Filling the Matrix: An ANOVA-Based Method to Emulate Regional Climate Model Simulations for Equally-Weighted Properties of Ensembles of Opportunity

Ole Bøssing Christensen (✉ obc@dmi.dk)
Danish Meteorological Institute  https://orcid.org/0000-0002-9513-2588

Erik Kjellström
Swedish Meteorological and Hydrological Institute: Sveriges meteorologiska och hydrologiska institut

Research Article

Keywords: Ensemble, Regional climate model, Model matrix, ANOVA, EURO-CORDEX, Model variability

Posted Date: March 31st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-366374/v1

License: ©  This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Filling the Matrix: An ANOVA-Based Method to Emulate Regional Climate Model Simulations for Equally-Weighted Properties of Ensembles of Opportunity

Ole B. Christensen¹, Erik Kjellström²,³

¹Danish Meteorological Institute (DMI), Copenhagen, Denmark
²Swedish Meteorological and Hydrological Institute (SMHI), Norrköping, Sweden,
³Department of Meteorology and Bolin Centre for Climate Research, Stockholm University,
Stockholm, Sweden

⁴Corresponding author: obc@dmi.dk

ORCID iD: Ole B. Christensen 0000-0002-9513-2588 Erik Kjellström 0000-0002-6495-1038

Keywords: Ensemble, Regional climate model, Model matrix, ANOVA, EURO-CORDEX, Model variability

Abstract

Collections of large ensembles of regional climate model (RCM) downscaled climate data for particular regions and scenarios can be organized in a usually incomplete matrix consisting of GCM (global climate model) x RCM combinations. When simple ensemble averages are calculated, each GCM will effectively be weighted by the number of times it has been downscaled. In order to facilitate more equal and less random weighting among downscaled GCM results, we present a method to emulate the missing combinations in such a matrix, enabling equal weighting among participating GCMs and hence among regional consequences of large-scale climate change simulated by each GCM. This method is based on a traditional Analysis of Variance (ANOVA) approach. The method is applied and studied for fields of seasonal average temperature, precipitation and surface wind and for the 10-year return value of daily precipitation and of 10-m wind speed for a completely filled matrix consisting of 5 GCMs and 4 RCMs. We quantify the skill of the two averaging methods for different numbers of missing simulations and show that ensembles where lacking members have been emulated by the ANOVA technique are better at representing the full ensemble than corresponding simple ensemble averages, particularly in cases where only a few model combinations are absent. The technique breaks down when the number of missing simulations reaches the sum of the numbers of GCMs and RCMs.

1. Introduction

Regional climate model (RCM) downscaling of global climate model (GCM) output is a most frequently used method to obtain geographically detailed and physically consistent information of local climate change effects (Giorgi, 2019). It is a foundation for the possibility of climate services to have robust information on local geographical scale of climate change (Hewitt et al., 2020).

Many different GCM simulations are available for downscaling, and many RCMs exist, which can perform such downscaling. In order to minimize effects of both model biases and internal climate
variability on conclusions drawn from downscaling simulations, many different GCM simulations need to be downscaled, preferably with many different RCMs.

RCM simulations require not just boundary conditions from GCMs but also a considerable computational effort, and therefore it is not feasible to perform simulations for all combinations of GCM simulations and RCM models. Traditionally, there has not been a synchronized strategy for how to combine GCMs with RCMs in order to obtain a matrix with a homogeneous use of plausible GCMs. Rather, such matrices are ensembles of opportunity, where pragmatic circumstances have decided which simulations have actually been performed (McSweeney et al., 2012). In spite of this, through a considerable international effort within CORDEX (the COordinated Regional climate Downscaling EXperiment, Giorgi and Gutowski (2015)), combination matrices between the most frequently used GCMs and RCMs are being constructed for all continents. One of the largest efforts has been concerned with a simulation domain covering Europe in around 12 km grid distance, in the EURO-CORDEX community initiative (Jacob et al., 2020). Currently a matrix for 3 RCP (Representative Concentration Pathways) forcing scenarios, 10 GCMs and 13 RCM versions containing a total of 124 simulations exist.

In conventional ensemble analyses, the existing ensembles of opportunity will give an unequal weight to the various GCMs caused by the quite arbitrary distribution of downscaling simulations for each GCM. In this study we examine a method, where “holes” in a GCM/RCM combination matrix can be emulated, based on information from the existing simulations in the matrix. The method is based on an ANOVA analysis (Christensen and Kjellström, 2020) and works for \( N_G \times N_R \) matrices where at least \( N_G + N_R - 1 \) simulations exist, \( N_G \) and \( N_R \) being the number of GCMs and RCMs, respectively. We study the accuracy of the emulation and the effect of holes on both simple ensemble means and emulated ensemble means for a complete matrix consisting of 5 GCMs and 4 RCMs for the RCP8.5 emission scenario and for the 12 km European EURO-CORDEX domain, where we analyse up to 1000 combinations of simulations for each number of holes.

