondering how powerful this manifold approach is. Is it possible that it is too powerful and can for example mimic systematic errors. If for example errors in the time base of the proxies are present or due to e.g. time averaging. Will the manifold method be able to incorporate such systematic errors? Do you see possibilities to avoid this?

I do believe the model incorporates systematic errors when they occur in the training set. For example averaged proxy data will be coupled to the real temperature measurements. When the model is used to reconstruct the temperature from another dataset that is obtained using the same analytical methods, the proxy data that are also averaged will be translated to real temperature data as well. So I personally believe that this 'mimicking of systematic errors' is actually a benefit. However, it may raise some problems, for instance when a model is extrapolated to data which were obtained under different analytical conditions (e.g. higher sampling resolution, less time averaging). However, it is difficult to quantify the impact of different analytical conditions, and more tests on a broader variety of data are needed to answer your question quantitatively. Therefore, I wrote that readers are encouraged to try out these methods on other datasets.

- Do you think that this method can easily be extrapolated to other species, like corals, etc... Can you use the manifold approach trained on mussels and apply them on corals? How good do they apply on e.g. mussels from different regions... In theory any data can be used to compute a new WDRM model. However, the reconstruction performances will depend on the environmental information available in the proxies. Some preliminary tests revealed that WDRM work best when the proxy-environment relationships (linear or nonlinear) are visible (e.g. in a scatter plot). The preliminary tests that we did were done on bivalve shells, but we encourage all readers to try the method for other climate archives such as corals, foraminifera or fish otoliths.

C460

Extrapolating the mussel model to corals may be a bit too farfetched at present. As mentioned by the reviewer, the model might contain systematic errors that are propagated differently on data sets that substantially differ (the case for data sets obtained for different species). To extrapolate a bivalve-model to corals you have to assume that the proxy incorporation processes are not species specific. I personally don't believe this is the case. Nevertheless, some preliminary tests showed that it may be possible to extrapolate the mussel model to oyster producing a calcitic shell.

From our results it appears realistic for the model to be extrapolated to different regions. Indeed, the extrapolation of a training set from Terneuzen to a validation set of Knokke, yielded good results. Although this two sites are located relatively close to each other, they strongly differ in their physico-chemical characteristics. The Terneuzen site is an estuarine environment with a lower salinity, while the Knokke site is a saltier coastal environment. Therefore, I do not expect extrapolation to regions different from those studied here to be problematic.

To clarify this I made some changes in "section 6.1.";

"Several studies reveal that trace element profiles in shells may vary significantly between successive years (Barats et al. 2009) and between different study sites (Gillikin et al. 2006a). Our study as well reveals year to year and site specific variations (see Figure 3). However, the accurate temperature reconstruction based on the shell from Knokke sampled during a different year, at a different site relative to the training site suggests that the models are relatively robust against year to year and site specific variations in trace element composition. Moreover, even though the distance between Terneuzen and Knokke is not more than 40 km the two sites strongly differ in environmental conditions: the Terneuzen site is an estuarine environment with a lower salinity compared to the Knokke site which is a more saline coastal environment, therefore the model is probably fit for application to a wider environment than studied here. However, we did observe differences in site specificity for different proxy combinations (e.g. the Ba/Ca problem that is observed for the Breskens site) and therefore the site specificity

C461

of every proxy has to be investigated. This needs to be done independently and in combination with other proxies before a model based on a specific proxy combination can be extrapolated to a broader environment. "

I add a sentence to "section 6.2."

"Although we do not expect this model to be directly applicable to other species because Mg/Ca (the main player in the temperature reconstruction) is assumed to be driven by a physiological temperature response that is probably species specific, some preliminary tests suggest that the models may be extrapolated to other bivalves with calcitic shells. Moreover it should be possible to generate a specific WDMR model for other substrates such also corals, trees, sediments ..."

- To explain anything, it is always better to model this by a more complex model. Non-linear models are usually more complex than linear ones. (How) does WDMR penalize for complexity?

