Evaluating systematic errors across recent configurations of the HadGEM3 climate model

John Rostron, David Sexton, Yoko Tsushima, Kalli Furtado

1. STUDYING SYSTEMATIC MODEL ERRORS USING PERTURBED PARAMETER ENSEMBLES (PPEs)

Each member ('variant') of a PPE uses a unique set of values for model parameters (Fig. 1).

PPEs are well suited to exposing and studying systematic errors, which will persist across all parameter settings.

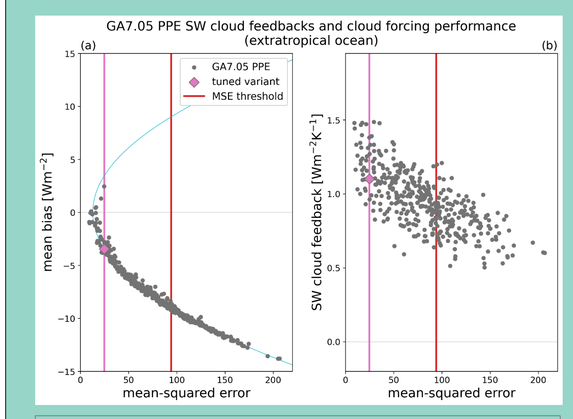


Fig. 2 GA7.05 PPE values for the mean bias vs mean-squared error for extratropical LW cloud forcing (**a**); SW cloud feedback vs mean-squared error (**b**). The tuned GA7.05 variant is shown in pink; the MSE threshold for constraining the PPE is shown in red.

The feedback constraints were also impacted by smaller errors for the tuned variant (used for CMIP6), which were not representative of the PPE.

We see systematic changes in feedbacks between recent HadGEM3 configurations (Fig. 3).

Aim:

Develop methods to use PPEs to track changes in systematic errors as the HadGEM3 model is updated.

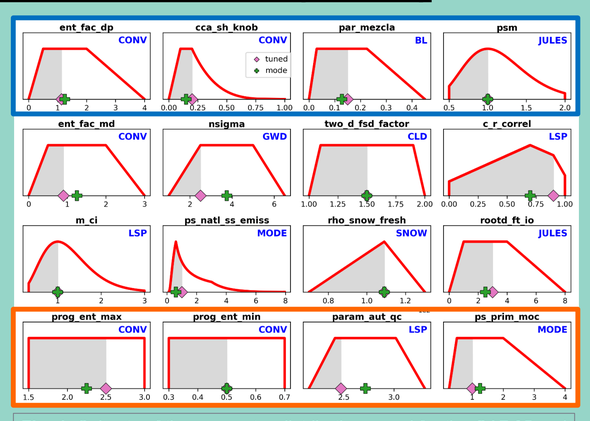


Fig. 1 Subsets of the parameter distributions used for the GA7.05 and GA8 PPEs. Parameters unique to these PPEs are highlighted in blue and orange, respectively. Parameter values for the tuned (pink diamonds) and modal (green crosses) variants are also indicated.

Systematic errors in LW cloud forcing caused a bias in performance-based constraints on climate feedbacks in the HadGEM3-GA7.05 PPE[1] (Fig. 2).

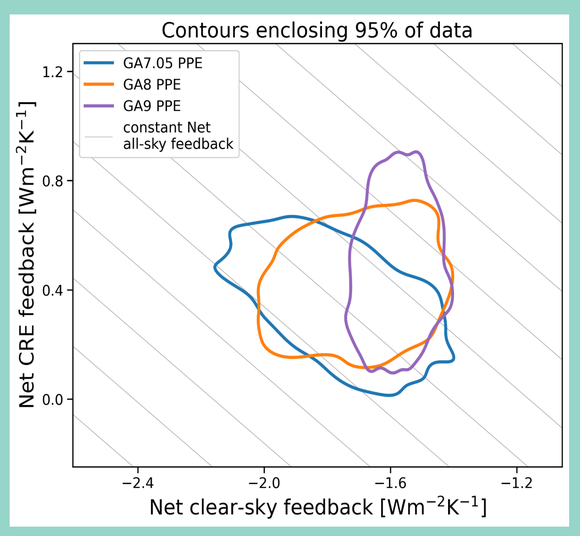


Fig. 3 Net cloud vs net clear-sky feedbacks in recent HadGEM3 PPEs based on: GA7.05 (blue), GA8 (orange) and GA9 (purple).

2. METHODS

We analyse PPEs based on two recent HadGEM3 atmosphere configurations: GA7.05 and GA8.

We use a 5-year amip experiment to evaluate present-day climatologies and model errors.

