1 Response to anomymous reviewer #2

We are thankful to the reviewer for his detailed, constructive and relevant comments. We are reproducing below the original review (abridged when necessary) and inlined our responses. Quotes form the article are shaded.

In this work, the authors introduce an emulation-based global sensitivity analysis to astronomical forcing of a climate model. After a review of the methodology on emulators for univariate and multivariate output and the mea- sures of global variance, they design a set of physically consistent experiments. Finally, they perform the analysis on several variables of climatological interest to understand which inputs mostly affect the outputs, and compare this uncer- tainty with the variability of the emulator.

This work could be publishable, but it needs to be vastly restructured to focus more on the science rather than the methodology, and it must address some concerns about the scientific conclusions. Further, the mathematical notation needs to be greatly revised and I urge the authors to check more carefully its consistency across the work before the next submission. General comments

- The manuscript discusses an application of the global sensitivity analysis to astronomical forcing in a climate model, yet the vast majority of the paper reviews material in emulator and global sensitivity measures literature. This is also clear from the Introduction, where only the last paragraph describe the scientific aim of the work. Since this paper does not present novelty in the methodology (or so it seems from how it is presented), the discussion of the results need to be expanded and it has to be more clear how the scientific findings add to the current science.

  > **AUTHORS RESPONSE:** This comment was made by both reviewers. Our methodological and scientific objectives are better explained in the revised version. We have also reduced, where appropriate, the repetition of previously material, summarised the more technical questions and moved some of the more technical developments (in particular, the experiment design) to an appendix. The introduction was also entirely revised.

- I don't find the characterization of experiments 20 and 27 as 'outliers' quite compelling. In the manuscript, there seems to be in two overlapping justifications: - different equilibrium climate for different initial conditions, - poor emulator performances. I don?t think either of the points can qualify these two experiments as outliers. If there is convergence to different equilibriums, then this must be further investigated. As the authors state in the supplement: It is unclear whether these patterns reveal distinct attractors of the ocean circulation states, reached from the different initial condition sets, or whether they correspond to weakly connected regions of the attractors that have randomly been sampled from the 500-year sampling and averaging procedure use for output processing. As for the second point, a poor performance of an emulator only means your statistical modeling is overly simplistic. Therefore this indicates that, as it is, the emulator does not work well for a comprehensive global sensi- tivity analysis which, by definition, has to cover all parameter space.

  > **AUTHORS RESPONSE:** The word "outlier" no longer appears in the manuscript. This said, we have faced unexpected behaviours of isolated experiments before (Ayara-Melo et al., in press in Climate of the Past) and we believe that this is a subject that requires careful consideration. Let us ignoring for a time experiment 27, which is not so badly simulated once experiment 20 is omitted from the experiment design, and concentrate on experiment 20 for the time being. Experiment 20 is not correctly predicted by the emulator. Even worse, accounting for Exp. 20 in the calibration significantly degrades the emulator performance on the other calibration points. In a word, Exp. 20 behaves against our *expectations*, in the sense that Gaussian process emulator using linear relationships to the forcing plus a smooth stochastic component reflects our prior beliefs about the simulator response. We agree with the reviewer that this situation calls for a careful inspection of experiment 20, and that we were falling short

of this discussion in the original version.