### 2. Methods and Data

Based on an ANOVA analysis (Christensen and Kjellström, 2020) it is possible under certain conditions to “fill holes” in a matrix, i.e., calculate emulated values for model combinations, which have not yet been filled by an actual simulation, based on an incomplete matrix of actual simulations. The terms of an ANOVA analysis determining the effect of period in a scenario, of GCM, and of RCM, are listed in Tab. 1. Given trustworthy values of the linear (single-index) terms \( (S_i, G_j, R_k) \) from Tab. 1) in an ANOVA analysis, we can emulate values for the holes. For this purpose we postulate that there is no specific effect of the combination in question, but that it is simply a sum of a period-specific, a GCM-specific and an RCM-specific term; since there is no way to estimate the single-simulation ANOVA contribution when it is missing, this contribution will be excluded. In other words, we find the best possible value based on individual contributions from GCM and RCM identities.
Emulated values of an entry $Y_{ijk}$ can then be calculated from the equation $SGR_{ijk} = 0$ for a hole corresponding to GCM $j$ and RCM $k$ and valid for both periods $i=1,2$; this corresponds to the explicit equation
\[ 0 = Y_{ijk} - Y_{i..} - Y_{.jk} + Y_{i..} + Y_{.jk} - Y_{...} \]  
(1)

where the dots indicate averaging over a dimension. Since both total means and single-RCM or single-GCM means enter this equation, a fully coupled linear system of equations will result from a situation with several holes. This equation system is not always solvable.

**Table 1.** Formulae for the various ANOVA terms, under the requirement that all terms sum to zero over any single index. Dots indicate mean over indices.

| Term | Formula |
|------|---------|
| Grand ensemble mean | $M$ | $Y_{...}$ |
| Scenario effect (climate change) | $S_i$ | $Y_{i...} - Y_{...}$ |
| GCM climate effect | $G_j$ | $Y_{.j..} - Y_{...}$ |
| RCM climate effect | $R_k$ | $Y_{..k} - Y_{...}$ |
| GCM climate change effect | $SG_{ij}$ | $Y_{ij..} - Y_{i...} - Y_{.jk} + Y_{...}$ |
| RCM climate change effect | $SR_{ik}$ | $Y_{i..k} - Y_{i...} - Y_{..k} + Y_{...}$ |
| GCM-RCM cross term for mean | $GR_{jk}$ | $Y_{.jk..} - Y_{.j..} - Y_{..k} + Y_{...}$ |
| GCM-RCM cross term for change | $SGR_{ijk}$ | $Y_{ijk..} - Y_{i..k} - Y_{.jk} + Y_{...} + Y_{i...} + Y_{.jk} - Y_{...}$ |

The technique described above is now applied to seasonal average fields obtained from 20 simulations from the EURO-CORDEX initiative (Jacob et al., 2014). The ANOVA analysis of these simulations, which form a complete matrix with 5 GCMs and 4 RCMs, has been described in Christensen and Kjellström (2020); the analysis was performed for the 19 models available at the time, the remaining simulation emulated as described in this study. The 19 model simulations have also been analysed in terms of performance for the historical climate (Vautard et al., 2020) and for climate change (Coppola et al., 2020). The models are listed in Tab. 2.

**Table 2.** Global and regional models analysed in this study.

| GCMs | CNRM-CM5 (Voldoire et al., 2013) | EC-EARTH (Hazeleger et al., 2012) | HadGEM2-ES (Collins et al., 2011) | MPI-ESM-LR (Giorgetta et al., 2013) | NorESM1-M (Bentsen et al., 2013) |
| RCMs       | HIRHAM5               | REMO2015              | RACMO22E              | RCA4               |
|-----------|-----------------------|-----------------------|-----------------------|--------------------|
|           | (Christensen et al., 2006) | (Jacob et al., 2012) | (van Meijgaard et al., 2008) | (Kjellström et al., 2016) |

We have applied a bootstrap strategy, where a relevant number of simulations are excluded from each calculation of emulated values: 1000 different matrices for each given number of holes have been generated (for 1 and 2 holes we take all possible choices: 20 possibilities for one hole, and 20*19/2=190 for two holes, corresponding to the number of different GCMxRCM matrices with 1 or 2 holes out of 20 simulations). For each of the resulting matrices, the hole-filling algorithm is applied. We assume that present and future periods either both exist or both do not exist.

### 2.1 Matrix properties

We find minimum requirements for properties of the GCMxRCM population matrix (one if a simulation exists, otherwise zero) for enabling unique calculations of emulated values for the empty holes. Examples of population matrices where the procedure does not work are instances with no simulations at all for a specific GCM or a specific RCM. Also, situations where the matrix can be split into two non-connected sub-matrices are invalid; see examples in Fig. 1. Conceptually: if one sub-matrix has generally much higher values than a remaining disconnected sub-matrix, there is no way to know if the set of GCMs or the set of RCMs are the reason; this is reflected in a redundancy in the set of equations, which makes the solution non-unique.

For the current situation of five GCMs and four RCMs the maximum number of holes turns out to be twelve. It means that for a 5x4 GCM-RCM matrix we need at least eight simulations to have a chance to build the entire matrix with this ANOVA technique. Of course it does not mean that the reconstruction of the missing element would be satisfactory in practice, but that it is a minimum number of simulations required.
Figure 1. Examples of four GCMxRCM matrices with twelve holes. The two in the top are solvable. The bottom two not: The first can be split into two disconnected sub-matrices as indicated by yellow and green colours; the second has a row without simulations (red).

The minimum necessary properties that each GCM and also each RCM is represented at least once, and that each simulation shares at least either the same GCM or the same RCM with others, and finally that no independent sub-matrices exist, lead to a minimum number of existing simulations of $N_{GCM} + N_{RCM} - 1$. At the same time, it is straightforward to generalise the examples above to show that this number is also sufficient to solve the equation. So, even if there may exist other rules than those above, which may disqualify particular combination matrices, it still holds that the minimum number of existing simulations for applying the method studied is the number above, the sum total of both kinds of models, minus one.

