

## Reviewer #1's comments and our responses

1. The reviewer commented that the method description is not enough for the general audience of *Climate of the Past*. The reviewer pointed out that the model comparison (Sec 3.3) with Figure 5 is confusing. The Reviewer #2 also suggested adding more description about the HMM (comment 5) and comparison to the classical correlation analysis (comment 1). We can add more general explanations about the method and also modify Section 3.3 as follow:

Statistical modeling involves developing relationships between one set of predictor variables and another set of predictands. Paleoclimate reconstructions likewise develop models to relate proxy information (predictors) to past climate variables (predictands). Thus, statistical modeling and paleoclimate reconstructions both seek the same goals, and approaches of varying complexity are found to infill missing data or to understand relationships among variables.

In paleoclimate studies, as in any set of observations, not all important variables can be observed or reconstructed. It is typical in such situations to hypothesize linkages among observed variables, but a more direct observation of the mechanism involved in the linkage are not recorded. So, one might expect that A causes B that causes C, but only A and C are observed. Statistical modeling can help identify or quantitatively assess relationships between A and C, even in the presence of hidden variables such as B.

The autoregressive (AR) approach adds value by allowing the state at previous times to be among the predictors of the present state predictands. Typically, causes precede effects, so the AR approach allows for an interpretation of causality--if a predictor precedes the predictand in time, then it is the cause rather than *vice versa*. Simultaneous correlations among variables are frequently interpreted as implying causality, but they can represent a number of relationships--cause and effect, effect and cause, or accidental correlations without causal relationships. The greater precision of the AR models allows for examination of causal relationships under the assumption of cause preceding effects.

Furthermore, the HMM provides a quantitative justification of transitions between different epochs governed by regime shifts in the surrounding climate. Even though these shifts might not be directly detectable in any of the recorded variables alone, the HMM provides a technique that allows all variables to contribute equally in identifying shifts in the relationships among the variables.

### Section 3.3:

The HMM is a special case of the AR-HMM; The AR-HMM with zero autocovariance term ( $\theta_{s(t)}$ ) is identical to the HMM. So, if the AR-HMM results in the proxies having weak autocorrelation,  $\theta_{s(t)}$  should be close to zero, and the other parameters of the AR-HMM (the noise covariance matrices ( $\Sigma_{s(t)}$ )) will resemble their equivalents in the HMM. Thus, were the HMM an adequate model to describe the proxy data, then allowing the extra degrees of freedom

in the AR-HMM would result in little extra predictive power, and this result would not change the interpretation of the data from the interpretation found using the HMM alone. However, in this particular dataset, the AR-HMM resulted in extremely large auto-correlation relationships (the entries of the estimated  $\theta_{s(t)}$  are close to one) and furthermore the other model parameters (the estimated noise covariance matrices) are quite different between the HMM and the AR-HMM. Fig. 5 visualizes and compares the estimated  $\theta_{s(t)}$  of the HMM and AR-HMM. The fact that the AR-HMM coefficients do not resemble the HMM in pattern, magnitude, or implied relationships means that a dependence of the data on values at a previous time is a critical aspect of the data. Thus, a key conclusion from the statistical models is that the past values of each proxy predicts its own proxy variability better than the different proxy-to-proxy cross-correlations at the same time (or indeed the cross-correlations among past and present values). This fact implies that the different proxies in this particular dataset are not causally related to one another, as is often assumed in multi-proxy paleoclimate analyses (e.g., Hu et al. 2017). This result probably does not apply to all multi-proxy records, indeed many are probably causally linked, but our methodology for testing that assumption by comparing HMM to AR-HMM is generic. Thus, in this location, the four proxies (*SST*;  $C_{37}$ ;  $\delta^{15}N$ ;  $\%N$ ) are not related to each other in the local sense that variability in any one dominates or contributes significantly to variability in another through a local physical or biological mechanism.

2. The reviewer commented that a) the Uk'37 proxy may be biased in the region toward warm temperatures and that b) associations between alkenone-based temperatures and productivity might be erroneously interpreted.

