Comments on "Global sensitivity analysis of Indian monsoon during the pleistocene" by Araya-Melo et al.

This manuscript uses a Gaussian process emulator – a type of statistic model to make maximum use of the GCMs' output to get different effects of different climate forcing (it is called a space) in the full range of the forcing. As the authors mention, this is the originality. Through this study, the authors show that this statistic modeling technique provides reliable numerical results. I totally agree with the authors that this emulator technique has a large potential use for climate modeling community. I highly recommend this manuscript to be published on Climate of the past to get more and more attentions of climate modelers.

Honestly to say, it is quite new technique for me too. This is the first time for me to hear. So I have to read a lot from Wiki to understand the technique. This means that I would like to see more clarity about the technique. I also generally agree with the conclusions on the effects of precession, obliquity, CO2, and glacier. It is quite consistent with our published modeling result based on a transient climate simulation over the last 280ka. Please see the references at the end of this comment. I think they will be good reference for the authors to explain the mechanism why continental temperature is in phase with June insolation, and why sea-surface temperature is in phase of May insolation.

The only concern on the conclusion is that the phase of continent precipitation. Without glacier, our simulation shows that the phase of continent precipitation is leading June insolation, between the phases of sea-surface temperature and continent temperature. As we know, the Asian monsoon system is controlled by the land-sea temperature contrast. It seems to me that it includes some kind of glacier effect. Because I can not fully understand the technique, this is just my guessing based on my best knowledge. Also another reason is that the region to represent the Indian monsoon. The current selected region includes most of Tibetan Plateau. It could not represent the Indian monsoon. I strongly recommend the authors change the region. It would be good idea to follow the region we used in our paper.

Comments in detail:

1. In the introduction part, please give some introduction to what is a global sensitivity analysis (VS local sensitivity analysis). It is very easy to make climate modelers to understand it as global of the world.
2. Please check the spelling of "Paleoclimate".
3. In the introduction part, please mention that currently with long transient simulation and cross-spectrum analysis, climate modelers do can distinguish the individual effect of some factors. Please see the reference at the end of comments.
4. Page 1614, Line 9: change to reference to (Gordon et al., 2000)
5. Following the Latin Hypercube sampling, the maximum combination of M divisions and N variables is $(M!)^{N-1}$. For example, M=4 divisions and N=3

variables, it would have 576 possible combination. So for this study, M = 10 and N =5. I could see there would be lot combinations. I would like to know how the authors decided to have 57 combinations.

6. Please make the following clarities for the equations (1) and (2):
   Where is X_j? What is the meaning of X^*?
   It would be good to give an example of X_j. Also please give an example of y (the vector of actual outputs, Page 1617, Line 24).
7. What is the n design points on Page 1617, Line 24? Currently, you have M=10 divisions and N=5 varaibles. What is the model? Is it a climate model or statistic model?
8. Please make the following clarities for the equation (3):
   Where is the $T(X)_j$? Please make the notion of i and j clear.
9. Page 1618, Line 12: What is the vague prior? In the equation, there is no this term. Why make this assumption?
10. Page 1618, Line 27: where does λ_i comes from?
11. In the section 2.3, I am more interested in how to build an emulator for the 61 experiments. Suppose I want to build an emulator from some experiments, it seems I need to choose a $h(X)$ and a Gaussian process correlation function $c(X, X^*)$. Also assume I choose the linear regression for $h(X)$ and the equation 4 for $c(X, X^*)$. According to my understanding, I assume X_j is the one of the combination of the five forcing. For example, it is one of the combinations of the five forcing for the 61 simulations. Am I done with building an emulator? Also I am very confused about what I need to input to the emulator and what I expect to have from the emulator?
12. Continuing the question 11, in section 3.1, I am totally lost how validate the emulation. What should I compare to? I think the authors should give a bird-view of how to build an emulator, how to validate the emulator, and then how to use the emulator expand the GCM's simulation.
13. For figures, I would recommend the authors rearrange all the figures according the order appeared in the text. I have pain to look at the figures back and forth.

Reference:
(Kutzbach et al., 2008)
(Chen et al., 2011a;Chen et al., 2011b)

Chen, G. S., Kutzbach, J. E., Gallimore, R., and Liu, Z. Y.: Calendar effect on phase study in paleoclimate transient simulation with orbital forcing, Clim Dynam, 37, 1949-1960, Doi 10.1007/S00382-010-0944-6, 2011a.
Chen, G. S., Liu, Z. Y., Clemens, S. C., Prell, W. L., and Liu, X. D.: Modeling the time-dependent response of the Asian summer monsoon to obliquity forcing in a coupled GCM: a PHASEMAP sensitivity experiment, Clim Dynam, 36, 695-710, 2011b.
Kutzbach, J. E., Liu, X. D., Liu, Z. Y., and Chen, G. S.: Simulation of the evolutionary response of global summer monsoons to orbital forcing over the past 280,000 years, Clim Dynam, 30, 567-579, 2008.