Common error patterns in the regional atmospheric circulation simulated by the CMIP multi-model ensemble

Swen Brands[1]

[1]MeteoGalicia

November 24, 2022

Abstract

The ability of global climate models to reproduce recurrent regional atmospheric circulation types is introduced as an overarching concept to explore potential dependencies between these models. If this approach is applied on a sufficiently large spatial domain, the similarity of the resulting error pattern can be compared from one model to another. By computing a pattern correlation matrix for a large multi-model ensemble from the Coupled Model Intercomparison Project Phases 5 and 6, groups of comparatively strong correlation coefficients are obtained for those models working with similar atmospheric components. Thereby, frequent shared error patterns are found within in the ensemble, which also occur for nominally different atmospheric component models. The error pattern correlation coefficients describing these similarities are nearly unrelated to model performance and can be used as statistical dependency weights.

Common error patterns in the regional atmospheric circulation simulated by the CMIP multi-model ensemble

Swen Brands [1]

[1]MeteoGalicia, Consellería de Medio Ambiente, Territorio y Vivienda - Xunta de Galicia, Santiago de Compostela, Spain

Key Points:

- Global Climate Models (GCMs) have common regional atmospheric circulation error patterns in the northern hemisphere extratropics.
- Similar error patterns are obtained for GCMs using the same AGCM family and, unexpectedly, also for some GCMs using different AGCM families.
- A set of weighting coefficients documenting statistical dependence is provided which is virtually unrelated to model performance.

Corresponding author: Swen Brands, `swen.brands@gmail.com`

Abstract

The ability of global climate models to reproduce recurrent regional atmospheric circulation types is introduced as an overarching concept to explore potential dependencies between these models. If this approach is applied on a sufficiently large spatial domain, the similarity of the resulting error pattern can be compared from one model to another. By computing a pattern correlation matrix for a large multi-model ensemble from the Coupled Model Intercomparison Project Phases 5 and 6, groups of comparatively strong correlation coefficients are obtained for those models working with similar atmospheric components. Thereby, frequent shared error patterns are found within in the ensemble, which also occur for nominally different atmospheric component models. The error pattern correlation coefficients describing these similarities are nearly unrelated to model performance and can be used as statistical dependency weights.

1 Introduction

As the number of nominally different Global Climate Models (GCMs) participating in the Coupled Model Intercomparison Project (CMIP) increases (Taylor et al., 2012; Eyring et al., 2016), so does the need to explore the degree to which they have been developed independently. This effort is important because the spread of the multi-model ensemble is assumed to provide reliable uncertainty estimates of the climate system's response to external forcing (Lee et al., 2021). Thus, similar development strategies, such as common parametrization schemes, reference datasets used for model verification and, most evidently, the sharing of entire component models representing the physical and biogeochemical properties of the climate system (Brands, 2022b; Brands et al., 2022) would weaken the ensemble's suitability for uncertainty estimation and also compromise the use of unweighted multi-model mean values (Masson & Knutti, 2011; Knutti et al., 2013; Abramowitz et al., 2019). To account for this, those GCMs presenting considerable dependencies with the other members of the multi-model ensemble are down-weighted or even eliminated (Boé, 2018; Brunner et al., 2020; Maher et al., 2021). Similar weighting strategies are applied as a function of the GCMs performance in reproducing key aspects of the observed climate (Brunner et al., 2020; Liang et al., 2020), which is particularly important if the model errors induce too strong or too weak feedback processes along the integration period which lead to an unrealistic climate response to external forcing (Hall & Qu, 2006; Nijsse et al., 2020; Tokarska et al., 2020; Simpson et al., 2021).

The present study focuses on the exploration of GCM *dependencies* and in this research field, two distinct approaches have been proposed so far (Boé, 2018). The *a posteriori* approach seeks dependencies by analysis of model output. In this approach, *distance* between the output variables from the different members of the multi-model ensemble, such as temperature, precipitation or sea-level pressure, is directly used to measure model dependency in a purely statistical way, the more distant models being less similar (Masson & Knutti, 2011; Knutti et al., 2017; Brunner et al., 2020). A variation of this approach is to apply model *errors* with respect to observations instead of inter-model distance, with larger errors indicating less similarity (Jun et al., 2008; Knutti et al., 2010; Pennell & Reichler, 2011; Bishop & Abramowitz, 2012). The principal limitation of the *a posteriori* approach is that decreasing distances or model errors are not necessarily indicative of increasing model dependencies (Annan & Hargreaves, 2017). In other words, a group of models might produce similar output in spite of being conceptually different, in which case more confidence should be put on them in the multi-model approach. If applied on its own, the a posteriori approach is therefore unable to distinguish similarity due to conceptual model dependencies from similarity due to convergence of independent models.

This is where the *a priori* approach comes into play. There, model similarity is estimated on the basis of expert knowledge about the models' architecture. Ideally, this should be done by the model developers themselves, comparing and discussing their own source codes. However, probably because of the codes' complexity and long history — some model developments started in the late 1960s (Volodin et al., 2010)—, and also because of legal code availability restrictions, the *a priori* approach is still in its infancy and only a few attempts to put light on the models in this context are reported in the literature. Annan and Hargreaves (2017) defined *a priori* model similarity by institutional belonging, i.e. GCMs from the same institution are more dependent than those from different institutions. Boé (2018) went into more detail by counting the number of common components used in the coupled model configurations participating in CMIP5, taking into account the four basic components atmosphere, land-surface, ocean and sea-ice that constitute the minimum climate system component coverage of the coupled model configurations currently used as global climate models (GCMs). Finally, Maher et al. (2021) define model dependencies by looking at common sub-models or versions thereof.

The present study proposes a synthesis of the *a priori* and*a posteriori* approaches. Exploiting the newly available global climate model metadata archive built by Brands et al. (2022), it is possible to identify the names and versions of up to ten climate system components for currently 61 nominally distinct coupled model configurations participating in CMIP5 and 6. Since the focus is here put on atmospheric circulation, the metadata of the atmospheric general circulation model (AGCM) is extracted for each of these GCMs. Then, those GCMs using the same AGCM or versions thereof are put into the same group, each one representing a specific "AGCM family". This *a priori* approach makes it possible to put GCMs from different institutions into the same group if they use the same AGCM, which is actually often the case in the CMIP ensemble.

