Downscaling the probability of heavy rainfall over the Nordic countries

Rasmus E. Benestad[1], Kajsa M. Parding[1], and Andreas Dobler[1]

[1]The Norwegian Meteorological Institute, Henrik Mohns plass 1, Oslo 0313, Norway

Correspondence: R.E. Benestad (rasmus.benestad@met.no)

Abstract. We used empirical-statistical downscaling to derive local statistics for 24-hr and sub-daily precipitation over the Nordic countries, based on large-scale information provided by global climate models. The local statistics included probabilities for heavy precipitation and intensity-duration-frequency curves for sub-daily rainfall. The downscaling was based on estimating key parameters defining the shape of mathematical curves describing probabilities and return-values, namely the annual wet-day frequency f_w and the wet-day mean precipitation μ. Both parameters were used as predictands representing local precipitation statistics as well as predictors representing large-scale conditions. We used multi-model ensembles of global climate model (CMIP6) simulations, calibrated on the ERA5 reanalysis, to derive local projections for future outlooks. Our analysis included an evaluation of how well the global climate models reproduced the predictors, in addition to assessing the quality of downscaled precipitation statistics. The evaluation suggested that present global climate models capture essential aspects of the covariance, and there was a good match between annual wet-day frequency and wet-day mean precipitation derived from ERA5 on the one hand, and local rain gauges in the Nordic region on the other. Furthermore, the ensemble downscaled results for annual f_w and μ were approximately normally distributed which may justify using the ensemble mean and standard deviation to describe the ensemble spread. Hence, our efforts provide a demonstration for how empirical-statistical downscaling can be used to provide practical information on heavy rainfall which subsequently may be used for impact studies. Future projections for the Nordic region indicated little increase in precipitation due to more wet days, but most of the contribution comes from increased mean intensity. The west coast of Norway had the highest probabilities of receiving more than 30 mm/day precipitation, but the strongest relative trend in this probability was projected over northern Finland. Furthermore, the highest estimates for trends in 10-year and 25-year return-values were projected over western Norway where they were high from the outset. Our results also suggested that future precipitation intensity is sensitive to future emissions whereas the wet-day frequency is less sensitive.

1 Introduction

Increasing atmospheric concentrations of greenhouse gases, such as carbon dioxide CO_2 and methane CH_4 from human activity, strengthen the greenhouse effect and bring on global warming as well as changes in the global hydrological cycle (IPCC, 2021). Such climate change can be expressed as the sum of changes in local temperature and rainfall statistics, which

may affect both nature and society. Global climate models (GCMs) and earth system models (ESMs[1]) are our primary tools for making projections of the future climate and represent main features of Earth's climate system, but they are not designed to describe the small scales and local climate change (Takayabu et al., 2015). Nevertheless, the local response to global warming can be estimated through downscaling (see Appendix A), and international efforts on downscaling have been coordinated under the World Climate Research Programme (WCRP) and its downscaling experiment (CORDEX) (Gutowski Jr. et al., 2016). The term *downscaling* in this case refers to the process of using large-scale information that GCMs are able to reproduce skillfully, on scales larger than their *minimum skillful scale* (Takayabu et al., 2015), and subsequently add additional information about inter-scale dependencies and systematic effects from fixed geographical factors. Hence, our definition of downscaling is different to both merely transforming the data to a finer grid mesh and bias adjustment that corrects model output so that they have similar statistical characteristics as observations without further considerations of the GCMs' minimum skillful scale[2]. Results from GCMs are often downscaled to provide projections for a future climate on a regional or local scale, but the omnipresence of pronounced non-deterministic regional-scale decadal variability (Deser et al., 2012, 2020) represents a challenge and a source of uncertainty (Hawkins and Sutton, 2009). The non-deterministic chaotic contribution from natural and internal regional variations complicates the assessment of the credibility and robustness of ensemble projections, and one question is how to synthesize them into user-relevant information. This is highly relevant for results from downscaling approaches on national climate service levels, for instance within the European downscaling efforts in EURO-CORDEX.

Another source of uncertainty in downscaled climate projections is connected to methodological choices and assumptions (Jacob et al., 2020). There are two main approaches in downscaling: (i) dynamical downscaling with regional climate models (RCMs) and (ii) empirical-statistical downscaling (ESD). The former has often been more visible within CORDEX, many climate service providers as well as impacts and adaptation communities (Rampal et al., 2024), and CORDEX data often refers to a set of RCM simulations excluding ESD results, e.g. the IPCC interactive atlas[3]. The one-sided focus may be a legacy of the past European projects PRUDENCE (2001-2004) and STARDEX (2002–2005) which had their distinct focus (Christensen et al., 2007; Christensen and Christensen, 2007; Goodess et al., 2003), however, results from STARDEX didn't indicate that RCMs were superior in terms of reproducing information about extreme rainfall (Haylock et al., 2006). Traditionally, ESD has been used to estimate small-scale (local) temperature or precipitation in terms of daily variability or aggregated statistics over months, seasons or years (Maraun et al., 2015), and downscaling of heavy precipitation has mainly involved dynamical downscaling with RCMs, while the merits of ESD perhaps have not been so widely recognised.

One advantage with ESD is that it requires little computational resources which makes it suitable for downscaling large multi-model ensembles (Benestad, 2011; Mezghani et al., 2017). Furthermore, ESD can be designed so that it's transparent and easily traceable, as the R-markdown script in this paper's supporting material tries to facilitate (Benestad, 2024). It is also possible to estimate various statistical aspects on precipitation through ESD, and Trenberth et al. (2003) argued that the

[1]Henceforth, we use the term 'GCM' when referring to both GCMs and ESMs.

[2]There are, however, ESD methods that are closer to bias correction, downscaling grid points separately and hence not taking minimum skillful scale into consideration. For example, NASA's NEX-GDDP data set (https://www.nccs.nasa.gov/services/data-collections/land-based-products/nex-gddp) is presented as downscaled climate scenarios but the method is a type of bias correction. Also see Appendix A for further discussion on this topic.

[3]https://interactive-atlas.ipcc.ch/regional-information/about

characteristics of precipitation are just as vital as the amount. The characteristics of rain may indeed be more apt to change as climate changes, and some key statistics on precipitation involve both the typical amount falling on a rainy day (wet-day mean precipitation μ), how often it rains (wet-day frequency f_w), how long it is between each rainfall (dry-spell duration or number of consecutive *dry* days n_{dd}), the duration of wet-spells (number of consecutive *wet* days n_{wd} to account for clustering of precipitation events in time), the spatial extent of the precipitation (Lussana et al., 2024), and its phase (rain/snow). Here we will show how ESD can be designed to extract information on precipitation statistics such as probabilities of exceeding a certain threshold and intensity-duration-frequency curves.

There have been many studies on mean trends or extreme precipitation, but less on moderate heavy rainfall. Extremes often involves either general extreme value theory (GEV), calibrated with block maxima, or the General Pareto distribution with peak over threshold, thus fitting the tails of the distribution (Coles, 2001). GEV also involves fitting the three parameters *location*, *scale* and *shape* which are often not well constrained for limited samples of block maxima. Statistical models for moderate intense events, on the other hand, may be calibrated from the bulk of the data sample with fewer parameters (f_w and μ), and may be easier to evaluate when time series only span a few decades. Furthermore, if the parameters have a straight-forward physical interpretation, they may also serve to enhance our understanding of shifts in the statistics. Moderate extremes, such as merely 'heavy rainfall' (e.g. 20–50 mm/day), may also trigger landslides, cause erosion, and affect the spread of water-borne disease or eco-toxins. Furthermore, since GCMs only provide a coarse large-scale representation of the real climate system, it is necessary to use downscaling methods that are not degraded too much by their lack of precision. Hence we aimed for a robust and approximate method for downscaling 24-hr precipitation statistics, to some extent scarifying its exactitude which perhaps could be obtained through a sophisticated representation in an ideal setting (e.g. GEV)[4]. Furthermore, multi-variable predictors (common in traditional downscaling and in machine learning) place great and unrealistic demands on GCMs because different variables simulated by a GCM may be strongly correlated with the predictand over a historical calibration period, but may evolve in different directions in the future (Parding et al., 2019). In other words, we expect a trade-off between exactitude and robustness, and hence we aimed for robust, reliable low precision, and approximate results for moderate extremes in our case (see Appendix A for more details).

2 Data and Methods

2.1 Data

The daily rain gauge data used in this analysis were collected from the ECA&D (Klein Tank et al., 2002) within the latitude range 55–71°N and longitude range 5–30°E. The initial selection comprised 2131 rain gauges as predictand covering the time interval 1950–2021 from Belarus (4), Denmark (14), Estonia (27), Finland (443), Germany (1), Latvia (29), Lithuania (13), Norway (669), Russia (11), and Sweden (920), located at a range of elevations, the highest point being 2062 m above sea level. Only rain gauge records with sufficient number of valid data were included in the subsequent downscaling, and

[4]This refers to how closely we can reproduce the shape of the mathematical curve describing probabilities rather than a bias/variance issues for the predicted outcomes.

rain gauge measurements from only 652 locations remained in our predictand after short station records had been removed. Figure 1 shows the geographical distribution of the rain gauges and their mean annual total rainfall. The analysis was based on key aggregated statistics: annual wet-day frequency f_w and annual wet-day mean precipitation μ. We used the threshold of 1 mm/day to distinguish between dry and wet days. Annual f_w and μ with the same threshold were also used as predictors and were estimated from both the ERA5 reanalysis (Hersbach et al., 2020) as well as GCMs.

The GCM data was taken from CMIP6 (Eyring et al., 2015, 2016) for historical runs (HIST) as well as various emission scenarios (SSP370, SSP126, SSP245, and SSP585) described in IPCC (2021). Only a subset of GCM runs were included here as daily precipitation was needed to estimate annual f_w and μ for use as predictors. To reduce the data transfer amount, server-side data processing facilities at the German Climate Computing Centre (DKRZ) were used to derive the annual values with the climate data operators (CDO) software (Schulzweida, 2021) installed on site. Nevertheless, a great deal of effort was required to derive f_w and μ from ERA5 and all CMIP6 runs, and hence we make a case for a standard protocol for reanalysis and CMIP data archives that includes monthly f_w and μ. The predictors f_w and μ from CMIP6 HIST simulations were evaluated against ERA5 following Benestad et al. (2023), testing the GCMs' ability to reproduce the mean seasonal cycle, interannual variability in annual f_w and μ and their historical trends (see Appendix B), and one simulation (CESM2-WACCM-FV2) was removed due to poor evaluation results. Our analysis focused on 29 model runs following SSP370, but the other emission scenarios are included in the supporting material (Benestad, 2024).

2.2 Downscaling methodology

Our analysis introduces a new aspect in terms of downscaling by using large-scale wet-day mean precipitation μ as predictors for estimating the predictand consisting of station-level μ, as well as using large-scale wet-day frequency f_w as predictors to downscale local f_w at a station level. Both these types of predictors were estimated from the ERA5 reanalysis and CMIP6 GCMs for the Nordic region 5°W–45°E/55–72°N, using common empirical orthogonal functions (henceforth 'common EOFs') as a framework for representing both the real world and modelled conditions (Benestad, 2001). This choice implied using a so-called '*hybrid PP-MOS*'[5] framework to represent the predictors and ensured that the covariance structures from ERA5 used for calibration matched those from GCMs used for projection. The introduction of the ERA5 reanalysis has been a step change in terms of progress within ESD, as there was a close match between f_w and μ from the reanalysis and rain gauge measurements respectively (see the Appendix B), enabling their use as predictors. More details and explanations about the downscaling set-up and analysis are provided in Appendix A.

Here we distinguish between empirical orthogonal functions (EOFs) and principal component analysis (PCA). We used the former for gridded data, as is the normal convention in the scientific literature (Lorenz, 1956; Wallace and Dickinson, 1972; North et al., 1982; Preisendorfer, 1988; Navarra and Simoncini, 2010), whereas PCA (Wilks, 2006) was used for data series that had an irregular spatial distribution such as rain gauge measurements. Moreover, we used PCA to represent the predictands as it tends to emphasise large-scale structures in groups of local measurements (Benestad et al., 2015a), and a step-wise multiple ordinary linear regression (OLR) was used to find an optimal connection between principal components from EOFs

[5]https://cordex.org/wp-content/uploads/2022/08/White-Paper-ESD.pdf

representing the large-scale predictors and the principal components from PCA representing local f_w and μ. The downscaled annual f_w and μ were subsequently used to estimate the probability that daily precipitation amount (X') exceeded a given threshold (x') using the simple and approximate relation $Pr(X' > x') \approx f_w \exp(-x'/\mu)$ based on Benestad et al. (2019). The analysis for daily precipitation amounts was extended to sub-daily timescales where the shape of intensity-duration-frequency (IDF) curves was downscaled based on their dependency on $x'_{\tau,L} = \alpha\mu(L/24)^{\zeta}\ln(f_w\tau)$, where α was a calibrated adjustment factor, L was the duration of wet-spells in hours, τ was the return period, and ζ described the fractal dimension for temporal scale inter-dependencies (Benestad et al., 2020). The downscaling was carried out using the R-package `esd` (Benestad et al., 2015b) and the downscaled results for the sites of the rain gauge measurements were subsequently gridded through kriging of the spatial weights with elevation as a co-variable using the R-package `LatticeKrig` (Nychka et al., 2016). More details about the methods are provided in Appendix A.