An example of the kind of equation system to solve is the following, where we have two holes, $Y_{jk}$ and $Y_{j'k'}$ and assume that $j' \neq j''$ and $k' \neq k''$, i.e., that the holes represent different GCMs as well as different RCMs. Let’s look at the control period $i=1$. In Eq. 1 only the last term, the total ensemble mean, connects the two holes. The two-dimensional system can then be expressed in the form

$$
\begin{bmatrix}
(N_G - 1) (N_R - 1) & 1 \\
1 & (N_G - 1) (N_R - 1)
\end{bmatrix}
\begin{bmatrix}
Y_{1j'k'} \\
Y_{1j''k''}
\end{bmatrix}
= 
\begin{bmatrix}
B_{1j'k'} \\
B_{1j''k''}
\end{bmatrix}
$$

(2)

where the number of GCMs is $N_G = 5$, the number of RCMs is $N_R = 4$, and the $B$ indicate expressions, which do not depend on the unknown hole values, only on various averages of existing simulation data for the point, season, and field in question. The exact definitions can be determined from Eq. 1.

It is important for the practical feasibility of using this method, that only one matrix inversion is necessary for each model combination. The only operation proportional to the number of points and seasons is the matrix multiplication necessary for obtaining the $Y$ values corresponding to matrix holes.

3. Results and discussion

3.1 Determining emulated values

We want to study to which degree sparsely filled matrices where holes are synthetically filled can replicate features of the full matrix. An example of a metric of the degradation of emulation quality as a function of the number of holes is shown in Fig. 2, where we look at the ensemble...
mean temperature averaged over both periods ($Y_{jk}$) as well as the mean climate change ($Y_{2jk} - Y_{1jk}$) separately at a seasonal basis. Only solvable matrices (cf. Fig. 1) have been considered. For each matrix, the holes have been filled as described above. Since any specific peculiarities of an individual simulation ($GR$ term) are impossible to emulate, we define the metric as a comparison of emulated values at each point $jk$ with the value emulated when only $jk$ was missing, i.e., the development of emulated values as more and more existing simulations are removed.

For a number $n$ of holes, we look at each $jk$ combination in turn, find the matrices where $jk$ is one of the holes, and calculate the average squared deviation from the 1-hole emulated value across all grid points. We average this quantity over all matrices with a hole at $jk$, and after that over all $jk$ combinations. In the end we have, for each field and season investigated, a measure of the mean squared deviation from the best emulated value, as a function of $n$. Taking the square root we are left with a measure of the deviation from the emulated value based on only one hole. The unit is the same as that of the quantity examined.

### 3.2 Effects of missing simulations

We will now present results from the bootstrapping analysis. Figs 3-5 show results for the three fields analysed: Seasonal mean temperature, mean precipitation, and mean 10m wind speed. For a given configuration, we compare the emulated field with the corresponding emulated field when only the combination in question is missing, i.e., the one-hole situation. The reason for this procedure is the intent to not include the specific characteristics of this combination but only look at the deterioration of the emulation as more and more actual simulations are removed.

The curves show the RMS average over missing simulations, over bootstrapped matrices, and over points. Full curves show results for mean climate, whereas dashed curves show results for climate change.
Figure 2 Average RMS deviation (deg. C) from one-hole emulated values of seasonal mean temperature as a function of the number of excess holes (the number of holes more than the one hole we compare to, for each hole in each configuration). Up to 1000 configurations have been examined for each number of holes (see text). More than 11 excess holes, i.e., a total of less than 8 existing simulations, cannot be treated with this method. Solid lines: mean climate. Dashed lines: Future minus mean, i.e., 50% of the climate change signal. DJF blue, MAM green, JJA orange, SON black.
Figure 3 Like Fig. 2, but for precipitation (mm/d).

Figure 4 Like Fig. 2, but for average 10-m wind speed (m/s).

The shapes of these curves are quite similar: A large jump in deviation appears when 1 extra hole is introduced. After that there is a slight upwards curve as a function of the number of holes, until we reach the maximum possible number of 11 extra holes, where we note that the steepness of
the curves increases. The average deviation over the full domain is of the order of 5-10 times larger for 11 extra holes compared to 1 extra hole both for the mean and for the climate change signal. We also note that there are differences in how large the errors are between the seasons. Winter stands out with the largest error for the mean climate for all three variables. Conversely, spring shows the smallest error. For the climate change signal the difference between the seasons are less consistent. However, winter stands out also here with larger errors for temperature and wind speed than in any of the other seasons.

A possible contribution to the large deviation for winter is illustrated in Fig. 5, bottom row: The differences between emulated and actual temperatures are quite large for sea-ice covered areas in the Barents Sea and north of Iceland. These regions were also recognised in Christensen and Kjellström (2020) as areas with a large $SGR$ cross-term in the ANOVA analysis. In other words, since the behaviour over sea ice is influenced by the particular combination of the GCM sea ice distribution and the RCM physics, the current emulation procedure will not be well suited to describe such areas due to the basic assumption that $SGR$ is set to zero for the missing simulations.

3.3 Estimating ensemble averages

A different way to approach the issue of added value from the hole-filling procedure is to look at how to best estimate the complete matrix mean including all 20 simulations from a set of fewer available simulations. This complete matrix mean is not necessarily closer to the physical truth than a simple average; it does, however, introduce a more “democratic” weighting of both the RCMs and the GCMs involved. In a pure ensemble of opportunity, each GCM will have an effective weight corresponding to the number of times it has been downscaled, and correspondingly for each RCM. Contrastingly, with the technique being developed here, there will be equal weight between the GCMs chosen to be represented in the matrix and also between the RCMs. We have therefore analysed two strategies for approximating the filled-matrix ensemble average from an incomplete matrix: The simple ensemble-of-opportunity “direct” average of the existing simulations in the incomplete matrix, and the mean obtained from matrix filling with emulated values filled into holes as outlined in this study.

