EURO-CORDEX CMIP6 GCM Selection & Ensemble Design: Best Practices and Recommendations

History of changes

Date	**Description**
24.02.2023	First citable version of the paper published to Zenodo

Authors: Stefan Sobolowski[1], Samuel Somot[2], Jesus Fernandez[3], Guillaume Evin[4], Douglas Maraun[5], Sven Kotlarski[6], Martin Jury[5], Rasmus E. Benestad[7], Claas Teichmann[8], Ole B. Christensen[9], Katharina Bülow[8], Erasmo Buonomo[10], Eleni Katragkou[11], Christian Steger[12], Silje Sørland[13], Grigory Nikulin[14], Carol McSweeney[10], Andreas Dobler[7], Tamzin Palmer[10], Renate Wilcke[14], Julien Boé[15], Lukas Brunner[16], Aurélien Ribes[2], Said Qasmi[2], Pierre Nabat[2], Florence Sevault[2], Thomas Oudar[17], Swen Brands[18]

Affiliations:

1. NORCE Norwegian Research Center AS, the Bjerknes Center for Climate Research, Bergen Norway
2. Centre National de Recherches Météorologiques (CNRM), Météo-France, CNRS, Toulouse, France
3. Instituto de Física de Cantabria, CSIC-Universidad de Cantabria, Santander, Spain
4. Institut des Géosciences de l'Environnement (IGE), Univ. Grenoble Alpes, INRAE, CNRS, IRD, Grenoble INP, Grenoble, France
5. Wegener Center for Climate and Global Change, University of Graz, Graz, Austria.
6. Federal office of Meteorology and Climatology (MeteoSwiss), Zurich, Switzerland
7. Norwegian Meteorological Institute, Oslo, Norway
8. Climate Service Center Germany (GERICS), Helmholtz-Zentrum Hereon, Hamburg, Germany
9. Danish Meteorological Institute (DMI), Copenhagen, Denmark
10. Met Office Hadley Centre, Exeter, UK
11. Aristotle University of Thessaloniki, School of Geology, Department of Meteorology and Climatology, Thessaloniki, Greece
12. Deutscher Wetterdienst, Frankfurter Str. 135, 63067 Offenbach, Germany
13. SWECO Norway AS, Bergen, Norway
14. Swedish Meteorological and Hydrological Institute (SMHI), Norrkjöping, Sweden
15. CECI, Université de Toulouse, CERFACS, CNRS, Toulouse, France
16. Department of Meteorology and Geophysics, University of Vienna, Vienna, Austria
17. Météo-France, Toulouse, France
18. MeteoGalicia, Consellería de Medio Ambiente, Territorio y Vivienda - Xunta de Galicia, Santiago de Compostela, Spain

A. Introduction

The following white paper and the accompanying tables are meant to help modellers in EURO-CORDEX, and elsewhere, make informed decisions when it comes to selecting GCMs for downscaling from the CMIP6 archive. While the approach to GCM selection during the CMIP5 downscaling phase cannot be wholly considered an “ensemble of opportunity” it is also true that the selection of GCMs was not as rigorous as it could have been. We aim to improve upon this situation and help us construct smarter, more reliable and more useful downscaled ensembles (dynamical, statistical and hybrid) and make the selection process more transparent. The community should consider this a living document that will evolve along with our scientific understanding. What follows are sections describing the background, the problem and our ambitions (section A). Following that, we outline the key categories and metrics that we deem important for evaluation of GCMs for downscaling and ultimately constructing an ensemble that covers the range of possible outcomes given the scenarios currently available (section B-F). Lastly, we discuss matrix design considerations (section G). We rely mostly on peer-reviewed literature for the assessment of CMIP6 models provided in the linked tables and have devised innovative ways to include relevant metadata and update the spreadsheet with new results [https://wcrp-cordex.github.io/cmip6-for-cordex]. We encourage the wider EURO-CORDEX community to become active in this activity, suggest ways to improve and alert us when new results become available or mistakes are noted in the evaluation tables[1]. We also hope that this EURO-CORDEX GCM selection work can inspire good practices or at least contribute to fruitful discussions in other CORDEX domains and CORDEX-FPS activities.

Avoiding the “curse of opportunity”

Grabbing the first and “best” GCM results available and using them as boundary conditions for an RCM is a suboptimal *ad hoc* approach for downscaling ensembles. First, such a strategy makes it likely that several downscaling communities pick the same GCM data which are conveniently available while neglecting others that are not so readily available. Second, the assessment of “best” is most often not based on a rigorous examination of a model’s qualities but rather an anecdotal or very cursory review. Here we refer to this ad hoc approach as the “curse of opportunity” since it may result in some GCMs being overrepresented while other GCMs with different characteristics are underrepresented. Furthermore, it is also well known that multi-model ensembles generally do not provide reliable estimates of probabilities (Collins, 2007), partly because GCMs differ in the ways they are set up and choices made in how some unresolved small-scale processes are represented through parameterization schemes. We nevertheless want to utilize the information provided by the multi-model ensemble simulations to provide some kind of representation of probable outcomes. In the US, the NARCCAP project used experimental design, also referred to as ‘factorial regression’, to account for the sparse matrix of GCM-RCM combinations (Mearns et al., 2013). The benefit of such a strategy is that it makes it possible to cover the full range of combinations without explicitly simulating every one. It is not trivial, however, to pick a representative ensemble with multiple models that perform differently (Benestad et al., 2017; Dalelane et al., 2018; Evin et al., 2021; Ferro et al., 2012; Knutti et al., 2009). Selecting a subsample of the available simulations may give biased results, but a well-designed ensemble may nevertheless provide a better representation of our understanding of the climate. One important question then is how to make an optimal subsample of a large multi-model ensemble of GCMs. There may be different strategies for selecting ensembles: (1) based on cluster analysis or spread of GCM results for some chosen variable (e.g. temperature or precipitation or changes thereof); (2) based on information about the model design and independence or (3) based on a-posteriori model run performance. The best strategy may depend on the purpose of simulations: do we want to get a more representative outlook for the future or try to understand why models give different results? The strategy based on a representative spread needs to consider whether the spread should reflect regional or global scales. French et al. (2017) considered the limited GCM-RCM matrix from NARCCAP and addressed the difficulty to assess whether differences in subregions are statistically significant or just due to sampling fluctuations. They found that results from pointwise inference often are misleading and that proper inference must account for multiple

---

[1] Community members may use the feedback@eurocordex.net address and/or we can track issues through git: https://github.com/WCRP-CORDEX/cmip6-for-cordex/issues?q=is%3Aissue+

comparisons. The size of ensemble is also important, as a small ensemble is subject to ‘the law of small numbers’ that is prone to greater random sampling fluctuations that may result in misleading statistics (Mezghani et al., 2019).

These issues highlight that the design of ensembles and selection of driving models will always be fraught with trade-offs and subjective decisions. Rather than aim for the perfect or best models and ensemble designs, we aim to help modellers make well-informed choices and also provide comprehensive information that will allow tailoring of ensembles to the purposes at hand (e.g. exploring storylines or worst case scenarios). The contributors to this exercise acknowledge these tensions, the inherently subjective nature of some of these choices and aim to make science-based decisions whenever possible.

Background (EURO-CORDEX v1.0)

The currently (January 2021) available CMIP5-based EURO-CORDEX GCM-RCM-RCP matrix[2] has been built up over a time span of about 10 years and, generally, follows the CORDEX simulation protocol[3]. Acknowledging the general idea to sample both global/regional model uncertainty and emission scenario uncertainty (and not explicitly sampling internal climate variability[4]) certain low-level systematics were applied when filling the matrix, especially related to the idea to sample a wide range of GCM and RCP forcings. There was no rigorous assessment of the driving GCMs with respect to their performance (e.g., storm tracks, jet, SSTs & sea ice) and/or plausibility (e.g., ECS, trend reproduction) as done a posteriori in e.g., McSweeney et al. (2012, 2015). Further, the effort to select driving GCMs from the spread of future change, focused on changes in temperature and precipitation over Europe. This future change criteria is perhaps most valid for temperature but likely not for precipitation, which will be substantially modified by the RCM over the region of interest.

