ARTICLE   OPEN

Check for updates

Using UNSEEN trends to detect decadal changes in 100-year precipitation extremes

T. Kelder [1,2 ✉], M. Müller [2,3], L. J. Slater [4], T. I. Marjoribanks [5], R. L. Wilby [1], C. Prudhomme [6,7], P. Bohlinger [2], L. Ferranti [6] and T. Nipen [2]

Sample sizes of observed climate extremes are typically too small to reliably constrain return period estimates when there is non-stationary behaviour. To increase the historical record 100-fold, we apply the UNprecedented Simulated Extreme ENsemble (UNSEEN) approach, by pooling ensemble members and lead times from the ECMWF seasonal prediction system SEAS5. We fit the GEV distribution to the UNSEEN ensemble with a time covariate to facilitate detection of changes in 100-year precipitation values over a period of 35 years (1981–2015). Applying UNSEEN trends to 3-day precipitation extremes over Western Norway substantially reduces uncertainties compared to estimates based on the observed record and returns no significant linear trend over time. For Svalbard, UNSEEN trends suggests there is a significant rise in precipitation extremes, such that the 100-year event estimated in 1981 occurs with a return period of around 40 years in 2015. We propose a suite of methods to evaluate UNSEEN and highlight paths for further developing UNSEEN trends to investigate non-stationarities in climate extremes.

npj Climate and Atmospheric Science (2020)3:47 ; https://doi.org/10.1038/s41612-020-00149-4

INTRODUCTION

Handling the non-stationarity of climate extremes is an active area of research[1–3] that is confounded by the brevity and sparsity of observational records[4–6]. Non-stationary precipitation analyses typically focus on detecting multidecadal to centennial changes in annual precipitation maxima[7–9]. However, annual maximum precipitation events do not necessarily cause high impacts and hence, a more pressing research challenge is the detection of changes in larger extremes[10,11], such as the 1-in-100-year event. Furthermore, the impacts of abrupt warming in recent decades may not yet be detectable in short precipitation records. Therefore, robust detection of short-term (decadal, rather than centennial) trends in climate extremes may provide actionable information.

An emerging alternative to traditional observation-based extreme value and weather generator analyses is to pool ensemble members from numerical weather prediction systems[12–22]—the UNprecedented Simulated Extreme ENsemble (UNSEEN) approach[15,16]. This technique creates numerous alternative pathways of reality, thus increasing the event sample size for statistical analysis. The larger sample size offers a broader view of present-day hazards and, therefore, has potential to improve design-levels. For example, the 2013/14 winter flooding in the UK had no observational precedent, but could have been anticipated with the UNSEEN approach[15]. Similarly, estimates of storm surge levels of the River Rhine[12,13], global ocean wind and wave extremes[17–19], and losses from extreme windstorms[21] have all been improved with the UNSEEN approach. UNSEEN can also enhance food security through better drought exposure estimates[14,22] and can assist policy makers and contingency planners by quantifying and explaining the most severe events possible in the current climate, such as heatwaves in China[16]. However, validating the UNSEEN method is a well-recognised challenge for

existing studies, and UNSEEN has not yet been used to facilitate detection of non-stationarity in climate extremes over short periods of a few decades.

Here, we provide a framework to systematically evaluate the robustness of the UNSEEN method and present the UNSEEN trends approach, where we provide more confident short-term trend estimates from the larger event sample to better constrain changes in climate extremes. We do this in a storyline context[23], taking observed flood episodes as a starting point for our analysis. We select the west coast of Norway and the Svalbard Archipelago as demonstration regions; two contrasting areas in terms of precipitation extremes. Western Norway experiences the highest precipitation extremes within Europe[24] and has a dense station network[25,26], whereas Svalbard is a semi-desert with very few meteorological stations[27], where climate change is expected to increase the frequency of precipitation extremes[27,28]. Both regions endured severe damages from recent extreme events, such as the September 2005[29] and October 2014[30] floods over Western Norway and the slush-avalanche induced by extreme precipitation over the Svalbard Archipelago in 2012[28]. These extreme events were driven by atmospheric rivers[29,31] (ARs), which cause heavy precipitation over a prolonged period. As AR-related floods predominantly occur in autumn and frequently strengthen over a period of several days[29,30], we select autumn (September to November) spatially averaged 3-day extreme precipitation (SON-3DP) as target events (Fig. 1).

Previous UNSEEN studies used the Hadley Centre global climate model, HadGEM3-GC2[14–16,22] and the European Centre for Medium-range Weather Forecasts (ECMWF) ensemble prediction systems[17–20] and earlier versions of the seasonal prediction system[12,13,21]. Here, we are the first to use the latest ECMWF seasonal prediction system SEAS5[32] with an open access, high-resolution, large ensemble, multidecadal, homogeneous hindcast

[1]Geography and Environment, Loughborough University, Loughborough, UK. [2]The Norwegian Meteorological Institute, Oslo, Norway. [3]Section for Meteorology and Oceanography, Department of Geosciences, University of Oslo, Oslo, Norway. [4]School of Geography and the Environment, University of Oxford, Oxford, UK. [5]School of Architecture, Building and Civil Engineering, Loughborough University, Loughborough, UK. [6]European Centre for Medium-Range Weather Forecasts (ECMWF), Reading, UK. [7]Centre for Ecology and Hydrology, Wallingford, UK. ✉email: T.Kelder@lboro.ac.uk
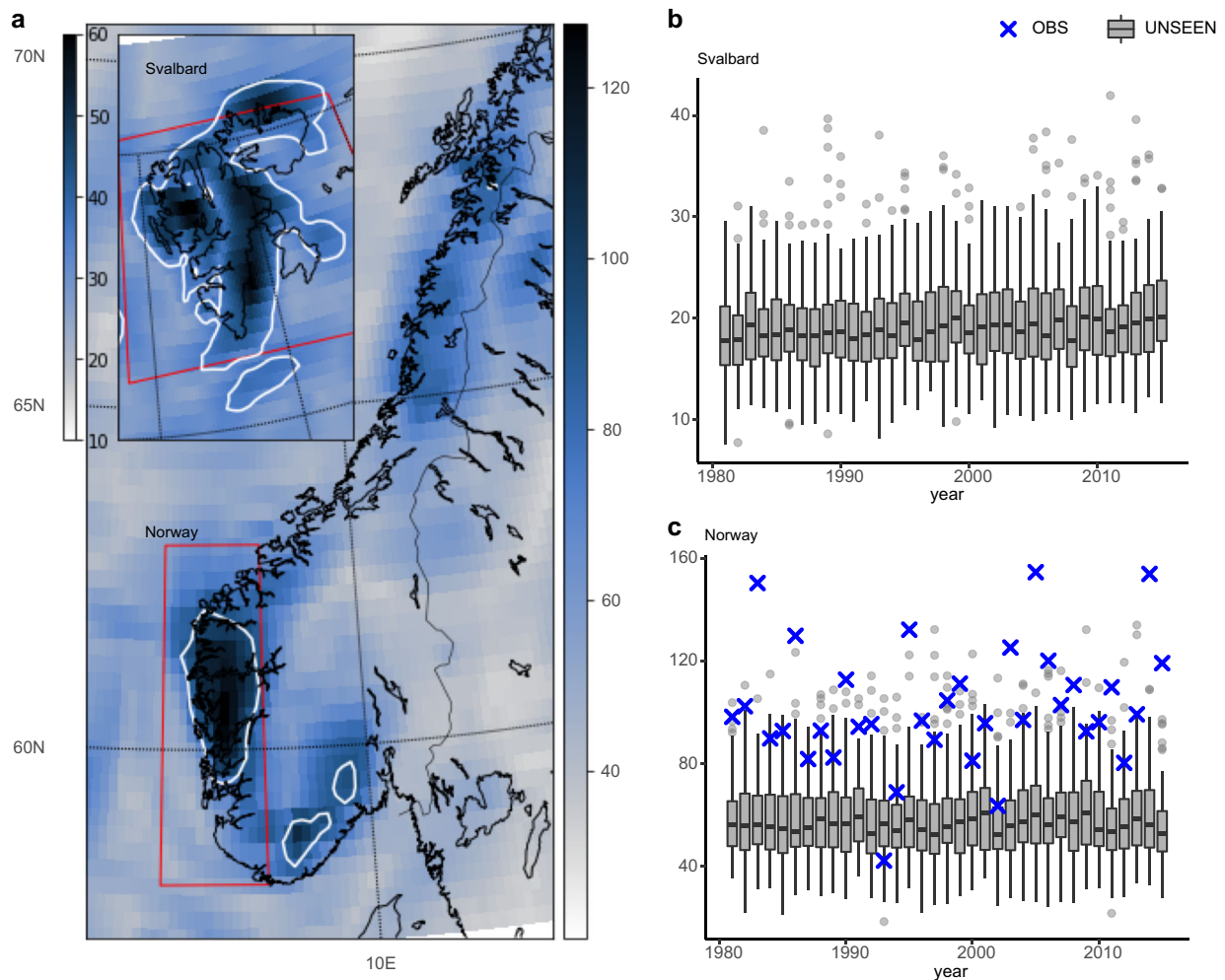
Fig. 1 Precipitation extremes over the study area. a Climatological values of autumn 3-day cumulative precipitation over the Norwegian and Svalbard region. Blue shades indicate the empirical 200-year precipitation values calculated from all August-initialised forecasts over 1981–2015. Two domains are indicated with red rectangles: a Norwegian west coast domain (WC, 4–7 E, 58–63N) and a Svalbard domain (SV, 8–30 E,76–80N). White contours denote the regions with high climatological values over which the precipitation is averaged within the domains: 90 mm for WC and 35 mm for SV. This restricts the domain to a homogeneous region with high precipitation. **b, c** Extracted seasonal maximum 3-day precipitation timeseries for Svalbard (**b**) and Norway (**c**). The grey boxplots show the UNSEEN ensemble (100 ensembles per year, illustrated as the median, interquartile range, 1.5x interquartile range and outliers) and the blue crosses show the events in the observed record. Note, for Svalbard no gridded precipitation record is available.

period (1981–2015). The ECMWF atmospheric model has skill in simulating ARs for Northern Europe[33], giving confidence in the realism of these extreme events in SEAS5, hence is a good candidate for the UNSEEN method. We use the 25-member ensemble across lead times of 2–5 months, resulting in a sample of 100 members (called the UNSEEN ensemble) and evaluate the independence and stability of the pooled sample for SON-3DP events across Western Norway and Svalbard. We then use the UNSEEN trends approach to identify unprecedented extreme precipitation events and to detect trends in 100-year precipitation events over a 35-year period. These findings give insights about the robustness of current design standards and may eventually advance understanding of physical processes driving climate extremes and their non-stationarity.