In this section we will concentrate on temperature for illustration purposes. In Fig. 5 we see an example for one arbitrary 5x4-member matrix with 12 holes and only 8 simulations. It is clear that the direct 8-member ensemble-of-opportunity mean is mostly much farther from the complete-matrix “truth” than it is possible to achieve with the matrix-filling method used here. Further, we note relatively poor performance for the direct-average method over sea in general, where the emulation technique gives equal weight to the SST values of each GCM, just as in the true full-matrix average. A notable exception is the sea ice covered areas north of Russia and of Iceland, where the matrix filling technique is further from the true average than the direct average. The emulated results are worse in some northern sea-ice covered areas, where Christensen and Kjellström (2020) saw large difference between individual simulations and their emulated counterparts, i.e., a large role of specific GCM-RCM combinations. This is probably related to
specifics of sea ice description in the 8 models used, compared to the total ensemble. This needs further investigation.

Figure 5 Top: Deviation of one direct 8-model average DJF temperature from true 20-model average (deg. C). Bottom: Deviation of emulated 20-value average based on the same 8 models from true 20-model average. The left column shows differences in mean climate, the right column shows differences in (full) climate change.

To systematise this also for other number of holes and other seasons, we plot in Fig. 6 the RMS average over all points and all bootstrapped matrices of deviations as those shown in Fig. 5 as a function of the number of holes. It is clear from the figure that the matrix-filling procedure creates a matrix that is much more similar to the original full matrix compared to the direct mean of any
ensemble of opportunity consisting of fewer members. This is true for all numbers of holes investigated both for the mean climate (top panel) and for the climate change signal (bottom panel).

Adding perspective, in Fig. 7 we add a curve spanning all possible numbers of holes, where 1000 different combination matrices (20 for 1 and 19 missing simulations, 190 for 2 and 18 missing simulations) have been chosen randomly, since there are no solvable matrices for more than 12 holes. Note the small discrepancy around 10-12 holes between the mean deviation of completely random but different combinations and mean deviation of only solvable different combinations – This must be because non-solvable combinations will frequently miss an RCM or a GCM entirely, hence deviating more from the true average than combinations where all models are represented; ensembles of solvable matrices are more homogeneous among models and therefore somewhat closer to the true average.

A comparison between the present hole-filling averaging and the simple averaging is detailed in Fig. 8, showing the ratio between the two measures for the same sets of combination matrices. Here, we show the results for temperature, precipitation, and 10-m wind speed; we will also investigate 10-year return values of daily precipitation and 10-m wind speed. It can be seen that the improvement taking the emulated mean versus the primitive mean is always largest when only a few simulations are missing. For temperature we can estimate the true ensemble mean of both mean climate and climate change around 3 times better with the ANOVA-based technique than we do with simple averaging of available models when the matrix is almost complete, decreasing to around a factor of 2 when we reach the limits of the current technique. For precipitation the situation is similar but with somewhat smaller ratios, falling from around 2.5 to around 1.5. For wind speed we see large improvements for the mean climate, similar to temperature, whereas there is hardly any improvement for climate change. For the extremes investigated we find that this method does not add any improvement over the primitive average.

3.4 Mathematical modelling of ensemble averaging methods

To put the results into perspective, let us estimate which dependence on the number of holes we may expect. We will for a moment assume that the ensemble can be viewed as following a statistical distribution with a constant spread around a mean, which we for the moment set to zero. Let there be \( N_G \times N_R = N \) members in total of the matrix, and let us examine \( m \) holes.

The emulated values will be located around the same mean, but have a variance, which is considerably lower, at least for large matrices, since they are determined by summation rules involving sums of many existing simulations. Simple calculations of the dependence on \( m \) show that the deviation is proportional to \( \sqrt{m} \) for the emulated mean, and to \( \sqrt{m / (N - m)} \) for the direct mean; the emulated mean will have a smaller deviation than the direct one, and the ratio between them will grow with \( m \). Let us look at a matrix of independent random numbers. Whenever a hole is created, the emulation will replace this random variable with a sum of random variables; for the first hole the variance of the emulated number will be a factor \( (N_G-1)/(N_R-1) \)
smaller than the simulation variance, or almost negligible. Ignoring interactions between holes, we
get to the first order that the variance will be proportional to the number of holes, and hence the
deivation will be the square root of that. For the direct mean, the deviation between the full and
the reduced matrix mean will be \((1/N - 1/(N-m))\) times the sum of existing simulations, plus \(1/N\)
times the sum of actual values of the holes. The variability of this can be reduced to \(m/(N(N-m))\)
times the single-simulation variability.

These dependence formulae are built on several assumptions. However, testing them on the
actual data shows them to be very accurate to model the full curves for the actual analysis of Fig. 7
as a function of \(m\) with the form \(A \sqrt{m/(N-m)}\) (Fig. 9). Also, the ratio between the two kinds of
deivation depends on \(m\) as \(1/\sqrt{(N - m)}\) as expected from this simple analysis (Fig. 10), though
with a slight downward bend for the largest possible \(m\).