As such, details of the matrix design did not follow a particular protocol and were to a large extent driven by the availability of GCM forcing data and further constraints of the individual modelling groups. The latter includes time constraints in individual research projects (which might have partly prevented a thorough evaluation and quality assessment of GCM forcings), different computer costs of the RCMs used, various levels of man- and computer- power availability over a 10-year period and requirements by the particular (often national-scale) research communities. Both issues are furthermore linked to the fact that EURO-CORDEX, in contrast to previous initiatives such as PRUDENCE and ENSEMBLES, did not receive dedicated funding for a consistent filling of the matrix with simulations, but strongly relied on in-kind and best-effort contributions of the individual groups. A consistent and well thought-out matrix design was, hence, hard to achieve, also because the simulations were produced over a period of about 10 years and not all requirements and contributions were known at the beginning. The Copernicus-funded PRINCIPLES project recently improved on this situation by financing and a more consistent completion of existing gaps in the simulation matrix through dedicated funding for additional simulations and an ANOVA-based approach; this technique emulates full-matrix averages and applies them to fields with limited variability like seasonal averages (Christensen & Kjellström, 2020).

Problem statement / Motivation

---

[2] Tables of available Eur-11 (12km) simulations are available here: https://github.com/euro-cordex/esgf-table

[3] While the discussion here is primarily concerned with the high resolution 12km ensemble, there is also an extensive ensemble of 50km simulations.

[4] This was true at least initially. However, more recently e.g., in PRINCIPLES (https://climate.copernicus.eu/c3s-production-european-climate-projections) internal variability has been included more explicitly. See also von Trentini et al. (2019).

As a consequence, the now available EURO-CORDEX matrix provides a simulation ensemble of impressive size and with a design that has at least partly been coordinated. However, in many respects the matrix still has to be seen as an ensemble of opportunity with little assessment of driving GCMs' fitness. This complicates, for instance, a complete use of the matrix in subsequent impact assessments. In particular, only 9 different GCMs have been downscaled, undersampling the total spread related to GCM choice, especially because these GCMs have not been selected with an "objective" approach as explained above. Additionally, despite, but also because of, the PRINCIPLES effort, the matrix remains largely unbalanced (Evin et al., 2021): some of the GCMs have been widely used (HadGEM2-ES, CNRM-CM5, EC-Earth) whereas others have been only used for a few runs (CanESM2, MIROC5, GDFL-ESM2G). Similarly, some RCMs have downscaled a wide range of GCMs (RCA4, RACMO22E, REMO, CCLM4-8-17) whereas others only count a few runs (ALARO-0, ALADIN53). We therefore need to aim for a more consistent and coordinated design of the next (CMIP6-based) generation of the EURO-CORDEX matrix as well as a more robust assessment of the driving GCMs and their strengths and weaknesses, in order to progress as a community in "taking ensembles seriously and valuing model independence" (Jebeile & Barberousse, 2021).

To address the above-mentioned issues, recent studies (published after the start of EURO-CORDEX) have explicitly proposed criteria to subsample the large CMIP5 GCM ensemble to serve a regional climate downscaling initiative (Brands et al., 2013; McSweeney et al., 2012, 2015; Parding et al., 2020) in addition to national climate service documents such as DRIAS, 2020[5]. Two main criteria have been emphasized by the authors:

- Past-climate performance: Assuming that a RCM driven by a "good" GCM will inherit the quality of the driver. Performance criteria have focused on fields relevant for RCM downscaling such as large-scale drivers (storm track position and intensity, weather regimes) and/or temperature seasonal cycle.
- Future-climate spread: Assuming that a RCM driven by a highly-sensitive GCM will be highly-sensitive itself.

To our knowledge, the optimal way to subsample the shared socio-economic pathways or the members for a given GCM has not been specifically addressed in the literature.

Ambition for CMIP6 (EURO-CORDEX v2.0)

The lessons learnt from downscaling the CMIP5 ensemble should inform the ensemble design for downscaling the CMIP6 simulations in EURO-CORDEX v2.0. In particular, we aim to achieve the following objectives:

Ideally:

1. Define a set of GCMs that should be downscaled for each scenario in any case (according to the criteria given below).
2. Have at least a core set of GCM-RCM combinations available for each scenario, to avoid having differences in climate change signals of different scenarios just because the ensembles of the different scenarios are different.
3. Produce ensembles for all scenarios or define "key scenarios" that have to be downscaled.

Realistically:

---

[5] In French, http://www.drias-climat.fr/accompagnement/sections/296

1. The accompanying tables provide a comprehensive[6] overview of model performance as well as future climate attributes based on available literature.
2. At the EURO-CORDEX General Assembly in January 2021 it was agreed to coordinate across scenarios in order to have intercomparable ensembles.
3. The scenario prioritization topic is ongoing but CORDEX SAT has released its CMIP6 guidance: https://cordex.org/wp-content/uploads/2021/05/CORDEX-CMIP6_exp_design_RCM.pdf and EURO-CORDEX will start from this [SSP1-2.6 & SSP3-7.0] and will likely further prioritize: https://docs.google.com/document/d/1ZZEiMz-yAtSEZFnyKAq26h2mlMlzsB5OoFDz6ZOm1YM/edit#heading=h.5a5giupmrrwo

Following the lessons learnt from the literature and from previous RCM initiatives, we therefore propose the following criteria for GCM selection for the CMIP6 EURO-CORDEX initiative:

- **Data availability** in order to list the GCMs effectively available to drive RCMs.
- **GCM plausibility** in order to only keep in the final ensemble the GCMs considered as plausible globally and over the area of interest, hereafter continental Europe. An exception can be made for GCMs with high climate sensitivity (higher than the likely range) to account for worst case warming storylines.
- **Future climate change spread** in order to explore at best the whole range of plausible futures. Note this criteria deals with scenario choice, GCM choice and GCM member choice.
- **GCM independence** in order to favor diversity of the drivers and avoid, as much as possible, redundancies.

Task Team Declaration of values

Given the ambitions noted above, which state “what” we aim to do, and “how” we intend to achieve said ambitions, it will serve us well to give some thought to “why” the community would want to do this. It is increasingly clear that values are integral to research and do not per-force threaten research integrity and/or objectivity (Pulkkinen et al., 2022). The ethical dimension of this activity is discussed in more detail in Section F. Here we explicitly state our values and acknowledge where we are likely to encounter tensions. The strength of EURO-CORDEX is its commitment to community and there are certain values which bind us and motivate us to continue to engage and collaborate. These include, but are not limited to, the following:

- Objectivity (follow the evidence where it leads, while acknowledging the subjective decisions and assumptions made along the way)
- Progress (a shared belief that we can always improve)
- Democracy (decisions are taken when consensus is reached and from a basis of mutual respect and tolerance)
- Diversity (robust results, both scientifically and culturally, are obtained when many voices are empowered and heard)
- Usability (we have a moral imperative to ensure our efforts are of benefit for our extended scientific community as well as society at large)

Along with this declaration of values we also acknowledge potential conflicts-of-interest. There are co-authors here who, through institutional associations, are very close to the developers for a number of the GCMs under consideration. These include: UKESM, NorESM, CNRM, EC-Earth, MPI and IPSL. While every effort is taken to evaluate the available simulations objectively it would be remiss not to acknowledge the obvious links. Related to this, there are

---

[6] To the best of our knowledge.

also institutional commitments (often linked to national basic funding) to downscale a particular model or group of models. This is also acknowledged and noted in the final recommendations in Section F.

