On Manifolds, Climate Reconstruction and Bivalve Shells

Henrik Ohlsson, Maite Bauwens, Lennart Ljung

Abstract—To estimate the past climate, for example the ocean temperature 1000 years ago, one has to turn to naturally occurring climate recorders. There exist a number of climate recorders in nature from which the past temperature can be extracted. However, only a few natural archives are able to record climate fluctuations with high enough resolution so that the seasonal variations can be reconstructed. One such archive is a bivalve shell. The chemical composition of a shell of a bivalve depends on a number of chemical and physical parameters of the water in which the shell was composed. Of these parameters, the water temperature is probably the most important one. It should therefore be possible to estimate the water temperature for the years the shell was built, from measurements of the shell’s chemical composition. In this paper, we explore this possibility. We do this by first observing that the chemical compositions lie on a one-dimensional manifold parameterized by the water temperature. This manifold is then utilized in the regression to obtain accurate estimates of past water temperatures.

I. INTRODUCTION

The climate debate has not only increased the interest for the recent changes in the climate but also for how it has been changing in the past. For the last hundreds of years, people have recorded the climate by writing down their observations. We have therefore rather good knowledge concerning the climate during this time. However, going further back in time, very little is known. To estimate the past climate, say the temperature in the ocean or in an estuarine 1000 years ago, one has to turn to naturally occurring climate recorders. A variety of natural climate recorders exist. One of the most obvious recorders is trees [1]. The ring-widths of a tree give an indication for how much the tree grew. A low ring density indicates rainy years. As trees, most natural climate recorders are not capable of recording fluctuations faster than in the orders of years. In marine systems, one natural climate recorder that is capable and grows fast enough to catch the seasonal fluctuations is bivalve shells.

The way bivalves build their shells is complicated and affected by several physical and biological parameters, like the salinity of the water [2], the chemical composition of the water [3] and the maturity of the organism [4]. The temperature of the water, however, is probably the most important parameter influencing the chemical composition of the shell [5].

The most commonly used temperature proxies (substances used to determine past temperature) are oxygen isotopes. The relation between $^{18}O/^{16}O$-ratio in biogenic calcite and temperature was first described in [6]. Wanamaker and colleagues, see e.g. [7], studied the relation between water temperature and the $^{18}O/^{16}O$-ratio in the shell of the bivalve *Mytilus edulis* (blue mussel). They obtained temperature estimates with a mean absolute error of 0.57°C. However, the $^{18}O/^{16}O$-ratio or salinity of the water had also to be known which made the method impractical.

Klein and colleagues [8] instead proposed to use the Mg/Ca-ratio as a temperature proxy and provided a linear estimate giving temperature estimates with a mean absolute error of approximately 1.5°C. Klein’s expression was based on data from the bivalve *Mytilus trossulus*. Wanamaker [2] provided two similar expressions for *M. edulis*. These methods gave temperature estimates with a mean absolute error of 2.4°C using Mg/Ca-ratios and 2.9°C using Sr/Ca-ratios.

Until now, only one proxy (of the shell) has been used at a time to estimate temperature. Many studies, however, underline the need for so called, multi-proxy models [8], [9], [10], [11]. For other species, like gastropods [12], corals [13] and foraminifers [14], the first step toward a linear multi-proxy approach has successfully been taken and given temperature estimates with a mean absolute error as low as 0.3°C [13].

In this paper we take a novel nonlinear multi-proxy approach to temperature reconstruction using *M. edulis*. More specifically, we make the assumption that the measures of the chemical compositions lie on a one-dimensional manifold in the measurement space. The motivation for a one-dimensional manifold comes from that the shell composition mainly depends on the temperature in the water. As the temperature changes, the shell composition will change as well and trace out a path, a one-dimensional manifold, in the space of chemical compositions. Quite intuitive and motivated by that most things in nature behaves in a continuous manner, the temperature is assumed to vary smoothly along this manifold. The manifold can therefore be seen as parameterized by temperature.

Manifold learning is an umbrella term for methods for parameterizing manifolds. A manifold learning method can hence be used to find a parameterization of the one-dimensional manifold in the chemical composition measurement space. However, this parameterization will not coincide with the temperature by itself. To adjust the parameterization, and imitate the temperature parameterization, Weight Determination by Manifold Regularization (WDMR, [15]) can be
WDMR, just like a manifold learning algorithm, finds a parameterization of a manifold. The parameterization computed by a manifold learning algorithm is usually constrained to have a unit variance, zero mean etc. (depending on the manifold learning algorithm used). This constraint is in WDMR replaced by fixating a number of values of the parameterization. If we for example know the associated temperature of some chemical composition measurements, this can be used to fix the parameterization to imitate the temperature parameterization.