As an alternative to the use of several atmospheric variables for calculating model distances or errors in an *a posteriori* manner, the present study applies a single, integrative variable: the 27 regional atmospheric circulation types defined by (Lamb, 1972) and Jenkinson and Collison (1977). Calculated upon 6-hourly instantaneous sea-level pressure values, these *Lamb Weather Types* (LWTs) are known to be linked with a number of key variables in atmospheric physics and chemistry (Trigo & DaCamara, 2000; Hertig et al., 2020) and can thus be considered an overarching concept to describe regional-scale climate variability (Huth et al., 2016). Here, LWT time series are calculated at each grid-box of a regular latitude-longitude grid covering the northern hemisphere extra-tropics (Jones et al., 2013) for each of the 61 GCMs mentioned above and for several reference reanalysis datasets used as quasi-observations. Then, the modeled and observed climatological relative frequencies of the 27 LWTs are compared and it is shown that the model-specific spatial patterns of the corresponding circulation error correlate considerably (r +0.65) for those GCMs belonging to the same AGCM group and, unexpectedly, even for some GCM pairs from different AGCM groups. This means that, if the right *a priori* and *a posteriori* methods are combined, then they mutually support each other. Finally, a set of statistical dependence weights is obtained from the error pattern correlation coefficients, which was found to be nearly unrelated to model performance.

2 Data and Methods

The present study makes use of the 6-hourly instantaneous LWT sequences for the time period from 1979 to 2005 computed in Brands (2022b), which were updated for the present study. The LWT approach is an automated circulation typing technique based the subjective classification made by Lamb (1972) for the British Isles. It is also known as the Jenkinson and Collison (1977) approach and provides 27 discrete regional atmospheric circulation types, each one representing a typical and recurrent synoptic situation affecting that region at a given point in time. These circulation types are entirely calculated upon zonal and meridional sea-level pressure gradients on a 16-point coordinate system covering 30 degrees in longitude and 20 degrees in longitude, centered on the respective region of interest. Figure 1 shows this coordinate system adjusted for the

Tokyo region as well as the type-specific SLP composite mean patterns to illustrate that the method also works fine other extratropical climate regimes.

Each circulation type is characterized by 1) the direction of the geostrophic flow at sea-level (or lack thereof) as indicated by the 8 main cardinal directions and 2) the predominance of cyclonic or anticyclonic conditions (or lack thereof) in presence or absence of the geostrophic flow. There are 8 *purely directional* (PD) types with no clear cyclonic or anticyclonic influence and a geostrophic flow blowing from the Northeast (NE), East (E), SE, S, SW, W, NW, or North (N) (see panels 10 to 17 in Figure 1). Furthermore, there are 16 *hybrid* types with either anticyclonic (panels 1 to 8) or cyclonic (panels 19 to 26) conditions *in combination* with a geostrophic flow from one of the aforementioned cardinal directions. Finally, there is one *purely anticyclonic* and another *purely cyclonic* type (panels 1 and 18, respectively), characterized by a negligible geostrophic flow, and one *unclassified* type characterized by weak pressure gradients and a lack of cyclonic or anticyclonic influence, corresponding to what is known as "barometric swamp" among weather forecasters (panel 27).

Here, the LWT approach is applied in a rolling manner (Otero et al., 2017), i.e. it is iteratively centered on each box of a regular latitude-longitude grid with a resolution of 2.5°, covering a zonal belt between 35° and 70°N. Along the time axis, the method loops through six-hourly instantaneous SLP values from 1979 to 2005, providing one discrete LWT per timestamp. A complete description of the method can be found in Jones et al. (1993) and also in Brands (2022b), the later being an open access study.

In Brands (2022b), the resulting 3-dimensional LWT arrays (dimensions: time, latitude and longitude) have been calculated for 2 distinct reanalyses and for the *historical* runs of 56 nominally different coupled model configurations contributing to CMIP5 and 6. For the present study, this catalogue has been extended by the ECMWF ERA5 reanalysis (Hersbach et al., 2020), here used as principal reference dataset, and by 5 additional GCMs, namely GFDL-ESM2G (Dunne et al., 2012), CMCC-CM2-HR4 (Cherchi et al., 2019), GFDL-ESM4 (Dunne et al., 2020), INM-CM5 (Volodin et al., 2017) and KACE1.0-G (Lee et al., 2019); the former participating in CMIP5 and the latter four in CMIP6, respectively. All applied LWT catalogues were permanently stored in Brands (2022a).

A detailed metadata archive for the GCMs used here is is provided in Brands et al. (2022) (see *get_historical_metadata.py* therein), including the names and versions of all component models used in these GCMs (up to 10), resolution details, reference articles and run specifications considered in the present study. For more details on the considered GCMs, the interested reader is also referred to Brands (2022b).

At each box of the aforementioned northern hemisphere grid, the Mean Absolute Error (MAE) of the $n = 27$ relative LWT frequencies for a given GCM, denoted m, is calculated with respect to the respective frequencies from the reanalysis, denoted o (Brands, 2022b; Wilks, 2006):

$$MAE = \frac{1}{n}\Sigma_{i=1}^{n}|m_i - o_i| \qquad (1)$$

, thereby obtaining 61 spatial error patterns (one for each GCM) covering the northern hemisphere extratropics. The corresponding maps can be found in the supplementary material to this article (see Open Research section below) and an illustrative example displaying the results for 4 nominally different GCMs is shown in Figure 2. Two of the GCMs shown therein are from the same development team (panels a and c) and the remaining two from others (b and d). Then, the error pattern correlation matrix is calculated in order to measure the spatial similarity of the error fields (see Figure 3) (Abramowitz et al., 2019). Since the correlation coefficient measures the linear similarity of two *anomaly* samples (error fields in this case), similarity is not necessarily related to model performance, here defined in Equation 1, and vice versa. We will come back to this point in Section 4.

Using the metadata collected in Brands et al. (2022), those coupled model configurations sharing the same atmospheric general circulation model (*AGCM*), or versions thereof, are put into the same group and placed next to each other in the correlation matrix. For ease of understanding, GCMs are printed normal and the AGCMs used therein are *highlighted* in the following.