B. Data availability / quality

Our first criteria is data availability, basic quality control and FAIR principles. For this, we are thankful for the efforts of CORDEX-MIP (Gutowski Jr. et al., 2016), which gathered commitments from CMIP6 GCM teams to provide lateral boundary conditions necessary for dynamical and statistical downscaling. This is already an improvement over the ad-hoc approach to obtaining GCM output for downscaling in CMIP5. However, it should be noted that at the time of the initial data request the focus was on SSP1-2.6 and SSP5-8.5. This was before the CMIP6 simulations and their related SSPs had been analyzed in depth. It has since emerged that a more likely "business as usual" scenario is SSP3-7.0 (Hausfather & Peters, 2020). This was impossible to know at the time of the data request and as such this mismatch likely had an effect on data availability. Below are the criteria under this heading.

Data availability criteria

1. CORDEX-MIP (necessary boundary conditions are provided)
2. Basic QA (missing values, suspect values, model levels, etc.)
3. FAIR MetaData, available on ESGF or similar (e.g., Climate Data Store)
4. GCMs provide data for a range of SSPs in particular, SSP1-2.6, SSP2-4.5, SSP3-7.0 and SSP5-8.5 (with initial prioritization to be given to 2.6 and 7.0, to be consistent with guidance from CORDEX-SAT)
5. GCMs provide data necessary for both dynamical and statistical downscaling

Despite the efforts of CORDEX-MIP and the wide range of realizations and model versions (40 unique version/realizations combinations that meet sub-criteria 4 & 5) available, only 16 unique GCMs meet the above criteria. Given the independence criteria described below, the number of truly independent GCMs is likely smaller than this. The data availability for the different scenarios can be assessed by accessing the columns under the "1. Availability" heading in:: https://wcrp-cordex.github.io/cmip6-for-cordex/CMIP6_studies_table_EUR.html

C. Plausibility criteria

The motivation for application of performance-based criteria is to select a group of global models that are capable of reproducing key climate processes over the region of interest for downscaling. For a midlatitude area such as Europe this necessarily includes larger scale features of the climate system such as circulation over the North Atlantic, SSTs in the surrounding seas and sea ice in the Arctic. Because of this strong dependence on upstream conditions we group our plausibility criteria under "Global/hemispheric" and "Regional" headings. The assumption is that realistic models will produce more realistic future projections, because they are able to represent processes correctly. For a detailed summary of the philosophy behind the application of such criteria see McSweeney et al. (2012) and Knutti et al. (2009).

It is important to note that the aim here is not to exclude models (unless a major deficiency is noted) but rather to inform the selection of GCMs for downscaling over Europe in as complete a manner as possible. Therefore, in the associated tables a final determination for each model is not provided. This is a community decision to be based on the desired construction of the ensemble matrix. This raises another issue. There is no escaping the fact that GCM selection can never be a wholly objective procedure. We accept this and rather strive to make well-informed, as objective as possible and transparent decisions.

What follows is a list of the criteria currently in place. There are many others which might be included but the task team has endeavored to limit the criteria to processes or quantities that are of highest relevance for regional downscaling for the EURO-CORDEX domain (whether it be done via dynamical, statistical or hybrid approaches). Availability of published studies assessing a large number of CMIP6 GCMs is also a limit for the plausibility criteria, knowing that new GCM assessment studies will likely emerge in the coming years. The following list will be updated via the linked tables, which may be thought of as the official subset of criteria.

Plausibility criteria

Global

1) Observationally constrained transient climate response (TCR) (Ribes et al., 2021; Tokarska et al., 2020, IPCC-AR6)
2) Global performance scores (Brunner et al., 2020)
3) Observationally constrained global future climate change range for mid-21st century (Qasmi and Ribes, 2022)

Regional (Europe)[7]

1) Large-scale circulation criteria over the North-Atlantic such Jet Stream North-South position (Oudar et al., 2020), storm track position (Pri20), blocking frequency (Dav20), Mean absolute error (MAE) for the frequency of European weather types (Brands, 2022a) and a CMIP6 GCM revisited version of the McSweeney et al. (2015) criteria.
2) Aerosol Optical Depth (AOD) Root Mean Square Error (RMSE) and past trend over Europe (P. Nabat, pers. comm.)
3) Regional Sea Surface Temperature (SST) RMSE for surrounding water areas: Mediterranean (F. Sevault, pers. comm.), Black Sea, Norwegian and Barents Sea, Baltic and North Sea, North Atlantic Ocean (A. Dobler, pers. comm.)
4) Regional Sea Ice Cover (SIC) RMSE for the relevant zones: Baltic Sea, Norwegian and Barents Sea (A. Dobler, pers. comm.)

Link to tables: https://wcrp-cordex.github.io/cmip6-for-cordex/CMIP6_studies_table_EUR.html
The headings under the “2. Plausibility” section of the tables link to a summary of the scores and plausibility limits, along with a link to the original publication. Note that many more studies might be available than those shown in the tables. Different criteria drove the decision to exclude particular studies. First of all, the main aim for the tables is to summarize the results, rather than being exhaustive. An exhaustive enumeration of all studies evaluating or presenting future outcomes of CMIP6 models would lead to unmanageably large tables. We tried to promote a diversity of metrics exploring different aspects. Also, studies providing metrics for a larger number of models and members are preferred, to avoid empty cells in the table. In any case, all studies considered are included in the GitHub repository and the causes for exclusion[8] from the summary table are recorded and the decision may be reversed after discussion or emergence of new evidence.

---

[7] We note that a few key metrics are missing from the literature, these are: seasonal cycle of near-surface temperature, precipitation and humidity.

[8] An example of a study (Fernandez-Granja et al., 2021) excluded from the table and how the causes for exclusion are registered can be found at https://github.com/WCRP-CORDEX/cmip6-for-cordex/blob/main/CMIP6_studies/Fer21.yaml

D. Selecting to cover the range of future outcomes

The third step in the selection of driving global climate model simulations for downscaling is future projection uncertainty coverage. Once the model simulations are available (Section B) and are shown to represent current climate reliably (Section C), we are left with a set of reliable enough model simulations from which we can explore a number of plausible future climates. We recommend sampling these plausible futures as extensively as possible:

- **Scenario uncertainty**

This type of uncertainty would be covered by selecting at least a low/high or low/medium/high concentration pathways. In order to obtain ensembles that are globally consistent, this choice has been agreed for the whole CORDEX-CMIP6 initiative. SSP1-2.6 and SSP3-7.0 have been selected to illustrate the high and low concentration pathways. SSP2-4.5 has less priority in the new CORDEX framework but could be used to simulate a medium pathway in agreement with current national efforts for reducing GHG emissions. Finally, the SSP5-8.5 is still of interest to illustrate extreme scenarios and worse-case trajectories and is especially of interest for those conducting risk assessments.

- **Model uncertainty**

For a given scenario, the model response to a given level of forcing is one key source of uncertainty to be sampled, especially near the end of the 21st century (Evin et al., 2021; Hawkins & Sutton, 2009). Exploring the model spread is therefore crucial to consider that the future climate lies in the simulated ensemble range. Here, the difficulty lies in the fact that the final ensemble spread will be assessed using the RCM runs whereas the GCM selection will be achieved before making the RCMs simulations. We then need to carefully select the GCM variables or characteristics that are most relevant for the coming RCM simulations. Large-scale features such as weather regimes or North-Atlantic storm tracks seem logical to explore whereas precipitation of the GCM is less clear. Note that exploring the model uncertainty in the GCM ensemble makes the assumption that the RCM is not free to invent its own climate change signal within the domain and is somehow constrained by the GCM sensitivity at least for some variables. For example, it assumes that an RCM driven by a GCM with high(low) regional sensitivity will itself show a strong(weak) climate change response. This implies that RCMs are not disturbing the large-scale climate change signal, at least for some variables within the domain of interest. Note, however, that this is likely not true for all GCM-RCM pairs and the topic of GCM-RCM inconsistency is an active area of research (Boé et al., 2020; Taranu et al., 2022).