We now better comment on this point. It is shown that experiment 20 oscillates at low frequency, at first sight similar to the dynamics of Dansgaard-Oeschger events. The oscillation period varies as the deep ocean temperature adjusts to the forcing and at the end of the experiment it is roughly 800 years. Consequently, a 500-year average (as used here) may randomly produce different averages. The emulator, founded on the hypothesis that the data supplied for calibration represent a stationary climate, has no chance to capture this. In fact, the climate system oscillates in this experiment 20 between two meta-stable states, which may be termed as "warm North Atlantic" and a "cold North Atlantic", and the emulator predicts the warm equilibrium (attached figure 1, also shown in the revised manuscript). After the original submission of the manuscript we tried a couple of more experiments with nearby input configurations, and realised that the oscillator behaviour occurs for these parameter choices as well. The response to be given this state of affairs partly is a matter of judgement. We could, for example, start an iterative experiment plan procedure, the objective of which would be to delineate the 3-D contours of the experiment design within which the oscillatory behaviour occurs, in order to build an emulator specific to that zone. This is a paper on its own, and the added value of this work for palaeoclimate science is questionable, for two reasons. The first one is that the region of the parameter space where this occurs is in fact almost never reached by Nature. In particular, the obliquity of 22° used for experiment 20 is in the lower 0.013th lower percentile (i.e., values lower that 22 ° occur less than 0.013 % of the time) according to the Laskar et al. 2004 solution over the last 10 million years. So it has in fact no weight on the global sensitivity measures. The second reason is that the oscillator behaviour has been documented before in Goosse and Renssen 2002, and Loutre et al. 2014 and generally judged to be 'non-robust' : i.e., the parameter region where this oscillation could occur is sensitive to physical parameter changes well within uncertainties. Therefore, we prefer to acknowledge the behaviour, consider that it *could be of significance* (i.e. we are definitely *not* ignoring the experiment or pretending that it never occurs) but use discard it from the emulator design procedure to the benefit of the performance of the emulator on the rest of the experiment domain, which covers quasi 100 % of the actual astronomical forcing region. The text reads as follows:

> Experiment 20 is the lowest configuration of obliquity (22°). This is lower that any actual obliquity during the Pleistocene (the minimum being 22.07° for Laskar et al. 2004). In this configuration, LOVECLIM develops a slow oscillation pattern that may be reminiscent of Dansgaard-Oeschger oscillations: millennial transitions between a warm and a cold North Atlantic phase, with fast warming and slow cooling (Figure 1). The phenomenon, a known feature of LOVECLIM (Gossse et al. 2002; Loutre et al. 2014), can be described as the apparition of a cold North-Atlantic phase that is being visited stochastically and increasingly frequently as obliquity decreases. According to a couple of experiments not further discussed here, this cold phase is being visited shortly once during the entire experiment at obliquity of 22.5 ° (lowest 7th percentile), though obliquity threshold is likely to depend on the configuration of precession. The oscillation itself could be of physical relevance for past climate variability, but the limits of the phase space region in which the oscillation occurs are likely to be sensitive to uncertain parameter changes. At this stage, one possible response would then be to identify, by sequential experiment design, the region of occurrence of the phenomenon and develop an emulator specifically aimed at characterise this oscillation. Given the likely sensitivity of the oscillation on model parameters, the significance of this enterprise for palaeoclimate interpretation is unsure. We rather choose to ignore the experiment for the time being (the following diagnostics ignore experiment 20), but briefly discuss the possible consequences of this choice in the final discussion.

- I found section 2.5 quite hard to read, and the mathematical notation not very clear (. . .) A better

notation would put the pedix p or $\bar{p}$ in expected values and covariances to make it explicit with respect to which of the three sources of uncertainty you are integrating.

> **AUTHORS RESPONSE:** This reviewer comment matches that of reviewer #1. We agree that there was considerable scope for improvement and completely rewrote this section.

2 Specific comments

- p.903 l.20. A variance in the input factors is a consequence of assuming that the inputs are random variables, which might or might not be a reasonable assumption. In any case, it is worth pointing out your choice, and why you made this choice.

  > **AUTHORS RESPONSE:** In fact this is yet another interpretation, i.e. that the variance is a natural variance arising from variations in the astronomical forcing through time. Our purpose is thus to provide a predictive variance analysis of what may be found in palaeoclimate records. It appeared reasonable to us to take advantage of the mathematics of global sensitivity analysis to end, as long as a hypothesis of time scale separation (i.e., the astronomical forcing varies slowly compared to the adjustment time of ocean-atmosphere dynamics) is acknowledged.

- p.904 l.12. An emulator is a computationally cheap stochastic approximation to the simulator. A deterministic approximation can just be a linearization of some primitive equations in the climate model.