2.3 Evaluation

The evaluation of the models and methods are documented in Appendix B and was applied to downscaled results through both conventional cross-validation and standard statistical tests of whether the observations belonged to the same statistical population as the downscaled multi-model ensemble. There was a close match between the aggregated rain gauge data and ERA5 for both f_w and μ, where the cross-validation was 0.93 for the leading PCA mode for annual f_w and where this leading PCA mode accounted for 50% of the variance. The downscaling exercise for the second PCA (29% of the variance) gave a cross-validation correlation of 0.92. Furthermore, the geographical weights of the calibrated ERA5 predictor matched spatial patterns of corresponding PCA mode, as should be expected when the same variable is used as both predictor and predictand. Similarly, the downscaling exercise between aggregated rain gauge and ERA5 for annual μ returned cross-validation correlations of 0.96 and 0.81 for first and second PCA modes respectively (representing 54 and 26% of the variance respectively), also with matching spatial weights between calibrated ERA5 data and PCA modes. In summary, both high cross-validation correlation and similar geographical distribution of spatial weights in the predictors and predictands indicate a good match between the ERA5 and rain gauge measurement annual precipitation statistics when both involve the same variable.

Our evaluation also involved testing the ability of the GCMs in reproducing the predictors in a skillful way, and is described in more details in Appendix B. It is important that the GCMs skillfully reproduce the same large-scale information that was found in the ERA5 reanalysis during calibration since we use it as predictors for making projections for the future. The test of simulated predictor quality used common EOFs (Barnett, 1999) to compare the spatio-temporal covariance structure captured by simulations with corresponding information derived from the ERA5 reanalysis, as in Benestad et al. (2023) but applied to f_w and μ respectively. The CMIP6 GCMs reproduced the mean seasonal cycle in f_w and μ aggregated from the ERA5 reanalysis, as well as the historical interannual mean variability in the annual f_w and μ (for the period 1959–2021). A comparison of historical trends in GCM historical runs and ERA5 further indicated that the GCMs were able to reproduce the observed historical changes in f_w and μ. The CMIP6 ensembles for f_w and μ were of limited size since they were generated from daily data and e.g. monthly f_w and μ values are not (yet) part of the CMIP standard output protocol. We thus limited our analysis to one particular configuration from each GCM (e.g. r1i1p1f1). The number of ensemble members of regional or local climate

projections can be interpreted as equivalent to statistical sample size, as each model simulation involves non-deterministic stochastic decadal variability (Deser et al., 2012, 2020). The normal distribution may provide useful information on statistical data samples with about 30 data points if the data are normally distributed, and hence, distributions of downscaled ensemble results were tested against a normal distribution as in Benestad et al. (2023). The results of these tests suggested that the ensemble mean and standard deviation can provide an approximate description of the ensemble.

The evaluation of both downscaling method and the GCM simulations established that local wet-day frequency f_w and wet-day mean precipitation μ can be skillfully estimated over the Nordic region from corresponding large-scale quantities from both the ERA5 reanalysis and CMIP6 simulations. The subsequent step was to use these results to make projections for future climatic outlooks and estimate changes in precipitation statistics, based on relationships established from previous studies (Benestad et al., 2019, 2020). Such steps are to the best of our knowledge the first efforts to downscale statistical properties for daily precipitation directly beyond downscaling extreme climate indices (Goodess et al., 2003; Haylock et al., 2006). Benestad et al. (2019) provided an evaluation of the statistical framework for estimating probabilities of moderate 24-hr precipitation, which involved 1875 rain gauge records from North America and Europe with more than 50 years of valid data over the period 1961–2018, and this evaluation will not be repeated here. To compensate for the thin upper tail of the exponential distribution, which is expected to significantly underestimate extremes, they introduced an empirical scaling factor α and restrained the analysis to 'moderate extremes' (20–50 mm/day).

3 Results

Figure 2 shows time series for the wet-day frequency f_w and wet-day mean precipitation μ extracted for Oslo-Blindern, and the black symbols show the annual statistics derived from historical measurements, whereas the green band shows corresponding statistics downscaled from the CMIP6 SSP370 multi-model ensemble. The comparison between model results (green band) and observations (black symbols) gives an indication of the precision of the downscaling, as it did not involve any further calibration beyond the original training of the downscaling model against the PCA-based predictand. Neither the observations nor the projections indicated any pronounced trend in the annual f_w for Oslo, however, statistics based on rain gauge measurements over all the Nordic sites nevertheless suggested a general weak increase in the number of wet days over the 1950–2021 period that was statistically significant at the 5%-level (supporting material). The downscaled projections for Oslo (green shading in Figure 2) and the Nordics (lower left panel in Figure 3), however, indicated a weak (geographically mixed and non-significant) general decrease in number of wet days for the period 2015–2099, based on the ensemble mean of the CMIP6 simulations following the SSP370 emission scenario. Other emission scenarios gave some variations in the outlook, and the SSP126 as well as the SSP585 results gave a more mixed picture of trends in future f_w, (supporting material). The trend estimates in f_w were expected to vary with the frequency of weather types, and the forces driving the atmospheric circulation that characterise different weather types tend to arise from variations in the distribution of atmospheric mass which is not necessarily strongly constrained by an increased greenhouse effect. However, there has been as slight trend in annual f_w in Oslo that was reproduced in a downscaling exercise using ERA5 as predictor (supporting material).

There has been a modest increase in annual wet-day mean precipitation μ that was more pronounced than the trends in f_w, which also is visible in Figure 2 (right panel) and Figure 3 (lower right panel). The trend estimates in μ were more spatially consistent within the various emission scenarios, although higher emissions were connected to stronger trends, and the results indicated increases for most of the region except in the vicinity of Troms municipality in northern Norway. Table 1 presents the ensemble mean and standard deviation for a small selection of locations projected for the period 2071–2100. The downscaled results suggested that projected trends in f_w were not sensitive to the emission scenario (SSPs), however, the magnitude of projected trends in μ ranked in increasing magnitude for SSP126, SSP245, SSP370, and SSP585 respectively.

Since the mean precipitation is the product of the wet-day frequency and wet-day mean precipitation[6] we estimated trends in total precipitation based on f_w, μ and the product rule, and used this information to explain total precipitation changes in terms of changing number of wet days or changing intensity. Figure 4 shows estimated future trends in precipitation (mm/day per decade in upper panel: $dx'/dt = \mu df_w/dt + f_w d\mu/dt$) as well as its contribution from changing number of wet days (lower left) and changes in mean precipitation intensity (lower right). The projections of the future climate in the Nordic region indicated a general increase in the total precipitation mainly due to increased wet-day mean precipitation μ and in spite of decreased wet-day frequency f_w, according to the selected CMIP6 simulations.

The wet-day frequency f_w and wet-day mean precipitation μ represent two key parameters for approximate estimation of the probability of heavy rainfall according to

$$Pr(X' > x') = f_w \exp(-x'/\mu), \tag{1}$$

proposed and evaluated by Benestad et al. (2019). Figure 5 shows observed fraction of days per year with more than 30 mm for Oslo-Blindern (black symbols) compared with such low-precision estimates based on this expression and the ensemble means for f_w and μ (using the expression $\overline{f_w} e^{-30/\overline{\mu}}$; red solid line) shown with error bars of one standard deviation (red dashed). In other words, the results presented here were the downscaled estimates for f_w and μ used as input in equation 1 without further calibration, and the statistics based on rain gauge measurements and information downscaled from the GCM ensembles indicated somewhat matching levels, however, the observations included some years with substantially higher numbers of days with heavy rainfall. These results nevertheless serve as an example where probabilities for heavy rainfall have been downscaled directly though the parameters f_w and μ, as opposed to aggregating data points from of a statistical sample containing traditionally downscaled time sequences of weather states. Another benefit with a parameterised expression for probability was that we could differentiate it according to the product rule: $dPr(X' > x')/dt = (df_w/dt)\exp(-x'/\mu) + f_w x'/\mu^2 \exp(-x'/\mu)(d\mu/dt)$. Figure 6 shows maps of both $Pr(X' > x')$ and percentage trends[7] for the SSP370 ensemble mean, and the results indicated highest probabilities for days receiving more than 30 mm of precipitation on the west coast of Norway, but the relative trends were greatest over northern Finland.

The parameterised expression for probabilities also enabled downscaling of approximate estimates of return-values based on $x'_\tau = \alpha\mu \ln(f_w \tau)$ where α is a calibration coefficient (Benestad et al., 2019). Figure 7 shows both 10-year (left panels) and 25-

[6] $\overline{x'} = \sum x'/n_w \times n_w/n = f_w \mu$ where $f_w = n_w/n$ and $\mu = \sum x'/n_w$.
[7] $100 \times dPr(X' > x')/Pr(X' > x')$

year (right panels) return-values as well as their estimated trends (lower panels) based on the ensemble mean SSP370 results. The greatest return-values were estimated over western Norway, with 10-year estimates ranging in 30–170 mm/day while 25-year estimates varied within the range 40–220 mm/day. The lowest estimates were downscaled for parts of northern Finland, Sweden and Norway. Projected future trends in x'_τ were estimated based on trends in the wet-day frequency df_w/dt and wet-day mean precipitation $d\mu/dt$ (lower panels in Figure 3), the above expression and the product rule, and increases in x'_τ were in general a result of increasing mean intensity rather than more wet days. The greatest trends in the return-values dx'_τ/dt were downscaled over western Norway with already high levels, but there were also notable increases over southwestern Finland and over parts of southwestern Sweden.

Downscaled f_w and μ also provided first-guess estimates for intensity-duration-frequency (IDF) curves, assuming there is a fractional dependence between temporal scales. We based our estimates of IDFs on Benestad et al. (2020), using the expression $x'_{\tau,L} = \alpha\mu(L/24)^\zeta \ln(f_w\tau)$ which describes mathematical curves whose shapes are approximately similar to IDF curves estimated through more traditional means, where α is a calibration coefficient, L is the duration in hours, τ is the return interval, and ζ describes the fractional dependency between temporal scales and was fitted to observational rain gauge measurement data. We estimated how the shape of IDF curves may change due to trends in f_w and μ (their trends are shown in the lower panels in Figure 3), and IDFs for Oslo for present and the future are shown in Figure 8. Different estimates for IDFs for the present $x'_{\tau,L}$ and the future $x^*_{\tau,L}$ provide an opportunity to estimate scaling factors for IDF curves $x^*_{\tau,L}/x'_{\tau,L}$ to account for further climate change: 1.13–1.14 for f_w and μ projected with SSP370 ensemble mean, not taking into account decadal variability. A crude measure for accounting for decadal variability was to use the ensemble spread $\pm\sigma$, and subtracting σ for the present and adding σ in the future gave scaling factors within the range 1.18–1.20 for SSP370. For higher emissions associated with SSP585, the scaling factors were 1.27–1.38, in this case only based on the ensemble mean and not accounting for decadal variability. All these estimates varied with the return-period τ, but the scaling factors were the same across time durations L in accordance with the expression above. In this case, we assumed that α and ζ were constant for a given site.

We also explored the connection between the wet-day frequency and duration of dry spells (number of consecutive dry days), which may provide some indication of meteorological drought risk (supporting material). The calibration of our ESD method indicated that there was a link between large-scale f_w from ERA5 and the mean duration of dry spells. The spell duration approximately followed a geometric distribution where the mean duration (number of consecutive dry days) was the inverse of the "success" probability, which implies that we approximately can estimate the probability of a dry spell lasting longer than a given threshold. A projected weak reduction in f_w over the Nordic region will therefore suggest slightly increased risks of meteorological droughts in the future.

4 Discussions

To our knowledge, this is the first time the shape of curves representing probabilities for heavy rainfall or IDF curves have been downscaled using a hybrid PP-MOS approach (which addresses the '*domain adaption*' aspect discussed in Rampal et al. (2024)) applied to multi-model GCM ensembles, albeit estimating the parameters defining their shapes. Those parametric

expressions nevertheless enabled us to analyse the causes for trends in precipitation, probabilities, return-values, probability of meteorological droughts, or for shifts in the shape of IDF curves. These statistics were calculated from formulas which used downscaled f_w and μ as input, and the results underscored that both the wet-day frequency and the wet-day mean precipitation are two key parameters for describing 24-hr precipitation. In our case, the results were more sensitive to the mean precipitation intensity μ than wet-day frequency f_w.

Our results suggested a slight reduction in the future wet-day frequency over the Nordic countries which may reflect predominant changes in the atmospheric circulation patterns, due to the location of storm tracks and blocking high-pressure systems. Present state-of-the-art GCMs still have biases when it comes to storm tracks and blocking frequencies, which is possibly due to a coarse representation of the polar jet stream and other processes in the Arctic (IPCC, 2021). The downscaling may underestimate long-term changes in the mean precipitation intensity μ, even if the evaluation of the CMIP6 models seemed to score well on the comparison between trends in GCMs and ERA5. A separate test where μ was downscaled solely based on ERA5 reanalysis didn't capture the historical changes observed in Oslo (supporting material). Furthermore, the projections of wet-day frequencies f_w didn't account for the risk that circulation patterns may change in ways not captured by present models. There may also be tipping points in the North Atlantic and sea ice cover, changes in the jet stream, effects from displaced storm tracks, and inaccurate simulation of blocking high-pressure system frequencies (IPCC, 2021). Nevertheless, a take-home message is that long-term trends in μ were sensitive to future emissions.

