Reply to comments by Referee #1

The study "TheDiaTo (v1.0) – A new diagnostic tool for water, energy and entropy budgets in climate models" by Lembo et al describes Thermodynamic Diagnostic Tool (TheDiaTo), a new diagnostic tool that evaluates the energy budget, the hydrological cycle, the Lorenz Energy Cycle and the material entropy production. This is a very useful tool for evaluating climate models at global scale, as well as over land and ocean. The strength of this tool is its flexibility, so it can be run over a large multi-model ensemble (such as CMIP) and give a robust overview of the performance of the models. The paper describes the components of the tool and their results well. The figures and tables are clear.

Specific comments:

- Introduction: For energy and water budget studies using observations, please add L'Ecuyer et al, 2015 and Rodell et al, 2015. For studies evaluating the energy and water budgets in climate models, please add Demory et al, 2014; Terai et al, 2017; Vanniere et al, 2018 (e.g. in P2 L29, P3 L22, P4 paragraph 1.2). These studies highlighted the need for more robust evaluation of these budgets in climate models.

Thank you for the suggestions. We have added these references in a new sentence in Sect. 1.2: "Water mass budget has been assessed in observations (L'Ecuyer et al., 2015; Rodell et al., 2015), as well as in climate models, focusing on the hydrological cycle alone (Terai et al., 2018) or evaluating it in conjunction with the energy budgets (Demory et al., 2014; Vannière et al., 2019)."

- section 2 data and software requirements: are calculations performed on the native grid of the models or on a common grid?

The calculations are performed on the native grid of the model, as provided in the CMIP repository. We specified this in Sect. 2 by adding the following sentence: "For model inter-comparisons (Section 5), computations are performed on the native grid of the model."

- section 3: can the tool consider observational data, where available, to validate the models?

The tool can be used to compare models and observational data, as long as the data are global-scale and on a regular grid. The ESMValTool v.2 framework allows for straightforward comparison with Reanalyses (as it will be discussed in the ESMValTool v.2 report that is soon to be submitted on this same journal). Since referee #1 also suggested to mention that, we have added a comment on the range of applicability to observational-based gridded datasets in Sect. 2 as follows: "Therefore, the tool is suitable for the evaluation of any kind of gridded datasets, provided that they contain the necessary variables on a regular grid, including blends of observations and Reanalyses. In our description of the software features, we focus on model evaluation and multi-model intercomparison."

- section 3: can the tool also work with seasonal means?

Technically, no constraint is given on the usage of a subset of months. Although this would cause in principle no problem on the computations of the budgets and the LEC strength, the material entropy production would be inconsistent, because there is no straightforward argument justifying that the indirect method would provide a reasonable estimate of this quantity (as mentioned when discussing Eq. 11 in Sect. 3.4). A thorough explanation of the assumptions is provided in Lucarini et al. 2011. Furthermore, the implied transports rely on the fact that the differential heating is balanced by the oceans and the atmosphere, but this is not certainly accurate *at the smaller than annual timescale (cfr. for instance Rose and Ferreira 2013 on the validity of the Bjerknes compensation mechanism).*

- section 3.2 hydrological cycle: Assessing the water budget in climate models is particularly important to determine their conservation error and its evolution throughout the simulation. A model may have a conservation error but that remains stable over time, while another one may constantly lose/gain mass over time (as noted in Liepert and Previdi, 2012). In this study, this would be particularly important to understand what is happening in BNU. Is it possible to see the time evolution of the conservation error with TheDiaTo (Fig. 2 of Liepert and Previdi), and if not, would it be possible to add it? This would be beneficial for a thorough evaluation of the climate models.

The diagnostic tool allows to obtain annual mean time series of energy and water mass budgets for the Norther Hemisphere, the Southern Hemisphere and the global mean. For sake of brevity, we decided not to include them. An example of what the time series for the BNU-ESM model would look like is shown in Figures 1-3. It can be noticed that the conservation error is quite massive, compared to the other models, but relatively constant.

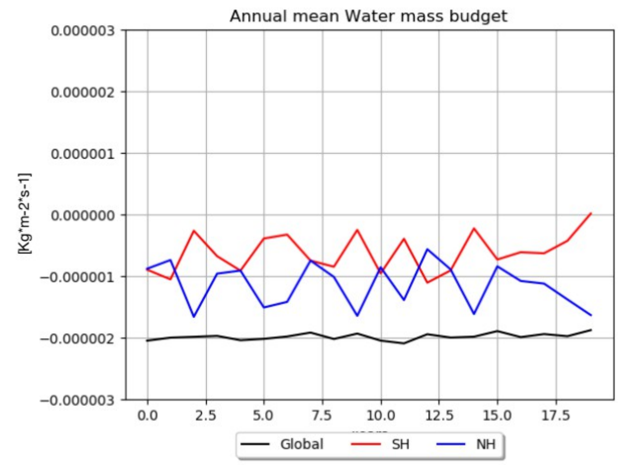

Figure 1: *annual mean time series of the energy budget for BNU-ESM pre-industrial run. Red denotes Southern Hemisphere, blue denotes Northern Hemisphere, black the global mean.*