RESULTS

Ensemble member independence and model stability

Independence of ensemble members is an important prerequisite for the UNSEEN approach, as dependent members would

artificially inflate the sample size, without adding new information. Previous studies have assessed the independence of ensemble members for lead times of 9–10 days[17–19], but as far as we are aware, no independence test has yet been performed in UNSEEN studies of seasonal prediction systems.

For the regions studied here, ensemble members with lead times beyond one month are not dependent on atmospheric initial conditions, because the synoptic patterns related to ARs are known to be unpredictable beyond two weeks[33,34]. However, predictability on a seasonal timescale may be found through slowly varying components of the ocean-atmosphere system. Therefore, although ensemble members might represent unique weather events because of the disconnect from initial atmospheric conditions, these could have a conditional bias due to favourable conditions in the slowly varying components of the ocean-atmosphere system.

To test the seasonal dependence of SON-3DP, we first select the seasonal maximum event for each forecast then concatenate these events to create a 35-year timeseries (Fig. 2a, b, c). To robustly assess the independence between each of the ensemble
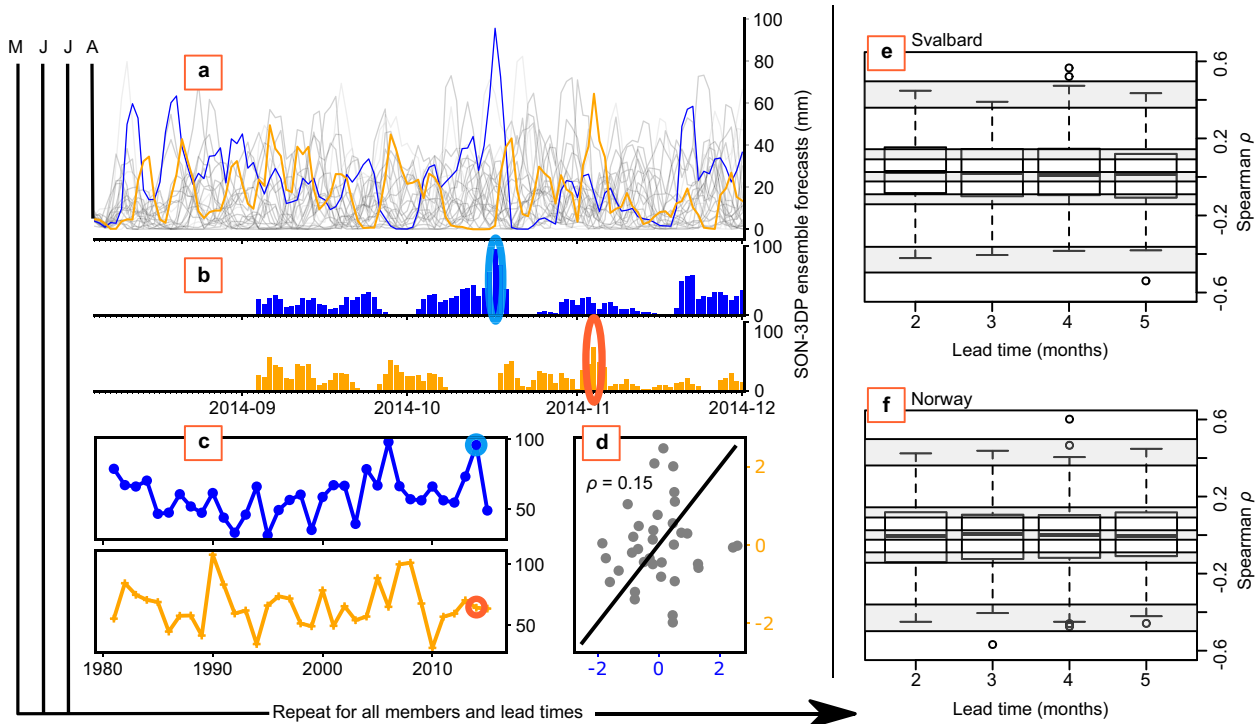
Fig. 2 A workflow for analysing ensemble member dependence. a August 2014 initialised 25-member seasonal forecasts of 3-day precipitation time series over the SON forecast horizon. Ensemble members 0 and 1 are shown in blue and orange, respectively. **b** From the forecast members 0 and 1, the September-November (SON) maximum value for the 2014 season is selected. **c** A series of the maximum 3-day precipitation values for the SON season for each year in the hindcast record is created for member 0 and member 1. The 2014 maximum, as illustrated in b, is encircled. **d** The standardised anomaly of the maximum 3-day precipitation series for the two members are correlated. Spearman's rho correlation is shown and the one-to-one line is indicated in black. This process is repeated for the 300 distinct ensemble member pairings for each of the four lead times (May–August). **e, f** Boxplots of the resulting 300 Spearman's rho correlations for each lead time over Svalbard (**e**) and Norway (**f**). Grey shading shows the confidence intervals of the boxplot statistics (whiskers: 1.5x interquartile range, box limits: interquartile range and centre line: median), based on a permutation test with 5% significance level.

members, we calculate the Spearman rank correlation coefficient (ρ) for every distinct pair of ensemble members (Fig. 2d), resulting in 300 ρ values for each lead time. The value of ρ ranges from ca. −0.6 to 0.6, and the median correlation is close to zero for all lead times for both Western Norway and Svalbard (Fig. 2e, f). A wide range in ρ values is expected due to chance from the large number of correlation tests, and none of the lead times fall outside the range that would be expected for uncorrelated data for the West Coast of Norway (Fig. 2f). For Svalbard, slightly higher ρ values are found, with the median correlation still within the expected range, but the interquartile range just exceeds the upper boundary of the confidence intervals for the first two lead times (Fig. 2e). The small correlations found for Svalbard might be driven by the trend that we detect for this region (UNSEEN-trends section), and thus, the UNSEEN ensemble members represent unique events that follow the slowly evolving climate signal, as desired.

A second potential issue for generating the UNSEEN ensemble could be a drift in the simulated climatology[35,36], which may alter precipitation extremes over longer lead times. Therefore, model stability is a requirement for pooling lead times. Model stability is assessed by comparing the distribution of predicted SON-3DP events across different lead times. For both regions, the probability density functions of the pooled SON-3DP events for the considered lead times are remarkably similar (Fig. 3a, b). Moreover, the empirical extreme value distributions of the individual lead times fall within the uncertainty range of the distribution of all lead times pooled together and thus, the model can be considered stable over lead times (Fig. 3c, d).

Fidelity of UNSEEN extremes

Confidence in simulated 'unprecedented extremes' in large ensembles is complicated by the inability to validate extremes, given the limited sample sizes of observations. Here, we evaluate the UNSEEN ensemble by bootstrapping the ensemble into datasets of 35 years and assessing whether observations fall within the range of the bootstrapped distribution, following previous UNSEEN studies[15,16] (see 'Methods'). We perform this analysis for the SEAS5 UNSEEN SON-3DP ensemble over Western Norway and Svalbard (domains in Fig. 1). As evaluation products, we use 1. SeNorge: a gridded precipitation product produced from the dense station network of Norway[25,26] and 2. The ERA5 reanalysis[37], which is more readily comparable to SEAS5 because they are both model-based products with the same resolution. Furthermore, ERA5 facilitates evaluation over Svalbard, where no gridded precipitation record exists. Note, the SeNorge high-resolution station-based observational dataset (1 km) is upscaled to the resolution of SEAS5 (36 km) to reduce the spatial-scale mismatch between forecasts and interpolated observations[38,39]. For a comprehensive global model validation of SEAS5, see Johnson et al.[32].

In this study, we use SEAS5 simulated large-scale precipitation (LSP) to represent precipitation extremes over the study regions (see Methods). This is because convective processes cannot be resolved in a global model of 36 km resolution[40] and because 3-day extremes over the domains are known to be driven by large-scale ARs[27,29,30]. Owing to this design decision, a discrepancy may exist between the simulated large-scale precipitation and
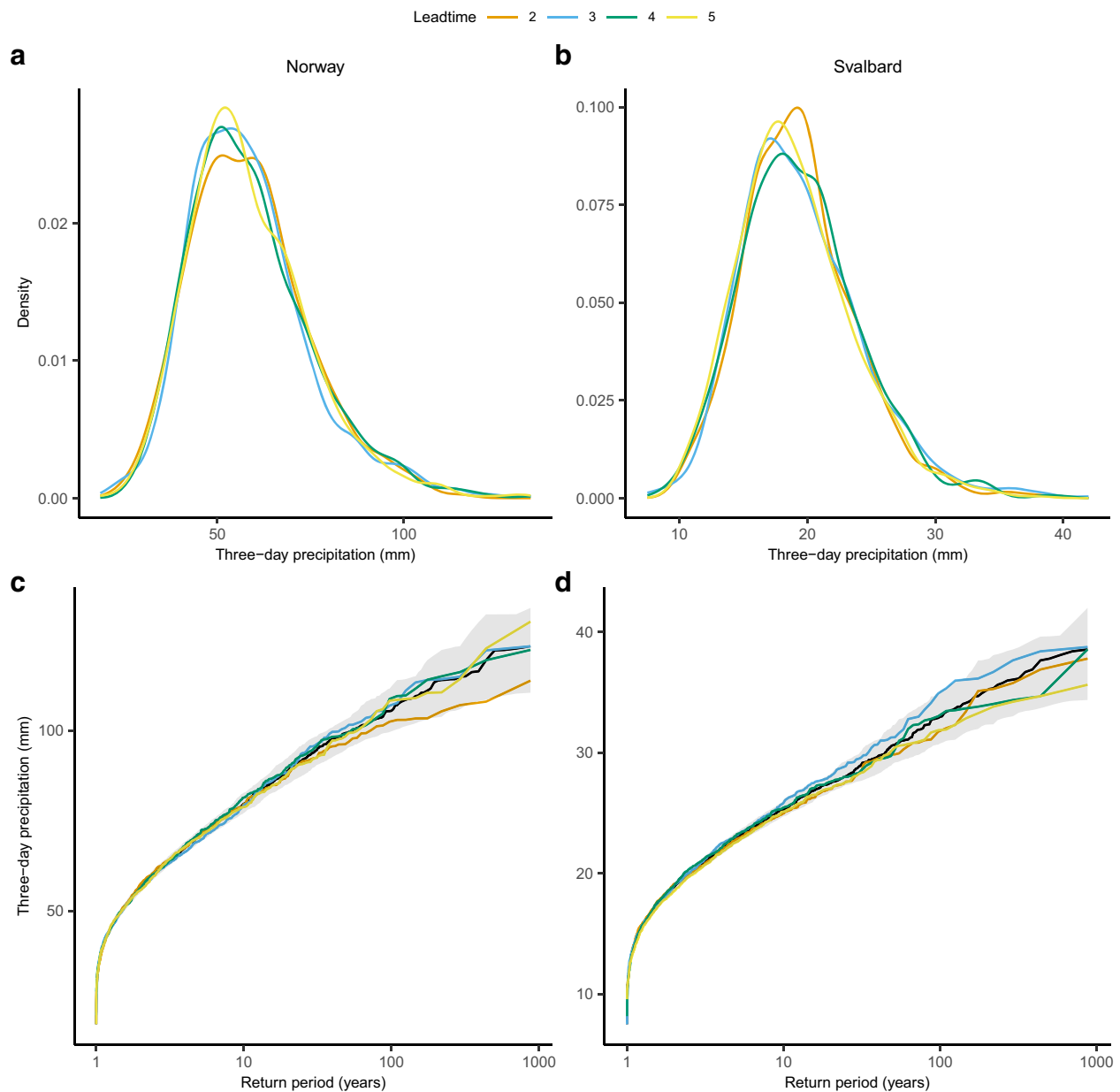
Fig. 3 SEAS5 model stability of extreme precipitation over Western Norway and Svalbard. The empirical probability density (**a**, **b**) and extreme value (**c**, **d**) distribution of SON-3DP for each lead time and for all lead times together (in black), for the West Coast (**a**, **c**) and Svalbard (**b**, **d**) domains. Grey shading in **c**, **d**, illustrates the 95% confidence intervals of the distribution of the pooled lead times, bootstrapped to timeseries of similar length to the individual lead times with $n = 10,000$.