One question is whether the fractal temporal scaling properties utilised in the approximate IDF representation in Figure 8 is stable or if we can expect it to change in time and space. It is also possible that there are diverging trends in f_w or μ during different seasons that cancel each other in the annual mean, e.g. associated by prevailing presence of different seasonal meteorological phenomena. Our results suggested that the annual wet-day frequency f_w was more coherent over space, as all the 20 leading EOFs combined accounted for 88% of the variance in the ERA5 reanalysis compared to 74% for the annual wet-day mean precipitation μ. Moreover, the leading EOF mode for the annual wet-day mean precipitation μ from ERA5 captured 19% of the variance as opposed to 30% for f_w, which suggests that μ to a greater degree reflects small scale processes and phenomena not being as strongly correlated over the region on annual time scales. Local and mesoscale processes and phenomena that may influence μ include surface-air fluxes, and local geographical effects such as orographic forcing. However, both f_w and μ are expected to reflect meteorological phenomena ranging from local microscale, mesoscale and synoptic scales that may produce precipitation with different characteristics, dynamics and mechanisms, including convection, cut-off lows, mid-latitude cyclones, frontal systems, atmospheric rivers, and orographic forcing. Both increased precipitation amount from higher surface temperature as well as changes in the distribution of the precipitation over the planetary surface play a role in the trends in extreme precipitation amounts. Benestad et al. (2024) found a link between increased intensity on the one hand, and increased rate of evaporation as well as changes in the global surface area receiving daily precipitation on the other. They also observed that changes in the global fractional surface area with daily precipitation were connected to the global statistics of the wet-day frequency f_w.

Using the same variables for predictors and predictands, as in this case, leaves it up to the GCMs to represent the underlying phenomena that generate precipitation. We could refer to this strategy as a hybrid 'SR-MOS' in the terms proposed by Rampal

et al. (2024) rather than 'PP-MOS', however, we stick to 'PP-MOS' for simplicity. Improved GCMs in the future may reproduce various meteorological phenomena and processes with improved skill which may lead to better estimates for future projections. It is also important that the reanalysis used for calibration matches the predictands closely.

Our results were produced with a hybrid PP-MOS strategy for downscaling climatic parameters represented through PCAs that may serve as a benchmark for machine learning and artificial intelligence (Rampal et al., 2024). There is value in combining this ESD approach with more advanced machine learning (ML) or artificial intelligence (AI) methods that produce results with very different constraints. However, since downscaling f_w or μ doesn't require as large data volume or as long time series as either ML/AI or traditional methods for studying extremes, such a comparison will be limited to cases with ample observational data or 'pseudo-realities' using model output. One merit of our strategy is that it provides an explainable method which enhances our understanding of projected changes and thus compliments many ML/AI methods. Hence, our downscaling strategy addresses some of the research questions stated in Rampal et al. (2024), and when the recipe of the entire analysis can be documented through an R-markdown script (supporting material), it's easier to provide transparency and the traceability sought in scientific discourse.

It's important to combine equivalent results from both ESD and RCMs when downscaling is used to produce regional or local climate projections for the future, since they are based on different assumptions and have different strengths and weaknesses but are expected to give similar results for aggregated precipitation and temperature. We leave a comparison with similar information from RCMs for future work, and it is also important to account for chaotic and stochastic variability on regional and decadal scales (Deser et al., 2012, 2020), for instance using large multi-model ensembles as a surrogate for statistical sampling and letting the ensemble spread give a crude representation of probable outcomes. This analysis suggested that the ensemble spread for both annual f_w and μ were approximately normal which implies that the ensemble mean as well as standard deviation may provide useful information about the ensemble spread.

All expressions used here in connection with ESD can also be combined with RCMs, as Oguz et al. (2024) used the EURO-CORDEX ensemble (RCMs) rather than ESD to estimate f_w and μ. They subsequently used the IDF curves as a basis for weather generators (Monte-Carlo simulations) to provide input for landslide modelling. Nevertheless, based on their utility, the wet-day frequency f_w and the wet-day mean precipitation μ should be listed among essential climate indicators[8], and they should be included in the standard output from reanalysis, GCMs (e.g. CMIP[9]) and RCMs (CORDEX[10]), e.g. as monthly mean f_w and μ. The CMIP ensemble here was limited to one simulation per GCM because f_w and μ had to be estimated from available daily output, making it difficult to explore uncertainties connected to initial conditions, natural variability as well as model choices (Mezghani et al., 2019). However, it may be possible to use factorial regression or ANOVA to assess how model choice affects the downscaled ensemble with larger multi-model ensembles that include multiple simulations with the same GCM (Benestad et al., 2017, 2016). With the available CMIP6 data in this case, it was only possible to carry out an assessment of the sensitivity to emissions through comparing downscaled results from SSP126, SSP245, SSP370 and SSP585.

[8]https://gcos.wmo.int/en/essential-climate-variables/

[9]https://esgf-data.dkrz.de/search/cmip6-dkrz/ or https://cds.climate.copernicus.eu/

[10]e.g. https://esgf-data.dkrz.de/search/cordex-dkrz/

5 Conclusions

We used the ERA5 reanalysis and local rain gauge measurements from the Nordic countries to calibrate empirical-statistical downscaling models, which were applied to CMIP6 projections using annual wet-day frequency f_w and wet-day mean precipitation μ respectively both as predictors and predictands. A good match between the ERA5 reanalysis and rain gauge measurements for these two key statistics over the Nordic region gave a good calibration of our downscaling method. Predictors from global climate models from CMIP6 were evaluated and scored well in terms of their ability to represent mean seasonal variations, interannual variability of annual aggregates and past trends of the large-scale predictors needed for the downscaling, and our downscaling used a hybrid PP-MOS approach for estimating parameters for mathematical curves providing actionable regional climate information. The downscaled f_w and μ were subsequently used to estimate local probabilities for heavy rainfall, return values and changes in the shape of intensity-duration-frequency curves. We used kriging with elevation as covariate to generate Nordic maps of f_w and μ and their projected changes. Projected changes in the future suggest increases in μ but very slight decreases in f_w, hinting at less frequent or similar level of wet days in the future but also more intense rainfall. The amplitude of projected trends in μ was sensitive to the emission scenario, but trends in f_w were not. The spread between the ensemble members was approximately normally distributed, which implies that essential information about the ensemble may be captured through the ensemble mean and standard deviation.

Code availability. The R-markdown script, on which this analysis is based, is provided in the supporting material and available from FigShare (Benestad, 2024)

Appendix A: Detailed information about the downscaling methodology

A1 Interpretations of the concept of downscaling

There are different definitions of *downscaling*, one being the mapping of data onto a finer grid and another for which the information on large-scale features, which climate models are able to reproduce, is combined with information about the dependency across spatial scales to derive small-scale information. The former may not always take into account the fact that numerical models have a *minimum skillful scale* and only provide a limited representation of reality (Takayabu et al., 2015). Downscaling is not restricted to producing gridded data with higher resolution, and there are examples where downscaling has been carried out for a single location (Wilby et al., 2002; Maraun et al., 2015; Benestad et al., 2008). Moreover, the main objective of the COST-Value project (Maraun et al., 2015) was to establish a standard evaluation scheme based on 85 different single locations scattered across Europe. On the other hand, a plain interpolation to finer a grid is usually not considered to be a downscaling approach, but bias-adjustment is sometimes referred to as downscaling. Neither an interpolation, spatial disaggregation nor bias-adjustment, or any combination thereof, emphasise the large-scale aspects that numerical models are able to reproduce with greater skill than grid-point estimates. Global climate models have a typical spatial resolution of 100 km and therefore only have a coarse representation of the land surface, and the mountain regions are represented by crude pixels with typically lower heights than in reality. Some of the said simple approaches for producing data on a finer grid may implicitly add information about elevation, e.g. through the inclusion of bias-adjustment or kriging with elevation as a covariate, but the models' minimum skillful scale is not the same as the model resolution. Moreover, it is acknowledged that the models' minimum skillful scale typically encompasses several grid boxes (Von Storch et al., 1993; Benestad, 2016). Various models in the CMIP6 ensemble have different spatial resolution, ranging from 50 km to 260 km, whereas the ERA5 has a resolution of approximately 31 km (this data is provided on a reduced Gaussian grid which has quasi-uniform spacing over the globe). Furthermore, model data typically represent the average value over a grid-box volume (e.g. temperature) or area (e.g. precipitation) with a spatial dimension of several cubic or square kilometres, whereas observations represent conditions with spatial scales of metres. The local rain gauge data can, for all intents and purposes, be considered as point source measurements (collected by funnels with a 20 cm diameter) and represent local (small-scale) climate information. In our analysis, downscaling provides the translation of large-scale information, that can be provided by global climate models, to local statistics for precipitation collected by rain gauges by adding information about their dependencies.

While RCMs and traditional ESD provide output for a sequence of atmospheric states (or "outcomes") on daily or sub-daily resolution, which we can refer to as weather conditions, our strategy has been to downscale the key *parameters* describing the shape of the mathematical curve for local probability, rather than estimating the statistics from samples made up of such data sequences. We can loosely refer to the former as '*downscaling weather*' whereas the latter can be termed '*downscaling climate*' if climate can be defined as weather statistics or probability density functions (pdfs) reflecting (sub-)daily precipitation amounts. Statistical properties of precipitation are expected to follow a more systematic geographical distribution than any random individual weather event, being influenced by prevailing large-scale conditions as well as fixed local geographical factors. Our objective was to downscale parameters describing the shape of a pdf or similar mathematical curves, and this

approach was first inspired by Pryor et al. (2005, 2006) and is based on a long-term effort and a series of projects (e.g. EU-SPECS[11], CixPAG[12], KlimaDigital[13], EU-SPRINGS[14]). The 'downscaling climate' approach can also be applied to e.g. summertime heatwaves or used to downscale the probability of the occurrences n_H as well as duration of hot spells $\overline{L_H}$ (Benestad et al., 2018), however, heatwaves were beyond the scope of the present analysis. Another example of the merit of this concept is the downscaling of storm track density (Parding et al., 2019), and future work in the EU-SPRINGS project will explore the possibility to downscale public health statistics for water-borne diseases that may lead to diarrhoea. The application of the 'downscaling climate' approach is not as wide-spread as downscaling of time sequences with individual atmospheric states.

A2 The predictors representing the large scales

Both the covariates from reanalyses used for calibrating the downscaling methods and corresponding covariates from global climate models used for making projections are referred to as *predictors* in the context of downscaling. Such predictors represent large-scale aspects that global climate models are able to reproduce with skill. Here we chose predictors that consisted of the same variables as the small-scale information that we sought through downscaling: the annual wet-day frequency f_w and the annual wet-day mean precipitation μ. This choice was motivated by the expectation of a systematic dependency between the large-scale and small-scale aspects of the same variable.

All of the CMIP6 models in our analysis were regridded to match the grid of ERA5 for the region 5°W–45°E/55–72°N. Since data produced by reanalyses and global climate models have a high degree of redundancy, the information contained therein can be reorganised as spatially coherent patterns which represent substantial fractions of the covariance structure. These patterns involve mathematical techniques within linear algebra (Strang, 1988) known as *empirical orthogonal functions* (Lorenz, 1956; Wallace and Dickinson, 1972; North et al., 1982; Preisendorfer, 1988; Navarra and Simoncini, 2010), commonly referred to as *'EOFs'*. EOFs (and PCA used to describe predictands in the next subsection) make use of this redundancy and organise the information so that the most salient aspects of its covariance structure is represented by its leading modes. Furthermore, the high degree of redundancy makes it possible to represent the most important covariance information in a much smaller volume of data than the original raw data, as illustrated by the schematic in Figure A1. Here we use X to represent the anomalies of the original data with a temporal dimension n_t and a spatial dimension n_r (for gridded data, $n_r = n_x \times n_y$, but here the particular geographical arrangement of the data points is not affecting the calculations). Both the EOFs (and the PCA for the predictand) were implemented though the means of a singular value decomposition (SVD) (Strang, 1988) where U represented the spatial weights ('geographical pattern'), Λ was a diagonal matrix that held the eigenvalues (variances) in decreasing order, and V contained the time series (principal components, PCs, used in the regression analysis) according to

[11]https://cordis.europa.eu/project/id/308378
[12]https://cicero.oslo.no/no/prosjekter/cixpag
[13]https://www.sintef.no/projectweb/klimadigital/
[14]https://www.springsproject.eu/

$$X = U\Lambda V^T. \tag{A1}$$

One important issue is that the same large-scale structures in the predictors found for the reanalysis during calibration of the downscaling methods must be found in the model simulations to make projections for the future. A simple way to ensure identical covariance structures in the two is to use so-called *common EOFs* as proposed more than 20 years ago by Benestad (2001), where anomalies of the GCM data are mapped onto the same grid ("regridded") as those from the reanalysis and the respective anomalies are combined so that the GCM data follows the ERA5 data in time, $X = [X_{ERA5}, X_{GCM}]$. Here, each GCM simulation was regridded to match the grid of ERA5 through bilinear interpolation, and ordinary EOFs were estimated for the joint data matrix. Since the spatial patterns U and the eigenvalues Λ were common for the joint data matrix, the two data sources were only distinguished through $V = [V_{ERA5}, V_{GCM}]$ in equation A1.