  > **AUTHORS RESPONSE:** Word "stochastic" added.

- p.904 l.14-18. I am not sure I agree with the term "feasible here". The smooth character of the response or the correlation in the outputs can be incorporated in some emulators, and this certainly is an appealing prop erty, but it has to do with the emulator flexibility, but not its feasibility.

  > **AUTHORS RESPONSE:** Word "feasible" changed by "reasonably simple to implement".

- p.904 l.28-29. If you want to produce geographical maps, you have to deal with spatially correlated output, which can be regarded as multivariate.

  > **AUTHORS RESPONSE:** This comment is interesting, and, if we understand it well, partly conflict with the recommendation of reviewer #1 to use indepedent emulators for all grid points. As reviewer #1 also points to missing references with respect to this subject, we now provide a slightly expanded discussion of this point. This said, the new version *does* also use the information on covariance provided by the co-variance emulator.

- p.905 all equations are missing the punctuation.

  > **AUTHORS RESPONSE:** Corrected.

- p.905 l.10-11. There is nothing Bayesian in assuming f has a probabilistic distribution. A Bayesian model elicit priors, while a frequentist model assumed them to be fixed and unknown.

  > **AUTHORS RESPONSE:** The approach is Bayesian in the sense that f is a deterministic but unknown function, and (key point:) we are representing our knowledge (or lack thereof) as a probability model.

- p.905 l.24-26, p.906 l.1-2. The use of the term "global" here is problematic. m is a mean response across all possible uncertainties assumed by the statistical model, and V encodes the deviation from this mean behav- ior. Why did you call it "global". Further, the stochastic component is does not have to be "smooth": the random field can be non-differentiable.

> **AUTHORS RESPONSE:** We agree with these comments and changes have been made accordingly

- p.906 l.7. y represents the data and it has a likelihood, not a prior. Only the statistical parameters involved in the analysis have a prior, assuming you are working in a Bayesian framework.

> **AUTHORS RESPONSE:** Likelihood is a property of parameters, so we disagree on the fact that y has a likelihood. It is standard usage in the Bayesian interpretation of Gaussian priors to talk about the prior (or predictive distribution) for $y = f(x)$ before observing the data, and then the posterior after updating these beliefs.

- p.906 l.8. The first matrix A should be in bold. Also why are you referring to it as a Gram ?

> **AUTHORS RESPONSE:** We are using bold to denote vectors, and capital letters for matrices. Consequently we have left A as it is. We are using the phrase 'Gram matrix' in the same way as used in Rasmussen and Williams' seminal book on GPs.

- p.906 l.17. I believe the correct expression is

$$m(x) = h(x)' + T(x)'A^{-1}(y - H\hat{\beta})$$

> **AUTHORS RESPONSE:** Corrected, except we don't use bold face for the matrix.

- p.907 l.3-4. See the point above about the use of the term ?global".

> **AUTHORS RESPONSE:** The term "global" is no longer used in this context. In addition, in response to the commenst of reviewer #1 we have rephrased all sentences where distinguishing "linear" from "non-linear" is presented as an objective.

- p.907 l.11. Please use a more appropriate reference. The Matern model is used in so many different areas that a reference to a standard textbook is perhaps more appropriate. See e.g. Stein (1999) or more recently the Handbook of Spatial Statistics. Also, the citation needs to be put in parenthesis.

> **AUTHORS RESPONSE:** We agree. Reference to Stein (1999) is now made.

- p.907 l.18. The use of nugget is necessary in the case of a squared exponen- tial correlation function, as otherwise the statistical model would assume an overly smooth change of the output with respect to the input. Under the squared exponential, a knowledge of the output for an arbitrarily small interval in the input implies that the output is uniquely determined every- where. So in this context, this can be seen as a form of mis-specification.

> **AUTHORS RESPONSE:** We agree, and the same is also true for other covariance functions, but we are unsure of what the reviewer requires here, as we already written that "[the nugget may be justified] as a way to account for the mis-specification in the correlation function".