The paper is organized as follows: Section II gives a brief presentation of the data sets used in the paper. In Section III we explore the manifold in the data. Section IV gives an introduction to manifold learning and Section V describes the tool used for fixating the parameterization of the manifold, WDMR. We end by showing the results and by a conclusion in Section VI and VII.

II. DATA

The two datasets used in this paper were provided by Vander Putten and colleagues [16] and Gillikin and colleagues [3], [17]. The experimental setup i.e., the monitoring and chemical analysis were similar for these two datasets. Both datasets consist of high resolution measurements of Mg/Ca, Sr/Ca, Ba/Ca and Pb/Ca in the common blue mussel (M. edulis) grown in the Scheldt estuary in the Netherlands and Belgium.

The measurements by Gillikin provide data from one shell from Knokke, monitored from February to September, 2002. For the blue mussel this corresponds to measurements of a complete growth season since the blue mussel stops growing during the winter. Vander Putten provides measurements from seven shells from Terneuzen, one shell from Ossenisse and one shell from Breskens. The measurements from these shells are from April to June, 1996. These data do not cover the whole growth season of the blue mussel, but only the spring during which the shells grow the most and the variation in the trace elements are the greatest.

Measurements of the chemical composition of the shells, the regressors, were made using a distance scale, starting at the hinge connecting the two valves and moving along the growth axis. Temperature measurements, on the other hand, were made using a time scale. Linking chemical composition measurement to the temperature measurements is thus not evident. In both studies the linkage was made by the so called anchor point-method described by Paillard and colleges [18]. This means that distance along the growth axis is assumed proportional to time i.e., a linear relation is assumed between time and the growth of the mussel. This assumption is an approximation since the growth of a shell is never linear [19]. To somewhat reduce this source of errors, chemical composition measurements were averaged together with their closest neighbors (along the growth axis of the mussel). This also had the affect of smoothing-out fast fluctuations due to fast environmental changes (of a different timescale than fluctuations in water temperature). The chemical composition measurements were normalized, shell by shell, by setting the variance to one and mean to zero.

The data was further divided into two parts: An estimation dataset consisting of measurements from 6 shells from Terneuzen, located in the middle of the Scheldt estuarine, and a validation dataset consisting of 4 shells from 4 different study sites along the Scheldt estuarine. The estimation dataset was used to estimate and tune the model and the validation dataset to evaluate the computed model’s performance. Using data from 4 study sites along the Scheldt estuarine, allowed testing the place specificity of the model. The fact that the shell of Knokke is sampled during another year also gave the possibility to check the time and site specificity of the computed model. Fig. 1 gives an overview of the geographical locations of the shell sites.

III. EXPLORING THE MANIFOLD IN THE REgressor SPACE

The shell measurements would lie on a one-dimensional manifold, parameterized by temperature, if the chemical composition only depended on temperature. However, the chemical composition of the shell gets affected by other environmental parameters, such as salinity and chemical composition of the surrounding water. The measurements can therefore not be expected to perfectly lie on the manifold. To show that this manifold structure is significant and usable is not easy and we will postpone the strongest argument until the end of the next section. For now we only justify the manifold assumption by plotting the Sr/Ca, Mg/Ca and Ba/Ca concentration ratios from five shells, see Fig. 2. The measurements’ associated temperatures are indicated by the color of the measurement point.

A very important observation from the figure is that the temperature seems to vary smoothly along the manifold but not smoothly in the measurement space. In other words, the
temperature varies smoothly in a geodesic sense but not in a Euclidean sense. Two chemical composition measurements can hence be close (in a Euclidean sense) but have very different temperature associated with them. A regression algorithm based on Euclidean distance is hence expected to perform poorly. For example, *K*-Nearest Neighbor Average (K-NN, see e.g. [20], p. 14) tuned for the best number of neighbors, will give a mean absolute error of 1.71°C for the Knokke data set. We will see that if we instead utilize the manifold structure of the regressor data and use a geodesic distance measure, we will be able to compute better estimates.

IV. MANIFOLD LEARNING

Manifold learning is a fairly new research area aimed at finding, as the name suggests, descriptions of data on manifolds. The area has its roots in machine learning, and is a special form of nonlinear dimensionality reduction. Some of the most known manifold learning algorithms are isomap [21], *Locally Linear Embedding* (LLE, [22]), Laplacian eigenmaps [23] and Hessian eigenmaps (HLLE, [24]).