3 Results

Along the diagonal of the correlation matrix in Figure 3, marked with black boxes, 12 distinct AGCM "groups" or "families" (used as synonyms here) can be distinguished, each one containing at least 2 GCMs (see Table 1). These groups house 58 out of the 61 considered GCMs, the remaining 3 GCMs using unique AGCMs. For 11 out of these 12 families, the within-group pattern correlation coefficients (r) do not fall below 0.65, except for EC-Earth2.3 and MIROC-ESM constituting outliers of their respective AGCM group (*IFS* and *MIROC-AGCM*). Excluding the latter two GCMs, these 11 families are hereafter referred to as "clusters" because in addition to forming an AGCM group, their members also meet the aforementioned within-group correlation threshold. From this it follows that FGOALS-g2 and g3 form an AGCM "group" (*GAMIL*, group 2 in Figure 3) but not a "cluster" because their within group correlation coefficient is too low (r= 0.47).

Concerning the pattern correlation between different AGCM clusters, placed away from the diagonal and depicted in blue for one illustrative comparison in Figure 3, the *ECAM* cluster correlates comparatively strong with the *HadGAM/UM*, *LMDZ*, *GSMUV/MRI-AGCM* and *INM-AGCM* clusters, yielding correlation coefficients in the range of 0.66–0.75, 0.60–0.79, 0.59–0.75 and 0.63–0.82, respectively, and even stronger associations with the *GFDL-AM* cluster (0.58–0.89). The *ECHAM* cluster is also closely associated with *CanAM4*, i.e. the AGCM used in CanESM2 (0.73–0.82).

The *HadGAM/UM* cluster yields correlation coefficients in the range of 0.66–0.80 and 0.62–0.73 with the *LMDZ* and *GFDL-AM* clusters, except for the somewhat weaker links with the *GFDL-AM* version used in KIOST-ESM (0.55–0.62). *HadGAM/UM* is also strongly linked with *CanAM4* (0.70–0.77) and with the *INM-AGCM* version used in INM-CM5 (0.73–0.80).

The two BCC-CSM versions are here assigned to the *CAM* group because BCC-CSM's atmospheric component *BCC-AGCM* was originally developed from *CAM3* (Wu et al., 2010). The *CAM* cluster correlates comparatively strong with one half of the *ECHAM* cluster (MPI-ESM-LR, MPI-ESM-MR, MPI-ESM1.2-LR and MPI-ESM1.2-HR, 0.61–0.81), as well as with GFDL-CM3 and GFDL-ESM2G (0.62–0.82), and with GISS-E2.1-G (0.73–0.81).

The *IFS* cluster is only moderately correlated with the remaining AGCM groups and the lowest pattern correlations with the other groups are obtained for the *MIROC-AGCM/CCSR-AGCM* cluster.

With $r < 0.40$ on average (see axis labels in Figure 2), MIROC-ES2L, MIROC5 and FGOALS-g2 are the most independent coupled model versions considered here, whereas MPI-ESM-LR, MPI-ESM-MR and MPI-ESM-1.2-LR are the most dependent or, if seen the other way around, most influential GCMs ($r > 0.70$). Among the institutions contributing a single model, IITM-ESM constitutes a rather independent GCM that relies on *GFS* in the atmosphere, which is not used by any other GCM. CSIRO-MK3.6 is also relatively poorly correlated with the other GCMs, but has not been further developed since CMIP5. As stated above, CanESM2's average correlation coefficient with the remaining GCMs is comparatively large.

Several sensitivity tests have been conducted to test whether the aforementioned results are robust to several well known uncertainty sources. These include 1) the use of alternative reference reanalysis datasets, 2) the application of alternative historical model integrations initiated from other starting dates of the corresponding pre-industrial control run and 3) the exclusion of those regions where the JRA-55 reanalysis compared with ERA-Interim does not rank first when treated as if it was another GCM, thereby indicating relevant reanalysis uncertainties (see Figure 1b in Brands (2022b)). The largest effect of reanalysis uncertainty is an increase in the average error pattern correlation coefficients for the *IFS* group of up to 0.20 when switching from the ECMWF products to JRA-55. This means that the error patterns of the EC-Earth versions are less similar to the remaining GCMs if validated against ECMWF reanalyses than for the validation against JRA-55 (compare Supplementary Figure 1a and b with c). A likely reason for this is that both EC-Earth and the ECMWF reanalyses use *IFS*, meaning that they are a priori dependencies in this case. Since EC-Earth's *performance* was also found to be slightly favoured by an evaluation against ECMWF reanalyses (Brands, 2022b), it may be argued that, from a model dependence point of view, JRA-55 is the more suitable reference reanalysis for multi-model evaluations in the northern hemisphere extratropics if EC-Earth is involved. In spite of this issue, the correlation matrices calculated upon the 3 mentioned reanalyses are similar to each other and the overall effects of reanalysis uncertainty on the results are small (see Supplementary Figure 1). The effects of internal model variability are even smaller (see Supplementary Figure 2), probably because the climatological mean state is studied here instead of inter-annual variability, the latter known to be more sensitive to this kind of variability in the extratropics (Maher et al., 2021). Removing the regions prone to reanalysis uncertainty from the study has more substantial effects on the results, but does not change the main conclusions either (see Supplementary Figure 3). The pattern correlation coefficients decrease only slightly (see boxplot next to the colorbar), the AGCM families are still visible along the diagonal and the similarity between the ECHAM and CAM families increases, tending to form a joint supercluster. Coming back to the full-domain analyses, if the error patterns for JRA-55 or ERA-Interim w.r.t ERA5 are correlated with those obtained for the GCMs, i.e. the reanalyses are treated as if they were GCMs, the coefficients are similar to those obtained for the majority of GCMs in Figure 3 (0.60 and 0.63 on average, not shown). This again points to common error structures in both the GCMs and reanalyses.

If subtracted from 1, the average error pattern correlation coefficient obtained from correlating a given GCM with all others can be used as model weight ($w = 1 - \overline{r}$, see Table 1). Interestingly, the correlation coefficient between these weights and the mean model errors derived and updated from Brands (2022b) is 0.20 only, which is insignificant for a two-tailed t-test conducted at a test level of 0.05.

