We thank the Anonymous Referee #2 for his/her constructive suggestions for improving the paper. Our specific reply is below with Referee's comments in *italics*.

1) The authors claim that this is the first time that has been demonstrated that MCMC is viable for calibration of GCMs. Clearly, the paper of Jackson et al. (2008) in Journal of Climate makes a first consideration to this issue with the CAM 3.1. and via a multiple/optimization type of sampling. For this context, Villagran et al. (2008) makes the point that versions of Adaptive Metropolis could provide a more appealing sampling strategy that reduces biases. There is no discussion in detail of these other references in the context of what the authors are trying to achieve for the ECHAM5 model.

We thank the Referee #2 for pointing out the work by Jackson et al. (2008) and Villagran et al. (2008). The paper of Jackson et al. (2008) was also pointed out by Referee #1 – please see our response to his/her first comment. We try to extend this discussion in the revised manuscript, and refer to the results of Jackson et al. (2008), for instance, on the performance comparisons between adaptive MCMC and MVFSA. Villagran et al. (2008) was referred to in the original manuscript, but the discussion was very limited and we will improve this, too.

2) One of the main contributions of this paper seems to be the formulation of several objective functions and to establish a comparison between them. However, the paper only provides narrow evidence about the question of how to choose an objective function. In general, the discussion around this (section 4.2) is rather unclear and vague compared to other sections of the paper.

This comment is entirely true and justified. Our reasoning was that the T21L19 resolution as used in the original manuscript, is rather limiting, especially due to large systematic model errors it has. Therefore, we didn't want to directly scale up the results to higher-resolution models. Thus the care used in our text.

In the revised manuscript, we will expand somewhat the discussion regarding the choice of the cost function (see our response to the third comment by Referee #1). In particular, regarding Section 4.2, we will provide a clearer reasoning why we consider J_G+ZONAL as the best cost function of those we have considered so far, and we will also make it clear that alternative and probably better formulations of the cost function can be designed (again, please see our response to Referee #1).

3) It is somewhat deceiving that the paper does not provide posterior distributions of parameters since there is an 'apparent drift' on parameter values in Figure 2. If the MCMC needs to be run longer and some convergence issue is apparent here, what is it really gained by the current analysis? Or should we think of this study as rather preliminary? For example, Villagran et al. (2008) consider the effect of sampling assuming only 500 algorithmic iterations were available. This should be discussed in the paper. I think it is really worth showing posterior distributions at least to understand better the limitations of the current results and to gain perspective on how much longer does the MCMC needs to be run to achieve trustworthy results. In regards to Table 2, why not show the corresponding posterior estimates and standard deviations for each of the objective *functions that were considered along with what is currently presented there? This could allow readers to understand better some of the different biases and eventually recognize the best objective functions, which is one of the main goals of this work.*

We have extended one of the MCMC chains (J_G+ZONAL) to about 4000 runs. The results indicate that there is an initial drift lasting for ~500-1000 runs, after which the posterior distribution remains fairly stable. Thus, in the revised manuscript, the convergence of posterior distributions will be discussed based on this longer chain. In addition, a table may be added that compares statistical characteristics of the parameter posterior distributions for the different cost functions. Although the posterior distributions are not fully converged for the shorter chains (1000 runs), this nevertheless helps to evaluate the differences associated with the choice of the cost function.

To conclude, below is an example of parameter posterior distributions as chain histograms calculated from some 4000 runs where the distributions had become stationary, i.e., they had converged. One can see that two parameters (CMFCTOP and ENTRSCV) are identified whereas the other two (CAULOC and CPRCON) have not – with this formulation of the cost function.