All manifold learning algorithms take as input a set of points sampled from some unknown manifold. The points are then expressed in a parameterization of the manifold (a set of points of the same dimension as the manifold) by searching for a set of new points preserving certain properties of the data. For example, Laplacian eigenmaps tries to preserve the Euclidean distance between neighboring points. Isomap tries to preserve the geodesic distances *i.e.*, the distance along the manifold, between points and locally linear embedding and Hessian eigenmaps make assumptions about local linearity and point neighborhoods which are aimed to be preserved. Most manifold learning algorithms will therefore not give an explicit expression for the map between high-dimensional points and their associated parameterization values.

We will here focus on locally linear embedding since WDMR, that we will use in the subsequent section to estimate water temperature, is an extension of locally linear embedding.

A. Locally Linear Embedding

*Locally Linear Embedding* (LLE, [22]) is a manifold learning technique which aims at preserving neighbors. For a given set of regressors \( \{x_1, \ldots, x_N\} \), in our case chemical composition measurements, residing on some \( n_r \)-dimensional manifold in \( \mathbb{R}^{n_r} \), LLE finds a new set of points \( \{z_1, \ldots, z_N\} \), \( z_j \in \mathbb{R}^{n_c} \) satisfying the same neighbor-relations as the original regressors. The \( \{z_1, \ldots, z_N\} \) are the regressors \( \{x_1, \ldots, x_N\} \) expressed in the parameterization of the manifold. The LLE algorithm can be divided into two-steps:

**Step 1: Define the weights \( w_{ij} :s \)**

Given data consisting of \( N \) real-valued regressors \( x_i \) of dimension \( n_x \), the first step minimizes the cost function

\[
\mathcal{E}(w) = \sum_{i=1}^{N} \left\| x_i - \sum_{j=1}^{N} w_{ij} x_j \right\|^2
\]  

under the constraints

\[
\begin{align*}
\sum_{j=1}^{N} w_{ij} &= 1, \\
w_{ij} &= 0 \text{ if } \|x_i - x_j\| > C_i(K) \text{ or if } i = j.
\end{align*}
\]  

Here, \( C_i(K) \) is chosen so that only \( K \) weights \( w_{ij} \) become nonzero for every \( i \). In the basic formulation of LLE, the number \( K \) and the choice of lower dimension \( n_c \leq n_x \) are the only design parameters, but it is also common to add a regularization

\[
F_r(w) = \frac{p}{K} \sum_{i=1}^{N} |w_{i1}, \ldots, w_{iN}| \left[ \sum_{j=1}^{N} w_{ij} \right] \sum_{j, w_{ij} \neq 0} \|x_j - x_i\|^2
\]

to (1a), see [22].

**Step 2: Express the regressors in the parameterization of the manifold**

In the second step, let \( z_i \) be of dimension \( n_z \) and minimize

\[
\Phi(z) = \sum_{i=1}^{N} \left\| z_i - \sum_{j=1}^{N} w_{ij} z_j \right\|^2
\]  

with respect to \( z = [z_1, \ldots, z_N] \), and subject to

\[
\frac{1}{N} \sum_{i=1}^{N} z_i z_i^T = I
\]  

using the weights \( w_{ij} \) computed in the first step. The solution \( z^{LLE} \) to this optimization problem is the regressors expressed in the parameterization of the manifold. By expanding the squares we can rewrite \( \Phi(z) \) as

\[
\Phi(z) = \sum_{i,j} \left( \delta_{ij} - w_{ij} - \sum_{k} w_{ik} w_{kj} z_j^T z_j \right)
\]  

\[
\equiv \sum_{i,j} M_{ij} z_j^T z_j = \sum_{k} \sum_{i,j} M_{ij} z_k z_j = Tr(z M z^T)
\]
with $\mathbf{M}$ a symmetric $N \times N$ matrix with the $ij$th element $M_{ij}$. The solution to (2) is obtained by using Rayleigh-Ritz theorem [25]. With $v_i$ the unit length eigenvector of $\mathbf{M}$ associated with the $i$th smallest eigenvalue,

$$
\left[ v_1, \ldots, v_{N_c} \right]^T = \arg \min_{\mathbf{z}} \Phi(\mathbf{z}) \quad \text{s.t.} \quad \mathbf{z} \mathbf{z}^T = \mathbf{N} \mathbf{I}.
$$

If LLE is applied to the chemical composition measurements of the 10 shells, it can hence give us the regressors expressed in a parameterization of the assumed one-dimensional manifold, which has been done in Fig. 3. In Fig. 3, the regressors expressed in the one-dimensional parameterization is plotted against the measured water temperature. As seen, a linear estimate in the LLE parameterization would achieve a reasonably good estimate of the temperature. This is a strong argument for the existence of a manifold in the regressor space. If it would not be possible to find a parameterization with a clear relation to temperature, the assumption of a manifold parameterized by temperature would not be good.