WDMR is a non-parametric method. The number of parameters therefore grows with the number of data (if you have  $N_e$  data-estimation samples and  $N_v$  data-validation samples the number of model parameters is  $N_e+N_v$ ). However, there also are a number of design parameters. The number of design parameters depends on what kernel is used in WDMR (usually 2-3 design parameters). Yes, there is a penalty for complexity, a so called regularization. The regularization term is "designed" to penalize estimates that are not smooth along the manifold. The WDMR criterion takes the form:  $\text{Measure-OfFitToTheEstimationData} + \lambda \cdot \text{Regularization}$  This criterion is minimized with respect to the  $N_e+N_v$  model parameters (which is equal the temperature predictions for climate reconstruction).

RESPONCE ON SPECIFIC COMMENTS:

p 1108, line 7: elaborate on this type of multiproxy model. E.g. dynamic effects are not taken into account; relation distance-time is assumed to be known.

C462

At this stage of the paper the equations (1) are not given to describe a certain type of multi-proxy model but rather to give the reader a feeling about the (algebraic) benefit of using multi-proxy models in general. However to clarify that this general idea (illustrated here on a linear model), can be extrapolated to other model types I added a paragraph about this to the introduction. In Section “2.2.1. Linking proxy data to environmental information” I give details about the distance time assumption. And full details about the two methods used in this paper are given in the new section “3. The methods”. In this paper no dynamical effects are taken into account. Although I do believe this is a very interesting research path, I think introducing the multi-proxy idea, the idea of using nonlinear models and the idea of using dynamical models in a single paper would lead to confusion.

I added the following to the introduction:

“Note that solving a non-linear model with a large number of parameters is much more complex, but the idea behind it would be the same. Although it is algebraically possible to reverse such multiple regression equations when there are as many proxies as environmental parameters, this would induce large errors on the estimated parameters. Therefore the multi-proxy models obtained in this paper are considered as black-box models that cannot be reversed to obtain a mechanistic understanding in the proxy incorporation.”

p 1108, line 20: do you mean weighing by measurement uncertainty or by prior knowledge about the proxy (Mg is good, Sr less,...).

I added:

“The weight given to a proxy depends on the quality of the proxy environmental relationship in the calibration or training set and less importance is given to proxies that show a less clear or noisy relationship with the environmental condition (temperature). Noise may result from the large influence of an additional environmental or biological condition or from measurement uncertainty. This means that proxies that have a large

C463

load of environmental information have the largest influence on the final reconstruction, even though other proxies are used to explain or confirm parts of the signal.”

p 1110, line 14: what are the underlying assumptions of each of these methods and is it possible to relate these to the success of these methods to predict temperature?

I added:

I extended the paragraph to: “Two of the three nonlinear regression methods explored in that paper reduce the multi-proxy problem into a single dimensional problem by observing that the proxies lie on a one-dimensional manifold. One of the two is based on intuition and tailored for temperature reconstruction using bivalve shells. The other is a new system identification approach, Weight Determination by Manifold Learning (WDMR), and based on manifold learning. The third approach, Support Vector Regression (SVR), does not rely on an assumption of a manifold in the proxy space; it rather increases the dimensionality of the problem by creating ‘new proxies’ from nonlinear combinations of the original proxy data. In Bauwens et al. (2010) it is concluded that manifold based methods are the most powerful tools for reconstructing paleo-environmental conditions based on proxy records in shells of short-lived bivalves, suggesting that the proxy-environmental relationships are straightforward and no extra information is gained by using a more complex SVR model. “

p 1111, line 12: matching temperature and proxies by e.g. anchor point method implies that the investigator has already some idea about the relationship to be expected. I assume a linear relation was used here to construct a time base for the proxy. Does this "choice" for the time base influence further results? Are these nonlinear multiproxy models able to mimic systematic errors (systematic measurement errors, time base errors, time averaging errors, etc...)? Is it possible to solve the calibration problem (relating proxies to environmental parameters) without solving the time problem?

Indeed, a linear relation between a proxy (mostly  $\delta^{18}\text{O}$ ) and its environment (mostly temperature) is often supposed when the anchor point method is used. However, this

C464

is not always the case. Many studies use a more objective way to date an anchor point (e.g. calcein marking, or monitoring of the shell growth). The shells used to train the model were sized at 2 occasions, providing 3 anchor points between which growth is assumed linear. Thus, bias of the proxy-environmental relationships for the training data will be minimized.