- p.907 eq (7). I believe the authors took the notation from equation (3) in Andrianakis and Challenor (2012). That, however, was a marginalized likelihood. What you wrote is a marginalized loglikelihood so please fix the notation accordingly (there is no proportionality on the log scale).

> **AUTHORS RESPONSE:** Corrected. Which made us realise that Ioannis Andrianakis, with whom we interacted about this likelihood function, had to be acknowledged.

- p.908 l.3. Why did you choose $\varepsilon = 1$ Also, since this is the same symbol used to define the obliquity at p.910, l.2, it would be probably best to write equation (8) without M and ϵ, which are not used in the rest of the paper anyway.

> **AUTHORS RESPONSE:** $\epsilon = 1$ (not ε) is a standard value in Andriankis and Challenor, but we indeed preferred to follow the seggestion of the reviewer and removed it. We kept the M.

- p.908 l.9. I believe there is a typo: $f : \mathbb{R}^d \to \mathbb{R}^m$

> **AUTHORS RESPONSE:** There was a typo, but we now use p as the number of outputs, to avoid overloading the use of m, already used for means.

- p.909 l.6. I would urge some caution in the use of the term "information" in this context. SVD finds the direction of maximum variance, and it could be a desirable property, but since we have a spatially index field, I am not convinced that the spatial information (i.e. the spatial correlation) of the data is minimized. SVD does not account for the spatial nature in this context and it has been shown to be potentially discarding substantial information.

> **AUTHORS RESPONSE:** Fair point. The reference to "information" is now removed. There is enough literature about SVD anyway.

- p.910 l.19. If i1 and i2 both depend on the eccentricity and perihelion, how can you claim they are independent?

> **AUTHORS RESPONSE:** No, they are not independent (and we don't say so) but they are *uncorrelated*. This is precisely because they are not independent that we have not attempted to provide sensitivity mesaures to $i1$ and $i2$ separately.

- p.912 l.3. Please use a less colloquial beginning of phrase.

> **AUTHORS RESPONSE:** Agreed and corrected.

- p.912 point 3. Dividing $[1, 1]$ in N equal width intervals and using permutations implies that you are assuming i1, i2 and i3 are all equally important for LOVECLIM, at least a priori, and that every part of the parameter space is equally important. Is that a reasonable assumption?

> **AUTHORS RESPONSE:** The analysis certainly shows that the ranges associated with all three factors induce more-or-less similar-size effects on climate. So, yes, this is a reasonable assumption. But, we are not sure that this really matters. This is a design that will be used to build the GP emulator. We want the emulator to work well across all of the prior space, which is why we use a space filling design here. The importance of different inputs and regions of space is expressed in the prior distributions $\rho(x)$.

- p.914 l.9. What is Xp? It was never defined.

> **AUTHORS RESPONSE:** Line 9 is the definition. $x \in \mathcal{X} = \mathcal{X}_p \cup \mathcal{X}_{\bar{p}}$ and $x_p \in \mathcal{X}_p$

- p.914 l.18. There is a comma missing.

> **AUTHORS RESPONSE:** Corrected

- p.914 l.22-23. If you assume a nugget effect in your emulator, even an infinite number of model runs would still give you an uncertain estimate of $f(x)$ for every x.

> **AUTHORS RESPONSE:** Not quite. It all depends of what we call $f(x)$. Our model is that there is a smooth continuous function $f(x)$ which is modelled as a GP. Here, this is the hypothetical average over an infinite number of years, that would effectively marginalise over all possible initial conditions (assuming ergodicity). Hence, the the observations are effectively $y = f(x) + e$, where e is the nugget term. If we had an infinite dataset then it would be possible to estimate $f(x)$ arbitrarily precisely.

- p.915 l.24. $\eta(xp)$ is a linear functional.

> **AUTHORS RESPONSE:** Agreed and corrected

- p.915 l.2. The notation looks wrong. Either you use Vp or Vpp everywhere.

> **AUTHORS RESPONSE:** There was indeed a misprint, but this part of the notation was changed anyway.