recorded (total) precipitation. For Western Norway, the bootstrapping test indicates that the observed mean and standard deviation fall outside the 95% confidence intervals of the SEAS5 LSP ensemble, indicating a low bias (black histograms compared to blue line; Fig. 4). This bias can be estimated as a ratio between the mean of the reference SON-3DP extreme events and the mean of the SEAS5 LSP events. The ratios indicate a high discrepancy for observed precipitation (SeNorge:SEAS5 LSP, 1.74) and reanalysis (ERA5 TP: SEAS5 LSP, 1.80), but a lower discrepancy if we use ERA5 large-scale precipitation as reference instead (ERA5 LSP: SEAS5 LSP, 1.29; supplementary Fig. 1). This shows that the large-scale SON-3DP simulations over Western Norway are more reliable than the comparison with observed data suggest, but that convective precipitation, which is not accounted for within SEAS5 LSP, makes a substantial contribution to total precipitation during the SON-3DP events. For Svalbard, where we only have the ERA5 reanalysis, there is close agreement between SEAS5 and ERA5 large-scale SON-3DP extremes (ratio of 1.08, supplementary Fig. 2). Note, low biases in precipitation extremes are not uncommon in global Earth System Models[41], especially for mountainous regions like Western Norway.

To further assess the discrepancy between SEAS5 and ERA5 large-scale precipitation output (the 1.29 ratio), we compare the large-scale circulation of the two most severe observed events in 2005 and 2014, as well as the composite of the 10% highest UNSEEN events for Western Norway (Fig. 5). This analysis indicates that the UNSEEN events are described by a south-westerly flow onto the West coast of Norway, consistent with the observed large-scale patterns and expected atmospheric circulation
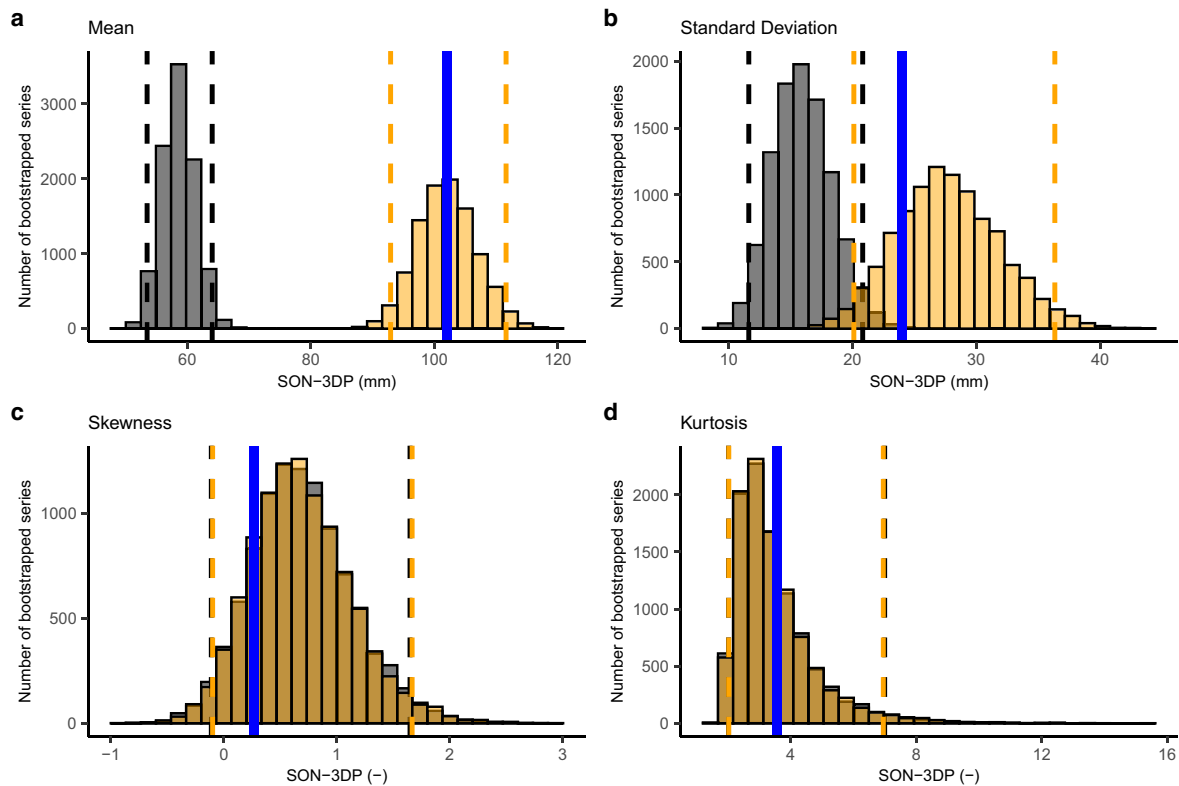
Fig. 4 UNSEEN distribution characteristics compared with the observed record. The histograms show the distribution of the **a** mean **b** standard deviation **c** skewness and **d** kurtosis for the UNSEEN ensemble before (grey) and after (yellow) bias correction, when bootstrapped to timeseries of 35 years ($n = 10,000$). The dashed lines indicate the 95% confidence intervals and the blue lines indicate the SON-3DP statistics derived from the observed precipitation time series over the period 1981–2015.

patterns associated with precipitation extremes over Norway within the ERA-Interim reanalysis[42].

Bias correction

Bias correction of the UNSEEN ensemble is necessary for impact modelling[43] and, as is the case here, to compare UNSEEN extremes to observed extremes. A range of methods are available for correcting biases in climate models[44], to statistically 'nudge' the quantiles of the model simulations closer to the observed values. However, the UNSEEN extremes of interest are by definition beyond the observed range and hence extrapolation of the bias correction is required. No perfect bias correction method exists and climate extremes within the observed range are already largely sensitive to uncertainties within the observations, and to resolution mismatch[45]. Extrapolation most commonly assumes a constant transfer function beyond the highest observed quantile or assumes an extreme value distribution[44]. Since both these assumptions are very sensitive to the largest observed values and might change the distribution without any physical reason, we consider that the simplest form of bias correction, the mean bias ratio between observed extremes and simulated extremes, is the least likely to introduce additional error. Note, with a single factor over the 35 years we only remove the systematic or time-independent error, since only this time-independent component can be removed with a statistical correction procedure[46]. This approach assumes that the model error is stationary over time, a feature that is hard to evaluate but has been suggested by model-based experiments[47].

We apply this constant bias correction (1.74, see 'Fidelity of UNSEEN extremes') to the UNSEEN ensemble for Western Norway to generate the bias-corrected UNSEEN ensemble (henceforth

referred to as UNSEEN-BC). We found little sensitivity to using the median (1.72), 5-year (1.69) or 20-year (1.70) values in the bias correction procedure. The effect of the bias correction on the statistical moments of the distribution is illustrated with the bootstrapping test (Fig. 4, see 'Fidelity of UNSEEN extremes'), showing that the mean and standard deviation are adjusted whilst the shape of the distribution (the skewness and kurtosis) is preserved. The test furthermore indicates that UNSEEN-BC is statistically consistent with the observed values.

Another important requirement of the bias correction procedure is that it does not inflate any detected trends[44]. We repeat the fidelity test for the non-stationary generalised extreme value (GEV) parameters (Fig. 6), instead of the empirical moments presented above. To avoid poorly fit timeseries, we set the quality control criteria: location ($0 < \mu_0 < 200$), log scale ($0 < \phi_0 < 5$) and shape ($-1 < \xi < \infty$). Applying these criteria excludes 58 out of the 10,000 bootstrapped samples. As in the case of the empirical moments (Fig. 4), the bootstrapped UNSEEN mean (location) and variability (log scale) do not pass the fidelity test, while the bootstrapped UNSEEN-BC distributions do. Furthermore, Fig. 6 indicates that the location trend parameter (μ_1) is sensitive to the bias correction, whereas the scale trend (ϕ_1) and shape (ξ) parameters are not. Although the absolute trend is sensitive to the bias correction procedure, the trend expressed as a percentage change between 1981 and 2015 (see Eq. (8) in 'Methods') is not. For example, if the simulated mean was 40 mm in 1981 and 60 mm in 2015, a bias-correction factor of 2.0 would result in 80 mm in 1981 and 120 mm in 2015. The absolute increase between 1981 and 2015 would double from 20 to 40 mm, but the percentage increase would still be 50%. Hence, the trend estimates and confidence intervals in the 100-year return intervals expressed as a
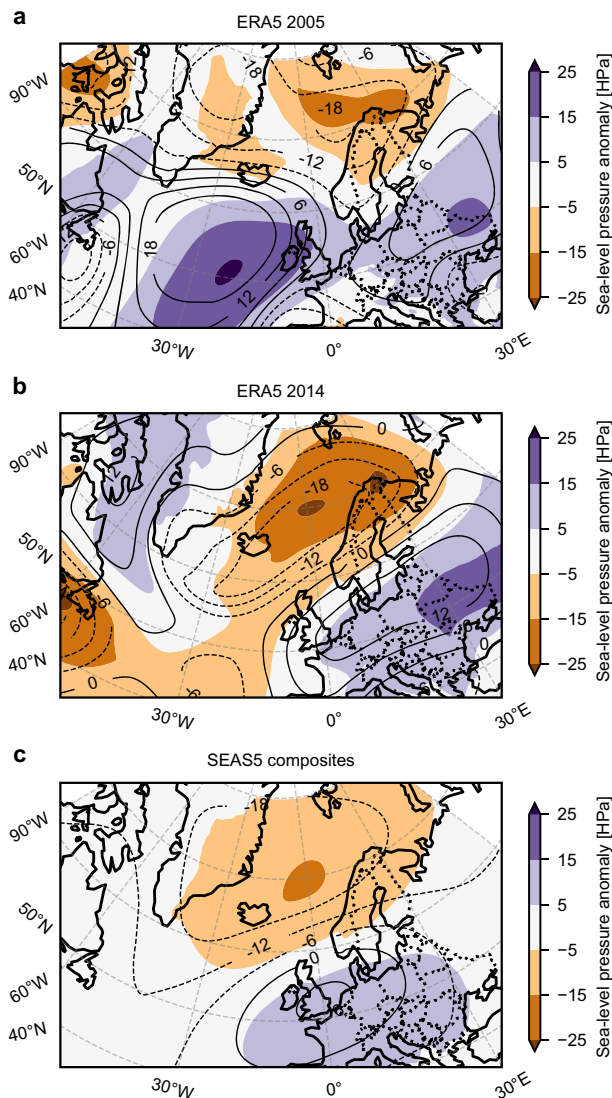
a
ERA5 2005

b
ERA5 2014

c
SEAS5 composites

Fig. 5 Large-scale circulation associated with observed and UNSEEN events. Filled contours show the mean sea-level pressure anomalies and contour lines show the 500 HPa geopotential height anomalies (in tens of metres) from the 1981–2016 climatology during **a** the 2005 event **b** the 2014 event and **c** for a composite of the 10% highest events within the UNSEEN ensemble.