We have now obtained a way to estimate the temperature from the chemical composition. However, we will obtain a better estimate using WDMR, as seen in the next section.

V. WEIGHT DETERMINATION BY MANIFOLD REGULARIZATION (WDMR)

LLE finds a parameterization satisfying the constraint $\frac{1}{N_c} \sum_{i=1}^{N_c} z_i^2 = I$, see (2b). Since the temperatures associated with the measurements probably do not satisfy this constraint, LLE will not give a parameterization equal to the temperature. To adjust the parameterization of LLE, Weight Determination by Manifold Regularization (WDMR [15], related formulations have also been developed by [26]) can be used. WDMR is a regression method that parameterizes a manifold, just like a manifold learning method. However, WDMR uses some estimation regressor-parameterization pairs to fixate the parameterization. The temperature, parameterizing the manifold in the space of chemical composition measurements, can therefore directly be found using WDMR.

We can at this point assume that the regressors (chemical composition measurements) are ordered such that the labeled estimation regressors (i.e., the regressors for which we fixed the parameterization using measured temperatures) are the $N_e$ first regressors $\{x_i\}_{i=1}^{N_e}$, and the remaining $N_e$ regressors are the unlabeled regressors which we seek a parameterization value for, $\{x_i\}_{i=N_e+1}^{N}$.

For the first step of the LLE algorithm, (1) to all regressors $\{x_i\}_{i=1}^{N}$ to obtain $\mathbf{M}$. This matrix contains the information of which regressors that are neighbors with which. In LLE we would now search for the regressors expressed in the parameterization using $\mathbf{M}$ and constrain these to satisfy $\frac{1}{N_c} \sum_{i=1}^{N_c} z_i^2 = I$. In WDMR, we remove this constraint and replace it by $N_e$ examples of regressors and their associated values in the parameterization of the manifold, the $\{x_i, \bar{z}_i\}_{i=1}^{N_e}$ pairs. This is done by

$$
\min_{\lambda} \text{Tr}(\mathbf{z}_e \mathbf{M} \left[ \begin{array}{c} \mathbf{z}_e^T \\ \mathbf{z}_v^T \end{array} \right] ) + (1 - \lambda) ||\mathbf{z}_e - \mathbf{z}_v||_F^2 \quad (3)
$$

with $|| \cdot ||_F$ being the Frobenius norm. The first term of (3) is the same as in LLE and will assure that if two regressors are close in geodesic sense, then they will also get similar $z$-values. The second term of (3) fix the parameterization of the estimation regressors to $\bar{z}_i$, $\lambda$ is a design parameter.

(3) is quadratic in $\mathbf{z}_e$ and $\mathbf{z}_v$ and has therefore a solution

$$
\mathbf{z}_v^{\text{WDMR}} \triangleq (1 - \lambda) S (\lambda \mathbf{M} + (1 - \lambda) J)^{-1} \left[ \begin{array}{c} \bar{z}_e \\ 0_{N_e \times N_v} \end{array} \right]
$$

with

$$
J \triangleq \left[ \begin{array}{c} I_{N_e \times N_e} \\ 0_{N_v \times N_e} \\ 0_{N_v \times N_e} \end{array} \right], S \triangleq [0_{N_e \times N_v} I_{N_v \times N_v}].
$$

$\mathbf{z}_v^{\text{WDMR}}$ is here the estimated parameterization values for the validation regressors, i.e., the estimated temperatures for us. We summarize the WDMR regression algorithm in Algorithm 1.

An alternative way to see WDMR is by seeing it as a method for computing weighting kernels. A weighting kernel tells how to weight together estimation temperature measurements to obtain estimates for the validation temperatures. The kernel computed by WDMR is defined by

$$(1 - \lambda) S (\lambda \mathbf{M} + (1 - \lambda) J)^{-1}.$$

There are many methods for computing weighting kernels. Most methods would simply weight together the temperatures of the closest estimation regressors. WDMR, however,
Algorithm 1 WDMR

Let $x_t$ be the $t$th element in $[x_e, x_v]$, $N_e$ the number of estimation regressors and $N_v$ the number of regressors we want to express in the parameterization of the manifold. Let $z_e$ be measured parameterization values for $x_e$ that should be used to fixate the parameterization of the manifold. Let $z_v$ be the searched parameterization values of $x_v$. For a chosen $K$, $r$ and $\lambda$,