4 Discussion and Conclusions

The present study has shown that the *a priori* grouping of the GCMs used in CMIP5 and 6 according to the applied atmospheric sub-model leads to clusters of similar spatial error patterns describing the models' capability to reproduce the regional atmospheric circulation as represented by the well known Lamb Weather Types. This way, 58 out of 61 considered GCMs can be be grouped into 12 distinct AGCM families, whereas the remaining 3 GCMs use unique AGCMs. For 11 of the thereby defined AGCM families, housing a total of 54 GCMs, the within-group error pattern correlation coefficients are sufficiently strong ($r > 0.65$) to depict *clusters* of statistical dependency visible along the diagonal of the error pattern correlation matrix. In some cases, the error patterns for *distinct* AGCM families also correlate strongly, e.g. for the ECHAM and GFDL-AM families. This probably indicates model convergence in spite of conceptual differences and should increase our confidence in the output of these models, particularly in those regions where they have shown to perform well (Brands, 2022b).

Complemented by the model performance estimates reported in the aforementioned study, the here presented dependence estimates provide consistent criteria for GCM weighting and selection, covering both CMIP5 and 6 (Lee et al., 2021). The correlation matrices w.r.t to various reanalysis datasets are provided in netCDF format (see supplementary material) and alternative average weights can be easily computed, e.g. by selecting only a single GCM per development team (Leduc et al., 2016), if this is preferred by the reader.

Making use of the metadata archive built by Brands et al. (2022), the GCMs can be alternatively ordered according to their sub-models for *other* climate system components, using appropriate alternative error measures. This effort, as well as the use of the proposed atmospheric circulation error to constrain future climate projections (Cox et al., 2018; Eyring et al., 2019), is left open for future studies.

Finally, it is noted that the model developers themselves have embarked on an effort to disentangle the complex dependencies in the CMIP ensemble from expert knowledge and source code, leading to impressive in-depth studies for specific climate system components (Séférian et al., 2020). At some point in the future, it might thus be possible to abandon the *a posteriori* or *ad-hoc* approach used to explore GCM dependencies solely with output data, in favour of a fully informed *a priori* approach (Annan & Hargreaves, 2017; Boé, 2018). The aforementioned metadata archive is a starting point for such an endeavor.

Open Research

The LWT catalogues and underlying Python code, including the applied GCM metadata archive, are publicly available from Brands (2022a) and Brands et al. (2022). The supplementary material to this article is available at `https://figshare.com/ndownloader/files/37598465`.

Acknowledgments

I would like to thank the research institutes participating in CMIP for sharing the GCM data analyzed here and also appreciate the public availability of the ECMWF ERA5 reanalysis dataset.

References

Abramowitz, G., Herger, N., Gutmann, E., Hammerling, D., Knutti, R., Leduc, M., … Schmidt, G. A. (2019). ESD reviews: Model dependence in multi-model climate ensembles: weighting, sub-selection and out-of-sample testing. *Earth System Dynamics*, *10*(1), 91–105. doi: 10.5194/esd-10-91-2019

Annan, J. D., & Hargreaves, J. C. (2017). On the meaning of independence in climate science. *Earth System Dynamics*, *8*(1), 211–224. doi: 10.5194/esd-8-211-2017

Bishop, C., & Abramowitz, G. (2012, 08). Climate model dependence and the Replicate Earth paradigm. *Climate Dynamics*, *41*, 885–900. doi: 10.1007/s00382-012-1610-y

Boé, J. (2018). Interdependency in multimodel climate projections: Component replication and result similarity. *Geophysical Research Letters*, *45*(6), 2771-2779. doi: https://doi.org/10.1002/2017GL076829

Brands, S. (2022a). A circulation-based performance atlas of the CMIP5 and 6 models for regional climate studies in the northern hemisphere [data set]. *Zenodo*. doi: https://doi.org/10.5281/zenodo.4452080

Brands, S. (2022b). A circulation-based performance atlas of the CMIP5 and 6 models for regional climate studies in the Northern Hemisphere mid-to-high latitudes. *Geoscientific Model Development*, *15*(4), 1375–1411. doi: https://doi.org/10.5194/gmd-15-1375-2022

Brands, S., Tatebe, H., Danek, C., Fernández, J., Swart, N., Volodin, E., … Tongwen, W. (2022). Python code to calculate Lamb circulation types for the northern hemisphere derived from historical CMIP simulations and reanalysis data [code]. *Zenodo*. doi: https://doi.org/10.5281/zenodo.4555367

Brunner, L., Pendergrass, A. G., Lehner, F., Merrifield, A. L., Lorenz, R., & Knutti, R. (2020). Reduced global warming from CMIP6 projections when weighting models by performance and independence. *Earth System Dynamics*, *11*(4), 995–1012. doi: https://doi.org/10.5194/esd-11-995-2020

Cherchi, A., Fogli, P. G., Lovato, T., Peano, D., Iovino, D., Gualdi, S., … Navarra, A. (2019). Global mean climate and main patterns of variability in the CMCC-CM2 coupled model. *Journal of Advances in Modeling Earth Systems*, *11*(1), 185-209. doi: 10.1029/2018MS001369

Cox, P., Huntingford, C., & Williamson, M. (2018, 01). Emergent constraint on equilibrium climate sensitivity from global temperature variability. *Nature*, *553*, 319-322. doi: 10.1038/nature25450

Dunne, J. P., Horowitz, L. W., Adcroft, A. J., Ginoux, P., Held, I. M., John, J. G., … Zhao, M. (2020). The GFDL Earth System Model version 4.1 (GFDL-ESM 4.1): Overall coupled model description and simulation characteristics. *Journal of Advances in Modeling Earth Systems*, *12*(11), e2019MS002015. doi: https://doi.org/10.1029/2019MS002015

Dunne, J. P., John, J. G., Adcroft, A. J., Griffies, S. M., Hallberg, R. W., Shevliakova, E., … Zadeh, N. (2012). GFDL's ESM2 Global Coupled Climate-Carbon Earth System Models. Part I: Physical formulation and baseline simulation characteristics. *Journal of Climate*, *25*(19), 6646–6665. doi: https://doi.org/10.1175/JCLI-D-11-00560.1

