PALAEO-DMC: DATA-MODEL_COMPARISON

Palaeo-DMC is a python package that performs PALAEOclimate Data-Model_Comparison using a Gaussian process (GP) to make a statistical estimate of continuous fields of the mean and standard deviation of a geological climate quantity derived from scattered point observations. The uncertainty of each data point is accounted for in the process and must be supplied with the input. The GP is then compared with regular gridded general circulation model (GCM, “climate model”) output and several comparison metrics are produced and plotted.

INSTALLATION

Untar Palaeo_DMC.tar in a directory location of your choosing. The contents of your directory should then be:

- Palaeo_DMC.py - main python script
- ./core/ - directory containing required supporting scripts
- ./Observation_Data/ - directory containing sub-directories of observation files. 2 examples are supplied:
  - PlioMIP_anom_SST - anomalies of sea surface temperature between the Pliocene (~3 Mya) and pre-industrial (Dowsett et al., 2012).
  - EoMIP_anom_SST - anomalies of sea surface temperature between the Eocene (~50 Mya) and pre-industrial (Lunt et al., 2012).
- ./Model_Data/ - a directory containing sub-directories of GCM data. 2 sets of examples are supplies:
  - PlioMIP - anomalies of sea surface temperature between the Pliocene (~3 Mya) and pre-industrial from the PlioMIP set of models (Haywood et al., 2013).
  - EoMIP - anomalies of sea surface temperature between the Eocene (~50 Mya) and pre-industrial from the EoMIP set of models (Lunt et al., 2012).
- ./Output_Examples/ - a directory containing examples of the output produced by the script.
  - O-EoMIP_anom_SST_M-EoMIP_Th-1/
  - O-EoMIP_anom_SST_M-EoMIP_Th-4/
  - O-PlioMIP_anom_SST_M-PlioMIP_Th-1/

Packages Required

The following packages are required and will need to be installed by the user if not already available in their local python setup.

Standard packages from PyPI

numpy

netCDF4

matplotlib

basemap

glob

cmocean

GPy

Package GPy is available from PyPI. It is the Sheffield University machine learning group Gaussian process toolbox which is used for the core statistical processes of this method.

INPUTS

Observation Data

Observation_Data/<Obs_set>/*.txt

Geological observations are stored in simple free format text files with 4 columns of data: x, y, data, SD. These data should be placed in a sub-directory of ./Observation_Data/. Multiple datasets can be included in a single sub-directory for e.g. sensitivity studies. All files to be analysed must be named with the suffix '.txt' to be found by the script. The filename (without the .txt suffix) appears in output plots and tables to identify the data.

- x - longitude (within the range -180 to +180 degrees)
- y - latitude (within the range -90 to +90 degrees)
- data - data value
- SD - standard deviation of data value at that location

Climate Model Data

Model_Data/<GCM_set>/*.nc

Climate model data are supplied as netcdf files containing the data variable with the longitude and latitude dimensions. Multiple datasets can be included in a single sub-directory for comparison with observations. All files to be analysed must be named with the suffix '.nc' to be found by the script. The filename (without the .nc suffix) appears in output plots and tables to identify the data. The following variable names must be used:

- lon - longitude (within the range -180 to +180 degrees)
- lat - latitude (within the range -90 to +90 degrees)
- var - the data variable to be analysed

Model_Data/<GCM_set>/mask.nc

As well as the data files, a single mask file (0-1) should be created in the same sub-directory specifying the region to be analysed e.g. all land, all ocean or any specified region: 1 for regions to be analysed and 0 for regions to be masked out). This file should be name must be mask.nc with the following variables:

- lon - longitude (within the range -180 to +180 degrees)
- lat - latitude (within the range -90 to +90 degrees)
- mask - the mask variable defining the region to be analysed (=1, 0 elsewhere)

RUNNING THE CODE

The script is called from the main directory level:

python DMC.py *<GCM_set>* *<Obs_set>* *<thinby>*

- o *<GCM_set>* is the name of the sub-directory of Model_Data containing the netCDF files of model data and the mask file.
- o *<Obs_set>* is the name of the sub-directory of Observation_data containing the text files of observation data.
- o <thinby> is an integer parameter allowing reduction is the size of the calculation if memory or cpu limits are reached on the local machine. Every 1 in *thinby* rows and columns are used for the analysis: a value of 1 uses all of the available data.

The examples supplied with the software would be analysed with:

- python Palaeo_DMC.py PlioMIP PlioMIP_anom_SST 1
- python Palaeo_DMC.py EoMIP EoMIP_anom_SST 1
- python Palaeo_DMC.py EoMIP EoMIP_anom_SST 4

OUTPUTS

./Output/O-<Obs_set>_M-<GCM_set>_Th-<thinby>

Output is sent to a directory named and created automatically by the script based on the input options used when calling it.

The following outputs are produced:

- Plots of all GCM models
- Plots of observation data points and uncertainties
- Plots of Gaussian process mean and SD fields generated from observations
- Difference between Gaussian process and climate model for each combination of data
- Plots of comparison metrics for each GCM/ observation combination
  - Pointwise RMSE
  - Gridded RMSE
  - Log-likelihood using full covariance matrix (LL-FC)
  - Log-likelihood using diagonal-only covariance matrix (LL-VO)
- Taylor diagrams using the standard pointwise data and the gridded GP data
- Text file table of comparison metrics

We thank Yannick Copin for the code that produces the Taylor Diagrams (https://gist.github.com/ycopin/3342888).

ACKNOWLEDGEMENTS

Palaeo_DMC.py written by Fran Bragg, School of Geographical Sciences, University of Bristol

All code in ./core written by Prof. Richard Wilkinson, University of Sheffield except taylorDiagram.py (Yannick Copin, https://gist.github.com/ycopin/3342888) and the plot_map function of Utilities.py written by Fran Bragg.

REFERENCES

Dowsett, H. J., Robinson, M. M., Haywood, A. M., Hill, D. J., Dolan, A. M., Stoll, D. K., Chan, W. L., Abe-Ouchi, A., Chandler, M. A., Rosenbloom, N. A., Otto-Bliesner, B. L., Bragg, F. J., Lunt, D. J., Foley, K. M. & Riesselman, C. R. 2012. Assessing confidence in Pliocene sea surface temperatures to evaluate predictive models. *Nature Climate Change,* 2**,** 365-371.

Haywood, A. M., Hill, D. J., Dolan, A. M., Otto-Bliesner, B. L., Bragg, F., Chan, W. L., Chandler, M. A., Contoux, C., Dowsett, H. J., Jost, A., Kamae, Y., Lohmann, G., Lunt, D. J., Abe-Ouchi, A., Pickering, S. J., Ramstein, G., Rosenbloom, N. A., Salzmann, U., Sohl, L., Stepanek, C., Ueda, H., Yan, Q. & Zhang, Z. 2013. Large-scale features of Pliocene climate: results from the Pliocene Model Intercomparison Project. *Climate of the Past,* 9**,** 191-209.

Lunt, D. J., Dunkley Jones, T., Heinemann, M., Huber, M., LeGrande, A., Winguth, A., Loptson, C., Marotzke, J., Roberts, C. D., Tindall, J., Valdes, P. & Winguth, C. 2012. A model-data comparison for a multi-model ensemble of early Eocene atmosphere-ocean simulations: EoMIP. *Climate of the Past,* 8**,** 1717-1736.