1) Find the weights $w_{ij}$ minimizing

$$
\sum_{i=1}^{N_e+N_v} \left| x_i - \sum_{j=1}^{N_e+N_v} w_{ij} x_j \right|^2 + F_r(w),
$$

subject to

$$
\sum_{j=1}^{N_e+N_v} w_{ij} = 1, \quad w_{ij} = 0 \text{ if } |x_i - x_j| > C_i(K) \text{ or if } i=j.
$$

2) With $M_{ij} = \delta_{ij} - w_{ij} - w_{ji} + \sum_{k=1}^{N_e+N_v} w_{ik} w_{kj}$, the regressors expressed in the estimated parameterization are given by

$$
z_v^{WDMR} = (1 - \lambda) [0 I (1 - \lambda) [I 0 0]^{-1} [z_e^T 0_{N_v \times N_e}].
$$

computes a weighting kernel that adjusts to an assumed manifold in the regressor space. WDMR therefore weight together temperatures of the, in a geodesic sense, closest estimation regressors. The estimated temperatures will therefore be a weighted sum of the measured estimation temperatures closest along the manifold. From the discussion in Section III this is a good strategy. Fig. 4 exemplifies the weighting kernel computed by WDMR.

![Figure 4](image)

The figure shows how the weighting kernel computed by WDMR adjusts to the manifold in a toy example. It can be seen that an estimation computed by WDMR will be based on geodesic rather than Euclidean distance. Black dots show estimation regressors and gray dots validation regressors. The two-dimensional regressors $[x_1, x_2]$ are placed along a u-shaped one-dimensional manifold. The kernel weights computed by WDMR for estimating the temperature of a validation regressor (marked with a red square) are shown with black bars.

Since all regressors provide information concerning the structure of the manifold, all chemical composition measurements, even though we do not have temperature measurements associated with them, can be used to improve the estimation of temperatures. We can therefore expect the WDMR estimate to improve even though only unlabeled regressors are added.

VI. Result

The result obtained by applying WDMR to the 6 estimation shells and the chemical composition measurements of the 4 validation shells are shown in Fig 5. The MAE i.e., the mean absolute error, between the estimated and measured temperature is also shown in the figure. For all 4 validation shells a mean absolute error less than one degree was obtained. These results justify the manifold assumption and also the benefit of using a nonlinear multi-proxy model.

The erratic behavior seen for late samples in the estimations from the shell from Ossenisse (second plot from the bottom) is interesting. WDMR has for this shell had a problem to decide to what part of the manifold the regressor samples belonged to. Some of the chemical composition measurements were wrongly thought to belong to a part of the manifold associated with a temperature of approximately $13^\circ$C.

VII. Conclusion

We have considered the problem of estimating water temperature from the chemical signature of the bivalve *M. edulis*. We argued that the composition measurements were concentrated around a one-dimensional manifold and that the temperature associated with the composition measurements would vary smoothly along this manifold. The temperate could hence be seen as a parameterization of the manifold.

To find this parameterization a regression method, WDMR, was used. WDMR works as a manifold learning method but has the ability to adjust the parameterization if given some example of how the desired parameterization values should be for some regressors.

The composition and associated temperature measurements of the 6 estimation shells were used in the WDMR algorithm to get the parameterization computed by WDMR to imitate the temperature parameterization of the manifold. The validation regressors expressed in the parameterization was shown to well estimate the measured water temperature.

The presented scheme performs well in comparison to existing methods for temperature reconstruction using *M. edulis*. However, more data is needed to further evaluate the method.

VIII. Acknowledgments

This research was funded by a grant from the Vrije Universiteit Brussel (HOA-9-OZR), a grant from the Flemish government (CALMAR’s), a grant from the Belgian government (METH1) and a grant from the European Science Foundation (ESF-EuroCLIMATE-PaleoSalt). The work was also supported by the Strategic Research Center MOVIII, funded by the Swedish Foundation for Strategic Research, SSF. The authors would also like to acknowledge Gillikin and Vander Putten for sharing their data.
Fig. 5. Water temperature estimations for validation data (solid line) and measured temperature (dashed line). From top to bottom figure: Terneuzen, measured temperature (dashed line). From top to bottom figure: Terneuzen, Fig. 5. Water temperature estimations for validation data (solid line) and measured temperature (dashed line). From top to bottom figure: Terneuzen, Breskens, Ossenisse, Knokke.

REFERENCES