If we assume the functional form to hold for the real simulation matrix as a function of \(m\), also for
\(m=1\), we can straightforwardly calculate the mean deviation \((D)\) between one-hole averages and
true averages both for emulated and direct means, averaged over all one-hole configurations, and
use the results to calculate proportionality factors. For simplicity of the equations below, we will
ignore the scenario index \((S)\) in these calculations and do it for the mean over scenarios; the
results apply similarly to each scenario individually; for climate change, the RMS mean of the \(SGR\)
term replaces the \(GR\) term in the formulae, \(SG\) and \(SR\) replace \(G\) and \(R\). The results are:

\[
D_{\text{emulated}}(1) = \frac{\sqrt{\langle GR^2 \rangle}}{(N_G - 1)(N_R - 1)}
\]

\[
D_{\text{direct}}(1) = \frac{\sqrt{\langle G^2 + R^2 + GR^2 \rangle}}{(N - 1)}
\]

where the ANOVA terms correspond to the full matrix analysis, and where angled brackets
indicate average over the matrix. Combining our approximations, we find the general formulae

\[
D_{\text{emulated}}(m) = \frac{\sqrt{\langle GR^2 \rangle}}{(N_G - 1)(N_R - 1)} \frac{\sqrt{m(N - 1)}}{N - m}
\]

\[
D_{\text{direct}}(m) = \frac{\sqrt{\langle G^2 + R^2 + GR^2 \rangle}}{(N - 1)} \frac{m(N - 1)}{\sqrt{N - m}}
\]

With these formulae we can, e.g., ask how large an improvement would be expected, using the
slightly more complex emulated mean compared to the simple mean, provided that the goal is to
be close to the total full-matrix mean. The ratio between the emulated and direct mean is

\[
\frac{D_{\text{emulated}}(m)}{D_{\text{direct}}(m)} = \frac{\sqrt{\langle GR^2 \rangle}}{\sqrt{\langle G^2 + R^2 + GR^2 \rangle}} \frac{(N - 1)^{3/2}}{(N_G - 1)(N_R - 1)\sqrt{N - m}}
\]
Remembering that the denominator in the square root is the total model variance for scenario means, we see that the expected improvement in estimating the total mean is directly proportional to the relative importance of the \( GR \) terms in the ANOVA analysis, i.e., how “well-behaved” the multi-model ensemble is towards being a sum of a GCM effect and an RCM effect (The \( G \) and \( R \) terms). The better behaved, the more reason to use an emulated estimate. The presence of noise at interannual to decadal scale will decrease the advantage of emulation, since individual simulations will be out of phase and therefore have relatively large individual \( GR \) values. These observations are reflected in the results of this study: Noisy fields like precipitation and wind, and especially extremes, have only a weak deterministic part, which can be described with the ANOVA terms, and have a large remaining noisy contribution to variability.

These formulae should work for each point and each season. In other words, it is possible to construct maps based on ANOVA parameters, which show the areas with the most likely improvement in results using emulation instead of direct averages. Of course, in a real-life situation with a given non-complete combination matrix, the full-matrix ANOVA parameters will have to be approximated from the parameters based on the emulation itself.

In Tab. 3 we show the ratio for \( m=1 \) for the 4 seasons and five fields: average temperature, daily precipitation, and 10m wind speed as well as the 10-year return value of daily precipitation and of maximum daily 10-m wind speed. In summary, the method is better suited for mean fields than for climate change; it is better for temperature and wind than for precipitation. There seems to be no systematic seasonal dependence, which is common for all fields. The error reduction, which is RMS averaged over points and over configurations, varies from a fourfold improvement for some temperature fields to a reduction of roughly one third for precipitation climate change.

For extreme precipitation and wind speed there seems to be no value at all using the emulation. According to Eq. 7, one missing simulation is the configuration where we expect the largest advantage of emulation, so we conclude that the spread between individual simulations due to climate variability, as manifested in the single-simulation \( GR \) and \( SGR \) terms, are too large for emulation to add value. In Fig. 11 we show the relative difference between the two deviations across Europe for winter averages of temperature, 10-m wind speed, precipitation, and 10-year return value of daily precipitation for configurations with a single missing simulation. It is clear that there is a large gain over the Atlantic Ocean for the averages, since this is very directly determined by the GCM with quite small variability; “GCM democracy” makes a relatively large difference in this case. For all seasonal-average fields, there is added value in all points with very few exceptions, using emulation instead of direct averaging. For the extreme case there are no areas with a systematic added value of the emulation technique at all. In this case, the individual simulation results at each individual point are influenced by internal variability, and the ANOVA analysis does not work well at all.
Table 3 The ratio between emulated-mean and direct-mean deviation from full-matrix mean for one-hole matrix configurations for temperature, precipitation, 10m wind speed, and 10-year return value of daily precipitation and of 10-m wind speed, for each season in per cent. Top panel: Mean climate; bottom panel: Climate change.

|                | DJF | MAM | JJA | SON |
|----------------|-----|-----|-----|-----|
| **Temperature** | 34  | 27  | 33  | 34  |
| **Precipitation** | 41  | 36  | 49  | 44  |
| **Wind speed** | 29  | 27  | 29  | 30  |
| **Extreme precipitation** | 123 | 115 | 115 | 101 |
| **Extreme wind** | 134 | 133 | 129 | 123 |

|                | DJF | MAM | JJA | SON |
|----------------|-----|-----|-----|-----|
| **Temperature** | 48  | 38  | 32  | 25  |
| **Precipitation** | 58  | 64  | 72  | 64  |
| **Wind speed** | 54  | 57  | 56  | 52  |
| **Extreme precipitation** | 120 | 121 | 121 | 123 |
| **Extreme wind** | 122 | 122 | 122 | 123 |
Figure 6 RMS deviations over points and bootstrapped simulations of deviation from true 20-model complete matrix average seasonal mean temperatures as a function of the total number of holes. Full lines: Deviation of direct average of ensembles from 20-model truth; dashed lines: Deviation of means over emulated full matrix from 20-model truth. Top panel: Mean climate. Bottom panel: climate change. DJF blue, MAM green, JJA orange, SON black.
Figure 7 Extension of Fig. 6 with simple averages of random ensembles added. Left column: Mean climate. Right column: Climate change. Top: Temperature (K). Middle: Precipitation (mm/day). Bottom: Wind speed (m/s). See text.
Figure 8 Ratio between the square root of mean over combinations of squared difference between incomplete-ensemble mean and true ensemble mean, and the corresponding squared difference between emulated-ensemble mean and true ensemble mean. Solid lines: mean climate. Dashed lines: Climate change signal. DJF blue, MAM green, JJA orange, SON black. Top panel: Temperature. Middle panel: Precipitation. Bottom panel: 10m wind speed.
Figure 9 The simple-average curves of Fig. 7 for temperature, divided by $\sqrt{m/(N - m)}$. Full lines: Mean climate. Dashed lines: Climate change. DJF blue, MAM green, JJA orange, SON black.