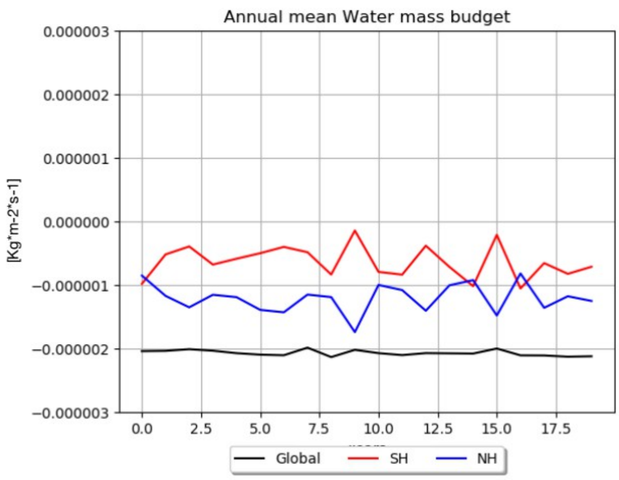

Figure 2: *same as Figure 1, for the last 20 years of the historical experiment.*

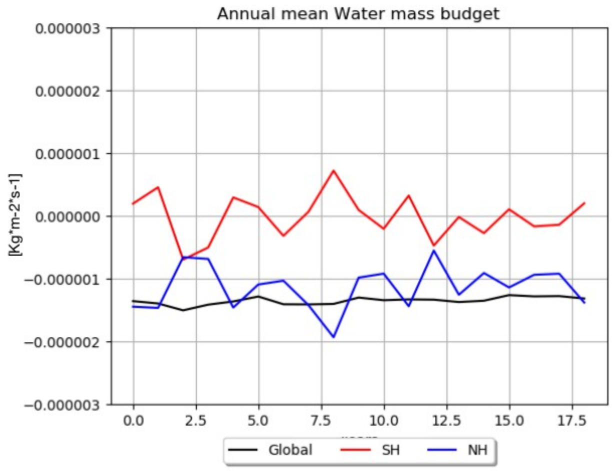

Figure 3: *same as Figure 1, for the last 20 years of the RCP8.5 experiment.*

- P15 L21: is there a reason for picking this period: 2441-2460?

The reason for that is that there is no other section of the pre-industrial CanESM2 run for which 20-years of all the needed variables is available on the repository. We explained this by adding the following sentence: "The choice of the sub-period is motivated by the fact that it is the only part of the run for which a 20-years subsequent dataset of the needed variables is available in the repository."

- P18 L28 & L32: 'bias' generally refers to errors relative to observations. Here, it may be more robust to refer to 'imbalance'.

Thank you. We replaced "bias" with the more appropriate term "imbalance".

- Tables 8 and 9: specify that rows are for PiC, hist and rcp85.

We have changed the table captions as follows: "For each model, the first row denotes the estimates from piC, the second row from hist, the third row from rcp8.5".

There are several typos throughout the paper. I noted a few here, but there are surely others:
- P1 L5: replace 'Top-of-Atmosphere (TOA)' by TOA, as defined just earlier

Done. Thank you.

- P2 L6: replace by 'cryosphere'

Done. Thank you.

- P2 L16: remove 'of things'

Done. Thank you.

- P2 L21: be more specific with 'the required standards mentioned above' so we don't have to jump back to recall these

We have replaced "prepared according to the required standards mentioned above" with "containing process-based and end-user-relevant diagnostics".

- P2 L23: same as above for ’enunciated above ideas’

We have added “reaching a steady state as a result of an interplay of multi-scale processes”.

- P5 L5: same as above C2 for ’mentioned aspects’

We have added “(i.e. the energy and water mass budgets and meridional transports, the LEC strength and the material entropy production)”.

- P7 L18 & P15 L26: replace ’EB’ by energy budget

Done. Thank you.

- P10 L 13: ’lhs’ -> left-hand side

Done. Thank you.

- P12 L10: ’rhs’ -> right-hand side

Done. Thank you.

- P17 L13 ’edddy’ -> ’eddy’

Done. Thank you.

- P20 L26: correct to ’orographically’

We believe that the correct spelling is “orografically”.

- P22 L25: ’rato’ -> ’ratio’

Done. Thank you.

- P22 L30: ’use’ -> ’used’

Done. Thank you.

- P23 L17: ’effected’ -> ’affected’

What we mean here is that baroclinic eddies “effect” baroclinic conversions. They can be both “synoptic-scale” and “planetary-scale” eddies, as shown in a recent paper of us (Lembo et al. 2019, 46, Geophys. Res. Lett.). We have added this reference, since it was not yet accepted by the time the manuscript was submitted.

- P24 L25: ’loking’ -> ’looking’

Done. Thank you.

- P25 L15: ’vertcial’ -> ’vertical’

Done. Thank you.