It is worth mentioning that characterizing the GCM future spread in process-related variables or dynamics-related variables and not just in simple surface variables such as precipitation and temperature remains challenging. In particular, future spread (for e.g. circulation variables like the jet stream shift) should be defined by various indices such as spatial shifts and frequency changes, and not simply by change in the mean variable values.

- **Internal variability**

Natural climate variability is another source of uncertainty in future climate projections. This is especially true for the regional to local scales, the precipitation and the next decades (Hawkins and Sutton, 2009; Evin et al., 2021). It is therefore important to take it into account, if possible, when designing GCM-RCM matrices and in particular when selecting specific member(s) for a given GCM. To our knowledge, dedicated work to the selection of GCM members has never been performed to prepare a RCM multi-model initiative. This is likely partly related to the data availability issue as data required to drive RCMs were usually available only for one member per GCM except for some exceptions (von Trentini et al., 2019) and in general using more models has been considered more important than adding additional realizations of the same model. This natural variability uncertainty is, however, intrinsically included into the RCM ensemble as random members of various GCMs are used. A few attempts to optimize the member selection have been performed[9] but they went largely undocumented. This specific member selection is nevertheless very relevant for the first decades of the scenario period in order to better cover the total climate change uncertainty range for this near-term temporal horizon.

For CMIP6 there are a limited number of GCMs which make multiple realizations available for driving RCMs. We believe it would be worthwhile, all other things being equal (i.e., the GCMs have satisfactory performance) to explore the role of internal variability by including these multiple realizations in the RCM-GCM matrix design of EURO-CORDEX, at least for some GCMs.

Inspired by the literature and by the CORDEX RCM experiment protocol, we have considered the following GCM characteristics to explore the range of plausible futures. These also appear in the linked tables.

Future change criteria

1) Jet stream position change (Oudar et al., 2020)
2) European near-surface temperature future change: 2070-2100 vs 1980-2010 for SSP5-8.5 (IPCC Atlas github repo.) and observationally-constrained warming classes for JJA, 2041-2060 vs 1850-1900, SSP245 (Qasmi and Ribes, 2022)
3) Aerosol Optical Depth future evolution over Europe, SSP585, end of the 21st century (P. Nabat, pers. comm.)
4) Mediterranean SST future evolution, SSP585, end of the 21st century (F. Sevault, pers. comm.)
5) TCR values (IPCC-AR6) and ECS values (Schlund et al., 2020)

E. Model independence and structural uncertainty

Here we introduce a final set of criteria, rarely discussed in the RCM-related literature to our knowledge but commonly used in practice, for example in impact studies: GCM independence (Boé, 2018; Brands, 2022a). Because GCMs are far from independent, the statistical properties (multi-model mean, standard deviation etc.) of the full ensemble may be quite biased if many interdependent models are considered in the driving GCM selection. We try here to assess the level of model dependency between the CMIP6 GCMs in order to eliminate obvious and less obvious near duplicates and to avoid introducing hidden biases in the ensemble and unnecessary duplication. Note that similar GCMs in CMIP6 are “similar” for a variety of reasons and often arise through shared components based on collaboration between different centres. For example, the MetOffice HadGEM3 shares its atmospheric model with Australia’s ACCESS-CM2, while Korea’s KACE-1-0-G takes both its atmospheric and land surface models from HadGEM3. Other examples are the shared components in NorESM (atmosphere and land surface models but not ocean) and TaiESM1 (all components, atmosphere and land surface are

---

[9] For instance, in EU-funded ENSEMBLES, ECHAM5-r3 was selected due to its better agreement with the observed trends. Also, the CNRM-CM5 member used in CORDEX-CMIP5 was initially the member r8i1p1 of the original ensemble and was renamed r1i1p1 by the GCM modelling group before diffusion on the ESGF and, in particular, for the provision of RCM LBCs. The specific choice of the member r8i1p1 was based on its better ability to reproduce 20th century past trend in global mean surface temperature (CNRM, pers. comm.).

modified) obtained from CESM. For some applications the fine details will be important while for others simply knowing models share interdependencies is enough. We see two ways to assess and treat model independence:

- Independence criteria based on a-priori model structure
- Independence criteria based on a-posteriori model output pattern

Classifying models in families depending on their building phase, for example the relationship between institutes or the number of common lines of code or the list of common sub-models is a promising way to deal with independence. This a-priori model uncertainty or model independence criteria could rely on the model version (same model with different level complexity, concerning the spatial resolution or the number of climate components represented) or on the model lineage (different models with shared components that is to say shared lines of code). However this a-priori approach has been rarely attempted (Boé, 2018; Brands, 2022b; Leduc et al., 2016), probably because of the difficulties to obtain published, easy-to-handle and robust information about the models. The rising use of ES-DOC (https://search.es-doc.org/) may facilitate such an approach in the coming years, the GCM metadata archive built by Brands et al. (2022) providing useful information in the meantime (see https://github.com/SwenBrands/gcm-metadata-for-cmip for a work-in-progress version)

Fortunately, model independence can also be assessed a-posteriori by comparing the model outputs. In particular, the spatial pattern of the error maps and the future climate change response maps appear to align well with model dependency. This feature has been used first in Knutti (2010) and Knutti & Sedláček (2012) and more recently for the CMIP6 ensemble by Brunner et al. 2020 and Brands 2022b. Brunner et al. (2020) defined a model tree based on a similarity criteria of the global spatial pattern of the surface temperature and mean sea level pressure for the 1980-2014 period of the historical runs. They obtained all obvious families already informed by the model naming but also less obvious but still known families such as the similarity between CESM models and NorESM or the MetOffice GCM-based family. Brands (2022b) found those GCMs using the same AGCM family to produce similar circulation error patterns in the northern hemisphere extra-tropics. The average error pattern correlation coefficient involving 61 GCMs from CMIP5 and 6 was found to be surprisingly strong (r = +0.6), the MIROC-AGCM family being most independent. Unexpected error pattern similarities between distinct AGCM families were also reported (e.g. between ECHAM and GFDL-AM).

Additionally we include an assessment of model complexity and also the spatial resolution of the available models as the effective resolution of these models is typically much larger than their grid-spacing. The criteria under this heading appear below and are also represented in the tables.

Model Independence / Structural uncertainty

1) Model complexity, by specifying the prescribed and interactive components considered (Brands 2022a)
2) Model independence (Brunner et al., 2020; Brands 2022b)[10]
3) Spatial resolution of GCM e.g., effective resolution (Klaver et al., 2020)

---

[10] A priori model complexity and independency in terms of submodels can be derived from: https://github.com/SwenBrands/gcm-metadata-for-cmip/blob/main/get_historical_metadata.py (to be moved to cmip6-for-codex), see also Brands, Swen et al., 2022

Taking into account model independence at the ensemble design step will aid in use and interpretation of the final ensemble. Users will not have to risk using an ensemble artificially weighted by GCM interdependencies nor will they need to compute "corrected" ensemble mean and spread, or thin out a-posteriori the available downscaled ensemble according to independence assessments. If such assessments could enter already at the stage of GCM selection the final reduction of the overall available ensemble would be less pronounced.

While model independence was applied in the final evaluation and proposed selection (see Tables 1 & 2 below), we decided not to use the level of model complexity and the model resolution criteria as critical a-priori selection procedure, except as a tie-breaker, as they may duplicate other performance criteria.