Figure 10 The deviation ratio of Fig. 8 for temperature, divided by $\sqrt{N/(N - m)}$. Solid lines: Mean climate. Dashed lines: Climate change. DJF blue, MAM green, JJA orange, SON black.
Figure 11 Relative difference between emulated ensemble average deviation from true mean and direct ensemble average deviation from true mean, i.e., the relative improvement by using emulation instead of direct averages, for winter (DJF) precipitation change. Top panel: Seasonal average temperature. Second panel: Seasonal average 10-m wind speed. Third panel: Seasonal
average precipitation. Bottom panel: 10-year return value of daily precipitation. Greenish colour means the emulation is worse than direct average.

4. Conclusions

The current work focuses on a systematic investigation of effects of making a GCM-RCM matrix sparser and sparser. This allows for a quantification of how much information may be gained by adding new simulations to existing sparse real-world simulation matrices. Of course, we can only aim for an emulation of matrices, which are already partly populated; it is an additional challenge to ascertain that the GCMs and RCMs in the matrix as far as possible are representative for larger multi-model ensembles. In situations where a well filled matrix is extended by addition of new simulations, e.g., a new GCM model downscaled by one or a few of the RCMs already in the matrix, the present technique can be used to fill in emulated values for the simulations with the new GCM not yet performed, and give an estimate of an ensemble average, where the new GCM has the same weight as the already present GCMs.

It turns out to be possible to get a general idea about the gains of using emulation to try to obtain a better ensemble average of a field with equal weighting of the participating GCMs directly from the full-matrix ANOVA parameters. This estimate does not specifically depend on the field in question, nor on season etc. It only depends on these through the values of the actual ANOVA parameters of the full matrix.

In a real-life situation, the combination matrix will be given, based on the actual simulations at hand. Averages based on emulation are expected to generally give better results than direct averages; based on the present analysis, the expected improvements can be estimated point by point and season by season from Eq. 7. The improvement is large when either the GCM or the RCM choice has a large influence on simulated results; it is smaller when the individual combination and/or inter-annual variability has a large influence compared to inter-simulation variability.

While the ANOVA-based hole filling technique works well when there is skill in the ANOVA contribution-splitting itself, our results for the 10-year precipitation return value shows that it does not work well when the ANOVA linear-term variabilities are much smaller than the total ensemble variability. This expected result is seen extremely clearly. The analytic formula for the two different ways of calculating the mean of an incomplete matrix shows that the ratio between emulated-matrix-mean error and direct-mean error is simply proportional to this term, for scenario mean matrices and for climate change matrices (not explicitly shown but exactly the same calculations).

A future perspective would be to investigate if it is possible to go beyond the current model-only world and learn something about biases. One obvious step would be to make a missing-simulation analysis of bias, i.e., investigating to which extent the biases of individual simulations can be written as the sum of a GCM-specific part and an RCM-specific part. This would supplement the
current analysis of mean fields and of climate change and also supplementing the evaluation of
the entire ensemble performed by Vautard et al. (2020).

A further perspective, which will also be pursued in the future, is to put these results into
perspective through further analyses of the role of internal variability, particularly of the GCM, in
significance determination. Even when looking at 30-year averages, longer-time variations exist in
GCM simulations, the details of which can be studied through downscaling of different ensemble
members of the same GCM.

5. Declarations

Acknowledgements
The authors would like to thank the Euro-CORDEX network and WCRP CORDEX for ensuring
availability of CORDEX data.

Funding
This study has been partly funded by the Copernicus Climate Change Service. ECMWF implements
this Service on behalf of the European Commission. Part of the funding is by the Danish state
through the National Centre for Climate Research (NCKF).

Conflicts of interest/Competing interests
Not applicable

Availability of data and material
All data used in this publication are publically available through the ESGF network, e.g.,
http://esgf-data.dkrz.de

Code availability
All data manipulation in this study are straightforward and described in the manuscript

Authors' contributions
Not applicable

6. References

Bentsen M, Bethke I, Debernard JB, Iversen T, Kirkevåg A, Seland Ø, Orange H, Roelandt C,
Seierstad IA, Hoose C, Kristjánsson JE (2013) The Norwegian Earth System Model, NorESM1-M—
Part 1: description and basic evaluation of the physical climate. Geosci Model Dev. 6:687–720.
https://doi.org/10.5194/gmd-6-687-2013

Christensen OB, Drews M, Christensen JH, Dethloff K, Ketelsen K, Hebestadt I, Rinke A (2007) The
HIRHAM Regional Climate Model. Version 5 (beta). Danish Climate Centre, Danish Meteorological
Institute. Denmark. Danish Meteorological Institute. Technical Report, No. 06-17
Christensen OB, and Kjellström E (2020) Partitioning uncertainty components of mean climate and climate change in a large ensemble of European regional climate model projections. *Clim. Dyn.*
https://doi.org/10.1007/s00382-020-05229-y

Christensen JH, and Christensen OB (2007) A summary of the PRUDENCE model projections of changes in European climate by the end of this century. *Climatic Change* 81, 7-30.
https://doi.org/10.1007/s10584-006-9210-7