We appreciate these concerns and responded by moving proxy information into the Method section and addressing the referees concerns there. In brief, we can cite two substantial data sets that look at the alkenone proxy in the Eastern Equatorial Pacific and argue that there is no indication of a SYSTEMATIC bias relative to mean annual SST in the region

We can evaluate the second claim, as we report Uk'37 unsaturation, bulk organic nitrogen and  $C_{37}total$ , an index of the sediment concentration of alkenones. As we demonstrate, the index of bulk phytoplankton production and  $C_{37} total$  are significantly correlated, suggesting that haptophyte production indeed follows total ecosystem production. And furthermore, the lack of a strong coupling between the Uk'37 index and either productivity proxy--as found by the statistical methods used in this paper--argues against the existence of the production-SST bias suggested by the reviewer. The fact that an inorganic proxy (opal, as reported by Chazen et al.) does not resemble the organic proxies can most likely be explained by variations in the preservation of opal, a notorious confounding influence on interpreting that proxy quantitatively

We now include additional text in 2.1 (Data Collection):

The four records examined are proxies for sea surface temperature (*SST*) through the alkenone proxy, biological productivity of a specific phytoplankton group ( $C_{37}$ ) through analyses of the

abundance of alkenones (representing haptophyte algal productivity), subsurface properties through analyses of  $\delta^{15}N$ , an index of subsurface oxygenation and denitrification, and the percentage of organic nitrogen (%N) which is a composite of all biological inputs to the sediment. We interpret the alkenone Uk'37 index as an approximation to mean annual sea surface temperature. Although anomalies Uk'37 values have been reported in the region (Prahl et al., 2010; Kienast, 2012), there is no convincing evidence for seasonal bias based on analyses of modern sediments over a broad region of the Eastern Equatorial Pacific with very strong gradients in the timing of maximum annual biological production (Kienast et al., 2012; Timmerman et al., 2014). Analyses of modern sediments in the region conducted at the Brown University laboratory show agreement with mean annual temperatures in the region of our core study to within the standard empirical proxy calibration (e.g. subset of data reported in Kienast et al., 2012). Our paleo-productivity interpretations are guided by the presence of a proxy that responds to total phytoplankton production (%N) and to a subset of the haptophyte production (C37total); we can therefore assess whether alkenone production is coupled or decoupled to a generalized biological response over time.

3. The reviewer suggested considering other factors, such as laminations bioturbation, and sediment mixing, to explain more about the decadal predictability.

The reviewer makes a good point about the potential down-core differences in variance being driven by variations in oxygenation and bioturbation. In some sense this is a chicken and egg question, because the existence of laminations is in fact coupled to some of the variables represented by our proxies, such as density stratification and organic matter flux. It is therefore difficult to assess whether the presence/absence of laminations is a confounding factor or part of the oceanographic signal represented in our time series.

However, when we compared a visual index of lamination/bioturbation, based on X-radiographs of the core, we see the following results. In the HMM, the “calm” state is associated with a significantly more negative d15N value, consistent with, although not proof of, a preferential smoothing of variance in non-laminated intervals. However, this association does not persist in the AR-HMM results, suggesting that this 2-state model does not reflect a preservational bias of variance.

We thank the reviewer for sharpening our analysis in this regard, and have modified the text in Section 3.3 by adding:

A caveat arises in assessing variance in the time series: changes in the extent of laminations down-core, which could introduce differential smoothing of the results. We can assess the possible influence of lamination versus bioturbation in two ways: a visual comparison of X-radiographs of the core, which show the presence/absence of laminations, and comparison to d15N, which is strongly indicative of lamination (high d15N signifies intense depletion of oxygen in the subsurface). The results of the HMM and AR-HMM differ significantly in this regard. The presence of State 1 versus State 2 correlates strongly with the degree of lamination/d15N

proxy in the case the HMM model (the “noisy” state occurring much more frequently in laminated intervals). This association is confirmed by the significant offset in the mean values of  $\delta^{15}\text{N}$  for State 1 and 2 (Table 1). However, the AR-HMM removes any significant dependence on the occurrence of the “noisy” versus “calm” states on the status of lamination down-core, and is confirmed by the negligible offset in the mean  $\delta^{15}\text{N}$  reported for the two states (Table 2).