

Sensitivity of extreme precipitation to climate change inferred using artificial intelligence shows high spatial variability

Leroy J. Bird ¹, Gregory E. Bodeker ^{1,2}✉ & Kyle R. Clem ²

Evaluating how extreme precipitation changes with climate is challenged by the paucity, brevity and inhomogeneity of observational records. Even when aggregating precipitation observations over large regions (obscuring potentially important spatial heterogeneity) the statistics describing extreme precipitation are often too uncertain to be of any practical value. Here we present an approach where a convolutional neural network (an artificial intelligence model) is trained on precipitation measurements from 10,000 stations to learn the spatial structure of the parameters of a generalised extreme value model, and the sensitivity of those parameters to the annual mean, global mean, surface temperature. The method is robust against the limitations of the observational record and avoids the short-comings of regional and global climate models in simulating the sensitivity of extreme precipitation to climate change. The maps of the sensitivity of extreme precipitation to climate change, on $\sim 1.5 \text{ km} \times 1.5 \text{ km}$ grids over North America, Europe, Australia and New Zealand, derived using the successfully trained convolutional neural network, show high spatial variability.

¹Bodeker Scientific, 42 Russell Street, Alexandra 9320, New Zealand. ²School of Geography, Environment and Earth Sciences/Te Kura Tātai Aro Whenua, Victoria University of Wellington/Te Herenga Waka, Kelburn, Wellington 6140, New Zealand. ✉email: greg@bodekerscientific.com

Extreme precipitation events (EPEs) are among the most damaging climate hazards^{1–3} and, as such, changes in their frequency and/or severity are of major societal concern. A heuristic used by many practitioners, such as civil engineers, based on the Clausius–Clapeyron relationship (CCR), is that a 1 °C increase in temperature leads to a 7% increase in atmospheric water holding capacity and hence a 7% increase in the severity of EPEs. This, however, is not what happens in reality as myriad other processes affecting precipitation are also affected by climate change, chief among them being local and large-scale dynamics, which can change the location of where rain falls^{4–6}. A more comprehensive discussion of variations from the CCR expectation to estimating the sensitivity of extreme precipitation to climate warming is provided in “Discussion”. The focus of this paper is on deriving a more empirical estimate of the sensitivity of EPEs to climate warming.

A common way to describe the statistics of EPEs is using Generalized Extreme Value (GEV) distributions⁷. In GEV analyses, an equation of the form:

$$G(z; \mu, \sigma, \xi) = \exp \left\{ - \left[1 + \xi \left(\frac{z - \mu}{\sigma} \right) \right]^{1/\xi} \right\} \quad (1)$$

is fitted to N -year precipitation block maxima (i.e., the maximum 1-day precipitation in an N -year period), where μ is the location parameter, σ is the scale parameter, and ξ is the shape parameter. While 10-year block maxima should ideally be used ($N=10$)⁸, given the paucity of multi-decadal reference-quality⁹ precipitation records it is typically impossible to obtain sufficient 10-year block maxima to achieve a robust statistical description of observed EPEs at a single site, and thus 1-year block maxima are most often used in GEV analyses.

In the absence of unbiased multi-decadal observational records, global climate models (GCMs), and in particular large ensemble simulations from GCMs, might be considered as a way to derive robust extreme precipitation statistics [e.g., ref. 10]. However, as a result of their coarse resolutions, GCMs are often biased in their representation of precipitation^{11–13} especially extreme precipitation and/or precipitation over complex terrain^{13,14}. Different downscaling methods might be considered to mitigate the effects of the coarse resolution of GCMs. While dynamical downscaling of GCM output using regional climate models^{15,16} can create high resolution data sets, the approach is computationally demanding. Furthermore, if the GCM providing the boundary conditions to the regional climate model cannot adequately simulate the dynamical and thermodynamical processes governing the evolution of changes in the statistical distribution of precipitation (e.g., shifts in storm tracks), the regional climate model will simulate incorrect precipitation fields, albeit at high spatial resolution. On the other hand, statistical downscaling, while computationally less expensive than dynamical downscaling, generally assumes that the climate is stationary and thus is unlikely to simulate correct changes in precipitation extremes, and therefore is subject to the same shortcomings inherent in the controlling GCM.

Whether applying GEV analyses to observations, GCM output, or regional climate model output, a common approach to inferring the sensitivity of extreme precipitation to climate change is to expand the GEV fit coefficients as a linear scaling of some climate change covariate, e.g.:

$$\mu = \mu_0 + \mu_1 \times X \quad (2)$$

where X might be the annual mean global mean surface temperature anomaly. However, inferring such non-stationarity in the GEV coefficients from relatively short historical observational records, often compromised by inhomogeneities introduced by changes in instrumentation or operating procedures^{17–19}, as well as being subjected to the effects of internal climate variability, is

seldom sufficiently statistically robust to reliably estimate the sensitivity of EPEs to changes in global temperature.

Here we explore a method whereby a convolutional neural network (CNN)^{20,21} is trained to learn the parameters of a GEV distribution, their dependence on latitude, longitude and elevation, and their dependence on a climate covariate (in this case the annual mean global mean surface temperature; T'_{Global}). We choose T'_{Global} as the climate covariate not so much as to perform an attribution of the observed changes in extreme precipitation, but to obtain a robust and simple measure of the sensitivity of extreme precipitation to climate change that can be compared across any region of the globe. If the goal is to explain the variability in extreme precipitation, additional regional climate covariates, which could even be opposite in sign to T'_{Global} , could also be included. However, in this study, we focus on quantifying the sensitivity of regional extreme precipitation to the global state of the climate system only, which drives, in large part, the total moisture carrying capacity of the atmosphere.