A3 The predictands representing the small scales

The predictand consisted of 652 local rain gauge measurements from the Nordic countries over the period 1951–2021, and one reason to use a principal component analysis (PCA) of annually aggregated statistics (f_w and μ) was that its gravest modes had a closer link to large-scale predictors than each local time series (Benestad et al., 2015a). The mathematics of PCA was similar to equation A1, but the original data and hence the matrices therein were distinct from that of the predictor and can be expressed as $X' = U'\Lambda' V'^T$. The downscaling only involved a representation of the predictands in the shape of PCA, where the local climate information was embedded in the spatial weights U' and eigenvalues Λ'.

The results from the downscaling were subsequently post-processed to provide maps as shown herein. The maps were generated though a kriging based on Markov random fields (Nychka et al., 2016) and made use of the R-package `LatticeKrig` which follows a "fixed rank Kriging" approach with a large number of basis functions. It was designed to provide spatial estimates that were comparable to standard families of covariance functions, and its Markov random field approach, combined with a basis function representation, was supposed to enable an implementation of different geometries. The kriging aspect here was merely used to provide spatial maps once local information had been derived for f_w and μ for the locations of the rain gauge measurements, and the main objectives here was to demonstrate how daily precipitation statistics can be derived through empirical-statistical downscaling and then be used for making local projections for a future climate. Furthermore, the kriging was only applied to the spatial patterns of the PCA for the leading modes U' to produce U'_{krig}, and the expression $X' = U'_{krig}\Lambda' V'^T$ was subsequently used to generate maps of f_w and μ with X' representing either f_w or μ. For downscaled estimates, the contents of V' was replaced with the results of the regression model presented in the next section below.

While the statistical parameters f_w and μ were subject to downscaling, we sought solutions for expressing probabilities $Pr(X' > x')$ and return periods of heavy precipitation ("moderate extremes", typically, $X' \in [10, \cdots, 50]mm$) based on a modified exponential distribution. We used approximated estimates for the probability of heavy precipitation based on $Pr(X' > x') = f_w e^{-x'/\mu}$ and return values according to $x'_\tau = \alpha\mu \ln(f_w \tau)$. Benestad et al. (2019) evaluated these expressions for 9817 locations in Europe and North America, and we will not repeat this evaluation here (the results are published in

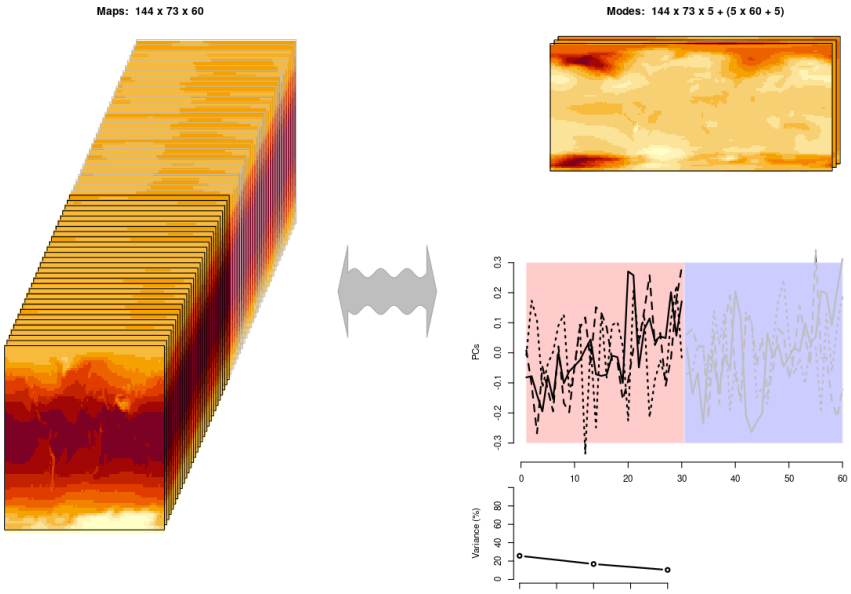

Figure A1. A graphical illustration of representing the predictors in terms of EOFs. The left hand side shows the data matrix with one map for each year, but since there are many reoccurring/similar ("typical") spatial patterns, it is possible to represent the most salient information of this data matrix in terms of three dominant patterns (right hand side) with temporal weights describing their presence and eigenvalues indicating their general prominence. This schematic furthermore illustrates the concept of 'common EOFs' where part of the data matrix holds reanalysis data and the other part holds GCM data. Their temporal weights are also distinguished with different background colour on the right, and the part representing the reanalysis are then used in the calibration against rain gauge data whereas the other is used for making projections. Typically, the common EOFs require much less computer memory and are easier to process than the original data. They also provide a framework for evaluating the predictors since the temporal weights associated with the reanalysis and the GCM should have similar statistical properties. Since our schematic only includes the three leading mode it reflects the expression $X \approx U\Lambda V^T$ rather than equation A1.

an open-access journal). The approach for estimating the parameters that determine the shape of intensity-duration-frequency (IDF) curves was evaluated by Parding et al. (2023) and Benestad et al. (2020) for sites in Norway, and this evaluation will not be repeated here either (the said papers are also in open-access journals). The main objective here was to show how parameters that specify the shape of mathematical curves for local precipitation statistics can be derived directly through empirical-statistical downscaling, given that the curves themselves provide useful information.

A4 Details about the downscaling method

To ensure that the same spatial covariance structure in ERA5 associated with variation in the rain gauge statistics is the same in the GCM, the regression analysis was carried out within a framework of the spatial patterns held in matrix U that are common

for both reanalysis and model. The calibration involved a set-wise multiple ordinary linear regression (OLR) which only used part of the principal components V_{ERA5}:

$$\hat{V_j'} = \beta_{0,j} + \sum_i \beta_{i,j} V_{ERA5,i}. \tag{A2}$$

In equation A2 the term $V_{ERA5,i}$ is principal component i of the EOFs representing the predictor from ERA5 used for calibration, whereas V_j' represents the order j principal component from the PCA representing the predictand and the aggregated statistics based on the local rain gauge data. In this case, V' and $V_{ERA5,i}$ were synchronised time series representing local and large-scale annual precipitation aggregates respectively. It is equation A2 that facilitates the transform from large to small scales and is referred to as the *downscaling method*, in this case involving a regression model, whereas the rest of the data processing provides the preparations, framing and the proper context for this analysis. The calibration provided estimates for the regression coefficients β_i which were then used to make projections for the future according to

$$X' = U' \Lambda' V_{DS}', \tag{A3}$$

where U' and Λ' are the spatial weights and eigenvalues from the PCA representing the predictand, and $V_{DS,j}' = \beta_{0,j} + \sum_i \beta_{i,j} V_{GCM,i}$ incorporates the results from equation A2. In other words, we used equation A3 together with the regression coefficients and the part of the common EOFs representing the global climate models ($V_{DS}' = [V_{DS,1}', V_{DS,2}', \cdots]$) to make projections.

In our downscaling attempts over the Nordic region, we used the 5 leading PCA modes ($j = [1, 2, \cdots, 5]$) to represent the most salient information of annual f_w and μ estimated from the rain gauge measurements (the predictands), representing 100% of the variance in the station-based statistics for both. To represent the predictors, we used the 7 leading EOFs ($i = [1, 2, \cdots, 7]$), estimated for f_w or μ from ERA5, in a step-wise multiple OLR to estimate each PCA mode for the predictand. In other words, the OLR was used to relate large-scale information from ERA5 to local information provided by the rain gauge data, and time series representing annual f_w and μ were generated based on the regression coefficients β_i and subsequently computed according to equation A3.

The first step of the model calibration involved a 5-fold cross-validation (Gutiérrez et al., 2018), where the data was split into 5 equal segments and one was withheld from the calibration of the remaining 4 segments and then compared with predicted values ('out-of-sample). This exercise was repeated for all combinations and the final cross-validation scores were estimated based on all iterations. The final calibration, however, was carried out for annual data over the entire period 1951–2021 (51 data points for each PCA mode).

Appendix B: Evaluation

B1 Cross-validation

It is a standard practice to evaluate downscaled results through a cross-validation exercise and tables B1– B2 show cross-validation correlations for each of the five PCA modes and for each type of GCMs. The scores vary slightly due to different spatial resolution and slight differences in their embedded covariance information.

Table B1. Cross-validation correlation of the principal components from PCA used to represent the predictand for f_w (columns). The rows represent the different results for the different ensemble members.

	"PC1"	"PC2"	"PC3"
ACCESS.CM2.r1i1p1f1	0.93	0.93	0.79
ACCESS.ESM1.5.r1i1p1f1	0.91	0.93	0.76
AWI.CM.1.1.MR.r1i1p1f1	0.92	0.92	0.8
BCC.CSM2.MR.r1i1p1f1	0.9	0.94	0.75
CanESM5.r1i1p1f1	0.9	0.93	0.77
CMCC.CM2.SR5.r1i1p1f1	0.92	0.93	0.79
CNRM.CM6.1.r1i1p1f2	0.91	0.93	0.78
CNRM.ESM2.1.r1i1p1f2	0.91	0.93	0.8
EC.Earth3.r1i1p1f1	0.9	0.93	0.79
EC.Earth3.AerChem.r1i1p1f1	0.9	0.93	0.8
EC.Earth3.Veg.r1i1p1f1	0.89	0.94	0.79
EC.Earth3.Veg.LR.r1i1p1f1	0.9	0.93	0.79
FGOALS.g3.r1i1p1f1	0.9	0.94	0.75
GFDL.ESM4.r1i1p1f1	0.9	0.94	0.78
INM.CM4.8.r1i1p1f1	0.92	0.92	0.79
INM.CM5.0.r1i1p1f1	0.92	0.93	0.8
IPSL.CM5A2.INCA.r1i1p1f1	0.93	0.93	0.8
IPSL.CM6A.LR.r1i1p1f1	0.93	0.92	0.81
KACE.1.0.G.r1i1p1f1	0.93	0.93	0.78
MIROC.ES2L.r1i1p1f2	0.92	0.92	0.8
MIROC6.r1i1p1f1	0.9	0.91	0.79
MPI.ESM1.2.HR.r1i1p1f1	0.92	0.92	0.8
MPI.ESM1.2.LR.r1i1p1f1	0.92	0.93	0.78
MRI.ESM2.0.r1i1p1f1	0.92	0.92	0.81
NorESM2.LM.r1i1p1f1	0.89	0.93	0.8
NorESM2.MM.r1i1p1f1	0.89	0.93	0.77
UKESM1.0.LL.r1i1p1f2	0.9	0.93	0.75

Table B2. Cross-validation correlation of the principal components from PCA used to represent the predictand for μ (columns). The rows represent the different results for the different ensemble members.