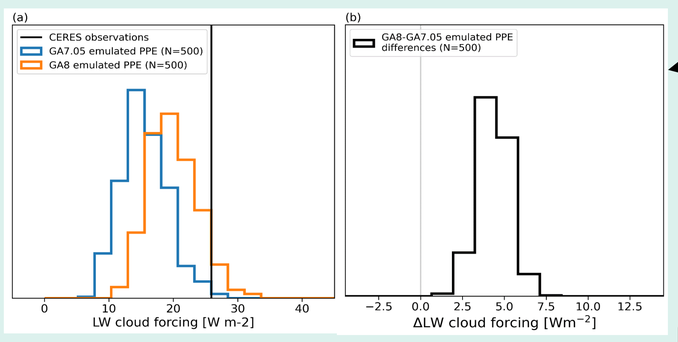
Table 1 The two HadGEM3 atmosphere-only PPEs used in this study.

Name	Parameters perturbed	PPE members
GA7.05 PPE[2]	47	406
GA8 PPE	71	503

We use emulators to predict climatologies for each PPE at the same parameter values for parameters which are 'common' to both PPEs (averaging over the effects of 'unique' parameters; Figs. 1 and 4). We can then calculate differences between the PPEs for each common parameter sample (Fig. 4).

We calculate spatial differences by emulating leading EOF amplitudes for the common parameter samples; reconstructing the spatial fields; and taking differences between the PPEs.

Fig. 4 Emulated distributions for LW cloud forcing (global annual means) using common parameter samples. Distributions for the GA7.05 (blue) and GA8 (orange) PPEs (**a**) and their differences (**b**). **c** Schematic of common (C_i) and unique (U_i, V_i) parameter sampling, and how differences in emulated predictions are calculated.



(c)

GA7 LW cloud forcing	U_1	U_2	U_3	...	Average
C_1	14.732	16.548	14.784	...	15.355
C_2	17.219	18.357	17.113	...	17.563
C_3	12.539	13.794	12.449	...	12.927
...	...	...	...	...	...
Average	14.830	16.233	14.782	...	15.282

Difference

GA8 LW cloud forcing	V_1	V_2	V_3	...	Average
C_1	19.916	19.727	18.744	...	19.462
C_2	17.809	18.182	19.277	...	18.423
C_3	18.963	18.812	18.493	...	18.756
...	...	...	...	...	...
Average	18.896	18.907	18.838	...	18.880

3. CHANGES IN CLOUD FORCING ERRORS

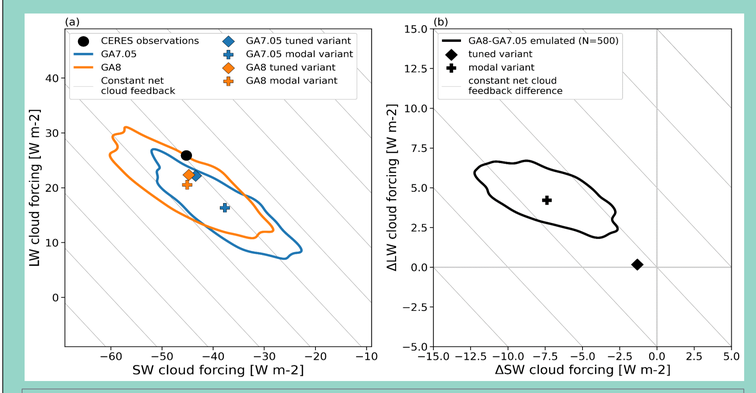


Fig. 5 As Fig. 4, but for LW vs SW cloud forcing. The tuned (diamonds) and modal (crosses) variants for each PPE are also shown. Contours enclose 95% of the data.

We find systematic errors in LW and SW cloud forcing for the GA7.05 PPE (Fig. 5)

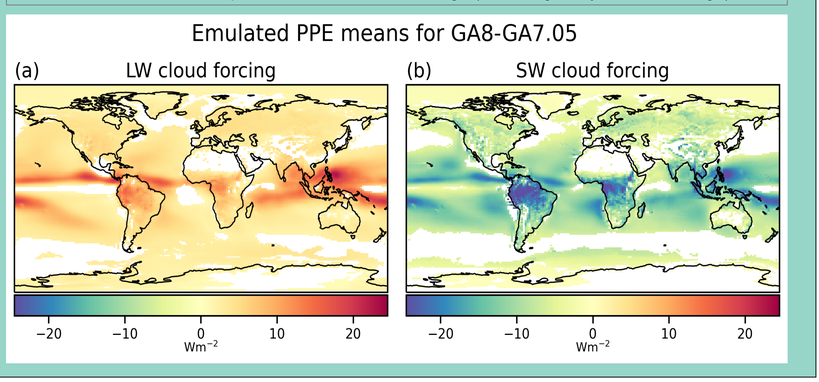
Improvements in marginal distributions for the GA8 PPE, which are robust across parameter settings.