The CNN is deep in that it consists of several stacked layers of neurons that allow a range of features at different spatial scales to be captured. We consider observations of daily total precipitation from 10,000 sites in North America, Europe, Australia and New Zealand extracted from the Global Historical Climatology Network (GHCN) and augmented with data from the New Zealand National Climate Database (CLIDB). Justification for choosing these four regions is provided in the next section. To demonstrate the value of the CNN, we also show some results where GEV fits were applied individually to the precipitation records at each of these sites, noting that the temporal coverage of the observations is likely to be heterogeneous and often of just a few decades in length. The results, shown below, indicate that the uncertainties on the GEV fit parameters derived individually for each site, and especially on the parameters describing the non-stationarity in the GEV fits, are so large as to make the derived sensitivities unusable and unphysically spatially heterogeneous.

Our intention is that by developing a single CNN, with bespoke training for each region, and that generalizes over thousands of locations to capture the geographical and non-stationary nuances of the GEV fit coefficients, we can determine the sensitivity of precipitation extremes to changes in global mean surface temperature over our four target regions in a way that is robust against any single site with inhomogeneous, or otherwise biased, observations and is able to generate maps at high spatial resolution (e.g., $0.015625^\circ \times 0.015625^\circ$) including across regions where there are no observational records.

The architecture, training, validation and evaluation of the CNN are described in “Methods”. We show that the CNN demonstrates skill in capturing the sensitivity of extreme precipitation to the global temperature change (T'_{Global}) and that maps of these sensitivities for different average recurrence intervals (ARIs) are highly spatially heterogeneous, providing further evidence that the use of CCR-based heuristics is too simplistic.

Results

GEV fits. The CNN was trained on daily total precipitation depths at 10,000 GHCN+CLIDB sites (shown in Fig. 1) from 1960 to 2019 during which T'_{Global} increased by ~ 0.9 °C. When selecting these 10,000 sites, it was found that regions of the globe outside of North America, Europe, Australia and New Zealand had severely inhomogeneous data coverage both in space and time. Reductions in the number of GHCN stations (see, e.g., Fig. 2 of Wan et al.,²² and Fig. 1 of Westra et al.²³), or the volume of data reported by stations, was found to be particularly problematic for training the CNN (see Methods). As a result, only GHCN and CLIDB sites within North America, Europe, Australia

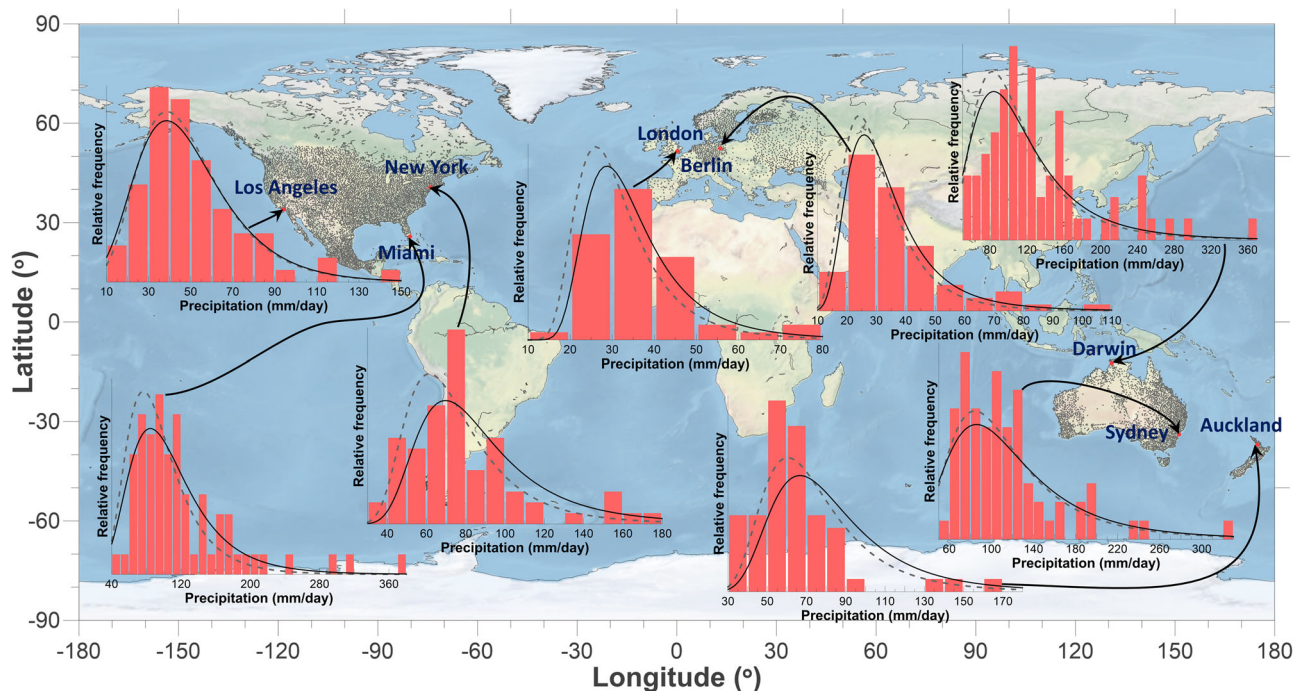


Fig. 1 The eight sites used to demonstrate the ability of the CNN to generalise spatially and the GEV fits derived by the CNN for those sites under two levels of climate change. The 10,000 sites used for the training are shown with grey dots while the eight sites are shown using red dots. Insets show histograms of observed 1-year block maxima of daily precipitation at the eight sites as well as the GEV fits for those sites, generated by the CNN, for T'_{Global} of 0°C (dashed grey curve) and 1.5°C (solid black curve).