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. doi: https://doi.org/10.5194/gmd-9-1937-2016

Eyring, V., Cox, P., Flato, G., Gleckler, P., Abramowitz, G., Caldwell, P., … Williamson, M. (2019). Taking climate model evaluation to the next level. *Nature Climate Change*, *9*, 102-110. doi: https://doi.org/10.1038/s41558-018-0355-y

Hall, A., & Qu, X. (2006). Using the current seasonal cycle to constrain snow albedo feedback in future climate change. *Geophysical Research Letters*, *33*(3), L03502. doi: https://doi.org/10.1029/2005GL025127

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., … Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999-2049. doi: https://doi.org/10.1002/qj.3803

Hertig, E., Russo, A., & Trigo, R. M. (2020). Heat and ozone pollution waves in central and south Europe—characteristics, weather types, and association with mortality. *Atmosphere*, *11*(12), 1271. doi: https://doi.org/10.3390/atmos11121271

Huth, R., Beck, C., & Kučerová, M. (2016). Synoptic-climatological evaluation of the classifications of atmospheric circulation patterns over Europe. *International Journal of Climatology*, *36*(7), 2710-2726. doi: https://doi.org/10.1002/joc.4546

Jenkinson, A., & Collison, F. (1977). *An initial climatology of gales over the North Sea* (Vol. 62; Tech. Rep. No. 18). Bracknell, UK: Meterololgical Office.

Jones, P. D., Harpham, C., & Briffa, K. R. (2013). Lamb weather types derived from reanalysis products. *International Journal of Climatology*, *33*(5), 1129-1139. doi: https://doi.org/10.1002/joc.3498

Jones, P. D., Hulme, M., & Briffa, K. R. (1993). A comparison of lamb circulation types with an objective classification scheme. *International Journal of Climatology*, *13*(6), 655-663. doi: https://doi.org/10.1002/joc.3370130606

Jun, M., Knutti, R., & Nychka, D. W. (2008). Local eigenvalue analysis of CMIP3 climate model errors. *Tellus A: Dynamic Meteorology and Oceanography*, *60*(5), 992-1000. doi: 10.1111/j.1600-0870.2008.00356.x

Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., & Meehl, G. (2010, 05). Challenges in combining projections from multiple climate models. *Journal of Climate*, *23*, 2739-2758. doi: 10.1175/2009JCLI3361.1

Knutti, R., Masson, D., & Gettelman, A. (2013). Climate model genealogy: Generation CMIP5 and how we got there. *Geophysical Research Letters*, *40*(6), 1194-1199. doi: https://doi.org/10.1002/grl.50256

Knutti, R., Sedláček, J., Sanderson, B. M., Lorenz, R., Fischer, E. M., & Eyring, V. (2017). A climate model projection weighting scheme accounting for performance and interdependence. *Geophysical Research Letters*, *44*(4), 1909-1918. doi: https://doi.org/10.1002/2016GL072012

Lamb, H. (1972). British Isles weather types and a register of daily sequence of circulation patterns, 1861-1971. *Geophysical Memoir*, *116*, 85pp. (HMSO)

Leduc, M., Laprise, R., de Elía, R., & Šeparović, L. (2016). Is institutional democracy a good proxy for model independence? *Journal of Climate*, *29*(23), 8301 - 8316. doi: 10.1175/JCLI-D-15-0761.1

Lee, J., Kim, J., Sun, M.-A., Kim, B.-H., Moon, H., Sung, H. M., … Byun, Y.-H. (2019, 08). Evaluation of the Korea Meteorological Administration Advanced Community Earth-System model (K-ACE). *Asia-Pacific Journal of Atmospheric Sciences*, *56*, 381–395. doi: https://doi.org/10.1007/s13143-019-00144-7

Lee, J., Marotzke, J., Bala, G., Cao, L., Corti, S., Dunne, J., … Panickal, T., S. Zhou (2021). Future global climate: Scenario-based projections and near-term information. In V. Masson-Delmotte et al. (Eds.), (pp. 553–672). Cambridge University Press. doi: 10.1017/9781009157896.006

Liang, Y., Gillett, N. P., & Monahan, A. H. (2020). Climate model projections of 21st century global warming constrained using the observed warming trend. *Geophysical Research Letters*, *47*(12), e2019GL086757. doi: https://doi.org/10.1029/2019GL086757

Maher, N., Power, S., & Marotzke, J. (2021, 02). More accurate quantification of model-to-model agreement in externally forced climatic responses over the coming century. *Nature Communications*, *12*, 788. doi: 10.1038/s41467-020-20635-w

Masson, D., & Knutti, R. (2011). Climate model genealogy. *Geophysical Research Letters*, *38*(8), L08703. doi: https://doi.org/10.1029/2011GL046864

Nijsse, F. J. M. M., Cox, P. M., & Williamson, M. S. (2020). Emergent constraints on transient climate response (TCR) and equilibrium climate sensitivity (ECS) from historical warming in CMIP5 and CMIP6 models. *Earth System Dynamics*, *11*(3), 737–750. doi: 10.5194/esd-11-737-2020

Otero, N., Sillmann, J., & Butler, T. (2017, 04). Assessment of an extended version of the Jenkinson-Collison classification on CMIP5 models over europe. *Climate Dynamics*, 1559–1579. doi: 10.1007/s00382-017-3705-y

Pennell, C., & Reichler, T. (2011). On the effective number of climate models. *Journal of Climate*, *24*(9), 2358–2367. doi: 10.1175/2010JCLI3814.1

Simpson, I. R., McKinnon, K. A., Davenport, F. V., Tingley, M., Lehner, F., Fahad, A. A., & Chen", D. (2021). Emergent constraints on the large-scale atmospheric circulation and regional hydroclimate: Do they still work in CMIP6 and how much can they actually constrain the future? *Journal of Climate*, *34*(15), 6355–6377. doi: 10.1175/JCLI-D-21-0055.1

Séférian, R., Berthet, S., Yool, A., Palmiéri, J., Bopp, L., Tagliabue, A., … Yamamoto, A. (2020, 09). Tracking improvement in simulated marine biogeochemistry between CMIP5 and CMIP6. *Current Climate Change Reports*, *6*, 95-119. doi: 10.1007/s40641-020-00160-0