Collins WJ, Bellouin N, Doutriaux-Boucher M, Gedney N, Halloran P, Hinton T, Hughes J, Jones CD, Joshi M, Liddicoat S, Martin G, O'Connor F, Rae J, Senior C, Sitch S, Totterdell I, Wiltshire A, Woodward S (2011) Development and evaluation of an Earth-System model – HadGEM2. *Geosci Model Dev.* 4:1051–1075, https://doi.org/10.5194/gmd-4-1051-2011

Coppola E, Nogherotto R, Ciarlo JM, Giorgi F, Somot S, Nabat P, Corre L, Christensen OB, Boberg F, van Meijgaard E, Aalbers E, Lenderink G, Schwinghamer C, Sandstad M, Sillmann J, Bülow K, Teichmann C, Iles C, Kadygrov N, Vautard R, Levavasseur G, Sørland SL, Demory M-E, Kjellström E and Nikulin G, 2020. Assessment of the European climate projections as simulated by the large EURO-CORDEX regional climate model ensemble. *J Geophys Res* https://doi.org/10.1029/2019JD032356

Giorgetta MA, et al. (2013) Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. *J Adv Model Earth Syst* 5:572–597. https://doi.org/10.1002/jame.20038

Giorgi F, Gutowski WJ (2015) Regional Dynamical Downscaling and the CORDEX Initiative. Annual Review of Environment and Resources, 40:1, 467-490. https://doi.org/10.1146/annurev-environ-102014-021217

Giorgi, F. (2019). Thirty years of regional climate modeling: Where are we and where are we going next? Journal of Geophysical Research: Atmospheres, 124, 5696–5723.
https://doi.org/10.1029/2018JD030094

Hazeleger W, Wang X, Severijns C, Ştefănescu S, Bintanja R, Sterl A, Wyser K, Semmler T, Yang S van den Hurk B, van Noije T, van der Linden E, van der Wiel K (2012) EC-Earth V2.2: description and validation of a new seamless earth system prediction model. *Clim Dyn* 39: 2611.
https://doi.org/10.1007/s00382-011-1228-5

Hewitt, C.D., Guglielmo, F. Joussame, S. Bessembinder, J., Djurdjevic, V. Doblas Reyes, F. Garrett, N. Jiménez, I.C., Kjellström, E., Krzic, A., St. Clair, A.L., Máñez, M., 2020. Recommendations for future research priorities for climate modelling and climate services. Accepted for publication in Bull. Am. Met. Soc.

Jacob D, Elizalde A, Haensler A, Hagemann S, Kumar P, Podzun R, Rechid D, Remedio AR, Saeed F, Sieck K, Teichmann C, Wilhelm C (2012) Assessing the transferability of the regional climate model
REMO to different coordinated regional climate downscaling experiment (CORDEX) regions.  

*Atmosphere* 3:181–199. https://doi.org/10.3390/atmos3010181

Jacob, D., J. Petersen, B. Eggert, A. Alias, O. B. Christensen, L. M. Bouwer, A. Braun, A. Colette, M. Déqué, G. Georgievski, E. Georgopoulou, A. Gobiet, L. Menut, G. Nikulin, A. Haensler, N. Hempelmann, C. Jones, K. Keuler, S. Kovats, N. Kröner, S. Kotlarski, A. Kriegsmann, E. Martin, E. van Meijgaard, C. Moseley, S. Pfeifer, S. Preuschoff, C. Radermacher, K. Radtke, D. Rechid, M. Rounsevell, P. Samuelsson, S. Somot, J.-F. Soussana, C. Teichmann, R. Valentini, R. Vautard, B. Weber and Pascal Yiou, 2014: EURO-CORDEX: new high-resolution climate change projections for European impact research. *Regional Environmental Change*. https://doi.org/10.1007/s10113-013-0499-2

Jacob, D., Teichmann, C., Sobolowski, S., Katragkou, E., Anders, I., Belda, M., Benestad, R., Boberg, F., Buonomo, E., Cardoso, R.M., Casanueva, A., Christensen, O.B., Christensen, J.H., Coppola, E., De Cruz, L., Davin, E.L., Dobler, A., Dominguez, M., Fealy, R., Fernandez, J., Gaertner, M.A., Garcia-Diez, M., Giorgi, F., Gobiet, A., Goergen K., Gomez-Navarro, J.J., Gutierrez, C., Gutierrez, J.M., Guttler, I., Haensler, A., Halenka T., Jerez., S., Jiménez-Guerrero, P., Jones, R.G., Keuler, K., Kjellström, E., Knist, S., Kotlarski, S., Maraun, D., van Meijgaard, E., Mercogliano, P., Montavez, J.P., Navarra, A., Nikulin, G., de Noblet-Ducoudré, N., Panitz, H.-J., Piazza, M., Pichelli, E., Pietikäinen, J.-P., Prein, A.F., Rechid, D., Rockel, B., Romera, R., Sanchez, E., Sieck, K., Soares, P.M.M., Somot, S., Smec, L., Sørland, S.L., Termonia, P., Truhetz, H., Vautard, R., Warrach-Sagi, K. and Wulfmeyer, V. 2020. Regional climate downscaling over Europe: perspectives from the EURO-CORDEX community. *Reg. Environ. Change*, 20:51, https://doi.org/10.1007/s10113-020-01606-9

Kjellström E, Thejll P, Rummukainen M, Christensen JH, Boberg F, Christensen OB, Fox Maule C (2013) Emerging regional climate change signals for Europe under varying large-scale circulation conditions, *Clim. Res.*, 56, 103–119, https://doi.org/10.3354/cr01146

Kjellström E, Bärring L, Nikulin G, Nilsson C, Persson G, Strandberg G (2016) Production and use of regional climate model projections – a Swedish perspective on building climate services. *Climate Services*, 2-3, 15-29, https://doi.org/10.1016/j.cliser.2016.06.004.