and New Zealand were used in training the CNN and only results for these four regions are reported on below.

Eight sites representing large human populations, excluded from the training, were selected to demonstrate the skill of the CNN in simulating GEV distributions for precipitation maxima unseen by the CNN (red dots in Fig. 1). Block maxima from these sites and the CNN generated distributions for two different values of T'_{Global} , are also shown in Fig. 1. Other than for Auckland, which had 55 1-year block maxima available, and London which had 51, all other sites had 70 1-year block maxima. Additional evidence of the skill of the CNN in generalising spatially is provided in “Evaluating the ability of the CNN to spatially generalise the GEV functions”.

1% annual exceedance probability precipitation levels. The 1% annual exceedance probability (AEP; 1-in-100 year event) precipitation levels, derived from the CNN-simulated GEV probability distribution functions evaluated on a grid of $0.015625^{\circ} \times 0.015625^{\circ}$ at a T'_{Global} of 1.1°C , are shown in Fig. 2. To provide a climatological context for these 1% AEP extremes, 100 GHCN sites (50 for New Zealand) were selected, each with a record of 20 years or more. The 1% AEP value at each site (coloured regions in Fig. 2) was then divided by the average (over the 20 years or more) annual maximum 1-day precipitation at each of those sites. These values are shown in small circles at each selected GHCN site in Fig. 2. They indicate that for North America the 1-in-100 year precipitation depth is $243 \pm 35\%$ (1σ) of a typical annual maximum 1-day precipitation depth, while for Australia this is $268 \pm 45\%$, for Europe $244 \pm 30\%$ and for New Zealand $246 \pm 38\%$. This suggests a potentially useful heuristic that 1% AEP precipitation depths are typically 2.5 times the magnitude of an annual maximum 1-day precipitation depth.

The CNN has learned the effects of topography and climate regimes, and has used this knowledge to predict higher

precipitation extremes over moist tropical latitudes, coastlines along warm oceans, and on the windward side of large mountain ranges, e.g., along the western coasts of North America, Norway, and New Zealand.

The 1% AEP field for North America is qualitatively very similar to that shown in Fig. 2(a) of Fix et al.²⁴ while the morphology of the 1% AEP field for Europe is very similar to that shown in Fig. 4(a) of Zittis et al.²⁵.

The results for New Zealand show good comparison with the High Intensity Rainfall Design System product developed by the New Zealand National Institute of Water and Atmospheric Research (<https://hirids.niwa.co.nz/>); CNN-derived 1% AEP values shown in Fig. 2(d) were compared against similar values extracted from the High Intensity Rainfall Design System database for 50 population centres around New Zealand, showing a low bias by the CNN of 1.8% and an R^2 value of 0.84.

To demonstrate the value our CNN approach, 1% AEP precipitation depths, derived individually for each of the GHCN sites with precipitation records of 35 years or longer, using the same 6-parameter non-stationary GEV fit as used by the CNN, are shown in Supplementary Fig. S1. While the overall morphology of the 1% AEP precipitation depths in each region is similar to that shown in Fig. 2, results are absent where there are no GHCN sites and, in places, show site-to-site variability that is physically unrealistic. The statistically robust GEV fits shown in Fig. 1 were only made possible by the CNN gaining insights into the broad underlying statistical structure of the precipitation extremes through training across 1-year block maxima from all 10,000 training sites. While we have not yet developed a statistically comprehensive method for extracting uncertainties on the GEV fit coefficients from the CNN, the spatial coherence of the CNN-generated GEV fits, the site-to-site consistency in their non-stationarity (and the evaluation of the non-stationarity - see “Methods”), and the ability of the CNN to generalise well for sites excluded from its training (Fig. 1 and “Evaluating the ability