	"PC1"	"PC2"	"PC3"
ACCESS.CM2.r1i1p1f1	0.94	0.8	0.77
ACCESS.ESM1.5.r1i1p1f1	0.94	0.79	0.74
AWI.CM.1.1.MR.r1i1p1f1	0.95	0.78	0.78
BCC.CSM2.MR.r1i1p1f1	0.95	0.76	0.75
CanESM5.r1i1p1f1	0.95	0.79	0.72
CMCC.CM2.SR5.r1i1p1f1	0.94	0.62	0.74
CNRM.CM6.1.r1i1p1f2	0.94	0.8	0.77
CNRM.ESM2.1.r1i1p1f2	0.94	0.77	0.71
EC.Earth3.r1i1p1f1	0.94	0.76	0.75
EC.Earth3.AerChem.r1i1p1f1	0.94	0.8	0.75
EC.Earth3.Veg.r1i1p1f1	0.95	0.78	0.76
EC.Earth3.Veg.LR.r1i1p1f1	0.96	0.77	0.76
FGOALS.g3.r1i1p1f1	0.96	0.78	0.76
GFDL.ESM4.r1i1p1f1	0.94	0.78	0.74
INM.CM4.8.r1i1p1f1	0.95	0.79	0.74
INM.CM5.0.r1i1p1f1	0.94	0.79	0.76
IPSL.CM5A2.INCA.r1i1p1f1	0.94	0.77	0.74
IPSL.CM6A.LR.r1i1p1f1	0.94	0.77	0.75
KACE.1.0.G.r1i1p1f1	0.92	0.76	0.74
MIROC.ES2L.r1i1p1f2	0.94	0.81	0.74
MIROC6.r1i1p1f1	0.95	0.78	0.72
MPI.ESM1.2.HR.r1i1p1f1	0.96	0.8	0.76
MPI.ESM1.2.LR.r1i1p1f1	0.94	0.8	0.75
MRI.ESM2.0.r1i1p1f1	0.94	0.8	0.76
NorESM2.LM.r1i1p1f1	0.94	0.8	0.76
NorESM2.MM.r1i1p1f1	0.95	0.8	0.76
UKESM1.0.LL.r1i1p1f2	0.94	0.79	0.76

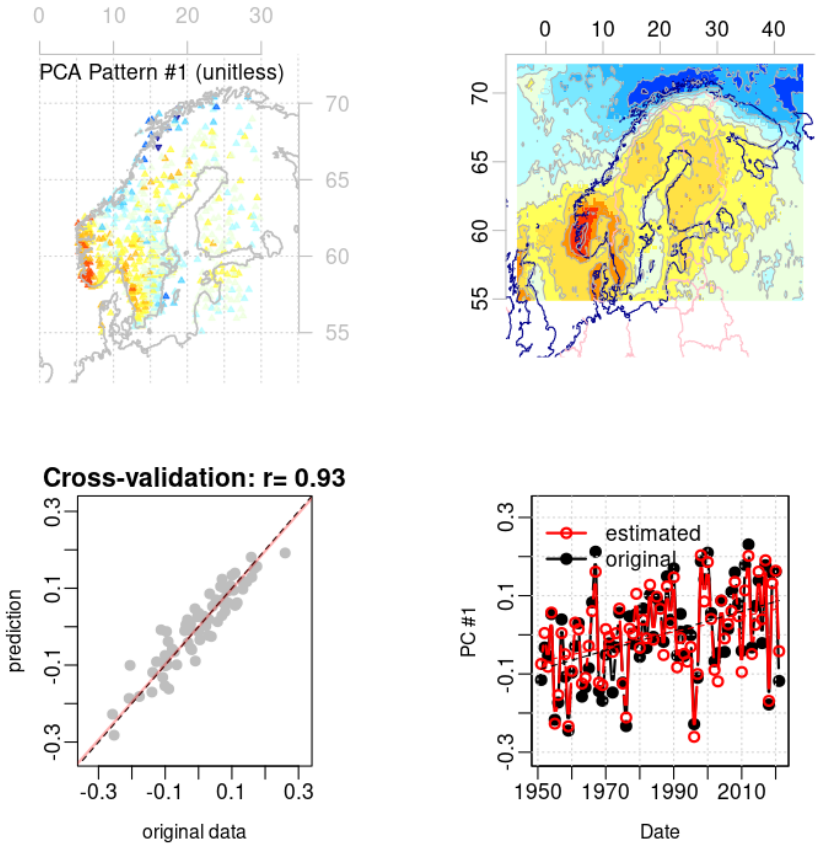

Figure B1. Diagnostics of the calibration of the multiple regression model for the leading PCA mode for annual f_w. The upper left panel shows the spatial weights of annual f_w from derived from rain gauge measurements and the upper right panel shows the spatial weights from the weighted combination of EOFs of corresponding ERA5 data weighted according to the regression coefficients from calibration exercise. The lower left panel provides the results from a 5-fold cross-validation and the lower right panel examines how well the multiple regression captures long-term trends. This is an example of a skillful calibration where the spatial weights match, the cross-validation score is high and the long-term trends are well reproduced.

B2 Evaluation of ERA5

A good match between annual rain gauge statistics and corresponding statistics derived from ERA5 also constitutes an evaluation of the ERA5 reanalysis. Hence, diagnostics of empirical-statistical downscaling can be used to evaluate reanalyses such as ERA5. Figure B1 gives a graphical presentation of diagnostics associated with the calibration of the regression coefficients for the leading PCA mode $\beta_{i,1}$ of f_w where $i \in [1, \cdots, 7]$. These figures indicate that the spatial weights with most impact on

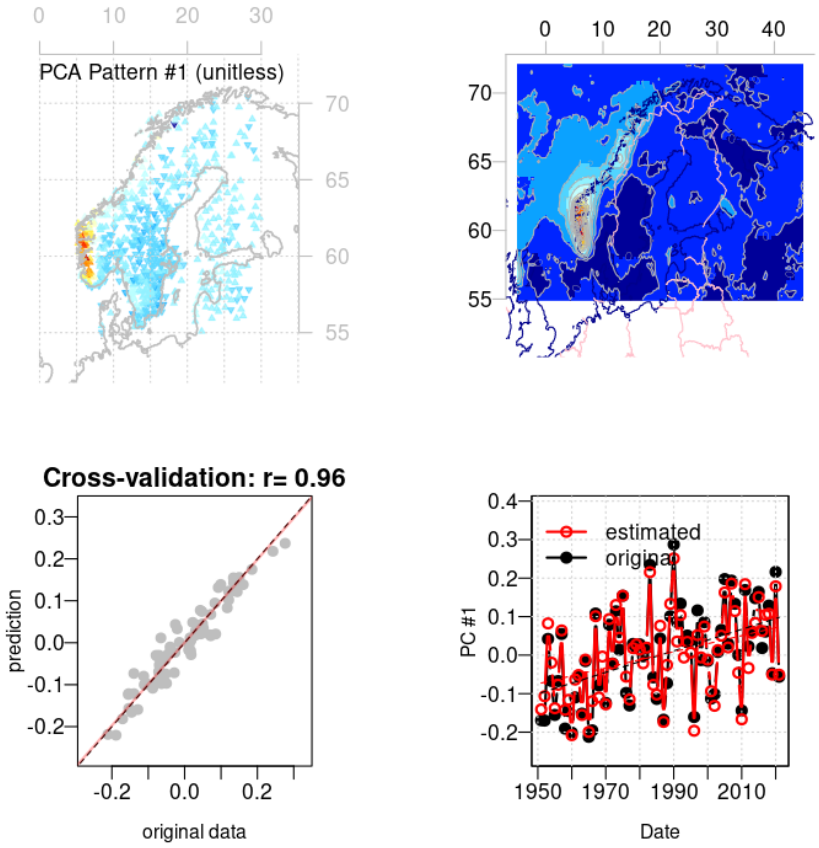

Figure B2. Same as Figure B1 but for μ.

annual f_w in ERA5 match the geographical distribution of the sites with greatest weight in U_1' (upper panels shows $\sum_i \beta_{i,1} U_i$), and the lower left panel shows the results of a cross-validation applied to pure ERA5 data.

Figure B2 shows similar results for μ and shows that there was a close match between the annual wet-day mean precipitation aggregated from rain gauge data and from the ERA5 reanalysis.

B3 Evaluation of the global climate models

Since large-scale aspects were used as predictors in the downscaling, we evaluated the skill of the selected global climate models in reproducing them. Large-scale aspects from ERA5 were used for the calibration of the downscaling models and therefore the climate simulations were compared with corresponding ERA5 data to make the evaluation relevant for the downscaling results. We started by assessing the mean annual cycle to provide a test of whether the representation of physical processes and conditions in the models capture the most salient variations such as the mean seasonal cycle. Further steps in our evaluation involved testing their ability to reproduce the characteristics of interannual variations and past trends in f_w and μ. Both interannual variability and assessment of past trends are relevant for when downscaling is used to make projections for the future, because the former reveals whether the models are able to reproduce the covariance information associated with Earth's climate. It is also important that the models are able to capture changes (interannual variability and long-term trends) in the past if they are to be trusted for predicting changes in the future. The results of these evaluations can be found in the supporting material, but are not presented here in more detail as our main objective was to demonstrate how it is possible to downscale statistical properties on daily precipitation directly.

B4 Ensemble evaluation

An evaluation of downscaled ensemble results may include an assessment of whether the data follows a normal distribution, and rank statistics can be used to test whether the model results belong to the same statistical population as the observed target data. We tested the downscaled data both in terms of their rank statistics based on individual years as well as the ratio of observed to modelled standard deviations associated with their reproduction of the interannual variability. It is important that the downscaled results reproduce the typical interannual variability and historical trends for the selected locations.

Figure B3 shows an evaluation of the statistical distribution of the downscaled ensemble results and suggests that the ensemble results was close to being normally distributed for both f_w and μ. Hence, information about the ensemble can be approximated by the ensemble mean and ensemble standard deviation.

The average rank of annual respective f_w and μ from the observations from Oslo-Blindern was estimated over the 1951–2014 period in terms of the downscaled results (Figure B4). If the ensemble results belonged to the same statistical population as the observations, then this rank statistics should follow a uniform distribution. For f_w the mean rank was 0.49 and well within the range 0–1 (p-value of 0.49). The observed standard deviation for f_w was 1.33 times that of the ensemble for the overlapping historical simulations. Likewise, the mean rank for μ was 0.44 with a corresponding ratio in standard deviation of 1.41. Figure B4 shows the case for Oslo-Blindern as an example of how the downscaled ensemble can be assessed, and in this case the downscaled ensemble gave a slight underestimate of the magnitude of the interannual variability.

An evaluation of trends indicated ranges for both f_w and μ which spanned the observed trends at the 652 locations, but the ensembles underestimated the interannual variability for both f_w and μ (supporting material).

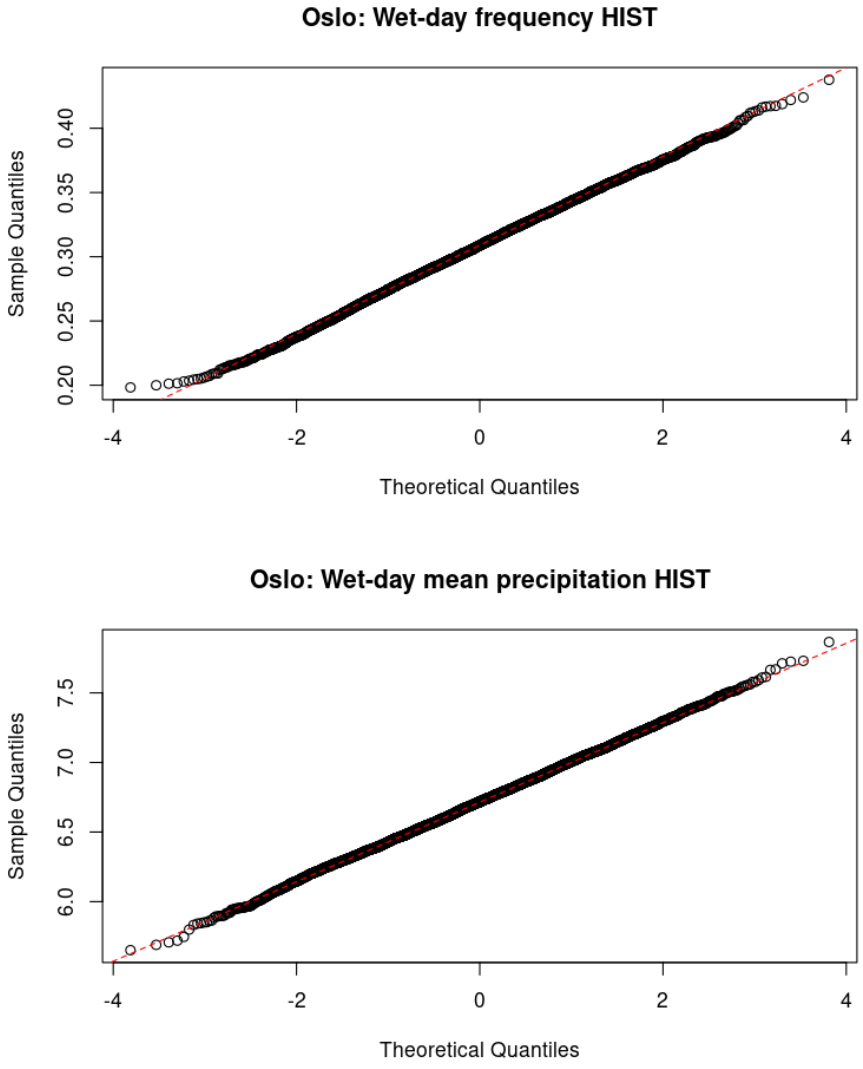

Figure B3. A comparison between the ensemble distribution (historical run) and the normal distribution for annual f_w (upper) and μ (lower) for their respective leading PCA. The near linear fit suggests that the distribution of the ensemble results is close to being normally distributed for the most important PCA mode.