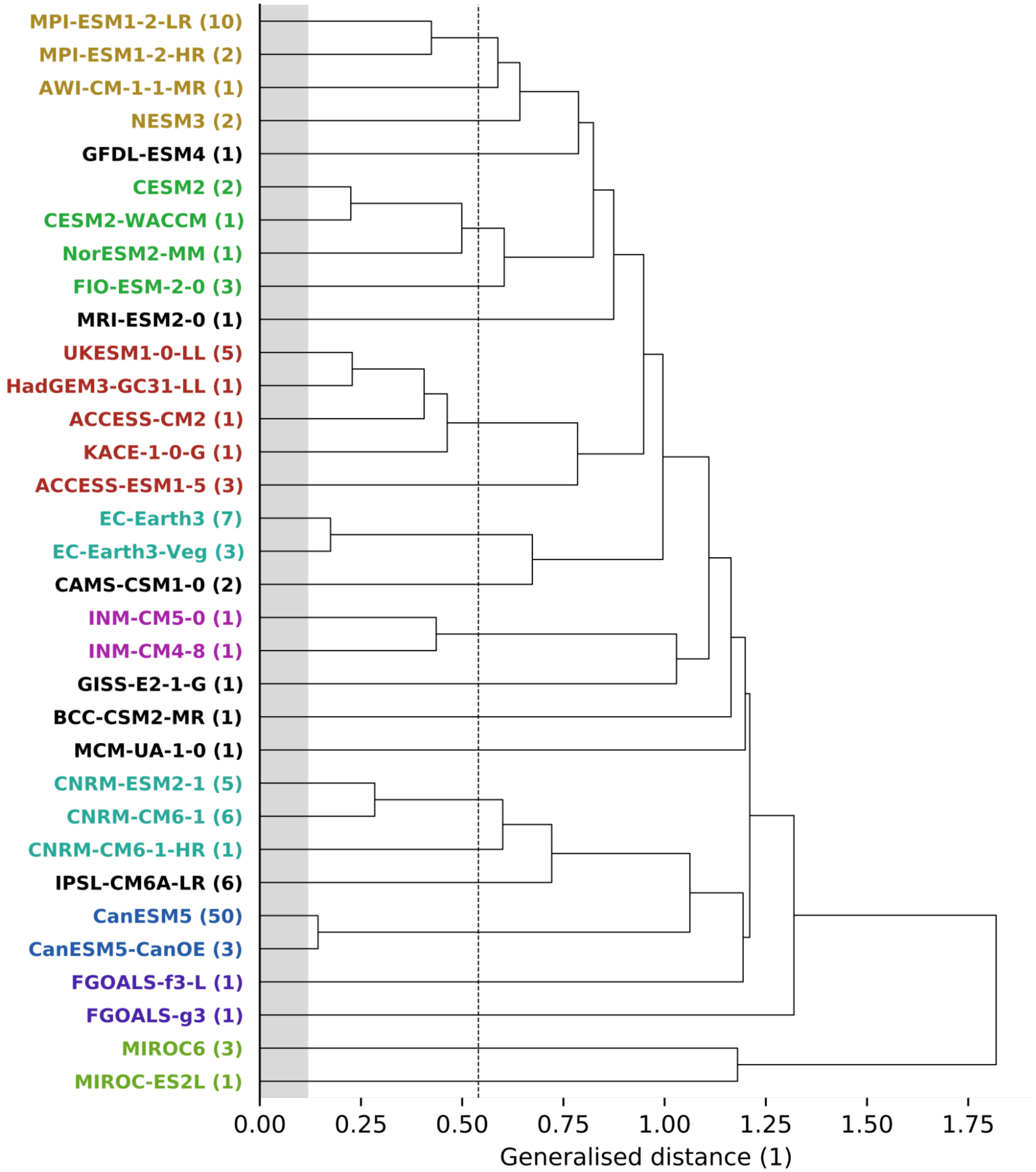

Figure 5 from Brunner et al. (2020) showing model families with shared dependencies in color and models with no shared dependencies in black.

F. Possible approaches for merging criteria for reaching the final GCM list

Model selection is an inherently subjective process. As such, it is important that every decision, even if subjective, is documented for later assessment in the light of the results and new information. Institutional (e.g. in-house GCM downscaling) and/or national (e.g. selecting models well-behaved for a given country/region) constraints have been traditionally part of the GCM selection process. This selection is usually associated with funds and interests, which push forward the downscaling of specific GCMs, regardless of any design we decide. While this kind of selection is purely subjective from the point of view of EURO-CORDEX as a whole, the final proposed GCM selection process is likely to be a choice among equally valid GCM subsets. In this sense, it is feasible to benefit from these committed simulations, and dress the ensemble of selected GCMs around them. For this purpose, an early collection of committed simulations is crucial, especially for the matrix design (Section G). This collection of downscaling plans from the different groups is currently being conducted globally, at the CORDEX level[11], and is taken into account in our summary tables, which provide a filter showing the availability, performance, spread and independence of these committed simulations.

Despite the aim to include all CMIP6 models in our assessment, and the acknowledgement that even some “implausible” models may be useful to downscale for particular purposes, ultimately decisions on which models to prioritize for downscaling must be made. The accompanying tables are meant to facilitate this decision making. However, a number of questions persist that can only be answered by the community of practice itself and not just a subset (i.e. this task team). Below are some considerations:

- What are the important aspects to look at?
- What are the thresholds for excluding models and how are they decided? (See subjectivity discussion above). ‘Hard’ and ‘soft’ thresholds can be more inclusive and facilitate a nuanced construction of the ensemble more than an ‘in’ or ‘out’ criteria (see e.g., McSweeney et al 2012 & 2015 and their use of a ‘traffic light system).
- Do the different evaluation criteria carry equal weight? There may be clusters of non-independent criteria (e.g., jet stream latitude is related to blocking frequency).
- How do we combine the plausibility criteria with the future spread criteria? E.g. McSweeney et al. 2012 use a clear 2-stage process whereby 1) models are evaluated and eliminated then 2) a diverse range of models is selected from those left. Conversely, it could be a more circular or subjective process in which we might allow some poorer models ‘in’ because we don’t feel sufficiently confident that the projections are so implausible as to exclude that part of future climate space (See McSweeney et al., 2015 ‘decision making framework’).

To facilitate the exploration of the decisions made, all background information and even the code used to create the tables are publicly available in a GitHub repository[12]. This includes human- and machine-readable files for every metric included in the tables and even those considered but not ultimately included (usually due to preferred, or more complete, sources for similar information). The source for the plausibility limits is also provided, since these are usually

---

[11] The modelling groups can report their plans here: CORDEX CMIP6 downscaling plans and are summarized by CORDEX domain at https://wcrp-cordex.github.io/simulation-status

[12] https://github.com/WCRP-CORDEX/cmip6-for-cordex

not stated in the published work and the authors were contacted to consider their input. This approach provides full transparency to the process of constructing the summary tables. Moreover, it provides a way to update them in collaboration with the community, which can use GitHub issues to discuss their concerns with particular aspects, provide alternative sources of information or even fork the whole repository to adapt it to their needs.

There is also an ethical dimension to the selection of GCMs. By providing a clear account of the process of selection we facilitate the further sub-selection by some users for more specific applications. This transparency includes the public dissemination of the table of evaluation metrics, the Github repository, as well as this document where we provide the necessary background and justifications for decisions rendered. Collectively these amount to a declaration of values that guided the selection process, which we acknowledge in the introduction (Pulkkinen et al., 2022). A further question is: Do they meet the values of end users? While this consideration has not been taken on explicitly, the EURO-CORDEX community has endeavored to obtain such information through informal discussions with national level users such as climate service centers and governmental agencies. In particular, we can consider that exploring a large range of future outcomes and exploring worst-case storylines even if less plausible do fit the values of some of the users of the coming simulations. We acknowledge that this is an area where we can improve, perhaps through a co-production procedure based on the use of surveys and questionnaires sent to the appropriate EURO-CORDEX simulation users.