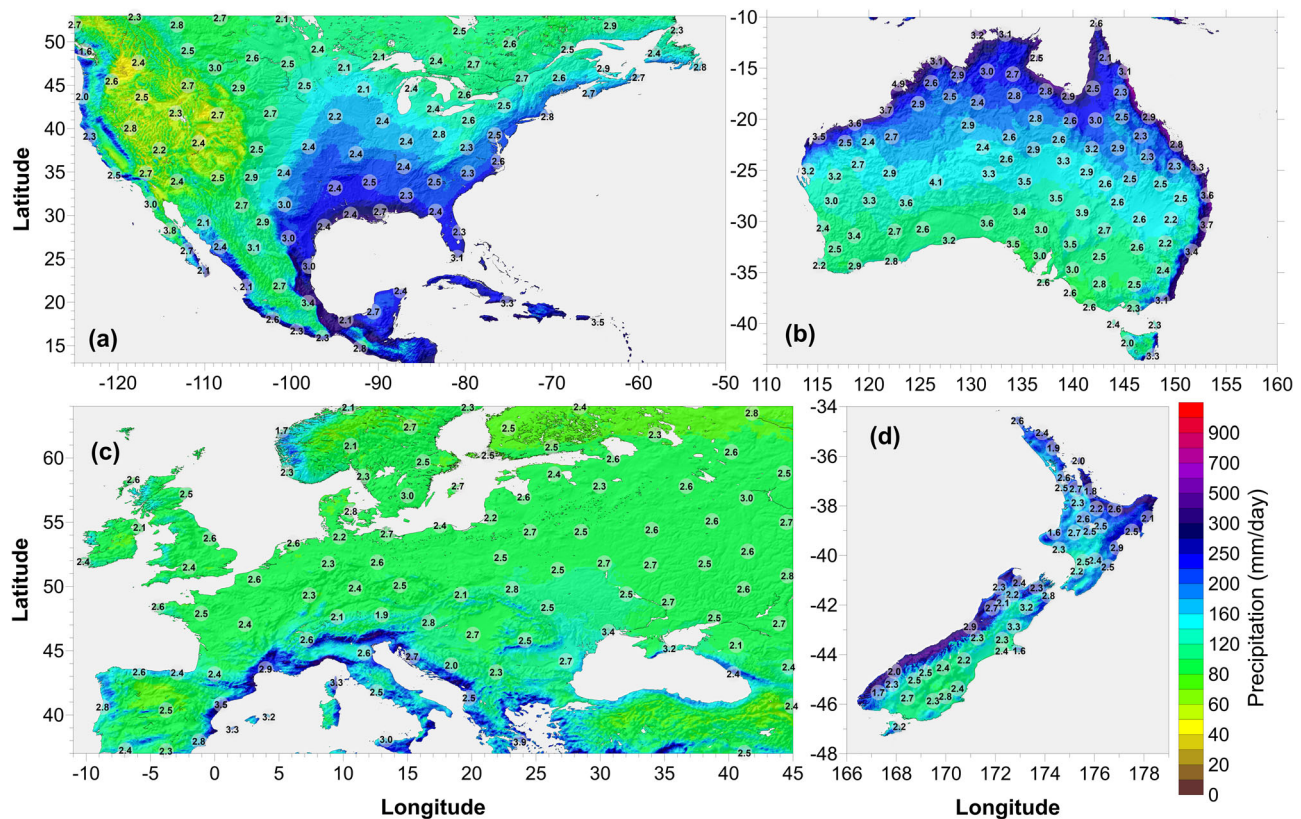


Fig. 2 1% annual exceedance probability precipitation depths. These have been evaluated at a T'_{Global} of 1.1°C for (a) North America, (b) Australia, (c) Europe, and (d) New Zealand. The underlying shaded relief shows the topography. Note the non-linear colour scale. Values in circles show the scaling with respect to the long-term (20 years or more) average of the annual maximum 1-day precipitation.

of the CNN to spatially generalise the GEV functions”), provides confidence that the CNN-generated GEV fits are statistically robust; a more complete evaluation of the CNN-generated GEV fits, and an evaluation of the ability of the CNN to correctly infer the non-stationarity in the GEV fits, is presented in “Methods”.

5% annual exceedance probability precipitation levels. 5% AEP (1-in-20 year event) precipitation maps, similar to those shown in Fig. 2, are shown in Supplementary Fig. S2. As expected, the magnitude of a 1-in-20 year event is uniformly lower than a 1-in-100 year event. The scaling values shown in the small circles (similar to those shown in Fig. 2) indicate that for North America the 1-in-20 year precipitation depth is $181 \pm 25\%$ of a typical annual maximum 1-day precipitation depth, while for Australia this is $187 \pm 26\%$, for Europe $177 \pm 20\%$ and for New Zealand $180 \pm 28\%$, suggesting a potentially useful heuristic that 5% AEP precipitation depths are typically 1.8 times the magnitude of an annual maximum 1-day precipitation depth.

Precipitation depth vs. ARI. A view of precipitation depth for a wider range of ARIs is provided in Fig. 3. Several sites show increases in precipitation depth with increasing ARI. The site in western Haiti (A in Fig. 3; 18.375°N , 73.875°W) shows the 1-day precipitation depth increasing from 333 mm for ARI = 100 years to 492 mm for ARI = 1000 years. Similarly, the site south-east of New Orleans (B; 29.688°N , 89.812°W) shows the 1-day precipitation depth increasing from 308 mm for ARI = 100 years to 432 mm for ARI = 1000 years. A few sites in Australia show even greater relative changes in precipitation depth with ARI; five of the 100 sites analysed (A-E in Fig. 3) show 1000-year ARI precipitation depth exceeding 700 mm, all along the northern

coastline, up from 445–480 mm for 100-year ARI. The European site close to Nice (A; 43.891°N , 7.391°E) shows the precipitation depth increasing from 302 mm for ARI = 100 years to 566 mm for ARI = 1000 years. The other European site with 1000-year precipitation exceeding 400 mm is that on the Croatian coastline (B) at 416 mm. The four sites in New Zealand with 1-in-1000 year precipitation exceeding 600 mm are all at high elevations in the Southern Alps along the south-west coast of the South Island (A–D in Fig. 3).