However, I am aware of the systematic errors that occur in the time base, since growth is never linear. . . I believe that the introduction of an artificial memory by using dynamical models can solve this problem. However, I think this problematic is behind the scope of the present paper: Introducing the multi-proxy idea, the idea of using non-linear models and the idea of using dynamical models in a single paper would lead to confusion.

p 1112, line 14: second disadvantage is unclear. I changed it into: "The disadvantages, however, are that (1) some potential useful information may be lost and (2) temperature reconstructions are not possible from individual measurements since the model extracts its information out of the overall shape of the proxy record, and not out of discrete data."

p 1114, line 12: what do you mean with laser ablation crater? Leave this out or elaborate. This was left out

p 1114, line 17 (paragraph 3.2): what is the conclusion? Is it worth using NL models? This becomes clearer by restructuring the text (please see reply to next comment) p 1114 (paragraph 3.3): provide more content. The aim is unclear. I think the readability is improved by reconstructing the text: put results immediately after the test. Now, you first describe the tests, where it is unclear what the outcome is and afterwards the results are discussed. I agree. I restructured the text, and hopefully the structure of the paper is clearer now that the results are given immediately after describing the test.

p 1115, line 4: change "4" to "four". Done

C465

p 1116, line 23: improvement factor: how to read this: is this a number between -1 and 1, where 1 is good and -1 bad? I admit that the explanation about the contribution factors was rather confusing although the principle is very simple.

I changed the paragraph and hope everything is clear now: "The RMSE values obtained on the validation data were used to quantify the model performances. For the nonlinear models seven unique contribution factors were defined in order to quantify the contribution of each proxy. Every contribution factor quantifies how much a specific proxy contributes to a specific model; in other words the contribution factor informs on how much the RMSE decreases by including the information of the investigated proxy into a specific model. For example one of the seven contribution factors for Mg is 1.62. This means that the RMSE of a MgSr model was 1.62 lower than the RMSE of a Sr-only model. Negative contribution factors, on the other hand, reflect that including a specific proxy in the model has a negative influence on model reconstruction performance. All contribution factors are defined as the difference between the RMSE between two model configurations (i.e. models run with different combinations of elemental ratios) (Table 1). This enables evaluating model performance change due to inclusion of additional proxies."

p 1117, line 8: a linear model is, generally spoken, a special case of a nonlinear one. So, by applying a model selection criterion, the optimal model complexity could be selected. That's tough, and very nicely formulated. I'll add this sentence to the paragraph.

p 1118, line 6: elaborate, the conclusion is unclear. I believe the conclusion is clearer with the text now being restructured. The previous paragraphs 'salinity robustness' and 'proxy evaluation' have one common conclusion now.

p 1119, line 24: SST, use temperature, as is done in the rest of the manuscript. Done

p 1120, line 5: additive noise and model errors: is it true that these terms are used without any explanation in this paragraph. So I extended the paragraph to clarify.

C466

Moreover I removed the term 'additive noise' since I think model errors are of major importance in the Barium context. "The failure of the Ba-model at the Breskens site is probably due to these model errors. Indeed Figure 3 shows that Breskens is the only site where a second Ba-peak is observed, although the temperature profiles at the three study sites monitored in 1996 are very similar. The model trained on shells of Terneuzen incorrectly couples the barium peak to temperature increase since all training shells independently showed a Ba peak coinciding with temperature increase in spring. As a result the model provides a similar interpretation for the second Ba peak observed for the Breskens shells although the origin of this second Ba peak is probably different. To avoid this kind of model errors it is not recommended to use Ba/Ca ratios as stand alone temperature proxy."

Please also note the supplement to this comment:

<http://www.geosci-model-dev-discuss.net/3/C457/2010/gmdd-3-C457-2010-supplement.pdf>

---

Interactive comment on Geosci. Model Dev. Discuss., 3, 1105, 2010.

C467