

The authors aimed to measure the physical similarity between several phenomenological dynamical paleoclimate models and the real world. They proposed that a model should have a link with known physical assumptions, rather than a statistical description of proxy data. Overall, the manuscript is interesting and may have important implications for the future development/evaluation of phenomenological models. As I am not an expert on this issue, I fully agree with the comments from Reviewer#1. However, I still have some comments that may be considered to improve the quality of the manuscript.

1. Figure 2 shows a specific case of the models used here. I am wondering how you choose the values of each parameter (e.g., why obliquity-period doubling is used?). Please clarify.
2. The physical similarity is measured in the $\varepsilon/\hat{\alpha}$ -V space. I am just wondering what are the limitations of the proposed method? For example, G23 model clearly differs from the VDP model in the $\varepsilon/\hat{\alpha}$ -V space. Are there possibilities that G23 model shares similarities with VDP model in some aspects?
3. The positive-vs-negative feedback loops are important to understand the periodicity of glacial-interglacial cycles. Could you add additional discussions on the differences in the loops between each model?