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, *93*(4), 485-498. doi: https://doi.org/10.1175/BAMS-D-11-00094.1

Tokarska, K. B., Stolpe, M. B., Sippel, S., Fischer, E. M., Smith, C. J., Lehner, F., & Knutti, R. (2020). Past warming trend constrains future warming in CMIP6 models. *Science Advances*, *6*(12), eaaz9549. doi: 10.1126/sciadv.aaz9549

Trigo, R. M., & DaCamara, C. C. (2000). Circulation weather types and their influence on the precipitation regime in Portugal. *International Journal of Climatology*, *20*(13), 1559-1581. doi: https://doi.org/10.1002/1097-0088

Volodin, E., Diansky, N., & Gusev, A. (2010). Simulating present-day climate with the inmcm4.0 coupled model of the atmospheric and oceanic general circulations. *Izvestiya, Atmospheric and Oceanic Physics*, *46*, 414-431. doi: https://doi.org/10.1134/S000143381004002X

Volodin, E., Mortikov, E., Kostrykin, S., Galin, V., Lykossov, V., Gritsun, A., … Iakovlev, N. (2017). Simulation of the present-day climate with the climate model INMCM5. *Climate Dynamics*, *49*, 3715–3734. doi: https://doi.org/10.1007/s00382-017-3539-7

Wilks, D. (2006). *Statistical methods in the atmospheric sciences* (2 ed.). Amsterdam: Elsevier.

Wu, T., Yu, R., Zhang, F., Wang, Z., Dong, M., Wang, L., … Li, L. (2010). The Beijing Climate Center atmospheric general circulation model: Description and its performance for the present-day climate. *Climate Dynamics*, *34*, 123-147. doi: https://doi.org/10.1007/s00382-008-0487-2

Table 1. Atmospheric general circulation model groups and coupled model configurations they are used in. GCMs belong to the same *cluster* if their AGCM is from the same group and if the error pattern correlation coefficients (r) with the remaining members of this group exceeds 0.65 (see Figure 3). Only 7 of the 61 considered GCMs cannot be assigned to one of the clusters, either because their within-group error pattern correlation coefficients are too low (this is the case for EC-Earth2.3, FGOALS-g2, FGOALS-g3 and MIROC-ESM) or because their AGCM is unique within the multi-model ensemble considered here (this is the case for CanESM2, IITM-ESM and CSIRO-MK3.6). Also shown are the number of GCMs pertaining to each AGCM group (parentheses in first column) and the weighting coefficients (1-$\bar{r}$, parentheses in second column) derived from the average error pattern correlation coefficients per GCM ($\bar{r}$) displayed in Figure 3. All results are w.r.t. ERA5.

AGCM group	**Coupled model configurations**
	GCMs fulfilling the clustering criteria (54)
HadGAM/UM (8)	ACCESS1.0 (0.36), ACCESS1.3 (0.36), ACCESS-CM2 (0.36), ACCESS-ESM1 (0.37)
	HadGEM2-CC (0.37), HadGem2-ES (0.36), Hadgem3-GC31-MM (0.36), KACE1.0-G (0.37)
ECHAM (8)	MPI-ESM-LR (0.28), MPI-ESM-MR (0.29), MPI-ESM1.2-LR (0.29), MPI-ESM1.2-HR (0.31)
	MPI-ESM-1-2-HAM (0.37), AWI-ESM-1-1-LR (0.36), NESM3 (0.32), CMCC-CM (0.36)
CAM (11)	CMCC-CM2-SR5 (0.39), CMCC-CM2-HR4 (0.39), CMCC-ESM2 (0.38), CCSM4 (0.41)
	NorESM1-M (0.37), NorESM2-LM (0.39), NorESM2-MM (0.37), SAM0-UNICON (0.39)
	TaiESM1 (0.38), BCC-CSM1.1 (0.36), BCC-CSM2-MR (0.35)
ARPECHE (4)	CNRM-CM5 (0.38), CNRM-CM6-1 (0.36), CNRM-CM6-1-HR (0.43), CNRM-ESM2-1 (0.37)
IFS (5)	EC-Earth3 (0.48), EC-Earth3-Veg (0.49), EC-Earth3-Veg-LR (0.38)
	EC-Earth3-AerChem (0.49), EC-Earth3-CC (0.54)
GFDL-AM (5)	GFDL-CM3 (0.29), GFDL-CM4 (0.37), GFDL-ESM2G (0.33), GFDL-ESM4 (0.36), KIOST-ESM (0.40)
GISS-E2 (3)	GISS-E2-H (0.47), GISS-E2-R (0.43), GISS-E2.1-G (0.37)
LMDZ (3)	IPSL-CM5A-LR (0.40), IPSL-CM5A-MR (0.39), IPSL-CM6A-LR (0.33)
MIROC-AGCM/CCSR AGCM (3)	MIROC5 (0.63), MIROC6 (0.49), MIROC-ES2L (0.71)
GSMUV/MRI-AGCM (2)	MRI-ESM1 (0.39), MRI-ESM2.0 (0.34)
INM-AM (2)	INM-CM4 (0.37), INM-CM5 (0.36)
	GCMs not fulfilling the clustering criteria (7)
GAMIL	FGOALS-g2 (0.62), FGOALS-g3 (0.40)
MIROC-AGCM/CCSR AGCM	MIROC-ESM (0.38)
CSIRO-AGCM	CSIRO-MK3.6 (0.54)
IFS	EC-Earth2.3 (0.42)
CanAM	CanESM2 (0.36)
GFS	IITM-ESM (0.44)