Change in the magnitude of 1% annual exceedance probability per degree warming. Here we explore the percentage change in the severity of 1% AEP events learned by the CNN for each 1°C change in T'_{Global} . The results presented in Fig. 4 are, essentially, the sensitivities of extreme precipitation to climate change quantified as T'_{Global} . Because the climate signal in observed extreme precipitation time series is often weak and spatially highly variable, especially at smaller scales, statistically robust derivation of the sensitivity of precipitation extremes to climate change at continental or smaller-scales, based solely on observational time series, has been challenging; Sarojini et al.²⁶ concluded ‘compelling evidence of anthropogenic fingerprints on regional precipitation is obscured by observational and modelling uncertainties and is likely to remain so using current methods for years to come’. Our CNN-based analysis of annual maximum daily precipitation demonstrated here suggests a way forward to derive these sensitivities at regional and sub-regional scales.

Sun et al.²⁷ applied GEV analyses to the HadEX2 data set²⁸ in which T'_{Global} was also used as the climate covariate, finding that the global median percentage change in extreme precipitation per $^\circ\text{C}$ increase in T'_{Global} is 6.6% (5.1% to 8.2%; 5%-95% confidence

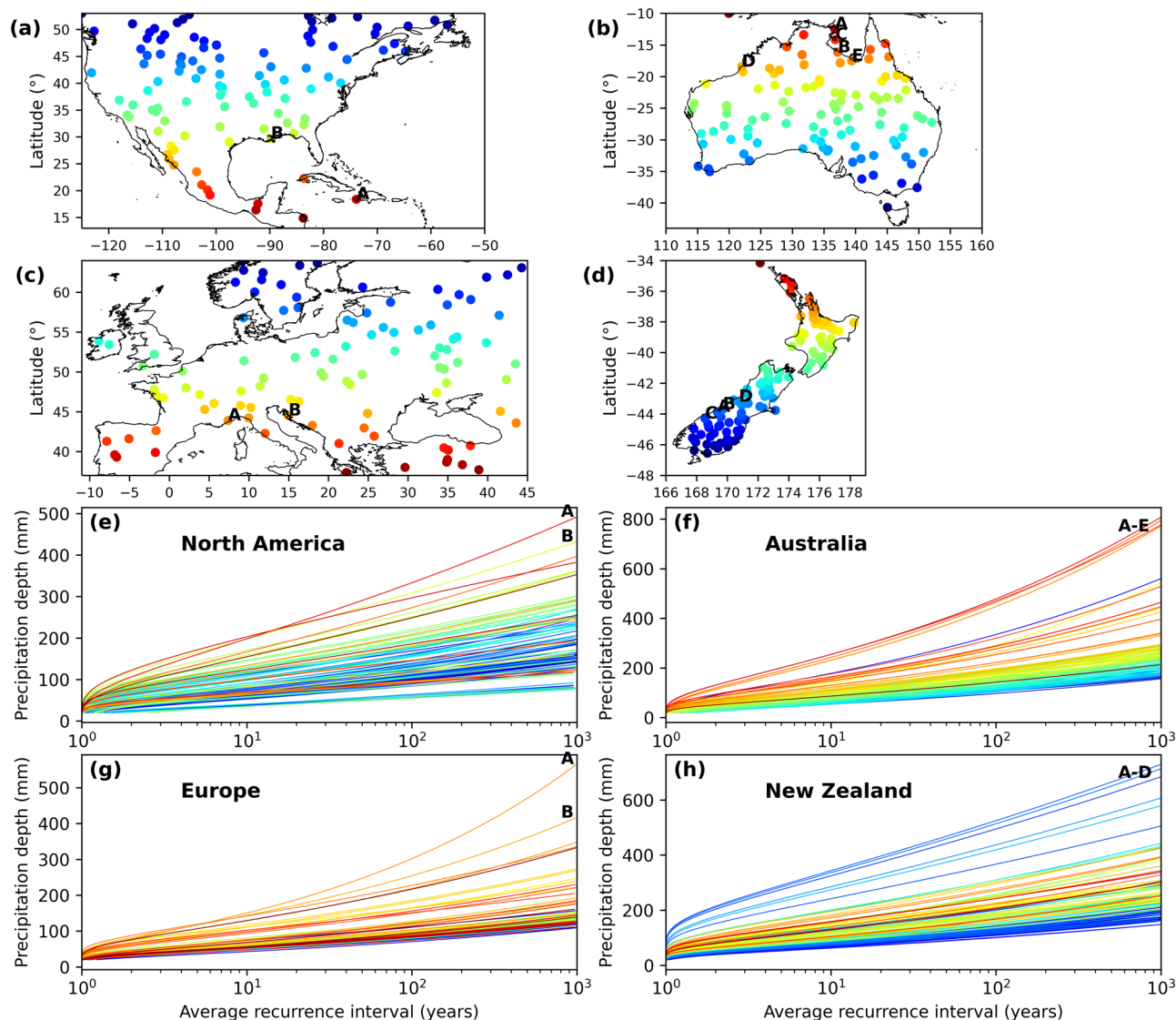


Fig. 3 Precipitation depth vs. average recurrence interval. Results are shown for 100 randomly selected sites in each of the four regions. Panels (a–d) show the location of the selected sites while panels (e–h) show the precipitation depth vs. ARI curves for each of those sites, matched by colour. Locations closer to the equator are shown in red while locations closer to the pole are shown in blue. Uppercase letter labels refer to specific sites discussed in the text and where labels are ordered from largest to smallest precipitation depth at ARI = 1000 years.

interval) for annual maximum daily precipitation (Rx1day). Over the four regions analysed here we find a mean towards the lower end of this range of 5.1% (with a much smaller 95% confidence interval of just 0.0026% given the huge number of data from all four grids). Westra et al.²³ found sensitivities of 5.9% to 7.7%. These differences likely result from differences in geographical coverage of the sites analysed and the analysis methods employed.