- p.915 l.6. I believe (12) should be

$$V(x, x^\star) = \int_{\mathcal{X}_{\bar{p}} \times \mathcal{X}_{\bar{p}}} V(\mathbf{x}, \mathbf{x}^\star) \rho(\mathbf{x}_{\bar{p}} | \mathbf{x}_p) \rho(\mathbf{x}^\star_{\bar{p}} | \mathbf{x}_p) \mathrm{d}\mathbf{x}_{\bar{p}} \, \mathrm{d}\mathbf{x}^\star_{\bar{p}} \tag{1}$$

  Besides, this is not a variance, as stated in line 11. This is a function of two arguments, so I don't understand why it is denoted as $\mathbb{V}ar(x_p)$

> **AUTHORS RESPONSE:** The notation was admittedly abusive. What we referred to is indeed the *covariance* associated with the Gaussian process, which is indeed a function of two points. This is clarified in the revised version.

- p.915 l.7-8. I don't understand why you claim that the expectation and variance are computed with respect to the Gaussian process model for f. (11) and (12) are integrating with respect to $\mathbf{x}_{\bar{p}}$, which reflects your uncertainty on your input parameters space, and not on the emulator.

> **AUTHORS RESPONSE:** The quantities η and m are indeed defined as integratals over the input space (which is not uncertain, remember our objective of characterising the relationship between inputs and outputs). Hence, these are deterministic quantities if it were for the simulator alone. The quantities m and V, which now enter the integrals, express our uncertain knowledge of the simulator output at the corresponding input. Thus we are uncertain about η and m, but we can characterize their expectation and variance. We feel that this was already explained, and we hope that changes in the notations have made this even clearer.

- p.915 l.16. How did you define $\mathbb{V}\mathrm{ar}(x_p)$)? You only defined $\mathbb{V}\mathrm{ar}\mathbf{f}(x_p)$. I believe that here you are computing the expectation of (12) with respect to the emulator uncertainty.

> **AUTHORS RESPONSE:** No, but we agree that this part of the manuscript was confusing. The variance was, in that notation, taken over the input space $\mathcal{X}_{\bar{p}}$ (following Oakley and O'Hagen). We hope that the change in notation makes things more clear.

- p.916 l.2. What is $\rho(xp)$ Before you only defined $\rho(x)$ and $\rho(x_{\bar{p}}|\bar{x}_p)$. Intuitively I'd say [...]

AUTHORS RESPONSE: Again, following this and other comments, the notation was thoroughly revised.

- p.919 l.1. Shouldn't the covariance matrix be $n' \times n'$, there are too many ms: one for the mean and one as a index for the sensitivity index.

AUTHORS RESPONSE: It was well $m \times m$ (now $p \times p$ in the new notation), in the sense that the emulator provides simulator estimates at every of the m (now p) grid points. However, we agree the other reviewer that there are to many m's and changed the dimension of output from m to p.

- p.919 l.1-4. What do you mean by "insightful enough for our purpose"? Understanding the interaction between multiple outputs in the global sensitivity analysis seems well within the scope of this work.

AUTHORS RESPONSE: The reviewer is correct in suggesting to use the covariance matrix information. We now write:

> The covariance matrix is thus of dimension $p \times p$ and provides information on the joint uncertainty of any two the simulator outputs. In practice, the computation of point-wise sensitivity indices only require to know the diagonal of this matrix. The PC emulator however allows us to go one step further compared to the independent emulator strategy. As the full covariance matrices V, T_i are available, the climate response linear fingerprints may be had obtained as the SVD decomposition of these covariance matrices (*now shown on a new figure*). Specifically, we refer to "fingerprints" of precession and obliquity, the eigenvectors of $T_{e\varpi}$ and T_ε, respectively. The first fingerprint of obliquity explains more than 90 % of the variance of all three variables considered here. The precession fingerprint aggregates two inputs ($e \sin \varpi$ and $e \cos \varpi$). It can therefore be expected to have at least a second significant fingerprint. This is the case, but this second component represents less than 30 % of the variance. We come to that shortly. Compared to the point-wise variance analysis above, the main advantage of the fingerprint analysis is to provide information on the in-phase or anti-phase relationships between climate variables, namely [...]