percentage change since 1981 presented in the next section are insensitive to the chosen bias correction procedure.

UNSEEN trends in 100-year precipitation

Climate models can be used to detect changes[48–51] and to attribute extreme events to human causes[52], but are less suited to detecting trends over the recent past, such as the last 35 years. By design, climate model simulations are initialised once at the beginning of a centennial run. In contrast, the seasonal forecasts used herein are initialised every month and are, thus, more constrained by real-world climate variability than climate model simulations. Consequently, seasonal forecasts sample a smaller range of climate conditions but are closer to reality than climate model simulations. This means that their use is consistent with analysing trends over the recent past described by the available forecast period (for SEAS5, currently 35 years). Furthermore, the model setup and version are the same for the entire hindcast simulation, ensuring that, with respect to the models and

initialisation, SEAS5 is a homogeneous dataset and thus suitable for climate analysis and detection of UNSEEN trends.

With 36 km resolution and 25 members, the ECMWF SEAS5 re-forecast set used here is based on a modelling system of high-resolution and has a large ensemble when compared with current high-resolution global climate models[53]. SEAS5 greenhouse gas radiative forcing captures the long-term trends in emissions[32], and the global mean temperature trend in SEAS5 follows ERA5[37] (Supplementary Fig. 3). There is a cold bias over the Norwegian study domain, but the trend is consistent with ERA5 for both Western Norway and Svalbard (Supplementary Fig. 3), confirming the capacity of SEAS5 to detect recent trends.

To illustrate the added value of UNSEEN trends, we fit a time-dependent GEV distribution to the observed and UNSEEN autumn 3-day precipitation extremes (SON-3DP, see Methods). The fidelity test indicates that the non-stationary GEV parameters fitted to the UNSEEN-BC ensemble for Western Norway are statistically consistent with the parameters fitted to the observed record (Fig. 6). Using observations, we find an increase in 100-year SON-3DP of 4% over the period 1981–2015 in Western Norway, but the uncertainty range is −27% to 34% (Fig. 7a). The uncertainty around the UNSEEN-trend estimate of 2% is better constrained due to the larger sample size, with confidence intervals ranging from −3% to 7%. A negative trend is, thus, statistically possible, indicating that the trend over Western Norway is not significant. For Svalbard, we find a significant positive UNSEEN-trend of 8%, with uncertainty bounds ranging between 4% to 12% (Fig. 7 b).

In addition to the trend in 100-year SON-3DP events, we illustrate the change in all return values by plotting the GEV distribution with the covariates 1981 and 2015 (Fig. 7c, d). The likelihood ratio test shows that the GEV distribution with time covariate improves the model fit for Svalbard (p-value = 2.7e-07). We find that the 100-year event estimated in 1981 occurs with a return period of 41 years in 2015 (Fig. 7c, d). For Western Norway, the GEV distribution including a time covariate does not improve the model fit for either the observed (p-value = 0.58) or the UNSEEN ensemble (p-value = 0.65), and thus, a stationary GEV distribution is deemed appropriate.

Stationary GEV analysis

We fit the GEV distribution to the observations, the UNSEEN and the UNSEEN-BC ensemble for Western Norway (see 'Methods' and Fig. 8). The fitted distributions indicate that the UNSEEN-BC ensemble diverges from observed values at return periods ~35 years and above. To evaluate the discrepancy, we test the sensitivity of the results to the choice of extreme value distribution (Supplementary Fig. 4). Although the Gumbel distribution (shape parameter $\xi = 0$) shows a relatively good fit to the observations and a similar distribution to the UNSEEN ensemble, the fit is not as good as the full GEV with fitted shape parameter, as suggested by Supplementary Fig. 4 and confirmed by the likelihood ratio test (p-value = 0.03 for the observed and p-value < 0.001 for the UNSEEN ensemble). In addition, results are also very sensitive to outliers, as can be seen when the observed extreme value distribution is fitted to a sample where the largest value is increased by 10% (Supplementary Fig. 4). The fidelity test on the stationary GEV parameters confirms the large uncertainty when constraining the shape of the distribution (ξ) based on short timeseries (35 years): the type of the GEV distribution (Weibull, Gumbel or Fréchet) cannot be constrained (Supplementary Fig. 5). This highlights the challenge associated with estimating the magnitude of events of long return periods (>20 years) from observed series of only 35 years, with greater statistical confidence in estimates from the larger UNSEEN sample.

We find that the 2005 and 2014 observed extreme events (two highest blue circles in Fig. 8) are similar in magnitude and
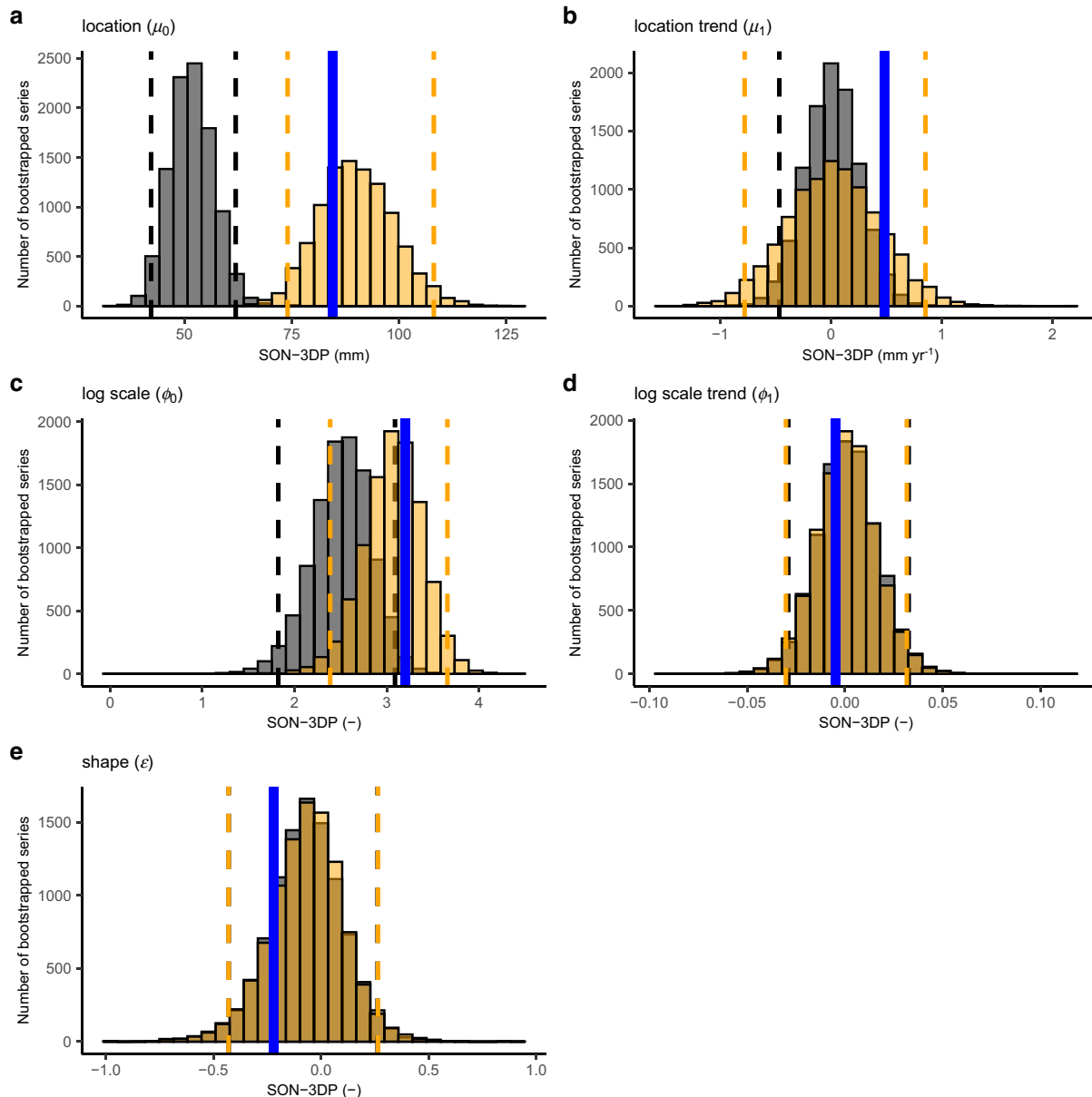
Fig. 6 UNSEEN GEV parameters compared to the observed record. a–e The histograms show the distribution of the parameters of the non-stationary GEV distribution for the UNSEEN ensemble before (grey) and after (yellow) bias correction, when bootstrapped to timeseries of 35 years ($n = 10{,}000$). The dashed lines indicate the 95% confidence intervals and the blue lines indicate the parameters derived from the observed precipitation time series over the period 1981–2015.

represent events with return periods of 21 years (CI of 19–24 years) when compared with the extreme value distribution of UNSEEN-BC. Based on the observed values, the return period estimate of 60 years for the events would be very uncertain, with the lower confidence interval never reaching the event magnitude (CI of 18–∞ years). Moreover, the highest UNSEEN-BC event (the highest orange circle) is 1.5 times higher than the highest observed event, with an estimated return period of ~2000 years (CI of 1150–4800 years). The estimated return period of this event based on the observations is completely dominated by the uncertainties (~5000, 600–∞ years) and can only be statistically modelled; the UNSEEN estimate is a physically simulated 'empirical' event within 3500 years of data. The 2005 and 2014 flood episodes caused flooding and landslides with severe damage[29,30] and yet UNSEEN-BC indicates that much more extreme events than seen in the observed record are feasible even under the present climate.