Over North America, changes in the severity of 1% AEP precipitation per °C increase in T'_{Global} vary from decreases of a little more than 10% to increases of around 18% with an interquartile range of 1.4% to 11.2%. These expected changes in extreme precipitation, derived only from the observational record, are consistent with previous analyses of precipitation observations over North America²⁹ that have shown increases in extreme precipitation, and with model-based analyses³⁰ which showed that external forcing, dominated by human influence, has contributed to the increase in frequency and intensity of precipitation extremes in North America. Changes along the southeast Gulf coast are around 14% increase per °C increase in T'_{Global} while along the east coast values frequently exceed 16%.

These increases may be related to expected increases in hurricane-induced rainfall³¹; noting however that this cannot be inferred from the CNN which has no internal representation of hurricanes, only the precipitation depth measured during hurricanes. Over the Rocky Mountains and other mountainous regions of North America, decreases approaching 10% are inferred. These negative sensitivities observed over Northwest North America are consistent with Sun et al.²⁷ who found that 56% of stations in this region experienced decreases in the severity of Rx1day precipitation between 1950 and 2018.

For Australia, the CNN infers changes of -20.2% to 5.9% with an interquartile range of -6.4% to -0.6%; most of the interior of the continent is simulated to experience negative sensitivities in extreme precipitation to global climate warming. In contrast, the eastern and northern coasts near warm oceans/currents are simulated to experience small positive sensitivities. Dey et al.³² attributed the increase in extreme precipitation in the north-west of Australia since 1950 to increased monsoonal flow due to increased aerosol emissions and not to an increase in atmospheric greenhouse gas loading. The sensitivities of extreme precipitation