There is also some more discussion on the response phase of African precipitation to precession, though we must keep in mind that this is more illustrative than anything else, given the poor performance of LOVECLIM in the tropics:

> Coming back to precession, we expect the simulated response phase to differ from place to place. To illustrate this point, we plot the emulated precipation the Atlantic as a function on the longitude of the perihelion for three points along the African monsoon flow. We assume obliquity and eccentricity typical of the Holocene (*new figure here*), and indicatively denote longitudes of perihelion corresponding to the time of the beginning of the Holocene, as well as that of 6,000 years ago, a reference period used for model intercomparison exercises (Braconnot et al. 2009). The timing of the maximum response is gradually shifted towards a late phase response as one travels northwards. This observation can be explained by considering the seasonal development of monsoon dynamics, along with the course of the zenithal sun.

- p.919 l.5. What is $S_{\bar{p}}$. Did you mean $\bar{S}$

> **AUTHORS RESPONSE:** Yes, misprint corrected (but no longer relevant given the new notation).

- p.920 l.1-14. Please rewrite this part using a more careful notation. What is $x|x_p$? Also, the inner integral is in the space of xp?, but where is xp?? Also, Xp is 1-dimensional, while Xp? is 2-dimensional, so I guess the integral in (22) should be 5-dimensional. Besides, what is g here?

> **AUTHORS RESPONSE:** T In fact, the notation should parallel that of equations 11 and 12. This was corrected. The integrals are indeed hugely dimensionals, hence the interest of Monte-Carlo approaches when analytical integrals are not available. On this point, analytical integrals can in fact be carried out when the distribution is uniform, and we now provide analytical solutions to compare Monte-Carlo estimates in this particular case.

- p.921 l.14. I guess the authors refer to "calibration" as "estimation". In the context of computer model experiment, calibration is a completely different problem, that has nothing to do with this statistical model.

> **AUTHORS RESPONSE:** We believe that calibration and estimation are somewhat exchangeable in this setting, but we changed to estimation.

- p.921 l.14. "hyperparameters" usually describe the parameters of a prior on your parameters. You have not given any prior on Λ and nu

> **AUTHORS RESPONSE:** This is a rather subtle point of semantics. In a sense, the choice of Λ and ν control the predictive distributions of model outputs and we wanted to provide a clear distinction with quantities like β, for example, that are more easily intepretable for the climate scientist. This said, we removed the "hyper-" prefix and refer to "correlation parameters".

- p.921 l.27-30. Could you elaborate on the claim that lower-order principal components represent variability modes of importance? Is there any reference on this?

> **AUTHORS RESPONSE:** This result from our own inspection of the principal components, and also from our own observation that using only a small number of PCs (e.g., 3) significantly degrades emulator performance. However, following the suggestion to reduce as much as possible the methodological section we have summarised the discussion as follows:
>
> We first attempt to use different correlation parameters obtained by maximising the penalised likelihood for each of principal component, independetly of the others. This is reflected by the decreasing trend in the components of Λ, the occurrence of high values of ν. The likelihood stabilises around PC #10 to a minimum value that indicates that the calibrated GP is not more informative than assuming independence of outputs on inputs. We therefore use $n' = 10$.

- p.922 poin t 3. How can you claim the results are "satisfactory overall" if you noticed, correctly, that the behavior is typical heavy-tailed and you used a Gaussian emulator? To me, Figure 7 underscores a fundamental misfit of all quantities.

> **AUTHORS RESPONSE:** The word "satisfactory" naturally reflects a subjective judgement that is best replaced with a more objective statement, or just removed (as we finally did). As whether we should consider that there is a fundamental misfit, again the word "fundamental" is subjective. It all depends a bit of what is our objective. For the statistician it is generally important to use the "right" distribution, i.e., fat tailed distributions would not be compatible with a Gaussian process. But solving this problem drives us away from the Gaussian process emulation. So, we can still be satisfied about the fact that the number of points predicted with

more than 3-sd is very small and that we are not grossly underestimating our uncertainties. This is probably enough to prevent us from reaching totally wrong conclusions about model response to astronomical forcing.