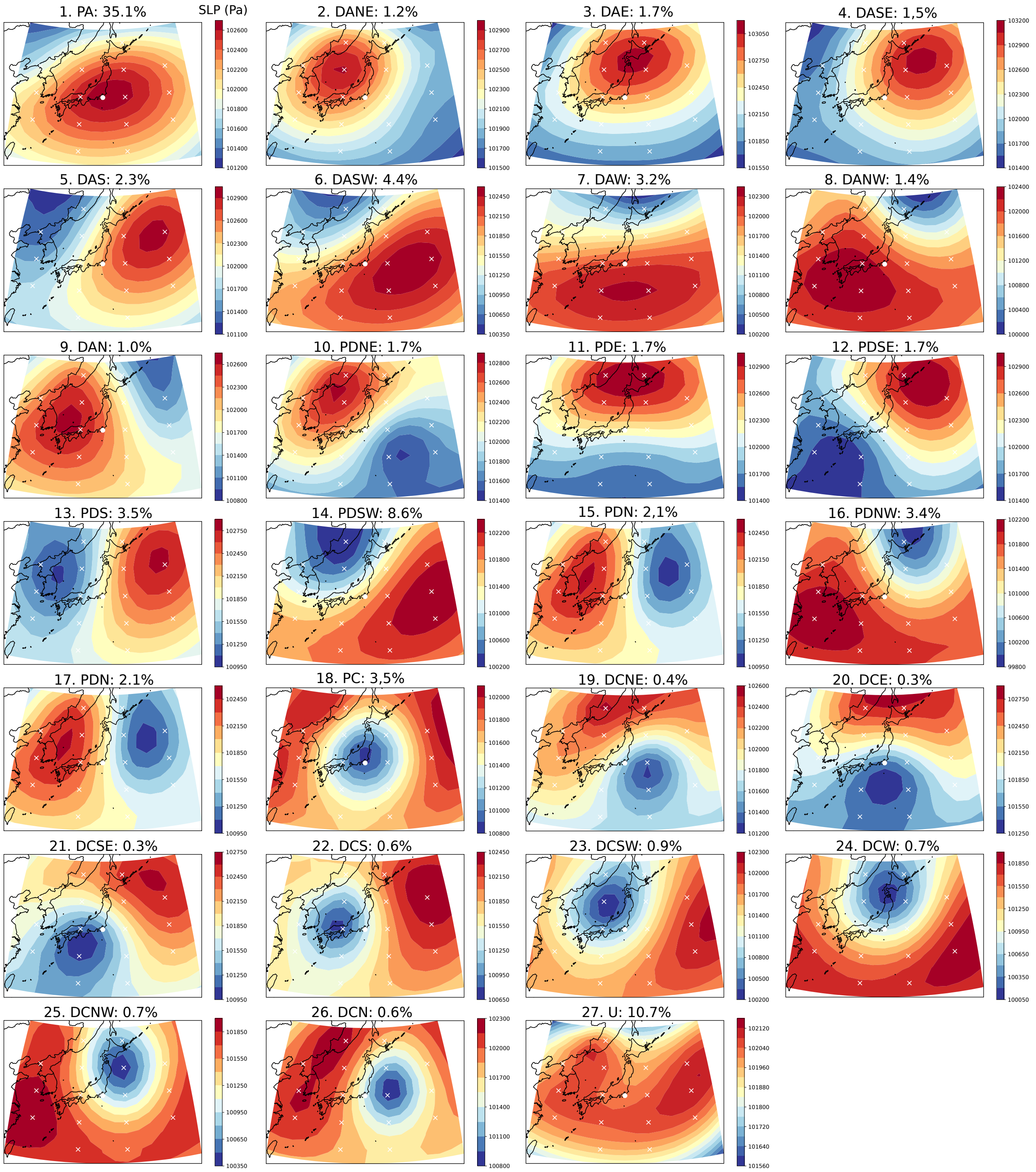

Figure 1. Composite mean sea-level pressure patterns (in Pa) for each of the 27 Lamb Weather Types over the Tokyo region. Also shown is the coordinate system the method is defined on and the relative occurrence frequencies of each type. Source: ERA5, period: 1979-2005