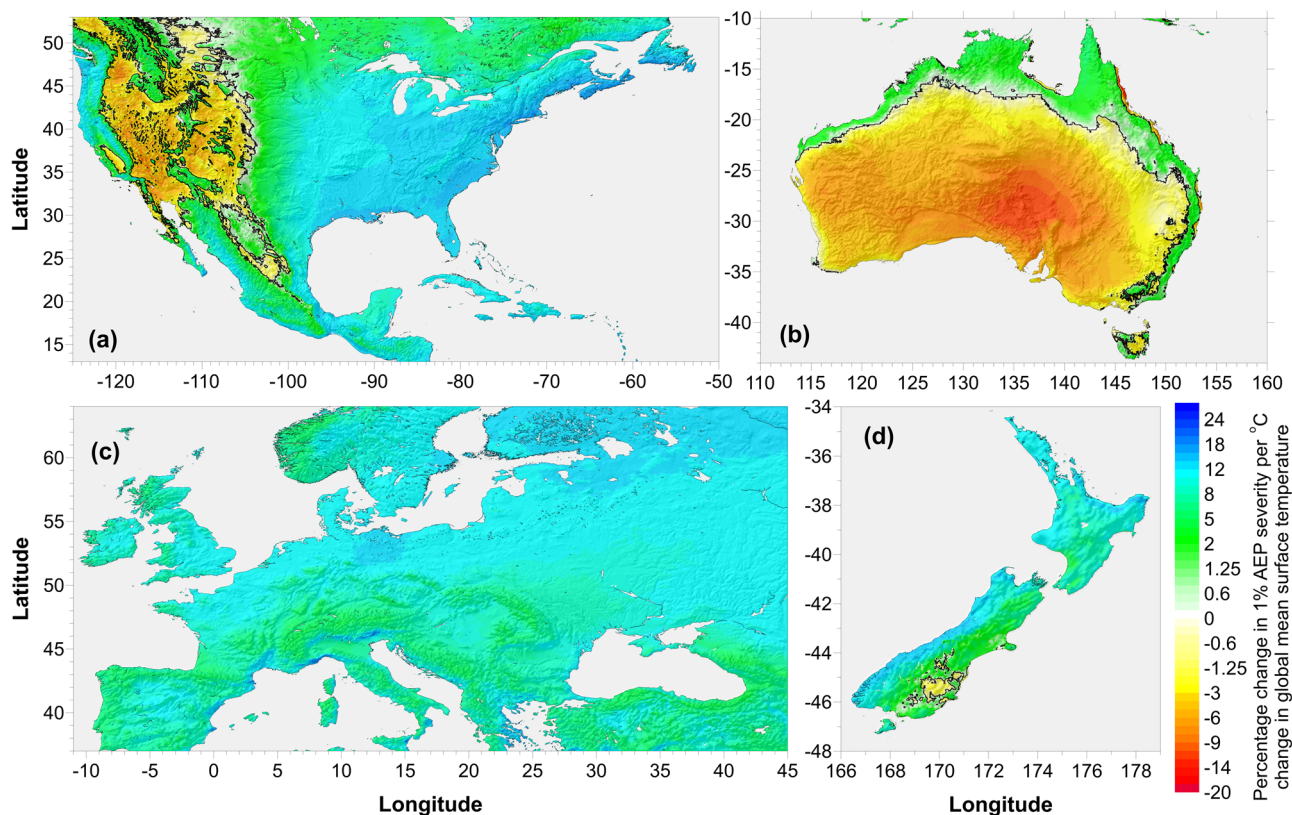


Fig. 4 Percentage change in the severity of 1% AEP precipitation per 1°C increase in T'_{Global} . These have been evaluated between a T'_{Global} of 0.2°C and 1.2°C for (a) North America, (b) Australia, (c) Europe, and (d) New Zealand. Note the non-linear colour scale with yellow, orange, and red showing decreases in 1% AEP severity and green, cyan and blue denoting increases. The thick black contour shows the 0% change line.

to climate warming that we have inferred are consistent with previous reports³³ of decreases in precipitation in austral autumn and winter over parts of southern and especially southwestern Australia in the past few decades. Likewise, Sun et al.²⁷ found that 56% of stations in Australasia showed negative sensitivities to T'_{Global} with average sensitivities of $\sim -6.5\%$ per $^{\circ}\text{C}$. Delworth and Zeng³⁴ used observations and climate models to show that some of this decline results from changes in large-scale atmospheric circulation, including a poleward shift of the westerly winds and an increase in atmospheric surface pressure over parts of southern Australia attributed to anthropogenic increases in atmospheric greenhouse gas concentrations and stratospheric ozone depletion. Pfahl et al.⁵, in diagnosing climate model simulations, found that while thermodynamics alone would lead to a spatially homogeneous positive sensitivity of extreme precipitation to warming, which is consistent across models and dominates the sign of the change in most regions, the dynamic contribution modifies regional responses, weakening them across the Mediterranean, South Africa and Australia. They found that over subtropical oceans, the dynamic contribution is strong enough to cause regional decreases in extreme precipitation, which may partly result from a poleward shift of the storm track. Such a shift has been a dominant attribute of Southern Hemisphere climate change over recent decades^{35,36}. Such dynamical offsetting of the expected thermodynamic increase in extreme precipitation under climate warming may suggest a potential mechanism underlying our observed negative response of extreme precipitation over the interior of Australia to warming.

For Europe, the CNN infers fairly homogeneous sensitivities of between 3.9% and 23.7% with an interquartile range of 8.2% to 11.1%, slightly larger than what would be expected from application of the CCR. The (predominantly) positive sensitivities

over Europe and North America are consistent with previously inferred anthropogenic influence on annual precipitation maxima over Northern Hemisphere land areas^{37,38}.

For New Zealand, the sensitivities fall between -2.7% and 16.0% with an interquartile range of 4.8% to 10.5% and a strong west-east gradient across the South Island driven by the dominance of the westerly flow and the topographic barrier of the Southern Alps. In the East Cape region of northeast New Zealand, where 1% AEP events are close to 480 mm day^{-1} and frequently cause damage³⁹, the CNN estimates sensitivities of up to 15% for every $^{\circ}\text{C}$ of global warming.

To demonstrate the value of our CNN approach, precipitation sensitivities were derived individually for each of the GHCN sites with precipitation records of 35 years or longer, using the same 6-parameter non-stationary GEV fit as used by the CNN. The resultant sensitivities are shown in Supplementary Fig. S3. The site-to-site variability is so large as to render the results unusable. This is not surprising given that the non-stationarity in the GEV fits can only be inferred from as few as 35 1-year block maxima over the recent past and, as such, few of these sites will observe even one 1-in-100 year event. This is highlighted in Supplementary Fig. S4 where 1-year block maxima from two GHCN sites in North America, just 9.34 km apart, show very different derived precipitation sensitivities.

It is clear from this figure that a single anomalous block maximum can bias the non-stationarity in the GEV fits; the block maximum at site B late in the period (when T'_{Global} is higher) is likely what causes the GEV fit at this site to generate a long tail to model this value ($>100\text{ mm day}^{-1}$ in the histogram). Given the direct GEV fit-derived cumulative distribution functions plotted in blue in the bottom row panels of Supplementary Fig. S4, it is not surprising that at Site A a

sensitivity in 100-year ARI precipitation of $-22.6\% \text{ } ^\circ\text{C}^{-1}$ is inferred, while at Site B a value of $58.3\% \text{ } ^\circ\text{C}^{-1}$ is inferred. The CNN-derived GEV fits (red), and their associated cumulative distribution functions, are far less variable as the GEV statistics are not derived using only the block maxima from the site, but from many neighbouring sites as defined by the convolutions used in the network.

Sensitivity vs. ARI. A view of the sensitivity of changes in extreme precipitation to T'_{Global} for a wider range of ARIs is provided in Fig. 5. Over North America, sensitivities are similar to what would be expected from the CCR over ARIs of 10, 20, 50, 100 and 200 years, though with some indication of an increase in the spread of sensitivities with ARI and, as was seen in Fig. 4, with more southern sites generally experiencing higher sensitivities across all ARIs.