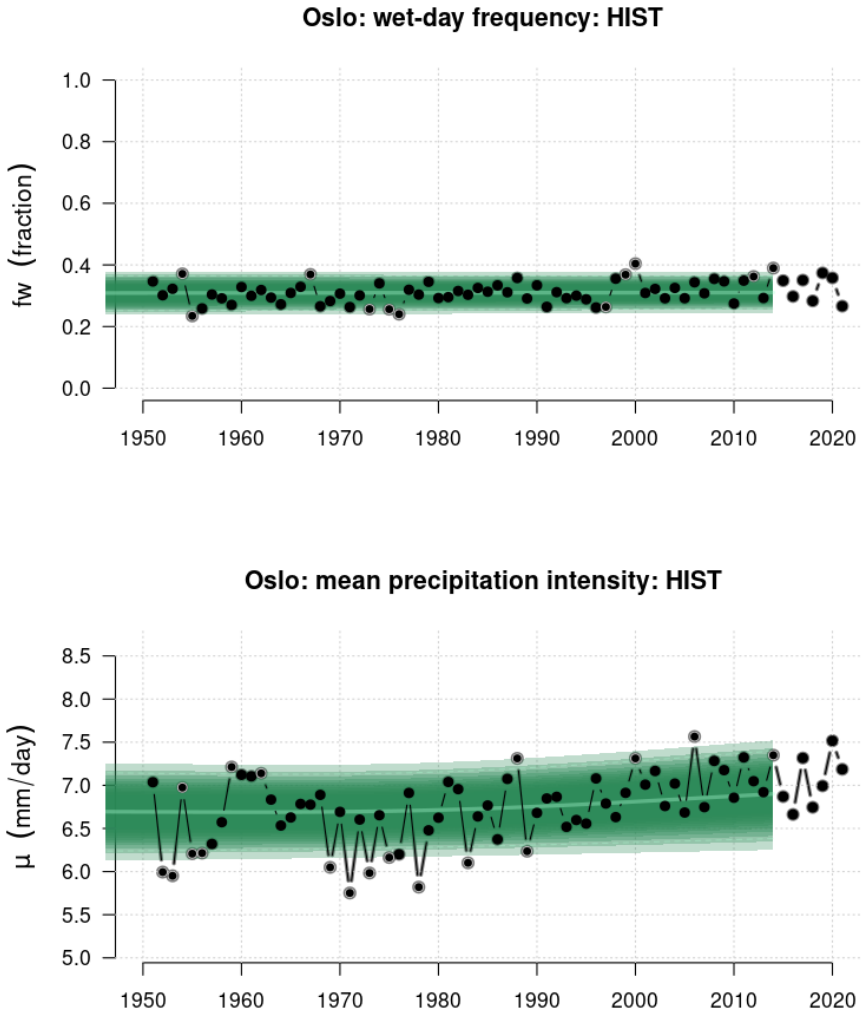

Figure B4. A comparison between the downscaled ensemble annual f_w (upper) and μ (lower) for Oslo and corresponding observations indicates that the model results reproduce both amplitude and long-term trends at a realistic level.

Appendix C: Projections of regional future precipitation statistics

We have in our analysis explored annually aggregated f_w and μ, but the presence of various meteorological phenomena tend to vary with the seasons and a mean annual trend may mask possible opposite trends in different regions. To assess this possibility we took a random sample from historical rain gauge measurements from Oslo and compared seasonal trends in both f_w and μ (supporting material). Our random test suggested that there were no pronounced opposite trends, but a more thorough exercise would entail downscaling seasonal mean precipitation statistics for the Nordic region. We leave the task of seasonal focus for the future, as a part of our objectives was to develop and evaluate downscaling approach for the EU-SPRINGS project and to provide the first projections for the planned national report 'Klima i Norge, 2100'. This strategy will also be explored in collaboration with Mozambique through CORDEX flagship pilot study (FPS) southeast Africa and the Norad-funded project SAREPTA[15]. This 'downscaling climate' approach for precipitation may work even if there is limited rain gauge data but it is important that reanalyses such as ERA5 correspond well with data on the ground.

[15]https://bistand.met.no/en/Sarepta

Author contributions. REB conceptualised, carried out most of the analysis, and drafted the paper; AD processed CMIP data and contributed to writing up the paper; KP has contributed to method development within the R-package `esd` and writing.

Competing interests. None

Disclaimer. The future projections are only as good as the GCMs on which they are based. They are the best information we have at the time of this analysis, are based on various assumptions, and there is always a risk that unaccounted for factors also play a role and may result in a different future climatic evolution.

Acknowledgements. This work aimed to benefit the following projects and networks: EU-SPRINGS (Project number 101137255 - HORIZON-HLTH-2023-ENVHLTH-02), Klima i Norge 2100 (Met Norway), SAREPTA (Norad), and CORDEX-ESD. The downscaling model and part of the analysis used the ECMWF fifth-generation ERA5 reanalysis hourly data downloaded from Copernicus C3S. We acknowledge the data providers in the ECA&D project by Klein Tank, A.M.G. and Coauthors, 2002. Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. Int. J. of Climatol., 22, 1441-1453. Data and metadata available at https://www.ecad.eu We acknowledge the CMIP6 community for providing the climate model data, retained and globally distributed in the framework of the ESGF. The CMIP6 data and server-side computing resources for this study were made available by the German Climate Computing Centre (DKRZ) under project ID 1088.

References

Barnett, T. P.: Comparison of Near-Surface Air Temperature Variability in 11 Coupled Global Climate Models, Journal of Climate, 12, 511–518, https://doi.org/10.1175/1520-0442(1999)012<0511:CONSAT>2.0.CO;2, 1999.

Benestad, R.: Downscaling Climate Information, vol. 1, Oxford University Press, https://doi.org/10.1093/acrefore/9780190228620.013.27, 2016.

Benestad, R.: Empirical-statistical downscaling of daily precipitation information in the Nordics, https://doi.org/https://doi.org/10.6084/m9.figshare.25809196.v2, 2024.

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., and Parding, K. M.: New vigour involving statisticians to overcome ensemble fatigue, Nature Climate Change, 7, 697–703, https://doi.org/10.1038/nclimate3393, 2017.

Benestad, R. E.: A comparison between two empirical downscaling strategies, Int. J. Climatology, 21, 1645–1668, dOI 10.1002/joc.703, 2001.

Benestad, R. E.: A new global set of downscaled temperature scenarios, Journal of Climate, 24, 2080–2098, https://doi.org/10.1175/2010JCLI3687.1, 2011.

Benestad, R. E., Hanssen-Bauer, I., and Chen, D.: Empirical-statistical downscaling, World Scientific, https://doi.org/10.1142/6908, 2008.

Benestad, R. E., Chen, D., Mezghani, A., Fan, L., and Parding, K.: On using principal components to represent stations in empirical-statistical downscaling, Tellus A, 67, https://doi.org/10.3402/tellusa.v67.28326, 2015a.

Benestad, R. E., Mezghani, A., and Parding, K. M.: esd V1.0, http://dx.doi.org/10.5281/zenodo.29385, 2015b.

Benestad, R. E., Senan, R., and Orsolini, Y.: The use of regression for assessing a seasonal forecast model experiment, Earth System Dynamics Discussions, pp. 1–20, https://doi.org/10.5194/esd-2016-14, 2016.

Benestad, R. E., Oort, B. v., Justino, F., Stordal, F., Parding, K. M., Mezghani, A., Erlandsen, H. B., Sillmann, J., and Pereira-Flores, M. E.: Downscaling probability of long heatwaves based on seasonal mean daily maximum temperatures, Advances in Statistical Climatology, Meteorology and Oceanography, 4, 37–52, https://doi.org/https://doi.org/10.5194/ascmo-4-37-2018, 2018.

Benestad, R. E., Parding, K. M., Erlandsen, H. B., and Mezghani, A.: A simple equation to study changes in rainfall statistics, Environmental Research Letters, 14, 084 017, https://doi.org/10.1088/1748-9326/ab2bb2, 2019.

Benestad, R. E., Lutz, J., Dyrrdal, A. V., Haugen, J. E., Parding, K. M., and Dobler, A.: Testing a simple formula for calculating approximate intensity-duration-frequency curves, Environmental Research Letters, https://doi.org/10.1088/1748-9326/abd4ab, 2020.

Benestad, R. E., Mezghani, A., Lutz, J., Dobler, A., Parding, K. M., and Landgren, O. A.: Various ways of using empirical orthogonal functions for climate model evaluation, Geoscientific Model Development, 16, 2899–2913, https://doi.org/10.5194/gmd-16-2899-2023, 2023.

Benestad, R. E., Lussana, C., and Dobler, A.: A link between the global surface area receiving daily precipitation, wet-day frequency and probability of extreme rainfall, Discover Water, 4, 10, https://doi.org/10.1007/s43832-024-00063-3, 2024.

Christensen, J. H. and Christensen, O. B.: 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, 2007.

Christensen, J. H., Carter, T. R., Rummukainen, M., and Amanatidis, G.: Evaluating the performance and utility of regional climate models: the PRUDENCE project, Climatic Change, 81, 1–6, https://doi.org/10.1007/s10584-006-9211-6, 2007.

Coles, S. G.: An Introduction to Statistical Modeling of Extreme Values, Springer, London, 2001.

Deser, C., Knutti, R., Solomon, S., and Phillips, A. S.: Communication of the role of natural variability in future North American climate, Nature Climate Change, 2, 775–779, https://doi.org/doi:10.1038/nclimate1562, 2012.

Deser, C., Lehner, F., Rodgers, K. B., Ault, T., Delworth, T. L., DiNezio, P. N., Fiore, A., Frankignoul, C., Fyfe, J. C., Horton, D. E., Kay, J. E., Knutti, R., Lovenduski, N. S., Marotzke, J., McKinnon, K. A., Minobe, S., Randerson, J., Screen, J. A., Simpson, I. R., and Ting, M.: Insights from Earth system model initial-condition large ensembles and future prospects, Nature Climate Change, 10, 277–286, https://doi.org/10.1038/s41558-020-0731-2, 2020.

Eyring, V., Bony, S., Meehl, G. A., Senior, C., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organisation, Geoscientific Model Development Discussions, 8, 10 539–10 583, https://doi.org/10.5194/gmdd-8-10539-2015, 2015.

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

Goodess, C., Osborn, T., and Hulme, M.: The identification and evaluation of suitable scenario development methods for the estimation of future probabilities of extreme weather events, Technical Report 4, Tyndall Centre, School of Environemtal Sciences, Univ. East Anglia, Norwich, 2003.

Gutiérrez, J. M., Maraun, D., Widmann, M., Huth, R., Hertig, E., Benestad, R., Roessler, O., Wibig, J., Wilcke, R., Kotlarski, S., Martín, D. S., Herrera, S., Bedia, J., Casanueva, A., Manzanas, R., Iturbide, M., Vrac, M., Dubrovsky, M., Ribalaygua, J., Pórtoles, J., Räty, O., Räisänen, J., Hingray, B., Raynaud, D., Casado, M. J., Ramos, P., Zerenner, T., Turco, M., Bosshard, T., Štěpánek, P., Bartholy, J., Pongracz, R., Keller, D. E., Fischer, A. M., Cardoso, R. M., Soares, P. M. M., Czernecki, B., and Pagé, C.: An intercomparison of a large ensemble of statistical downscaling methods over Europe: Results from the VALUE perfect predictor cross-validation experiment, International Journal of Climatology, https://doi.org/10.1002/joc.5462, 2018.

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., and Tangang, F.: WCRP COordinated Regional Downscaling EXperiment (CORDEX): a diagnostic MIP for CMIP6, Geoscientific Model Development, 9, 4087–4095, https://doi.org/10.5194/gmd-9-4087-2016, 2016.

Hawkins, E. and Sutton, R.: The potential to narrow uncertainty in regional climate predictions, Bull. Amer. Meteor. Soc., 90, p1095, 2009.

Haylock, M. R., Cawley, G. C., Harpham, C., Wilby, R. L., and Goodess, C. M.: Downscaling Heavy Precipitation Over the United Kingdom: A Comparison of Dynamical and Statistical Methods and their Future Scenarios, International Journal of Climatology, 26, 1397–1416, 2006.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horanyi, A., Munoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Holm, E., Janiskova, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thepaut, J.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

IPCC: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Tech. rep., Cambridge University Press, 2021.

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., Domínguez, M., Fealy, R., Fernandez, J., Gaertner, M. A., García-Díez, M., Giorgi, F., Gobiet, A., Goergen, K., Gómez-Navarro, J. J., Alemán, J. J. G., Gutiérrez, C., Gutiérrez, J. M., Güttler, 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., Montávez, J. P., Navarra, A., Nikulin, G., de Noblet-Ducoudré, N., Panitz, H.-J., Pfeifer, S., Piazza, M., Pichelli, E., Pietikäinen, J.-P., Prein, A. F., Preuschmann, S., Rechid, D., Rockel, B., Romera, R., Sánchez, E., Sieck, K., Soares, P. M. M., Somot, S., Srnec, L., Sørland, S. L., Termonia, P., Truhetz, H., Vautard, R., Warrach-Sagi, K., and Wulfmeyer, V.: Regional climate downscaling over Europe: perspectives from the EURO-CORDEX community, Regional Environmental Change, 20, https://doi.org/10.1007/s10113-020-01606-9, 2020.

Klein Tank, A. J. B. W., Konnen, G. P., Böhm, R., Demarée, G., Gocheva, A., Mileta, M., Pashiardis, S., Hejkrlik, L., Kern-Hansen, C., Heino, R., Bessemoulin, P., Müller-Westermeier, G., Tzanakou, M., Szalai, S., Pálsdóttir, T., Fitzgerald, D., Rubin, S., Capaldo, M., Maugeri, M., Leitass, A., Bukantis, A., Aberfeld, R., Engelen, A. F. V. v., Førland, E., Mietus, M., Coelho, F., Mares, C., Razuvaev, V., Nieplova, E., Cegnar, T., López, J. A., Dahlström, B., Moberg, A., Kirchhofer, W., Ceylan, A., Pachaliuk, O., Alexander, L. V., and Petrovic, P.: Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment, International Journal of Climatology, 22, 1441–1453, 2002.