CMIP6 Model recommendations

Below we present tables which may be considered as an initial recommendation of which GCMs exhibit an appropriate level of performance for downscaling by the EURO-CORDEX community. However, this is a living process that can, and should, evolve in time (e.g. new evaluations are performed, additional simulations are added to ESGF). The most strict of these tables (Table 1) is an illustration of the fact that being too restrictive on the criteria likely leads to too few models to ensure a robust downscaled ensembles. It should be noted that we applied a threshold so that models which score highly simply because they have not been evaluated are not included (a minimum of one plausibility score for each family of scores is requested). We also note that being evaluated for multiple criteria should be viewed favorably (e.g. Global warming level, large-scale circulation, surface forcings such SST, SIC or AOD). There are many other possible configurations that can be constructed using the filtering features in the github tables but here we present two which we believe represent an appropriate starting point for CMIP6 downscaling in EURO-CORDEX.[13]

Table 1. *Most strict; GCMs which are available for all four scenarios (ssp126, ssp245, ssp370,ssp585) and are deemed "plausible" for each evaluated criteria. To qualify models must be evaluated for at least one criterion per score family. The third column shows the number of failed criteria over the total number of criteria for each model. Models that are also part of institutional commitments are highlighted. The fourth column shows an illustration of future spread categories for the selected GCMs (here based on TCR values).*

---

[13] https://wcrp-cordex.github.io/cmip6-for-cordex/CMIP6_studies_table_EUR.html.

GCM name	Run	Marks/Criteria	TCR Plausible range (1.2K-2.4K)[14]
MPI-ESM1-2-LR	r1i1p1f1	0/18	1.84

Table 2. *Less strict; same as Table 1 except for GCMs which are "available" for all four scenarios. Scores are based on all evaluated members of a model even if only one member is "available". Only one model per family is kept in most cases and in the event of a tie criteria such as complexity and resolution may play a role as tie-breakers. Explanations appear in footnotes.*

GCM name	Run	Marks/Criteria	TCR Plausible range (1.2K-2.4K)
NorESM2-MM[15]	r1i1p1f1	1/17	1.33
MIROC6[16]	r1i1p1f1	1/20	1.55
MPI-ESM1-2-HR	r1i1p1f1	1/20	1.66
CNRM-ESM2-1	r1i1p1f2	1/19	1.86
CESM2[17]	r11i1p1f1	1/18	2.06
CMCC-CM2-SR5[18]	r1i1p1f1	1/15	2.09
IPSL-CM6A-LR[19]	r1i1p1f1	2/16	2.32
EC-Earth3-Veg[20]	r1i1p1f1	2/15	2.62
UKESM1-0-LL[21]	r1i1p1f2	2/19	2.79

Two additional models that are part of institutional commitments are *not* currently "available". These are: EC-EARTH3-Veg(0) r6i1p1f1 and EC-EARTH3(1) r1i1p1f1. It is likely these will appear on the ESGF well before EURO-CORDEX simulations are completed. As noted above this list will evolve as additional analyses are conducted. Models that currently score well in terms of performance may be downgraded as more analyses are performed. This is especially a concern for those which have only been evaluated for a few criteria. As such we focus on those evaluated for many criteria. Likewise models/simulations that are currently unavailable on ESGF may become available. Further, marks against plausibility should not be automatically

---

[14] TCR 90% range as provided by the IPCC AR6 https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_Chapter_07_Supplementary_Material.pdf

[15] Same family as CESM but helps with covering future spread as it has low TCR. NorESM2-MM performs much better than NorESM2-LM in terms of regional atmospheric circulation (Brands 2022a)

[16] MIROC-ES2L might also be considered, but has a 500km nominal resolution, scores ⅗ and is the same family as MIROC6.

[17] TaiESM1 would also qualify but shares all components with CESM2 (https://gmd.copernicus.org/articles/13/3887/2020/).

[18] CMCC-ESM2 might also be considered but it is evaluated for fewer criteria, is the same model family and is not evaluated for TCR. CMCC-CM2-SR5 shares its AGCM (CAM5.3), LSM (CLM4.5) and sea-ice model (CICE4.0) with CESM1.

[19] ISPL-CM6A-LR shares its OGCM (NEMO3.6) and ocean biogeochemistry model (PISCESv2) with CNRM-ESM2-1. However, it is likely that there will be an institutional commitment to downscale the IPSL model (t.b.c).

[20] EC-EARTH3 would also qualify but shares same family.

[21] ACCESS-CM2 also qualifies but is from the same family and does not include the Black Sea which could present issues for the EURO-CORDEX domain.

disqualifying. Rather, they should be considered as flags that warn of known shortcomings that downscaling teams should consider carefully. As an example we point to the ACCESS models, which perform well but one of their marks against, arises due to the absence of the Black Sea. This omission may have serious implications for the southeast Mediterranean and eastern Europe. While it might be tempting to turn the Marks/Criteria ratios into a skill-score many of the criteria can not be considered independent and such a skill-score would need to be carefully constructed to avoid spurious results. Nevertheless, we can see that the models in Table 2. generally perform quite well. They also cover well the range of plausible TCR[22] with a few warm outliers, which may be of interest for climate service providers aiming to provide risk assessments based on worst-case scenarios (EC-EARTH and UKESM).

G. Matrix design

A final issue to consider for the robust construction of downscaled ensembles of CMIP6, is how best to fill the 3-D matrix of simulations (either RCM-GCM-SSP or ESD-GCM-SSP pairings). As noted in the introduction the matrix of simulations from CMIP5 remains sparse, heterogenous and unbalanced despite some impressive efforts to ameliorate these shortcomings. It is nevertheless still very large and presents challenges for impact modellers and climate service providers who must then perform a sub-selection of the ensemble, often without sufficient guidance. This challenge will persist in the downscaling of CMIP6 due to differences in institutional strategies, resources, timing etc. In some ways the EURO-CORDEX ensembles are, like CMIP ensembles, often unbalanced as they grow over time and are “dynamic” in practice, if not by design. However, we believe we can take some steps at the outset that can help reduce the impact of typical matrix design hazards and agree to an “official”, balanced matrix for use in adaptation planning and impacts assessments, that can be supplemented by extra runs.

We suggest four criteria to help guide filling the matrix, acknowledging that these are preliminary and may change after further debate:

1) Balance the ensemble as best as possible in order that no SSP, GCM or RCM are over-represented.
2) Physical consistency between GCM and RCM, in order that the RCM does not deviate too much from the GCMs at large-scales.
3) Keeping an eye on the range of uncertainty. For example, we should not artificially decrease the range of outcomes provided by the GCMs and should allow the possibility to explore interesting but rare outcomes such as plausible worst-case scenarios.
4) Facilitate a-posteriori filling of the matrix via statistical approaches (see next subsection).

Table 3 shows the selection of ongoing or planned EURO-CORDEX simulations to date that contribute to a balanced matrix that fulfills the criteria mentioned above. It can be seen that major efforts have been made to avoid under/overrepresented GCMs and represent, as much as possible, a fractional factorial design where at least 3 runs by RCM and 4 runs by GCM are considered. However, there are also institutional or project constraints that must be respected, which leads to some GCMs that are more represented than other ones. For example, the same set of simulations are planned for COSMO-CLM and ICON-CLM. REMO also follows the same plan (with the same 3 GCMs) due to the NUKLEUS project[23].

---

[22] TCR is just one possibility, other spread criteria could also be used here.

[23] https://wcrp-cordex.github.io/simulation-status/CORDEX_CMIP6_status_by_experiment.html#EUR-11-NUKLEUS

Table 3. *The EURO-CORDEX balanced matrix experiment comprises simulations filling a GCM-RCM-SSP combination matrix with a fractional factorial design where at least 3 runs by RCM and 4 runs by GCM are considered (two additional simulations with CMCC-CM2-SR5 and MIROC6 need to be planned to fit this criteria). Blue crosses indicate the simulations which are planned, green crosses indicate simulations which are currently (Feb. 2023) running or are completed*[24]*. All simulations are planned to run with the scenarios SSP1-2.6 and SSP3-7.0.*

	Institution / RCM						
Driving GCM / run	**CNRM / ALADIN6x**	**CLMcom / COSMO - CLM**	**HCLIMcom / HCLIM43 - ALADIN**	**KNMI / RACMO23E**	**GERICS / REMO**	**CUNI,ICTP / RegCM**	**AUTH, CESAM,.. / WRF**
NorESM2-MM / r1i1p1f1	✘			✘		✘	✘
MIROC6 / r1i1p1f1		✘	✘		✘		
MPI-ESM1-2-HR / r1i1p1f1		✘			✘	✘	✘
CNRM-ESM2-1 / r1i1p1f2	✘		✘	✘		✘	
CMCC-CM2-SR5 / r1i1p1f1	✘	✘					✘
EC-Earth3-Veg / r1i1p1f1			✘	✘	✘	✘	