DISCUSSION

In this study, we build on the UNprecedented Simulated Extreme ENsemle (UNSEEN) approach[15] by applying well-established extreme value analysis[54] to the large UNSEEN ensemble to boost confidence in decadal trend detection. We apply UNSEEN trends to autumn 3-day precipitation in Western Norway and Svalbard – two regions that have faced severe damage from recent extreme precipitation events. A dense station network exists for Western Norway, which we use to compare against our trend estimates. In contrast, the sparse network in Svalbard limits trend detection and we show that our UNSEEN trends approach can provide additional information about the characteristics of extreme precipitation in the rapidly changing Arctic[28,31,55,56].

Our UNSEEN trends approach substantially reduces the uncertainty range of the spatial averaged trend estimate for 1981–2015 in Western Norway from −27% to 34% (estimated
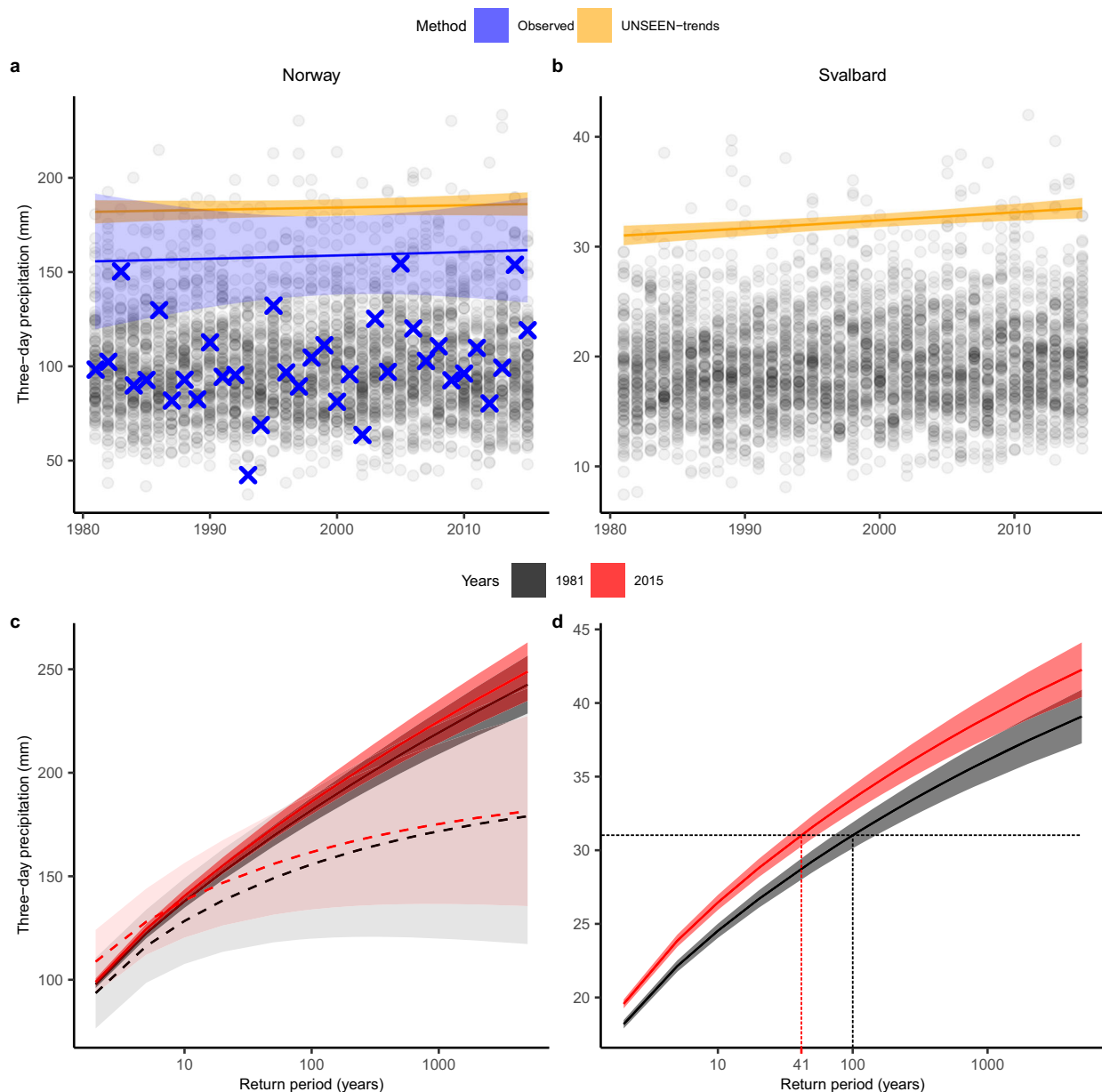
Fig. 7 Observed and UNSEEN trends in extreme precipitation. a, b The change in 100-year autumn 3-day precipitation (SON-3DP) over 1981–2015 is shown for **a** Western Norway and **b** Svalbard. The data points show the SON-3DP events in the observed record (blue crosses) and in the UNSEEN ensemble (black circles). Note, for Svalbard no gridded precipitation record is available and for Norway the bias-corrected UNSEEN-BC is used. **c, d** In addition to the change in 100-year precipitation, the entire GEV distribution is plotted for the covariates 1981 and 2015 over **c** Western Norway and **d** Svalbard. Solid lines and dark shading indicate the trend and uncertainty of the UNSEEN trends approach and dashed lines with light shading (in **c**) indicates the trend and uncertainty range based on observations (95% confidence intervals based on the normal approximation). In **d**, the magnitude of the event with a return period of 100 years in 1981 is illustrated with black dotted lines and the event of similar magnitude corresponding to a return period of 41 years in 2015 is illustrated with a red dotted line.

using the precipitation record), to only −3% to 7%. In both cases, we find a small and insignificant trend estimate of 4% (observed) and 2% (our approach). For Svalbard, we detect a much stronger and significant signal of 8% (4:12%). As a result, the 100-year precipitation event in 1981 became a ~40-year event in 2015. The trend in extreme precipitation over Svalbard could not be detected from observation-based studies due to the sparse observation network in this area[27]. Despite very few precipitation extremes being recorded in the Svalbard Archipelago, it is expected that their frequency and magnitude are increasing in a warming climate[27,28], which is consistent with our UNSEEN-trends analysis. Those precipitation extremes are connected to the inflow

of relatively warm air and, thus, can cause severe landslides and so-called rain-on-ice events[28]. Both could have significant impacts on people living in the Arctic and on local ecosystems.

The September 2005 and October 2014 flood episodes over Western Norway were identified as high-impact events in previous end-user engagement sessions within the Translating Weather Extremes into the Future (TWEX) project[30]. Thus, estimating their frequency of occurrence is of high relevance to decision makers and end-users. We show the very large uncertainties in estimating the return periods of the observed events and illustrate how the large UNSEEN-BC sample improves the statistical confidence in the return period estimates. We find that instead of the unreliable
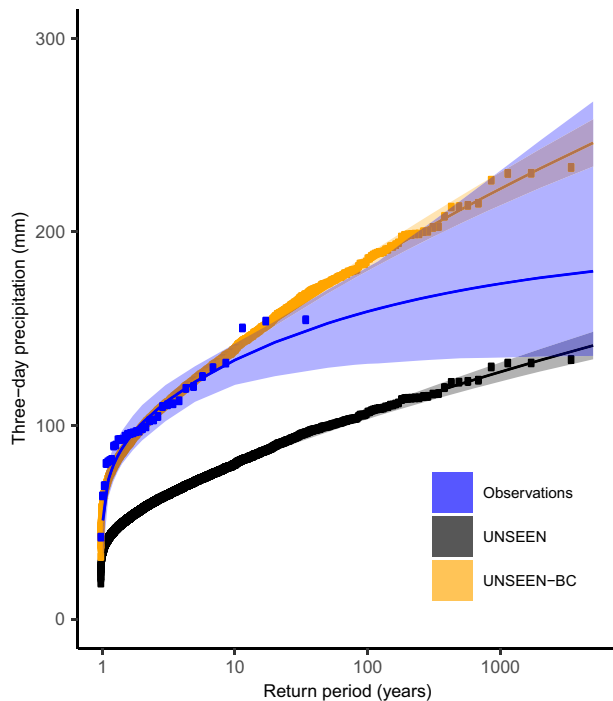
Fig. 8 Observed and UNSEEN extreme precipitation distributions for Western Norway. The data points show the SON-3DP events and the solid lines show the GEV fitted to the data, including 95% confidence intervals based on a parametric bootstrap.

60-year estimate from the records, UNSEEN-BC suggests that the flood episodes are not rare exceptions, but might be expected to occur once in 20 years under a stationary climate. Furthermore, the UNSEEN-BC ensemble shows that an event of 1.5 times the magnitude of the highest observed event could arise. This application of the UNSEEN approach is similar to previous research on the 2013/14 winter floods in the UK[15] and for the 1990 windstorm losses over Germany and the UK[21]. A difference to the previous studies is that we run the analysis on a 3-day resolution based on the latest ECMWF seasonal prediction system SEAS5, whereas monthly averages and other prediction systems were used elsewhere.

An important point of discussion is that the presented confidence ranges are statistical intervals that do not incorporate physical credibility. For example, the uncorrected UNSEEN ensemble for Western Norway has high statistical confidence in estimating return periods from its large sample size, yet, there is little credibility in these estimates because they are not consistent with observed events; so bias correction is necessary (Fig. 8). To evaluate the credibility of the UNSEEN ensemble, we perform three tests: ensemble member independence, model stability and model fidelity. We find that ensemble members are independent and the model is stable over selected lead times, hence, the effective sample size of autumn 3-day precipitation (SON-3DP) events in Western Norway and Svalbard can be increased by a factor of 100 compared with observations. The fidelity test shows that using a simple constant bias correction value, the statistical moments in UNSEEN-BC are consistent to the observed and that the simulated trend estimate (in percentage change) is preserved. This analysis acknowledges that the observed record is probabilistic in itself, in the sense that besides uncertainties within the records[45], the record could have been different had we faced other extremes (induced by stochastic processes in the atmosphere).