Lorenz, E. N.: Empirical Orthogonal Functions and Statistical Weather Prediction, Sci. rep. 1, Department of Meteorology, MIT, USA, Cambridge, Massachusetts, https://eapsweb.mit.edu/sites/default/files/Empirical_Orthogonal_Functions_1956.pdf, 1956.

Lussana, C., Benestad, R., and Dobler, A.: Changes in regional daily precipitation intensity and spatial structure from global reanalyses, International Journal of Climatology, 44, 1135–1153, https://doi.org/10.1002/joc.8375, 2024.

Maraun, D., Widmann, M., Gutiérrez, J. M., Kotlarski, S., Chandler, R. E., Hertig, E., Wibig, J., Huth, R., and Wilcke, R. A.: VALUE: A framework to validate downscaling approaches for climate change studies, Earth's Future, 3, 2014EF000259, https://doi.org/10.1002/2014EF000259, 2015.

Mezghani, A., Dobler, A., Haugen, J. E., Benestad, R. E., Parding, K. M., Piniewski, M., Kardel, I., and Kundzewicz, Z. W.: CHASE-PL Climate Projection dataset over Poland – bias adjustment of EURO-CORDEX simulations, Earth System Science Data, 9, 905–925, https://doi.org/https://doi.org/10.5194/essd-9-905-2017, 2017.

Mezghani, A., Dobler, A., Benestad, R., Haugen, J. E., Parding, K. M., Piniewski, M., and Kundzewicz, Z. W.: Sub-sampling impact on the climate change signal over Poland based on simulations from statistical and dynamical downscaling, Journal of Applied Meteorology and Climatology, https://doi.org/10.1175/JAMC-D-18-0179.1, 2019.

Navarra, A. and Simoncini, V.: A guide to empirical orthogonal functions for climate data analysis, Springer, Dordrecht ; New York, oCLC: ocn462919781, 2010.

North, G. R., Bell, T. L., and Cahalan, R. F.: Sampling Errors in the Estimation of Empirical Orthogonal Functions, Monthly Weather Review, 110, 699–706, 1982.

Nychka, D., Hammerling, D., Sain, S., and Lenssen, N.: LatticeKrig: Multiresolution Kriging Based on Markov Random Fields, https://doi.org/10.5065/D6HD7T1R, place: Boulder, CO, USA, 2016.

Oguz, E. A., Benestad, R. E., Parding, K. M., Depina, I., and Thakur, V.: Quantification of climate change impact on rainfall-induced shallow landslide susceptibility: a case study in central Norway, Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards, pp. 1–24, https://doi.org/10.1080/17499518.2023.2283848, 2024.

Parding, K. M., Benestad, R., Mezghani, A., and Erlandsen, H. B.: Statistical Projection of the North Atlantic Storm Tracks, Journal of Applied Meteorology and Climatology, 58, 1509–1522, https://doi.org/10.1175/JAMC-D-17-0348.1, 2019.

Parding, K. M., Benestad, R. E., Dyrrdal, A. V., and Lutz, J.: A principal-component-based strategy for regionalisation of precipitation intensity–duration–frequency (IDF) statistics, Hydrology and Earth System Sciences, 27, 3719–3732, https://doi.org/10.5194/hess-27-3719-2023, 2023.

Preisendorfer, R. W.: Principal Component Analysis in Meteorology and Oceanology, Elsevier Science Press, Amsterdam, 1988.

Pryor, S., School, J. T., and Barthelmie, R. J.: Empirical downscaling of wind speed probability distributions, Journal of Geophysical Research, 110, D19 109, https://doi.org/doi:10.1029/2005JD005899, 2005.

Pryor, S., School, J. T., and Barthelmie, R. J.: Winds of change? Projections of near-surface winds under climate change scenarios, Geophys. Res. Lett., 33, 2006.

Rampal, N., Hobeichi, S., Gibson, P. B., Baño-Medina, J., Abramowitz, G., Beucler, T., González-Abad, J., Chapman, W., Harder, P., and Gutiérrez, J. M.: Enhancing Regional Climate Downscaling through Advances in Machine Learning, Artificial Intelligence for the Earth Systems, 3, 230 066, https://doi.org/10.1175/AIES-D-23-0066.1, 2024.

Schulzweida, U.: CDO User Guide: Climate Data Operator, Version 2.0.0, October 2021, Tech. rep., MPI for Meteorology, https://code.mpimet.mpg.de/projects/cdo/embedded/cdo.pdf, 2021.

Strang, G.: Linear Algebra and its Application, Harcourt Brace & Company, San Diego, California, USA, 1988.

Takayabu, I., Kanamaru, H., Dairaku, K., Benestad, R., Storch, H. v., and Christensen, J. H.: Reconsidering the quality and utility of downscaling, Journal of the Meteorological Society of Japan, 94A, 31–45, https://doi.org/10.2151/jmsj.2015-042, 2015.

Trenberth, K. E., Dai, A., Rasmussen, R. M., and Parsons, D. B.: The Changing Character of Precipitation, Bulletin of the American Meteorological Society, 84, 1205–1218, https://doi.org/10.1175/BAMS-84-9-1205, 2003.

Von Storch, H., Zorita, E., and Cubasch, U.: Downscaling of Global Climate Change Estimates to Regional Scales: An Application to Iberian Rainfall in Wintertime, Journal of Climate, 6, 1161–1171, https://doi.org/10.1175/1520-0442(1993)006<1161:DOGCCE>2.0.CO;2, 1993.

Wallace, J. and Dickinson, R. E.: Empirical orthogonal representation of time series in the frequency domain: I. Theoretical considerations, J. App. Meteor., 11, 1972.

Wilby, R., Dawson, C., and Barrow, E.: sdsm — a decision support tool for the assessment of regional climate change impacts, Environmental Modelling & Software, 17, 145–157, https://doi.org/10.1016/S1364-8152(01)00060-3, 2002.

Wilks, D. S.: Statistical methods in the atmospheric sciences, no. v. 91 in International geophysics series, Academic Press, Amsterdam ; Boston, 2nd ed edn., 2006.

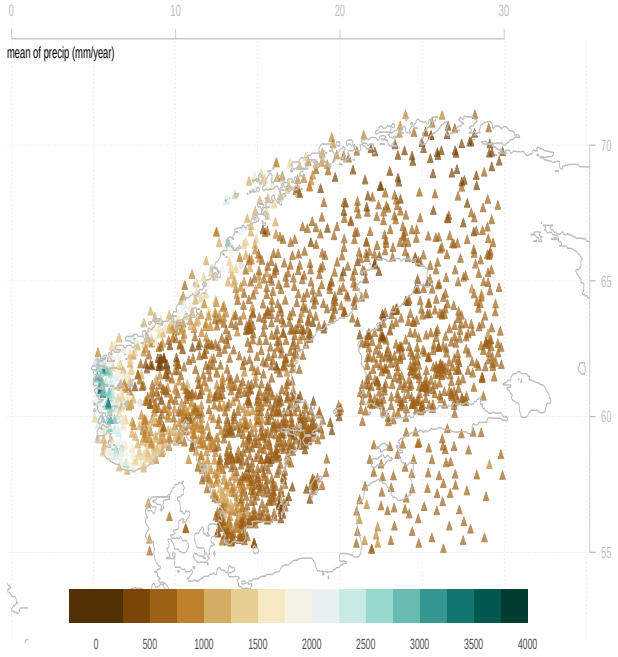

Figure 1. Map showing the rain gauge station network from ECA&D used as predictands in the empirical-statistical downscaling of 24-hr precipitation statistics. The colour legend shows the mean annual total precipitation.

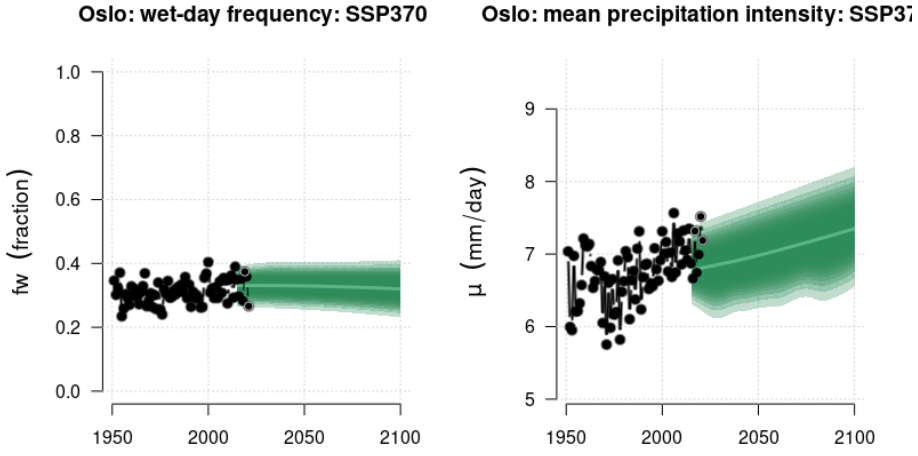

Figure 2. Ensembles of downscaled wet-day frequency f_w and wet-day mean precipitation μ for Oslo based on the SSP370 emission scenario. Black symbols show annual such aggregated statistics estimated from rain gauge measurements from Oslo-Blindern and the green shading marks the ensemble spread of corresponding downscaled results.

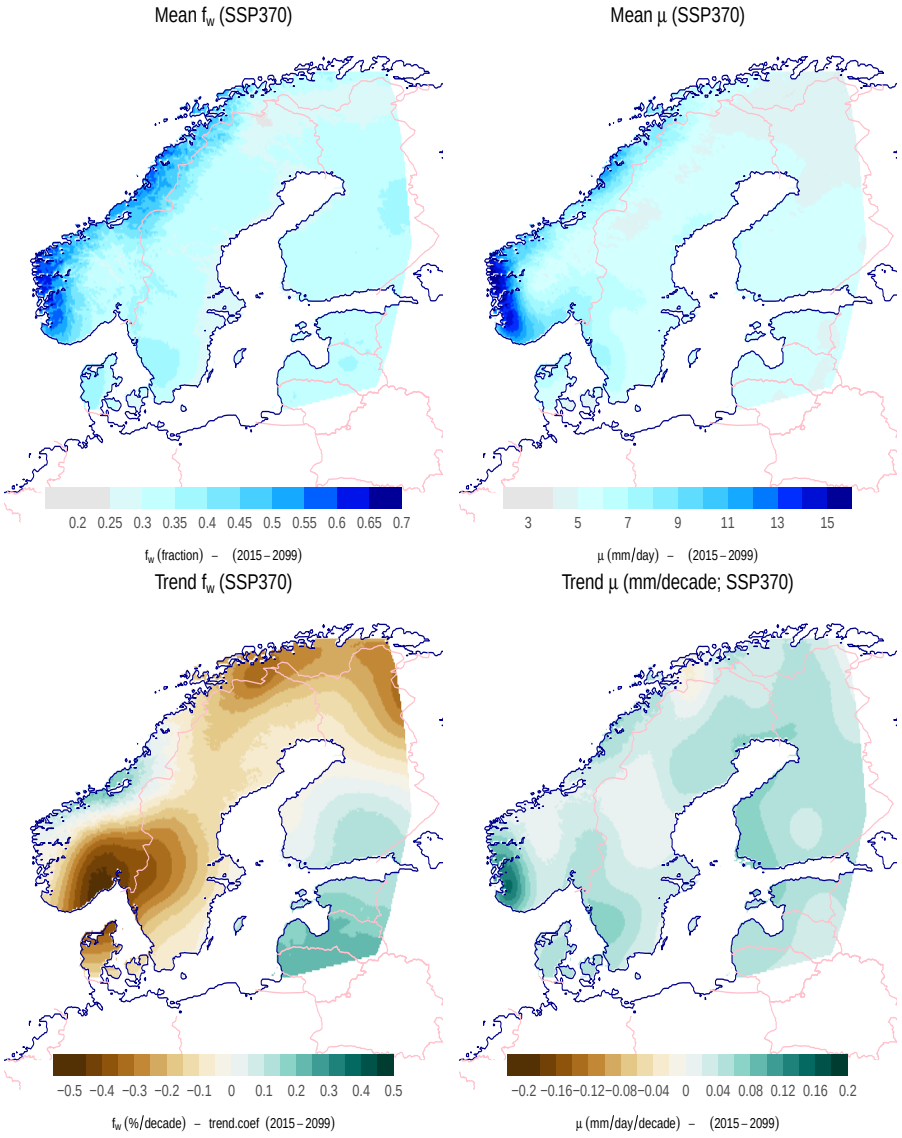

Figure 3. Maps of downscaled mean f_w (upper left) and μ (upper right) as well as trend estimates (lower).