H. Statistical exploitation of the EURO-CORDEX ensemble

We have noted that we try to avoid an ensemble of opportunity in the EURO-CORDEX downscaled ensemble. However, the input GCM ensemble (CMIP6) is itself an ensemble of opportunity in the sense that there are some model lineages/versions which are over-represented in the ensemble. Even if ESD approaches could, in principle, downscale the full CMIP ensemble, a weighting scheme or thinning of the full ensemble would be required to have a balanced ensemble of downscaled results. In an unbalanced multi-model ensemble, not weighting is weighting (Fernández et al., 2019). Statistical approaches could be used to assess the effects of different ensemble thinning strategies a posteriori (Christensen & Kjellström, 2021; Fernández & Frías, 2020) to statistically obtain balanced climate change signal and the associated uncertainties with an a-posteriori statistical filling of the matrix (Christensen & Kjellström, 2020; Déqué et al., 2012; Evin et al., 2021). A complementary approach involves so-called "hybrid" downscaling techniques (Doblas-Reyes et al., 2021). For aggregated variables, such as the mean temperature or rainfall totals, it may be possible to use hybrid downscaling to emulate GCM-RCM pairs and hence extend the results of a selected RCM to a large ensemble of GCMs (Erlandsen et al., 2020). This involves using existing GCM-RCM pairs for calibrating statistical models which are subsequently applied to different GCMs. Statistical emulation is now even possible at the daily scale, allowing one to potentially mimic time series of 2D maps of the missing GCM-RCM pairs (Doury et al., 2022). Furthermore, such hybrid modelling may serve as a useful method for analyzing the performance of the RCMs and how

---

[24] Updated status information can be found in https://wcrp-cordex.github.io/simulation-status/CORDEX_CMIP6_status_by_experiment.html#EUR-11-EURbalanced

they represent the dependency of small-scale processes to large-scale conditions. In many cases, the emulation can be evaluated against independent GCM-RCM combinations to assess the presence of non-linear dependencies. This use of emulation is also closely connected with so-called ‘pseudo-reality’, where empirical-statistical downscaling (ESD) is applied on simulated results.

I. Conclusions

This white paper, its accompanying meta-analysis and recommendations (see Table 2) reflects an effort by the EURO-CORDEX community to provide a more robust and transparent basis for the selection of driving GCMs and approaches to filling the matrix of simulations. There is no attempt here to point to a “best” performing set of models. Rather, we seek to provide guidance via a comprehensive assessment, backed by the available literature, of the CMIP6 GCMs under consideration for downscaling. As discussed above, we are aware that any selection of GCMs is subjective, but we want to make sure that the criteria that are applied are supported by as many experts from the regional and global climate modelling communities as possible i.e., to make the final decision as “objective” as possible. Furthermore, the criteria and the decision process should be documented to guarantee full transparency, traceability and verifiability. We believe that a science and information driven selection/guidance approach is in any case an improvement compared to the uncoordinated and, to a certain degree, coincidental selection procedure that has been applied in the past.

This document and the accompanying tools will be updated regularly as new results are published by the scientific community. Finally, the task team welcomes suggestions for improvements and additions to the tools and documentation. The accompanying tools have been designed in a way that allows their straightforward application to other CORDEX domains and other RCM initiatives.

References

Benestad, R., Sillmann, J., Thorarinsdottir, T. L., Guttorp, P., Mesquita, M. d S., Tye, M. R., Uotila, P., Maule, C. F., Thejll, P., Drews, M., & Parding, K. M. (2017). New vigour involving statisticians to overcome ensemble fatigue. *Nature Climate Change*, *7*(10), 697–703. https://doi.org/10.1038/nclimate3393

Boé, J. (2018). Interdependency in Multimodel Climate Projections: Component Replication and Result Similarity. *Geophysical Research Letters*, *45*(6), 2771–2779. https://doi.org/10.1002/2017GL076829

Boé, J., Somot, S., Corre, L., & Nabat, P. (2020). Large discrepancies in summer climate change over Europe as projected by global and regional climate models: Causes and consequences. *Climate Dynamics*. https://doi.org/10.1007/s00382-020-05153-1

Brands, S. (2021). A circulation-based performance atlas of the CMIP5 and 6 models for regional climate studies in the northern hemisphere. *Geoscientific Model Development Discussions*, 1–48. https://doi.org/10.5194/gmd-2020-418

Brands, S. (2022a). 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. https://doi.org/10.5194/gmd-15-1375-2022

Brands, S. (2022b). Common Error Patterns in the Regional Atmospheric Circulation Simulated by the CMIP Multi-Model Ensemble. *Geophysical Research Letters*, *49*(23). https://doi.org/10.1029/2022GL101446

Brands, S., Herrera, S., Fernández, J., & Gutiérrez, J. M. (2013). How well do CMIP5 Earth System Models simulate present climate conditions in Europe and Africa?: A performance comparison for the downscaling community. *Climate Dynamics*, *41*(3–4), 803–817. https://doi.org/10.1007/s00382-013-1742-8

Brands, Swen, Tatebe, Hiroaki, Danek, Christopher, Fernández, Jesús, Swart, Neil C., Volodin, Evgeny, Kim, YoungHo, Collier, Mark, Bi, Dave, & Tongwen, Wu. (2022). *Python code to calculate Lamb circulation types for the northern hemisphere derived from historical CMIP simulations and reanalysis data* (Version 4). Zenodo. https://doi.org/10.5281/ZENODO.6390256

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. https://doi.org/10.5194/esd-11-995-2020

Christensen, O. B., & Kjellström, E. (2020). Partitioning uncertainty components of mean climate and climate change in a large ensemble of European regional climate model projections. *Climate Dynamics*, *54*(9), 4293–4308. https://doi.org/10.1007/s00382-020-05229-y

Christensen, O. B., & Kjellström, E. (2021). Filling the matrix: An ANOVA-based method to emulate regional climate model simulations for equally-weighted properties of ensembles of opportunity. *Climate Dynamics*. https://doi.org/10.1007/s00382-021-06010-5

Collins, M. (2007). Ensembles and probabilities: A new era in the prediction of climate change. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *365*(1857), 1957–1970. https://doi.org/10.1098/rsta.2007.2068

Dalelane, C., Früh, B., Steger, C., & Walter, A. (2018). A Pragmatic Approach to Build a Reduced Regional Climate Projection Ensemble for Germany Using the EURO-CORDEX 8.5 Ensemble. *Journal of Applied Meteorology and Climatology*, *57*(3), 477–491. https://doi.org/10.1175/JAMC-D-17-0141.1

Déqué, M., Somot, S., Sanchez-Gomez, E., Goodess, C. M., Jacob, D., Lenderink, G., & Christensen, O. B. (2012). The spread amongst ENSEMBLES regional scenarios: Regional climate models, driving general circulation models and interannual variability. *Climate Dynamics*, *38*(5), 951–964. https://doi.org/10.1007/s00382-011-1053-x

Doblas-Reyes, F. J., Sörensson, A. A., Almazroui, M., Dosio, A., Gutowski, W. J., Haarsma, R., Hamdi, R., Hewitson, B., Kwon, W. T., Lamptey, B., Maraun, D., Stephenson, T. S., Takayabu, I., Terray, L., Turner, A., & Zuo, Z. (2021). *Linking Global to Regional Climate Change. In Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Doury, A., Somot, S., Gadat, S., Ribes, A., & Corre, L. (2022). Regional climate model emulator based on deep learning: Concept and first evaluation of a novel hybrid downscaling approach. *Climate Dynamics*. https://doi.org/10.1007/s00382-022-06343-9