McSweeney, Carol F., Jones, Richard G., and Booth, Ben B. B. (2012) Selecting Ensemble Members to Provide Regional Climate Change Information" *Journal of Climate* Vol. 25, No. 20, pp 7100, 1520-0442 https://doi.org/10.1175/JCLI-D-11-00526.1

van Meijgaard E, van Ulft LH, van den Berg WJ, Bosveld FC, van den Hurk BJJM, Lenderink G, Siebesma AP (2008) The KNMI regional atmospheric climate model RACMO version 2.1. KNMI Tech. Rep. TR-302, 43 pp

van Meijgaard E, van Ulft LH, Lenderink G, de Roode SR, Wipfler L, Boers R and Timmermans RMA (2012) Refinement and application of a regional atmospheric model for climate scenario calculations of Western Europe, *Climate changes Spatial Planning* publication: KvR 054/12, the
Vautard R, Kadygrov N, Iles C, Boberg F, Buonomo E, Bülow K, Coppola E, Corre L, van Meijgaard E, Nogherotto R, Sandstad M, Schwingshackl C, Somot S, Aalbers E, Christensen OB, Ciarlo JM, Demory M-E, Giorgi F, Jacob D, Jones RG, Keuler K, Kjellström E, Lenderink G, Levavasseur G, Nikulin G, Sillmann J, Solidoro C, Sørland SL, Steger C, Teichmann C, Warrach-Sagi K, Wulfmeyer V (2020) Evaluation of the large EURO-CORDEX regional climate model ensemble. *J. Geoph. Res.* https://doi.org/10.1029/2019JD032344

Voldoire A, Sanchez-Gomez E, Salas y Mélia D et al. (2013) The CNRM-CM5.1 global climate model: description and basic evaluation. *Clim Dyn* 40:2091. https://doi.org/10.1007/s00382-011-1259-y
**Figures**

|       |       |       |
|-------|-------|-------|
| 1     | 1     | 1     |
| 1     | 1     | 1     |
|       | 1     |       |
| 1     | 1     |       |
|       |       | 1     |
| 1     | 1     |       |

|       |       |       |
|-------|-------|-------|
| 1     | 1     |       |
| 1     |       | 1     |
|       | 1     |       |
| 1     | 1     |       |

**Figure 1**

Examples of four GCMxRCM matrices with twelve holes. The two in the top are solvable. The bottom two not: The first can be split into two disconnected sub-matrices as indicated by yellow and green colours; the second has a row without simulations (red).

**Figure 2**
Average RMS deviation (deg. C) from one-hole emulated values of seasonal mean temperature as a function of the number of excess holes (the number of holes more than the one hole we compare to, for each hole in each configuration). Up to 1000 configurations have been examined for each number of holes (see text). More than 11 excess holes, i.e., a total of less than 8 existing simulations, cannot be treated with this method. Solid lines: mean climate. Dashed lines: Future minus mean, i.e., 50% of the climate change signal. DJF blue, MAM green, JJA orange, SON black.

Figure 3

Like Fig. 2, but for precipitation (mm/d).
Figure 4

Like Fig. 2, but for average 10-m wind speed (m/s).
Figure 5

Top: Deviation of one direct 8-model average DJF temperature from true 20-model average (deg. C).
Bottom: Deviation of emulated 20-value average based on the same 8 models from true 20-model average. The left column shows differences in mean climate, the right column shows differences in (full) climate change. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.
Figure 6

RMS deviations over points and bootstrapped simulations of deviation from true 20-model complete matrix average seasonal mean temperatures as a function of the total number of holes. Full lines: Deviation of direct average of ensembles from 20-model truth; dashed lines: Deviation of means over emulated full matrix from 20-model truth. Top panel: Mean climate. Bottom panel: climate change. DJF blue, MAM green, JJA orange, SON black.
Figure 7

Extension of Fig. 6 with simple averages of random ensembles added. Left column: Mean climate. Right column: Climate change. Top: Temperature (K). Middle: Precipitation (mm/day). Bottom: Wind speed (m/s). See text.
Figure 8

Ratio between the square root of mean over combinations of squared difference between incomplete-ensemble mean and true ensemble mean, and the corresponding squared difference between emulated-ensemble mean and true ensemble mean. Solid lines: mean climate. Dashed lines: Climate change signal. DJF blue, MAM green, JJA orange, SON black. Top panel: Temperature. Middle panel: Precipitation. Bottom panel: 10m wind speed.
Figure 9

The simple-average curves of Fig. 7 for temperature, divided by \( \sqrt{m/(N-m)} \). Full lines: Mean climate. Dashed lines: Climate change. DJF blue, MAM green, JJA orange, SON black.

Figure 10
The deviation ratio of Fig. 8 for temperature, divided by $\sqrt{(N/(N-m))}$. Solid lines: Mean climate. Dashed lines: Climate change. DJF blue, MAM green, JJA orange, SON black.

**Figure 11**

Relative difference between emulated ensemble average deviation from true mean and direct ensemble average deviation from true mean, i.e., the relative improvement by using emulation instead of direct averages, for winter (DJF) precipitation change. Top panel: Seasonal average temperature. Second panel: Seasonal average 10-m wind speed. Third panel: Seasonal average precipitation. Bottom panel: 10-year return value of daily precipitation. Greenish colour means the emulation is worse than direct average. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any
country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.