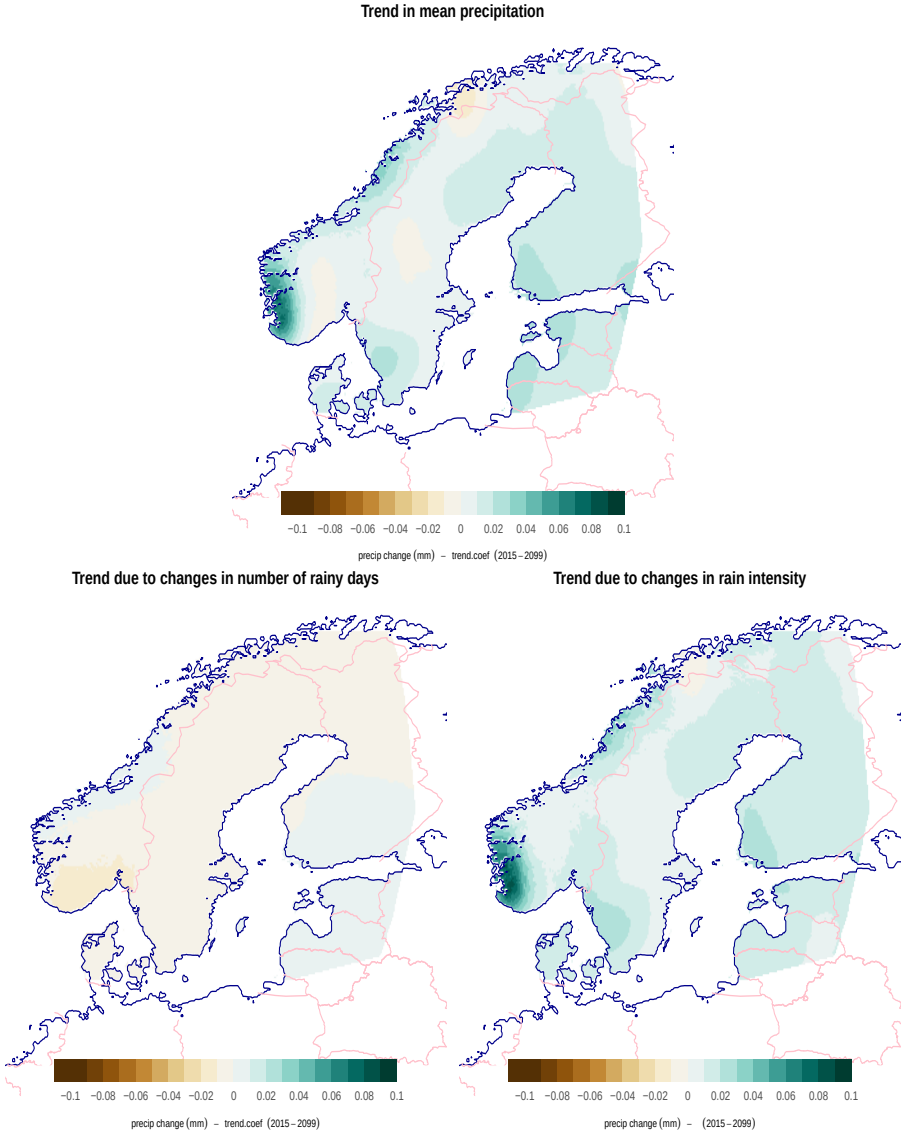

Figure 4. Estimated trend in mean precipitation $\overline{x'} = f_w \mu$ (upper) and the contribution due to wet-days f_w (lower left) and mean intensity μ (lower right).

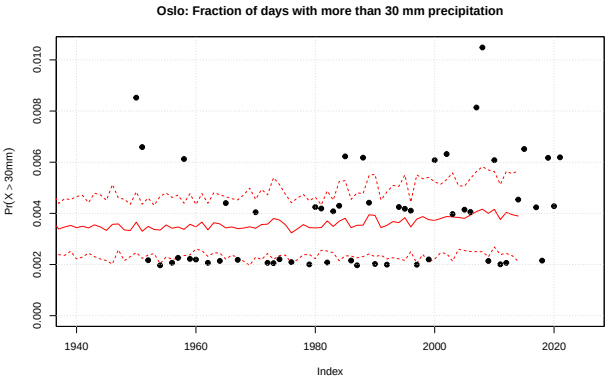

Figure 5. Observed and estimated fraction of days per year in Oslo with more than 30 mm. Solid line shows the ensemble mean and dashed lines the ensemble mean plus or minus the ensemble standard deviation.

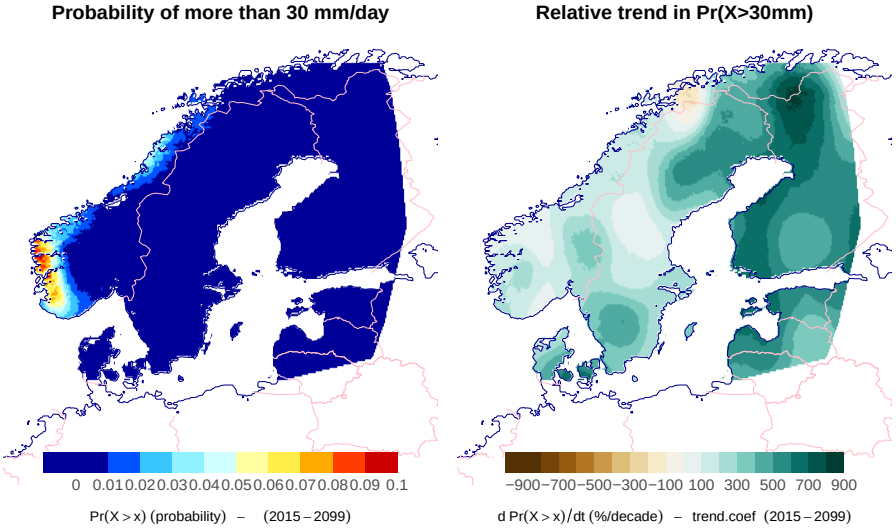

Figure 6. Estimates of the mean probability of more than 30 mm precipitation in 24 hours according to $Pr(X' > x') = f_w \exp(-x'/\mu)$ (left), and the proportional trend in the probability estimated using the product rule (right). These results are based on downscaled f_w and μ from the CMIP6 ensemble following the SSP370 emission scenario.

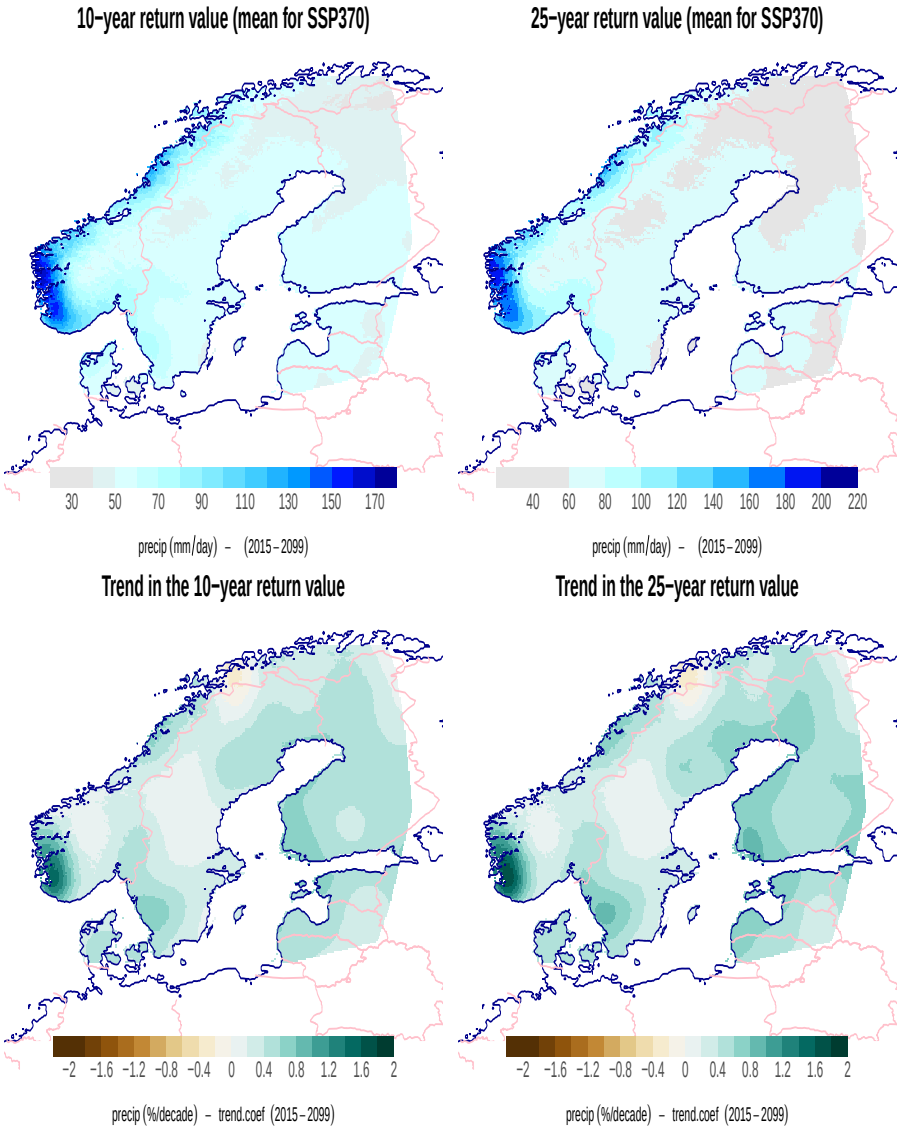

Figure 7. Estimates of the 10-year and 25-year return-values based on the expression $x'_\tau = \alpha\mu\ln(f_w\tau)$ (Benestad et al., 2019), and their future trend estimates (lower). The results are based on the SSP370 emission scenario and the CMIP6 ensemble mean downscaled f_w (right) and μ.

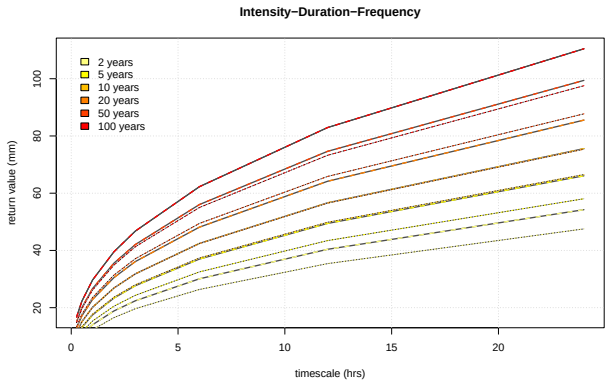

Figure 8. Estimate of intensity-duration-frequency curves for Oslo-Blindern based on downscaled f_w and μ (thin solid-dotted) and their future trend estimates (thick solid-dashed). These results are based on the SSP370 emission scenario and the expression $x'_\tau(L) = \alpha\mu(L/24)^\zeta \ln(f_w\tau)$ (Benestad et al., 2020).

Table 1. The ensemble mean and standard deviation of the wet-day frequency f_w and wet-day mean precipitation μ projected for 2071–2100 for a selection of locations.

Location	emission scenario	$\overline{f_w} \pm \sigma_f$	$\overline{\mu} \pm \sigma_\mu$
Geiranger	SSP370	0.44 ± 0.06	9.24 ± 0.78
	SSP126	0.44 ± 0.06	9.1 ± 0.78
	SSP245	0.44 ± 0.06	9.12 ± 0.8
	SSP585	0.44 ± 0.05	9.28 ± 0.9
Halden	SSP370	0.34 ± 0.05	7.18 ± 0.43
	SSP126	0.34 ± 0.05	7.06 ± 0.35
	SSP245	0.34 ± 0.05	7.16 ± 0.46
	SSP585	0.33 ± 0.05	7.24 ± 0.44
Helsinki	SSP370	0.32 ± 0.04	5.89 ± 0.34
	SSP126	0.32 ± 0.04	5.62 ± 0.28
	SSP245	0.32 ± 0.04	5.75 ± 0.36
	SSP585	0.31 ± 0.04	6.05 ± 0.43
Malmö	SSP370	0.3 ± 0.02	5.72 ± 0.25
	SSP126	0.31 ± 0.02	5.47 ± 0.22
	SSP245	0.3 ± 0.03	5.6 ± 0.29
	SSP585	0.3 ± 0.03	5.86 ± 0.34
Oslo	SSP370	0.32 ± 0.04	7.24 ± 0.41
	SSP126	0.32 ± 0.04	6.99 ± 0.35
	SSP245	0.32 ± 0.04	7.15 ± 0.45
	SSP585	0.32 ± 0.04	7.38 ± 0.47
Stockholm	SSP370	0.28 ± 0.03	5.31 ± 0.21
	SSP126	0.29 ± 0.03	5.15 ± 0.17
	SSP245	0.29 ± 0.03	5.24 ± 0.22
	SSP585	0.28 ± 0.03	5.39 ± 0.27
Tallinn	SSP370	0.34 ± 0.04	5.67 ± 0.31
	SSP126	0.35 ± 0.04	5.39 ± 0.28
	SSP245	0.34 ± 0.04	5.52 ± 0.36
	SSP585	0.33 ± 0.04	5.82 ± 0.4
Vestervig	SSP370	0.37 ± 0.04	6 ± 0.18
	SSP126	0.37 ± 0.04	6.01 ± 0.16
	SSP245	0.37 ± 0.04	6.01 ± 0.18
	SSP585	0.36 ± 0.05	5.99 ± 0.2