Erlandsen, H. B., Parding, K. M., Benestad, R., Mezghani, A., & Pontoppidan, M. (2020). A Hybrid Downscaling Approach for Future Temperature and Precipitation Change. *Journal of Applied Meteorology and Climatology*, *59*(11), 1793–1807. https://doi.org/10.1175/JAMC-D-20-0013.1

Evin, G., Somot, S., & Hingray, B. (2021). Balanced estimate and uncertainty assessment of European climate change using the large EURO-CORDEX regional climate model ensemble. *Earth System Dynamics*, *12*(4), 1543–1569. https://doi.org/10.5194/esd-12-1543-2021

Fernández, J., & Frías, M. D. (2020). *Balanced subsampling of future regional climate ensembles of opportunity* [Other]. oral. https://doi.org/10.5194/egusphere-egu2020-20052

Fernández, J., Frías, M. D., Cabos, W. D., Cofiño, A. S., Domínguez, M., Fita, L., Gaertner, M. A., García-Díez, M., Gutiérrez, J. M., Jiménez-Guerrero, P., Liguori, G., Montávez, J. P., Romera, R., & Sánchez, E. (2019). Consistency of climate change projections from multiple global and regional model intercomparison projects. *Climate Dynamics*, *52*(1), 1139–1156. https://doi.org/10.1007/s00382-018-4181-8

Ferro, C. A. T., Jupp, T. E., Lambert, F. H., Huntingford, C., & Cox, P. M. (2012). Model complexity versus ensemble size: Allocating resources for climate prediction. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *370*(1962), 1087–1099. https://doi.org/10.1098/rsta.2011.0307

French, J. P., McGinnis, S., & Schwartzman, A. (2017). Assessing NARCCAP climate model effects using spatial confidence regions. *Advances in Statistical Climatology, Meteorology and Oceanography*, *3*(2), 67–92. https://doi.org/10.5194/ascmo-3-67-2017

Gutowski Jr., W. J., Giorgi, F., Timbal, B., Frigon, A., Jacob, D., Kang, H.-S., Raghavan, K., Lee, B., Lennard, C., Nikulin, G., O'Rourke, E., Rixen, M., Solman, S., Stephenson, T., & Tangang, F. (2016). WCRP COordinated Regional Downscaling EXperiment (CORDEX): A diagnostic MIP for CMIP6. *Geosci. Model Dev.*, *9*(11), 4087–4095. https://doi.org/10.5194/gmd-9-4087-2016

Hausfather, Z., & Peters, G. P. (2020). Emissions – the 'business as usual' story is misleading. *Nature*, *577*(7792), 618–620. https://doi.org/10.1038/d41586-020-00177-3

Hawkins, E., & Sutton, R. (2009). The Potential to Narrow Uncertainty in Regional Climate Predictions. *Bulletin of the American Meteorological Society*, *90*(8), 1095–1107. https://doi.org/10.1175/2009BAMS2607.1

Jebeile, J., & Barberousse, A. (2021). Model spread and progress in climate modelling. *European Journal for Philosophy of Science*, *11*(3), 66. https://doi.org/10.1007/s13194-021-00387-0

Klaver, R., Haarsma, R., Vidale, P. L., & Hazeleger, W. (2020). Effective resolution in high resolution global atmospheric models for climate studies. *Atmospheric Science Letters*, *21*(4), e952. https://doi.org/10.1002/asl.952

Knutti, R. (2010). The end of model democracy? *Climatic Change*, *102*(3), 395–404. https://doi.org/10.1007/s10584-010-9800-2

Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., & Meehl, G. A. (2009). Challenges in Combining Projections from Multiple Climate Models. *Journal of Climate*, *23*(10), 2739–2758. https://doi.org/10.1175/2009JCLI3361.1

Knutti, R., & Sedláček, J. (2012). Robustness and uncertainties in the new CMIP5 climate model projections. *Nature Climate Change*, *3*(4), nclimate1716. https://doi.org/10.1038/nclimate1716

Leduc, M., Laprise, R., Elía, R. de, & Šeparović, L. (2016). Is Institutional Democracy a Good Proxy for Model Independence? *Journal of Climate*, *29*(23), 8301–8316. https://doi.org/10.1175/JCLI-D-15-0761.1

McSweeney, C. F., Jones, R. G., & Booth, B. B. B. (2012). Selecting Ensemble Members to Provide Regional Climate Change Information. *Journal of Climate*, *25*(20), 7100–7121. https://doi.org/10.1175/JCLI-D-11-00526.1

McSweeney, C. F., Jones, R. G., Lee, R. W., & Rowell, D. P. (2015). Selecting CMIP5 GCMs for downscaling over multiple regions. *Climate Dynamics*, *44*(11), 3237–3260. https://doi.org/10.1007/s00382-014-2418-8

Mearns, L. O., Sain, S., Leung, L. R., Bukovsky, M. S., McGinnis, S., Biner, S., Caya, D., Arritt, R. W., Gutowski, W., Takle, E., Snyder, M., Jones, R. G., Nunes, A. M. B., Tucker, S., Herzmann, D., McDaniel, L., & Sloan, L. (2013). Climate change projections of the North American Regional Climate Change Assessment Program (NARCCAP). *Climatic Change*, *120*(4), 965–975. https://doi.org/10.1007/s10584-013-0831-3

Mezghani, A., Dobler, A., Benestad, R., Haugen, J. E., Parding, K. M., Piniewski, M., & Kundzewicz, Z. W. (2019). Subsampling Impact on the Climate Change Signal over Poland Based on Simulations from Statistical and Dynamical Downscaling. *Journal of Applied Meteorology and Climatology*, *58*(5), 1061–1078.

https://doi.org/10.1175/JAMC-D-18-0179.1

Oudar, T., Cattiaux, J., & Douville, H. (2020). Drivers of the Northern Extratropical Eddy-Driven Jet Change in CMIP5 and CMIP6 Models. *Geophysical Research Letters*, *47*(8), e2019GL086695. https://doi.org/10.1029/2019GL086695

Parding, K. M., Dobler, A., McSweeney, C. F., Landgren, O. A., Benestad, R., Erlandsen, H. B., Mezghani, A., Gregow, H., Räty, O., Viktor, E., El Zohbi, J., Christensen, O. B., & Loukos, H. (2020). GCMeval – An interactive tool for evaluation and selection of climate model ensembles. *Climate Services*, *18*, 100167. https://doi.org/10.1016/j.cliser.2020.100167

Pulkkinen, K., Undorf, S., Bender, F., Wikman-Svahn, P., Doblas-Reyes, F., Flynn, C., Hegerl, G. C., Jönsson, A., Leung, G.-K., Roussos, J., Shepherd, T. G., & Thompson, E. (2022). The value of values in climate science. *Nature Climate Change*, 1–3. https://doi.org/10.1038/s41558-021-01238-9

Ribes, A., Qasmi, S., & Gillett, N. P. (2021). Making climate projections conditional on historical observations. *Science Advances*, *7*(4), eabc0671. https://doi.org/10.1126/sciadv.abc0671

Schlund, M., Lauer, A., Gentine, P., Sherwood, S. C., & Eyring, V. (2020). Emergent constraints on equilibrium climate sensitivity in CMIP5: Do they hold for CMIP6? *Earth System Dynamics*, *11*(4), 1233–1258. https://doi.org/10.5194/esd-11-1233-2020

Taranu, I. S., Somot, S., Alias, A., Boé, J., & Delire, C. (2022). Mechanisms behind large-scale inconsistencies between regional and global climate model-based projections over Europe. *Climate Dynamics*. https://doi.org/10.1007/s00382-022-06540-6

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. https://doi.org/10.1126/sciadv.aaz9549

von Trentini, F., Leduc, M., & Ludwig, R. (2019). Assessing natural variability in RCM signals: Comparison of a multi model EURO-CORDEX ensemble with a 50-member single model large ensemble. *Climate Dynamics*, *53*(3), 1963–1979. https://doi.org/10.1007/s00382-019-04755-8