Answer to Anonymous Referee #2

We thank the anonymous referee for the helpful suggestions which will lead to an improved manuscript. In the following, we provide a point-by-point answer to the reviewer whereby we show the reviewer's comment in black and our response in blue.

General comments
This study developed a modular framework for statistically emulating the CMIP5 ESM global spatially and temporally correlated yearly land temperature field time series. It is shown that this emulator framework can reasonably well emulate the median of the single-ESM initial-condition ensembles of yearly temperature at a negligible computational cost, and with separate emulators calibrated for 40 individual ESMs of the CMIP5, can generate a large ensemble that closely resemble a multi-ESM initial-condition ensemble.
However the visual and quantitative verification of the emulations from this modular framework have demonstrated that further improvement in the emulators' design is needed before it can be used for the main focus of this study, of using the emulated large ensemble for initializing the climate impact and integrated assessment models used for quantifying the uncertainties in the possible climate system's responses to specific GHG emission pathways. In particular the main concerns are related to the non-dependency of the emulated future projections on the historic ESM performance, the need for more detailed evaluation of the emulation performance, and that the local grid-point scaling reduces the inter-model spread or suppresses the internal variability compared to the ESMs of CMIP5.

Each one of the concerns raised by the reviewer in this introduction is explained in more detail by the reviewer in the specific comments below. Hence, we directly answer these points below.

Specific comments
1. The computationally cheap statistical emulator developed in this study aims to produce realizations which closely resemble initial-condition ensemble members of the considered ESMs. It will be useful to make a statement on whether it is expected that the future warming projected by each emulation member will be affected by the known differences between observed and simulated historical climate in the ESM considered for calibrating each modular emulator framework. An earlier study cited by the authors (Goodwin, 2016), using large ensemble from an efficient ESM, had suggested that part of the upper range of twenty-first century CMIP5 warming projections may reflect historical simulation–observation inconsistencies.

In this study, we provide a computationally cheap machinery which can take a run from any ESM it is provided with and generate additional realizations closely resembling an initial-condition ensemble. Thus, it is explicitly not part of the tasks of our emulator to judge the ESM simulations. Instead, the emulator is designed to be flexible enough to emulate whatever ESM run it is provided with. The idea behind this is that users may identify a subset of ESMs most suitable for their task at hand and then have the opportunity to generate additional realizations which approximate an initial-condition ensemble of this subset of ESMs. Consequently, validation of ESMs does not lie within the scope of the present paper. To clarify this, we will include statements stressing that we only focus on emulation here, not on model performance and its implications for the realism of the projections.

2. The ability of the ensemble of emulations to capture the distribution of ESM runs is evaluated in this study. It would have been useful to additionally evaluate the performance of these emulations in comparison with the pattern scaling approach used to relate the grid-scale local temperature linearly

with the global mean temperature. An earlier study cited by the authors (Castruccio et al., 2014), demonstrated using a lack-of-fit statistic that the emulated local annual temperatures outperformed pattern scaling at grid point scales, and captures the nonlinear evolution of spatial patterns of climate anomalies inherent in transient climates. A similar evaluation could have been used in the present study to compare the measured emulation performance with a simple global pattern scaling.

We thank the reviewer for this suggestion. We will consider including a supplementary comparison of true ESM time series to time series obtained by simple pattern scaling and by our emulator. This will highlight the most important distinction between pattern scaling and our emulation approach, namely that traditional pattern scaling focuses solely on the mean response and does not contain a local variability module. Hence, additional realizations obtained from pattern scaling are by definition far more underdispersive than output from the emulator introduced in this study.

3. The quantitative verification in section 6.2 brought out that although the CMIP5 ESM median is emulated well, the emulations are underdispersive compared to the training run, particularly at regional scales. This issue is identified to be related to the emulators' design. Although the authors conclude that this underestimation is the result of a trade-off between parameter estimation robustness and ensemble reliability, it will be challenging to expect that the climate impact and integrated assessment models initialised with these emulations will be able to infer properties of the existing CMIP5 multi-ESM initial-condition ensemble runs which have not been generated yet.

We would like to thank the reviewer for noting that the added value of the emulated ensemble members has not been shown in sufficient detail. To accommodate for this (and in connection with the call for more space-time validation by other referees), we will include further assessments of the space-time characteristics in the revised manuscript. Among other things, this will show that the emulations reliably reproduce the variability of ESM simulations at the grid-cell level.

Technical comments

1. Section 2 introducing the proposed modular emulation framework in the context of previous studies may be merged into Section 4 describing this additive framework with three sub-modules so that the extent to which the previously developed methods were utilised in the implementation of this modular framework can be brought out clearly.

Please note that we deliberately kept the framework introduction in the context of previous studies (Sect. 2) separate from the detailed study-specific technical implementation (Sect. 4) because we believe that this makes it easier to follow the manuscript. In the revised manuscript, Sect. 7.1 will be expanded to discuss similarities between existing literature and our emulator implementation in more detail.

2. Line 353, please correct "underdispersion".

This is noted and will be done.