We find that part of the discrepancy between UNSEEN and the observed SON-3DP events over Western Norway arises from the study design decision to use only simulated large-scale precipitation (see 'Methods'). As convective processes cannot be resolved in a global model of 36 km resolution[40], only the simulated large-scale precipitation output is chosen as the target variable. However, ERA5 suggests that the contribution of local convection to the total precipitation is substantial for SON-3DP events (see 'Fidelity of UNSEEN extremes'). We apply a constant correction ratio between the observed total precipitation and the simulation large-scale precipitation (1.74, see 'Fidelity of UNSEEN extremes') and, therefore, implicitly assume that convection can be corrected by a constant value. Further investigating the reliability of convective processes within SEAS5 and incorporating these mechanisms in the UNSEEN ensemble may improve confidence in the physical simulations of SON-3DP events over Western Norway.

Analysis of the atmospheric drivers of the extreme events may be used to further evaluate the credibility of UNSEEN extremes or might highlight plausible atmospheric conditions, leading to extremes that have not yet been observed[15]. For Western Norway, the large-scale circulation patterns during UNSEEN events are consistent with the large-scale circulation patterns during the 2005 and 2014 events (See 'Fidelity of UNSEEN extremes'). An in-depth analysis of the dynamics at play, such as the integrated vapour transport over this region during the events and their related teleconnections and sea-surface temperatures, merits further research but is beyond the scope of this paper. To assess the drivers of non-stationarity in extreme precipitation, covariates that might be more appropriate than time[57,58] could be selected, such as ocean temperatures, modes of climate variability[59], or indicators of large-scale synoptic weather systems[60,61]. Such analyses may improve our physical understanding of the non-stationary processes driving climate extremes and could provide insights into potential model biases, thereby improving confidence in the detected trends over Svalbard and in assuming a stationary GEV distribution for Western Norway. Century-long seasonal hindcasts, such as the ASF-20C global atmospheric seasonal hindcasts[62], might prove useful in assessing the sensitivity of UNSEEN trends to different time windows within a longer period.

The insights presented in this study are specific to Western Norway and Svalbard SON-3DP but the UNSEEN-trends approach is transferrable to other regions, temporal resolutions and spatial extent of events, seasons and climate variables. Global validation of the independence, model stability and model fidelity will show in which regions the approach may challenge the robustness of design level estimates, with a potentially high utility in supporting data-scarce regions[63]. Furthermore, the large sample size may allow estimation of extremes using empirical approaches that avoid assumptions about underlying distributions and their non-stationarity, thereby offering the possibility of improved design estimates[10] and empirical attribution of physical mechanisms. A wide range of scientific disciplines might benefit from the UNSEEN method by connecting seasonal prediction systems with impact models to assess unprecedented impacts and improved understanding of the physical mechanisms leading to such events.

Our results for Western Norway highlight the strength of UNSEEN in evaluating design-levels and present-day climate hazards, backed by a growing body of literature[12–22]. For example, UNSEEN might be helpful in reviewing design standards used to rebuild infrastructure after the severe damage caused by the 2005 and 2014 flood episodes. Searching questions could be asked, such as: what is the estimated economic risk of that infrastructure failing in the coming 20 years, based on the observed record as compared to UNSEEN estimates? UNSEEN trends may be incorporated in risk-based decision making, where the performance of infrastructure can be tested by estimating the risk of failure under stationary and non-stationary conditions[58]. For example, in Svalbard, the detected changes might be used to

evaluate whether design standards are still adequate, where ideally the physical drivers of the changes in the precipitation extremes are attributed. We assert that further applications could (1) help estimate design values, especially relevant for data-scarce regions; (2) improve risk estimation of natural hazards by coupling UNSEEN to impact models; (3) detect trends in rare climate extremes, including variables other than precipitation; and (4) increase our physical understanding of the drivers of non-stationary climate extremes, through the possible attribution of detected trends.

METHODS

Data

We use the fifth generation of the ECMWF seasonal forecasting system SEAS5 to generate the UNSEEN ensemble. SEAS5 is a global coupled ocean, sea-ice, and atmosphere model, which was introduced in autumn 2017[32]. The atmospheric component is based on cycle 43r1 of the ECMWF Integrated Forecast System. The spatial horizontal resolution is 36 km with 91 vertical levels. The ocean (Nucleus for European Modelling of the Ocean, NEMO[64]) and sea-ice (Louvain-la-Neuve Sea Ice Model, LIM2[65]) models run on a 0.25-degree resolution. The atmosphere is initialised by ERA-Interim[66] and the ocean and sea-ice components are initialised by the OCEAN5 reanalysis[67]. ECMWF provides a re-forecast (also known as hindcast) dataset for calibration of the operational forecasting system SEAS5. The data are initialised monthly with 25 ensemble members, each with 7-month forecast length on a daily resolution, covering the years 1981–2016[32]. Ensemble members are generated by perturbing initial ocean and atmosphere conditions and from stochastic model perturbations.

In the UNSEEN approach, ensemble members and initialisation dates are pooled to increase the sample size of the variable of interest. Here, we demonstrate an UNSEEN ensemble for the west coast of Norway and for the Svalbard Archipelago to evaluate recent atmospheric river (AR) related severe events[27,29,30]. ARs have been connected to precipitation extremes in the observed records for both Norway[42,68] and Svalbard[27] during September to March. AR-related floods mostly occur in autumn, because snowfall during winter precipitation events results in storage rather than runoff. One-day and five-day precipitation are a common diagnostic for extreme analysis[6,69]. ARs frequently strengthen over a period of several days[29,30] and, therefore, multi-day diagnostics prevent splitting events. Following the 2014 flood episode[30], we use 3-day precipitation totals in this study. We thus select autumn (September to November), 3-day extreme precipitation (SON-3DP) as target events.

Since the hindcasts are initialised every month on the first of the month and run over 7-months duration, there are five initialisation months (May-September) available to forecast the entire target autumn season (September-November). The first month is removed to avoid potentially dependent events. This leaves 100 hindcasts, based on 25 ensemble members with 4 initialisation dates to forecast the autumn season of each year (Fig. 2a–c). The window of 35 years between 1981 and 2016 leads to a total of 3500 forecasts of autumn weather conditions that could have occurred (Fig. 1b–c). We extract the maximum 3-day cumulative precipitation within autumn from the 3500 forecasts (SON-3DP), using the xarray package[70] in Python. To focus on the weather systems as experienced in recent severe events, we use only the large-scale precipitation output of the model. The west coast of Norway is mountainous and characterised by considerable topographic variation. Catchment-scale processes in these mountainous areas cannot be resolved by a global model with 36 km resolution. Therefore, the precipitation timeseries presented in this study are spatial averages where the 200-year precipitation exceeds 90 mm for the west coast of Norway (4–7° E, 58–63° N) and 35 mm for Svalbard (8–30° E, 76–80° N) (Fig. 1a).

To evaluate the precipitation extremes simulated by SEAS5, we use a 1×1 km gridded station-based precipitation product for Norway[25]. The data have recently been corrected for underestimation caused by wind-induced under catch and uses more information in the interpolation scheme for data-scarce areas, resulting in higher precipitation in those areas[26]. We upscale this gridded dataset to the same resolution as SEAS5 and extract SON-3DP values for the same spatial domain over 1981–2016. Note, for the Svalbard Archipelago no gridded precipitation data are available as a reference dataset. We use ERA5[37] large-scale and total precipitation for the fidelity test, mean sea-level pressure and geopotential height at 500HPa

for the large-scale circulation comparison, and seasonal mean (September-November) temperature for the global and regional temperature evaluation of SEAS5. The atmospheric component is based on cycle 41r2 of the ECMWF Integrated Forecast System (IFS), one IFS cycle older than SEAS5 (43r1). IFS large-scale precipitation represents processes resolved at scales of the grid box size or larger, such as cloud formation or changes in pressure, whilst convective precipitation represent smaller-scale convective processes (see the full IFS documentation and a summary of the differences between IFS cycles at: https://www.ecmwf.int/en/forecasts/documentation-and-support/changes-ecmwf-model).

Ensemble member independence testing

The method for independence testing applied in this study is inspired by previous research on potential predictability: the ability of the model to predict itself[33,71]. The potential predictability of a model is calculated by using one of the forecast ensemble members as the observations and the mean of the other ensemble members as the forecast. The correlation between the 'observed' ensemble member and the mean of the other ensemble members is calculated for every ensemble member and this range gives an estimate of the ability of the model to forecast itself. As this method assesses the correlation between ensemble members, it can be used to find the degree of dependence among ensemble members. In seasonal forecasting, this method is used to identify any predictability in the seasonal prediction system. In contrast, here we seek to demonstrate that there is no potential predictability in the system for the ensemble members to represent independent, unique events.

An illustration of our method for testing independence is shown in Fig. 2. A potential predictability test is performed but instead of correlating an ensemble member to the mean of the other ensemble members, a pairwise correlation test is applied between all ensemble members to robustly assess the individual ensemble member dependence. Indeed, we concatenate the seasons together member by member, even though they do not necessarily originate from the same run. This approach was chosen because the underlying initialisation method remains the same for each member over different seasons.

For the 25 ensemble members, there are 300 distinct pairings in the correlation matrix for each of the four lead times being analysed (May–August). We calculate the Spearman rank correlation ρ for the standardised SON-3DP anomalies (deviation from mean divided by the standard deviation) for each distinct pair. From the 300 ρ values for each lead time, boxplot statistics are calculated: the whiskers, the interquartile range and the median. When testing for significance of the 300 ρ values, care must be taken not to falsely detect significant correlations because of the large number of tests. For example, with a confidence interval of 5%, 15 out of the 300 correlations would be expected to be significant by chance alone. To avoid these problems, a permutation test is performed. The dataset, which previously consisted of 25 timeseries (members) of 35 datapoints (years) for four initialisations months (lead times), is resampled into 100 timeseries of 35 datapoints, with datapoints randomly picked from all members, years and lead times to remove potential correlations. This randomised dataset is split into four pseudo lead times of 25 timeseries, in order to calculate the boxplot statistics from the same amount of correlation coefficients (300) as before. The data are resampled 1000 times (without replacement), resulting in 4000 boxplot statistics (4 pseudo lead times * 1000 resampled series), from which the confidence intervals are calculated based on a 5% significance level (the 2.5 and 97.5 percentiles).

Model stability

The extreme precipitation distribution must be similar over lead times in order to generate the UNSEEN ensemble. We use four initialisation months (May–August) forecasting the target autumn season with lead times 2 to 5 months. For each lead time, 25 ensemble members over 35 years result into an 875-year long dataset and the pooled ensemble into 3500 years. To compare the distributions, we first plot the probability density function for each of the lead times using ggplot2[72]. Second, we plot the extreme value distributions, focussing more on the tails of the distribution. We calculate empirical quantiles of the extreme precipitation ensemble without assuming any distribution a priori, to avoid problems regarding statistical modelling of the extremes[10,73]. The quantile (Q) of a distribution is the inverse of the distribution function ($F(x)$):

$$Q(p) = F^{-1}(p) = \inf \{x : F(x) \geq p\}, \quad 0 < p < 1, \tag{1}$$

where the return value is associated with the quantile of percentile (p):

$$p = 1 - \frac{1}{T} \tag{2}$$

with T being the return period. We use the quantile function in R to compute the empirical return values and we refer to Hyndman and Fan[74] for more detail.