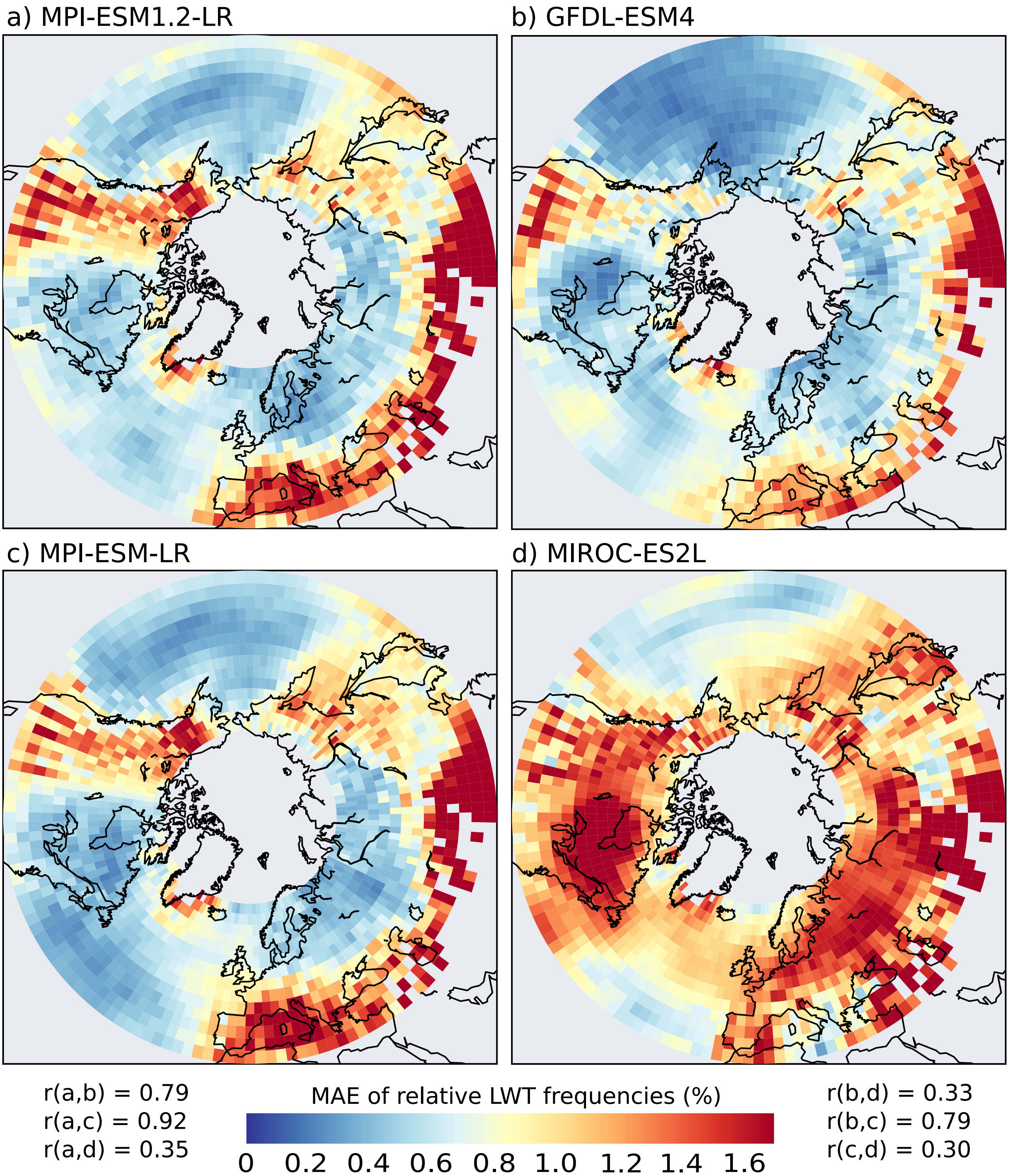

Figure 2. Spatial pattern of the Mean Absolute Error (MAE) in the relative frequencies of the 27 *Lamb Weather Types* for two GCMs pertaining to the same AGCM family (a and c) and for two GCMs pertaining to a distinct AGCM family each (b and d). Despite nominally distinct AGCMs are in use (ECHAM and GFDL-AM), the error pattern correlation (r) between the MPI and GFDL models is comparatively large.

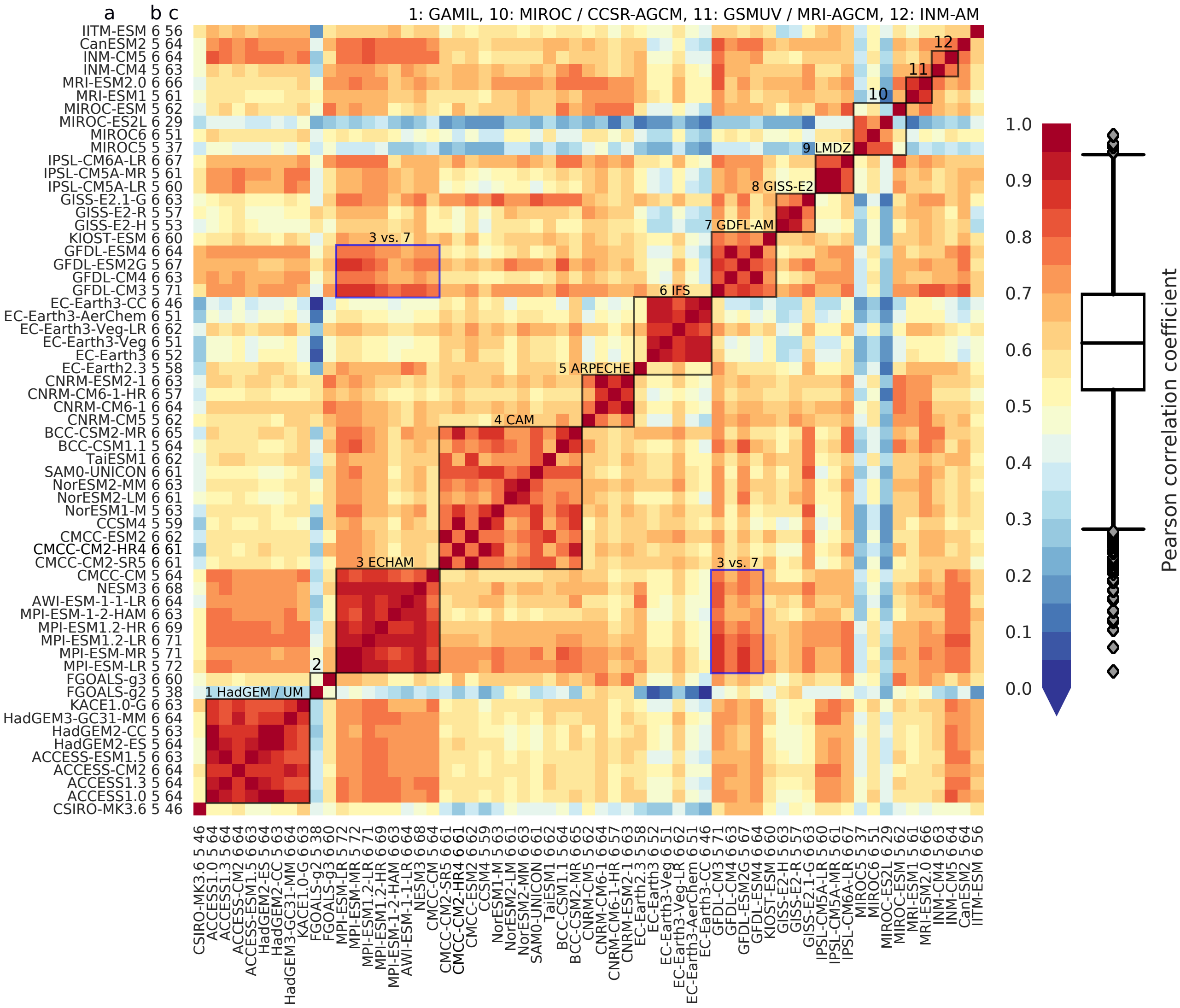

Figure 3. Spatial correlation of the northern hemisphere mean absolute error pattern in the relative frequencies of the 27 *Lamb Weather Types* for 61 distinct GCMs from CMIP5 and 6 evaluated against ERA5. The corresponding maps are provided in the supplementary material. The acronym (a), CMIP generation (b) and average spatial correlation coefficient $\times 100$ (c) of each GCM are provided along the axes. The boxplot describes the distribution of the correlation coefficients without repetitions and unity values. It is constructed with the median, interquartile range (IQR) and whiskers of this sample, the latter placed at the 25^{th} percentile - 1.5 $\times$ IQR and at the 75^{th} percentile + 1.5 $\times$ IQR. AGCM families are marked with black boxes. Blue boxes indicate an illustrative example for two distinct AGCM families correlating strongly.