- p.923 l.6. In the previous Section you claim that your final choice of PC components is 10, yet here you use only 2. Even if parameter estimation is less demanding than Monte Carlo integrals, I am worried that your estimated sensitivity might be underestimated, as you are removing too many modes of variability from the analysis. Is the estimated sensitivity robust with respect to the increasing number of PC considered?

  AUTHORS RESPONSE: This comment is entirely addressed by the fact that in the revised version we now compare MC estimates with analytical integrals, using the full number of PCs. It is confirmed that uncertainties associed with the Monte-Carlo estimates of integrals are negligible.

- p.929 l.2-4. If the emulator cannot reproduce experiment 20 as the simulator's response, to me it is a sign that the emulator is wrong, not that the experiment (or the simulator) is.

  AUTHORS RESPONSE: This comment is addressed in the response to the general comments.

- Figures 8-9. If possible, increase the size.

  AUTHORS RESPONSE: We agree that the figures are too small, but we have been constrained by the format of the the online format for Earth System Dynamics Discussion. Hopefully a better solution may be found with the publisher for the final version.

- Figure 8. "across". Also, the choice of the colorbar is quite strange. It is consistent for each row, and the the bounds can differ from column to column, but choice of colors should be much more standard (i.e. blue for low values, bright red for high values).

  AUTHORS RESPONSE: After experimenting a bit with colors we found that 'white' for low values allows the reader to immediately identify regions of significant variances and yet clearly distinguish the continental boundaries. The sucession 'yellow', 'blue', 'purple' is also effective to distinguish mid- from high values, especially for colorblind peoples (yellow and purple may still be a bit difficult to read for tritanopia-suffering patients, but they will still clearly identify the zones with higher variance).

- Supplement, p.2, 3rd paragraph. "reveals"

  AUTHORS RESPONSE: These pattern "reveal" (not reveals). Anyway, this bit of text was removed to give way to a more complete discussion of the dynamics of exp. 20.

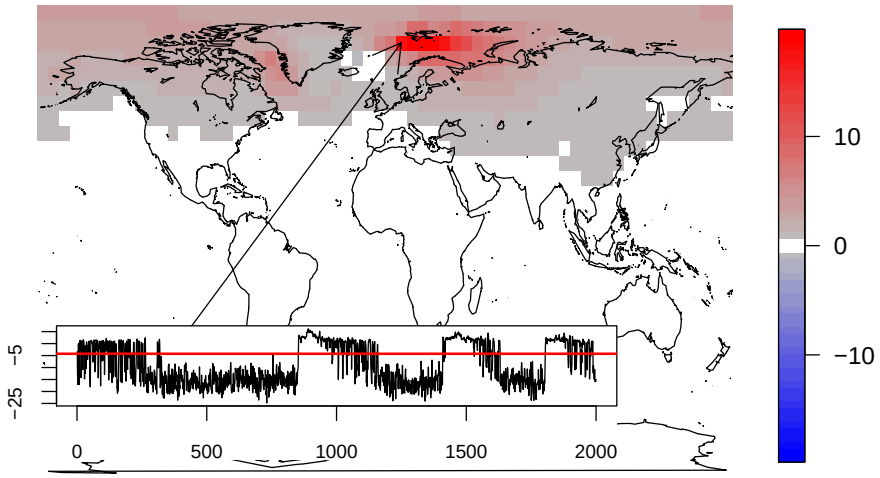

Figure 1: Slow oscillations developing in experiment 20 ($e = 0.040$, $\varpi = 334.6°$, $\varepsilon = 22°$). The surface annual temperature over one of the North Atlantic grid points is shown (inset) along with the geographic distribution of the difference between the warm and the cold phases. The horizontal red line in the inset is the emulator prediction.