Fidelity of the UNSEEN ensemble for Western Norway

To compare UNSEEN to the observed record, we apply a bootstrap test presented in previous studies[15,16]. We bootstrap 10,000 timeseries of 35 years with replacement from all ensembles (100×35 years) and calculate the mean, standard deviation, skewness and kurtosis for each. We test whether the four distribution statistics derived from the observed precipitation time series over the period 1981–2015 fall within the 95% confidence intervals for the statistics derived from the bootstrapped timeseries. We repeat the fidelity tests for the stationary and non-stationary GEV parameters, described in following sections.

Stationary GEV distribution

The GEV distribution is described by a location ($-\infty < \mu < \infty$), scale ($\sigma > 0$), and shape ($-\infty < \xi < \infty$) parameter[54]:

$$F(x) = \exp\left[-\left(1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-\frac{1}{\xi}}\right], \; \left(1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right) > 0. \tag{3}$$

We test the sensitivity to using the Gumbel distribution with $\xi = 0$, simplifying the distribution to:

$$F(x) = \exp\left[-\exp\left(-\left(\frac{x-\mu}{\sigma}\right)\right)\right], \; -\infty < x < \infty. \tag{4}$$

The quantiles of the distribution can again be obtained by inverting the distribution:

$$x_p = \begin{cases} \mu - \frac{\sigma}{\xi}\left[1 - \{-\log(1-p)\}^{-\xi}\right], \; \text{for } \xi \neq 0 \\ \mu - \sigma\log\{-\log(1-p)\}, \; \text{for } \xi = 0 \end{cases}, \tag{5}$$

where the return value x_p corresponds to the return period (T) 1/ probability (p). For all statistical model fits in this study (including non-stationary fits described in the next section), we apply maximum likelihood estimation (MLE) to estimate the parameters of the distributions, utilising the extRemes package[75] in R. The 95% confidence intervals of the distributions are calculated based on a parametric bootstrap, that can better highlight the uncertainties associated with the extrapolation of the extremes than the normal approximation (Fig. 8 and Supplementary Fig. 6).

UNSEEN trends

In this study, we present the idea of performing trend analysis on seasonal hindcasts, as the seasonal hindcasts provide a larger sample than observations and a higher resolution than global climate models (see the UNSEEN-trends section for more details). We apply well-established extreme value theory[54,76,77], by allowing the location (μ) and scale (σ) parameters of the GEV distribution (given in Eq. 3) to vary linearly with time (t). As the scale parameter needs to be positive, a log-link function is used:

$$\mu(t) = \mu_0 + \mu_1 t. \tag{6}$$

$$\ln \sigma(t) = \phi_0 + \phi_1 t. \tag{7}$$

This approach selects one block maximum per year, leading to 35 data points over the years 1981–2015 based on observed records. With UNSEEN trends, we have 100 times more values for each year and thus increase confidence in the regression analysis (see Fig. 7a, b for illustration). As for the stationary method, we use MLE to estimate the parameters of the distributions. For numerical reasons, we vary time (t) over 1:35 rather than 1981:2015, e.g., we have set $\mu_0 = \mu(1980)$.

Instead of the parametric bootstrap for the stationary method, we apply the normal approximation to find the 95% confidence intervals of return values. Implementing a parametric bootstrap for the non-stationary analysis (the uncertainty intervals for Fig. 7) is complicated because we have 35 regression points (each year in 1981–2015). For each of the regression points, a parametric bootstrap would have to be performed, which is computationally expensive. For the UNSEEN confidence intervals,

we do not expect much difference between the two methods (Fig. 8 and Supplementary Fig 6), so we applied a normal approximation for the non-stationary analysis.

We focus on the changes in the 100-year quantiles, because these are associated with the design-levels most widely used in flood defence[78]. The trend in the 100-year return value is defined as the percent change between 1981 and 2015:

$$\Delta x_T = 100 * \left(\frac{x_T(\mu(2015), \ln\sigma(2015), \xi) - x_T(\mu(1981), \ln\sigma(1981), \xi)}{x_T(\mu(1981), \ln\sigma(1981), \xi)} \right), \tag{8}$$

where x_T is defined by Eq. (5).

The robustness of the trends to experimental decisions such as block size, event duration and regression method can be further investigated but are beyond the scope of this research. For example, 6-month blocks can be selected at the expense of the ensemble size. This would result in 25 realisations, in comparison with 3-month blocks, which contain 100 realisations. A block size of three months (September-November) and three-day precipitation extremes were used in this study to represent flood-inducing events (see the data section above). A sensitivity test to different event durations could be performed to check the max-stable property of the GEV distribution[54]. A linear trend in time is assumed. With the large amount of data, more complex regression methods can be explored. The ECMWF SEAS5 seasonal prediction system is used, but other seasonal prediction systems with available hindcasts could also be assessed to test the sensitivity of the return value and trend estimation to the model.

DATA AVAILABILITY

SEAS5 re-forecast data was accessed through the MARS Catalogue. This catalogue has restricted access, it is available for National meteorological services of ECMWF Member and Co-operating States. Other users can request access here: https://www.ecmwf.int/en/about/contact-us?subject=Gain%20access%20to%20archive%20data. Alternatively, SEAS5 re-forecast data on 1-degree resolution, as well as ERA5 data, are openly available from the Copernicus Climate Change Service (C3S) Climate DataStore (https://cds.climate.copernicus.eu/). The SeNorge daily total precipitation data are available at https://doi.org/10.5281/zenodo.2082320. The extracted SON-3DP UNSEEN ensembles as well as the extracted SON-3DP observations are available on GitHub: https://github.com/timokelder/UNSEEN-trends.

CODE AVAILABILITY

All code to reproduce the analysis in this paper is available on GitHub under the open source MIT license: https://github.com/timokelder/UNSEEN-trends.

REFERENCES

1. IPCC. Summary for policymakers. in global warming of 1.5 °C. *An IPCC Special Report on the Impacts of Global Warming of 1.5 °C above Pre-industrial Levels and Related Global Greenhouse Gas Emission Pathways, in the Context of Strengthening the Global Response to the Threat of Climate Change*, 32 (World Meteorological Organization, 2018).
2. IPCC. Summary for policymakers. in *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation: Special Report of the Inter-governmental Panel on Climate Change* 3–22 (Cambridge University Press, 2012).
3. IPCC. Summary for policymakers. in *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* 1–32 (Cambridge University Press, 2014).
4. Wilby, R. L. et al. The 'dirty dozen'of freshwater science: detecting then reconciling hydrological data biases and errors. *Wiley Interdiscip. Rev.* **Water 4**, e1209 (2017).
5. Zwiers, F. W. et al. Climate extremes: challenges in estimating and understanding recent changes in the frequency and intensity of extreme climate and weather events. in climate science for serving society 339–389 (Springer Netherlands, 2013).
6. Alexander, L. V. Global observed long-term changes in temperature and precipitation extremes: a review of progress and limitations in IPCC assessments and beyond. *Weather Clim. Extrem.* **11**, 4–16 (2016).

7. Klein Tank, A. M. G. & Können, G. P. Trends in indices of daily temperature and precipitation extremes in Europe, 1946–99. *J. Clim.* **16**, 3665–3680 (2003).

8. Westra, S., Alexander, L. V. & Zwiers, F. W. Global increasing trends in annual maximum daily precipitation. *J. Clim.* **26**, 3904–3918 (2013).

9. Donat, M. G., Lowry, A. L., Alexander, L. V., O'Gorman, P. A. & Maher, N. More extreme precipitation in the world's dry and wet regions. *Nat. Clim. Chang.* **6**, 508 (2016).

10. Wiel, K., Wanders, N., Selten, F. M. & Bierkens, M. F. P. Added value of large ensemble simulations for assessing extreme river discharge in a 2 °C warmer World. *Geophys. Res. Lett.* **46**, 2093–2102 (2019).

11. Berghuijs, W. R., Aalbers, E. E., Larsen, J. R., Trancoso, R. & Woods, R. A. Recent changes in extreme floods across multiple continents. *Environ. Res. Lett.* **12**, 114035 (2017).

12. van den Brink, H. W., Können, G. P., Opsteegh, J. D., van Oldenborgh, G. J. & Burgers, G. Improving 10 4 -year surge level estimates using data of the ECMWF seasonal prediction system. *Geophys. Res. Lett.* **31**, L17210 (2004).

13. van den Brink, H. W., Können, G. P., Opsteegh, J. D., van Oldenborgh, G. J. & Burgers, G. Estimating return periods of extreme events from ECMWF seasonal forecast ensembles. *Int. J. Climatol.* **25**, 1345–1354 (2005).

14. Kent, C. et al. Maize drought hazard in the northeast farming region of China: unprecedented events in the current climate. *J. Appl. Meteorol. Climatol.* **58**, 2247–2258 (2019).

15. Thompson, V. et al. High risk of unprecedented UK rainfall in the current climate. *Nat. Commun.* **8**, 107 (2017).

16. Thompson, V. et al. Risk and dynamics of unprecedented hot months in South East China. *Clim. Dyn.* **52**, 2585–2596 (2019).

17. Breivik, Ø., Aarnes, O. J., Bidlot, J.-R., Carrasco, A. & Saetra, Ø. Wave extremes in the northeast Atlantic from ensemble forecasts. *J. Clim.* **26**, 7525–7540 (2013).

18. Breivik, Ø., Aarnes, O. J., Abadalla, S., Bidlot, J.-R. & Janssen, P. Wind and wave extremes over the world oceans from very large ensembles. *Geophys. Res. Lett.* **41**, 5122–5131 (2014).

19. Meucci, A., Young, I. R. & Breivik, Ø. Wind and wave extremes from atmosphere and wave model ensembles. *J. Clim.* **31**, 8819–8842 (2018).

20. Osinski, R. et al. An approach to build an event set of European windstorms based on ECMWF EPS. *Nat. Hazards Earth Syst. Sci.* **16**, 255–268 (2016).

21. Walz, M. A. & Leckebusch, G. C. Loss potentials based on an ensemble forecast: how likely are winter windstorm losses similar to 1990? *Atmos. Sci. Lett.* **20**, e891 (2019).

22. Kent, C. et al. Using climate model simulations to assess the current climate risk to maize production. *Environ. Res. Lett.* **12** (2017). https://doi.org/10.1088/1748-9326/aa6cb9.

23. Shepherd, T. G. et al. Storylines: an alternative approach to representing uncertainty in physical aspects of climate change. *Clim. Change* **151**, 555–571 (2018).

24. Lavers, D. A. & Villarini, G. The contribution of atmospheric rivers to precipitation in Europe and the United States. *J. Hydrol.* **522**, 382–390 (2015).

