EGU24-1181, updated on 24 Jan 2025
https://doi.org/10.5194/egusphere-egu24-1181
EGU General Assembly 2024

Evaluation of atmospheric circulation of CMIP6 models for extreme temperature events using Latent Dirichlet Allocation

Nemo Malhomme[1,2], Bérengère Podvin[3], Davide Faranda[1], and Lionel Mathelin[2]
[1]ESTIMR, Université Paris-Saclay, CNRS, CEA, UVSQ, Laboratoire des sciences du climat et de l'environnement, 91191, Gif-sur-Yvette, France.
[2]LISN, CNRS, Université Paris-Saclay, 91405, Orsay, France.
[3]Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire EM2C, 91190, Gif-sur-Yvette, France.

Climate models aim at representing as closely as possible the statistical properties of the climate components, including extreme events on which fine-tuning data may be less available. This is a fundamental requirement to correctly project changes in their dynamics due to anthropogenic forcing.
In order to evaluate how closely models match observations, we need algorithms capable of selecting, processing and evaluating relevant dynamical features of the climate components. This has to be reiterated efficiently for large datasets such as those issued from the Coupled Model Intercomparison Project 6 (CMIP6). In this work, we use Latent Dirichlet Allocation (LDA), a statistical soft clustering method initially designed for natural language processing, to extract synoptic patterns from sea-level pressure data and evaluate how close the dynamics of CMIP6 climate models are to ERA5 reanalysis, both in the general case and in the case of extreme temperature events.

LDA allows for learning a basis of decomposition of maps into objects called "motifs". From the ERA5 sea-level pressure data, the method robustly extracts a basis of motifs that are interpretable objects at synoptic scale, i.e. cyclones or anticyclones. Pressure data can be projected onto this basis, yielding motif weights that contain local information about the large-scale atmospheric circulation. LDA decomposition is efficient and sparse: most of the information of a given map is contained in few motifs. It is therefore possible to decompose any map in a limited number of easy-to-interpret synoptic objects. This allows for a variety of new angles for statistical analysis.

The weights statistics can be used to characterize the general and extreme dynamics in reanalysis and model data. By comparing the statistics obtained from reanalysis data with those obtained from a selection of CMIP6 models, we can quantify errors on each localized circulation pattern and identify model-specific and model-agnostic errors. We found that, on average, large-scale circulation is well predicted by the models, but model errors are increased for extreme events such as heatwaves and cold spells. Additionally, Mediterranean motifs were found to be associated with significant model errors for all the considered models in all cases.