Little improvement in the systematic error of the joint distribution.

Spatial changes reflect global means, with improvements in LW and SW cloud forcing components (Fig. 6).

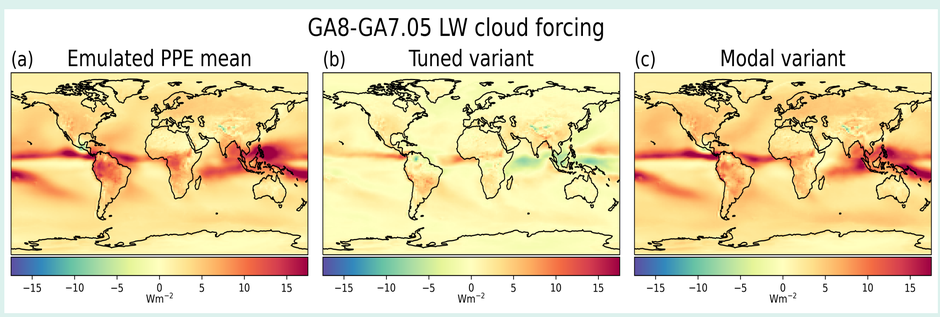
Large changes for deep convective regions; smaller changes in extratropics.

Fig. 6 PPE mean spatial maps of emulated GA8-GA7.05 differences using common parameter samples. Maps are shown for LW (**a**) and SW (**b**) cloud forcing. Masking is applied where fewer than 97.5% of the emulated predictions have the same sign (indicating no systematic change).



4. A CHEAPER ALTERNATIVE TO THE PPE AND TUNED VARIANT TO UNDERSTAND SYSTEMATIC CHANGES?

Fig. 7 a Same as Fig. 6a. Also shown are spatial fields of differences in LW cloud forcing for the tuned (**b**) and modal (**c**) variants. (These are reconstructed from projections onto the leading EOFs for each PPE.)



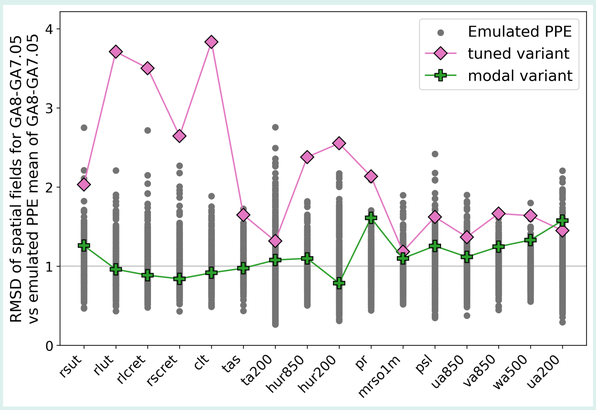


Fig. 8 Root mean-square differences (RMSDs) of spatial fields for GA8-GA7.05 for the emulated PPE mean vs the tuned variant (pink diamonds), and vs the modal variant (green crosses), across different variables. For context, RMSD values are given for each sample in the emulated PPE.

Differences between the GA8 and GA7.05 PPEs for cloud forcing are not well represented by the tuned variant (Figs. 5 and 7).

An alternative 'modal' variant (Fig. 1) is more representative of changes, across a number of variables (Figs. 5, 7 and 8).

Such an alternative to the tuned variant could be useful for tracking changes in errors in successive model configurations.

SUMMARY

- We are developing methods to evaluate changes in systematic errors in successive configurations of HadGEM3 using PPEs.
- Systematic errors in cloud forcing are improved in marginal distributions for the GA8 PPE vs the GA7.05 PPE, but they remain in the joint distribution.
- A 'modal' variant provides a better representation of changes in model errors between these PPEs than the tuned variant. Such a variant may be a useful tool for tracking changes in systematic errors during model development.

Met Office Hadley Centre, FitzRoy Road, Exeter, Devon, EX1 3PB United Kingdom
Tel: +44 1392 885680 Fax: +44 1392 885681
Email: john.rostron@metoffice.gov.uk

1. Rostron et al. (2020, Clim Dyn., https://doi.org/10.1007/s00382-020-05281-8)
2. Sexton et al. (2021, Clim Dyn., https://doi.org/10.1007/s00382-021-05709-9)