25. Lussana, C. et al. seNorge2 daily precipitation, an observational gridded dataset over Norway from 1957 to the present day. *Earth Syst. Sci. Data* **10**, 235–249 (2018).

26. Lussana, C., Tveito, O. E., Dobler, A. & Tunheim, K. seNorge_2018, daily precipitation and temperature datasets over Norway. *Earth Syst. Sci. Data Discuss.* **2019**, 1–27 (2019).

27. Serreze, M. C., Crawford, A. D. & Barrett, A. P. Extreme daily precipitation events at Spitsbergen, an Arctic Island. *Int. J. Climatol.* **35**, 4574–4588 (2015).

28. Hansen, B. B. et al. Warmer and wetter winters: characteristics and implications of an extreme weather event in the High Arctic. *Environ. Res. Lett.* **9**, 114021 (2014).

29. Stohl, A., Forster, C. & Sodemann, H. Remote sources of water vapor forming precipitation on the Norwegian west coast at 60°N - A tale of hurricanes and an atmospheric river. *J. Geophys. Res. Atmos.* **113**, D05102 (2008).

30. Schaller, N. et al. The role of spatial and temporal model resolution in a flood event storyline approach in western Norway. *Weather Clim. Extrem.* **29**, 100259 (2020).

31. Serreze, M. C. & Stroeve, J. Arctic sea ice trends, variability and implications for seasonal ice forecasting. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **373** (2015).

32. Johnson, S. J. et al. SEAS5: the new ECMWF seasonal forecast system. *Geosci. Model Dev.* **12**, 1087–1117 (2019).

33. Lavers, D. A., Pappenberger, F. & Zsoter, E. Extending medium-range predictability of extreme hydrological events in Europe. *Nat. Commun.* **5**, 5382 (2014).

34. Baggett, C. F., Barnes, E. A., Maloney, E. D. & Mundhenk, B. D. Advancing atmospheric river forecasts into subseasonal-to-seasonal time scales. *Geophys. Res. Lett.* **44**, 7528–7536 (2017).

35. Gupta, A. S., Jourdain, N. C., Brown, J. N. & Monselesan, D. Climate drift in the CMIP5 models. *J. Clim.* **26**, 8597–8615 (2013).

36. Hermanson, L. et al. Different types of drifts in two seasonal forecast systems and their dependence on ENSO. *Clim. Dyn.* **51**, 1411–1426 (2018).

37. Hersbach, H. et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **146**, 1999–2049 (2020).

38. Chen, C. T. & Knutson, T. On the verification and comparison of extreme rainfall indices from climate models. *J. Clim.* **21**, 1605–1621 (2008).

39. Osborn, T. J. & Hulme, M. Development of a relationship between station and grid-box rainday frequencies for climate model evaluation. *J. Clim.* **10**, 1885–1908 (1997).

40. Blenkinsop, S. et al. The INTENSE project: using observations and models to understand the past, present and future of sub-daily rainfall extremes. *Adv. Sci. Res.* **15**, 117–126 (2018).

41. Sillmann, J., Kharin, V. V., Zhang, X., Zwiers, F. W. & Bronaugh, D. Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *J. Geophys. Res. Atmos.* **118**, 1716–1733 (2013).

42. Benedict, I., Ødemark, K., Nipen, T. & Moore, R. Large-scale flow patterns associated with extreme precipitation and atmospheric rivers over Norway. *Mon. Weather Rev.* **147**, 1415–1428 (2019).

43. Teutschbein, C. & Seibert, J. Bias correction of regional climate model simulations for hydrological climate-change impact studies: review and evaluation of different methods. *J. Hydrol.* **456–457**, 12–29 (2012).

44. Maraun, D. Bias correcting climate change simulations-a critical review. *Curr. Clim. Chang. Rep.* **2**, 211–220 (2016).

45. Casanueva, A. et al. Testing bias adjustment methods for regional climate change applications under observational uncertainty and resolution mismatch. *Atmos. Sci. Lett.* **21**, e978 (2020).

46. Cannon, A. J., Piani, C. & Sippel, S. *in Climate Extremes and Their Implications for Impact and Risk Assessment.* 77–104 (Elsevier, 2020).

47. Krinner, G. & Flanner, M. G. Striking stationarity of large-scale climate model bias patterns under strong climate change. *Proc. Natl Acad. Sci. USA* **115**, 9462–9466 (2018).

48. Kharin, V. V. et al. Risks from climate extremes change differently from 1.5 °C to 2.0 °C depending on rarity. *Earth's Futur.* **6**, 704–715 (2018).

49. Kharin, V. V., Zwiers, F. W., Zhang, X. & Wehner, M. Changes in temperature and precipitation extremes in the CMIP5 ensemble. *Clim. Change* **119**, 345–357 (2013).

50. Sillmann, J., Kharin, V. V., Zwiers, F. W., Zhang, X. & Bronaugh, D. Climate extremes indices in the CMIP5 multimodel ensemble: Part 2. Future climate projections. *J. Geophys. Res. Atmos.* **118**, 2473–2493 (2013).

51. Kharin, V. V. & Zwiers, F. W. Changes in the extremes in an ensemble of transient climate simulations with a coupled atmosphere-ocean GCM. *J. Clim.* **13**, 3760–3788 (2000).

52. Angélil, O. et al. Comparing regional precipitation and temperature extremes in climate model and reanalysis products. *Weather Clim. Extrem.* **13**, 35–43 (2016).

53. Haarsma, R. J. et al. High resolution model intercomparison project (HighResMIP v1.0) for CMIP6. *Geosci. Model Dev.* **9**, 4185–4208 (2016).

54. Coles, S. *An Introduction to Statistical Modeling of Extreme Values.* vol. 208 (Springer London, 2001).

55. Stroeve, J. & Notz, D. Changing state of Arctic sea ice across all seasons. *Environ. Res. Lett.* **13**, 103001 (2018).

56. Box, J. E. et al. Key indicators of Arctic climate change: 1971-2017. *Environ. Res. Lett.* **14**, 45010 (2019).

57. Agilan, V. & Umamahesh, N. V. What are the best covariates for developing non-stationary rainfall Intensity-Duration-Frequency relationship? *Adv. Water Resour.* **101**, 11–22 (2017).

58. Salas, J. D., Obeysekera, J. & Vogel, R. M. Techniques for assessing water infrastructure for nonstationary extreme events: a review. *Hydrol. Sci. J.* **63**, 325–352 (2018).

59. Tabari, H. & Willems, P. Lagged influence of Atlantic and Pacific climate patterns on European extreme precipitation. *Sci. Rep.* **8**, 1–10 (2018).

60. Casanueva, A., Rodríguez-Puebla, C., Frías, M. D. & González-Reviriego, N. Variability of extreme precipitation over Europe and its relationships with tele-connection patterns. *Hydrol. Earth Syst. Sci.* **18**, 709–725 (2014).

61. Lavers, D., Prudhomme, C. & Hannah, D. M. European precipitation connections with large-scale mean sea-level pressure (MSLP) fields. *Hydrol. Sci. J.* **58**, 310–327 (2013).

62. Weisheimer, A., Schaller, N., O'Reilly, C., MacLeod, D. A. & Palmer, T. Atmospheric seasonal forecasts of the twentieth century: multi-decadal variability in predictive skill of the winter North Atlantic Oscillation (NAO) and their potential value for extreme event attribution. *Q. J. R. Meteorol. Soc.* **143**, 917–926 (2017).

63. Courty, L. G., Wilby, R. L., Hillier, J. K. & Slater, L. J. Intensity-duration-frequency curves at the global scale. *Environ. Res. Lett.* **14** (2019). https://doi.org/10.1088/1748-9326/ab370a.

64. Madec, G. et al. NEMO ocean engine. Note du Pôle de modélisation de l'Institut Pierre-Simon Laplace No 27 (2016). https://doi.org/10.5281/zenodo.1464816.

65. Fichefet, T. & Maqueda, M. A. Sensitivity of a global sea ice model to the treatment of ice thermodynamics and dynamics. *J. Geophys. Res. Ocean.* **102**, 12609–12646 (1997).

66. Dee, D. P. et al. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **137**, 553–597 (2011).

67. Zuo, H., Alonso-Balmaseda, M. A., Mogensen, K. & Tietsche, S. OCEAN5: the ECMWF ocean reanalysis system and its real-time analysis component. *ECMWF Tech. Memo.* (2018) https://doi.org/10.21957/la2v0442.

68. Azad, R. & Sorteberg, A. Extreme daily precipitation in coastal western Norway and the link to atmospheric rivers. *J. Geophys. Res. Atmos.* **122**, 2080–2095 (2017).

69. Zhang, X. et al. Indices for monitoring changes in extremes based on daily temperature and precipitation data. *Wiley Interdiscip. Rev. Clim. Chang.* **2**, 851–870 (2011).

70. Hoyer, S. & Hamman, J. J. Xarray: N-D labeled Arrays and Datasets in Python. *J. Open Res. Softw.* **5** (2017) https://doi.org/10.5334/jors.148.

71. Wilks, D. S. *Statistical Methods in the Atmospheric Sciences*. vol. 100 (Academic press, 2011).

72. Wickham, H. *ggplot2- Elegant Graphics for Data Analysis*. vol. **77** (Springer International Publishing, 2017).

73. Press, S. J. Applied multivariate analysis: using Bayesian and frequentist methods of inference. (Courier Corporation, 2005).

74. Hyndman, R. J. & Fan, Y. Sample quantiles in statistical packages. *Am. Stat.* **50**, 361–365 (1996).

75. Gilleland, E. et al. extRemes 2.0: an extreme value analysis package in R. *J. Stat. Softw.* **72**, 1–39 (2016).

76. Katz, R. W., Parlange, M. B. & Naveau, P. Statistics of extremes in hydrology. *Adv. Water Resour.* **25**, 1287–1304 (2002).

77. Katz, R. W. *Statistical methods for nonstationary extremes. in Extremes in a Changing Climate* 15–37 (Springer, 2013).

78. Rojas, R., Feyen, L. & Watkiss, P. Climate change and river floods in the European Union: Socio-economic consequences and the costs and benefits of adaptation. *Glob. Environ. Change* **23**, 1737–1751 (2013).

AUTHOR CONTRIBUTIONS

T.K. and M.M. conceived and T.K, M.M, L.J.S., T.M., R.L.W., C.P., P.B. designed the study. T.K. drafted the paper with extensive contributions from L.J.S., T.M., R.L.W., C.P., and M.M. T.K. analysed the data with input from all authors. M.M. acquired the data. T.K. produced the figures; L.F. produced Supplementary Fig. 3.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information is available for this paper at https://doi.org/10.1038/s41612-020-00149-4.

Correspondence and requests for materials should be addressed to T.K.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.