In Australia, the sensitivities are below the CCR expectation across all ARIs with a tendency, especially for the more northern sites, for the sensitivities to become more negative with increasing ARI.

Over Europe there is a clear tendency for the sensitivity of extreme precipitation to T'_{Global} to increase with ARI, especially for more northern locations. At an ARI of 200 years, most of the 100 randomly selected sites show sensitivities above the CCR expectation.

In contrast, sensitivities over New Zealand generally decrease with increasing ARI, with sensitivities spanning the CCR expectation at 10-year ARI and, especially for more northern locations, with sensitivities falling below the CCR expectation for northern locations at an ARI of 200 years.

Sensitivity vs. precipitation depth. Important questions that can also be addressed through this analysis are: How do precipitation sensitivities vary with average precipitation depth? Will climate change drive greater relative increases in extreme precipitation for locations that are already experiencing high precipitation? Plots of precipitation sensitivity (% change in precipitation depth per $^\circ\text{C}$ change in T'_{Global}) are shown in Fig. 6 for all four regions and for ARIs of 10, 20, 50, 100 and 200 years). Over North America, sites with higher average extreme precipitation (across all ARIs) tend to show

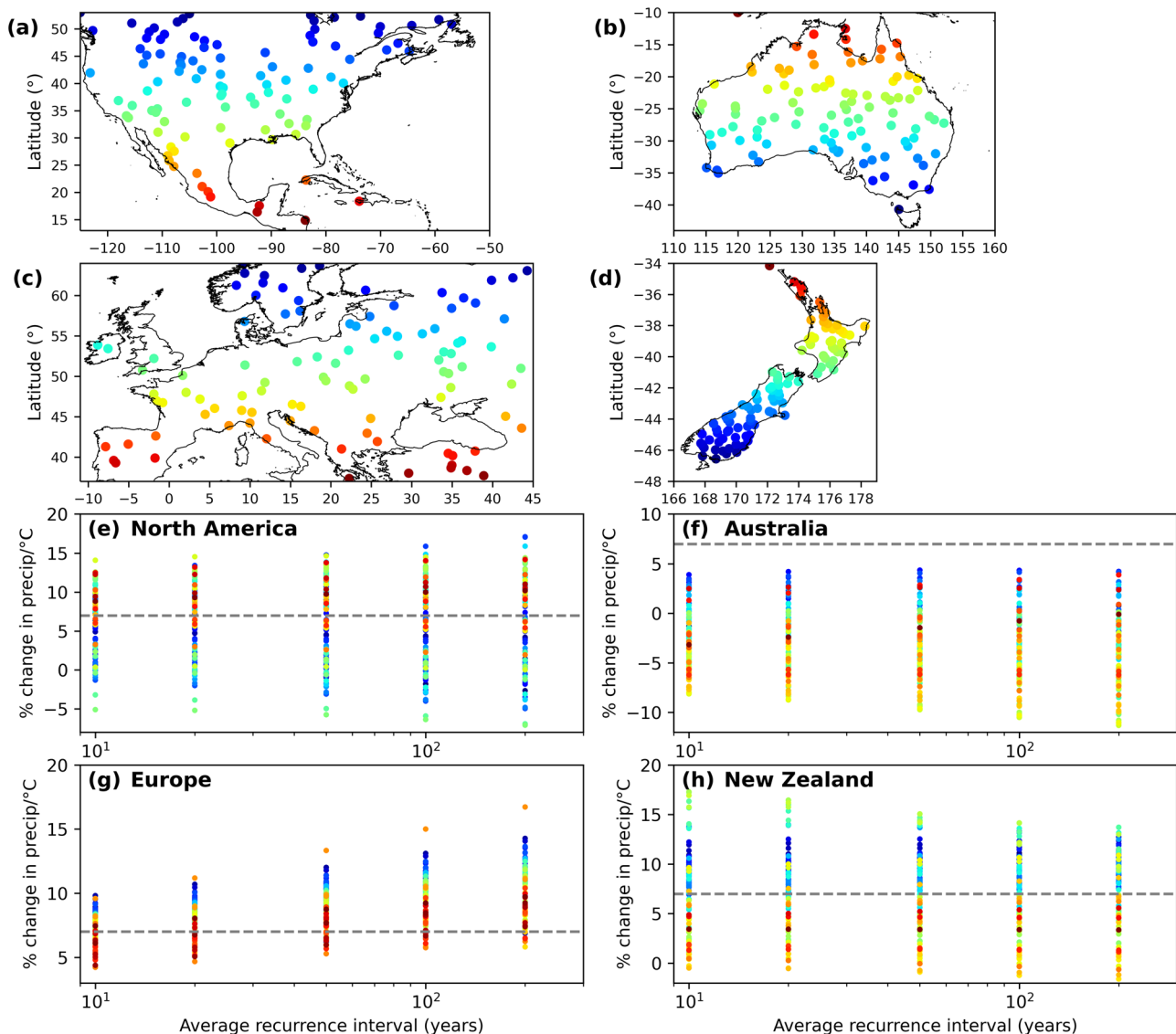


Fig. 5 Sensitivity of extreme precipitation depth to T'_{Global} vs. average recurrence interval. Site selection and colour coding (a-d) is as in Fig. 3 where panels (e-h) now show the precipitation depth sensitivity vs. ARI curves for each of those sites; the percent change expected from the CCR (7%) is plotted as a dashed line.

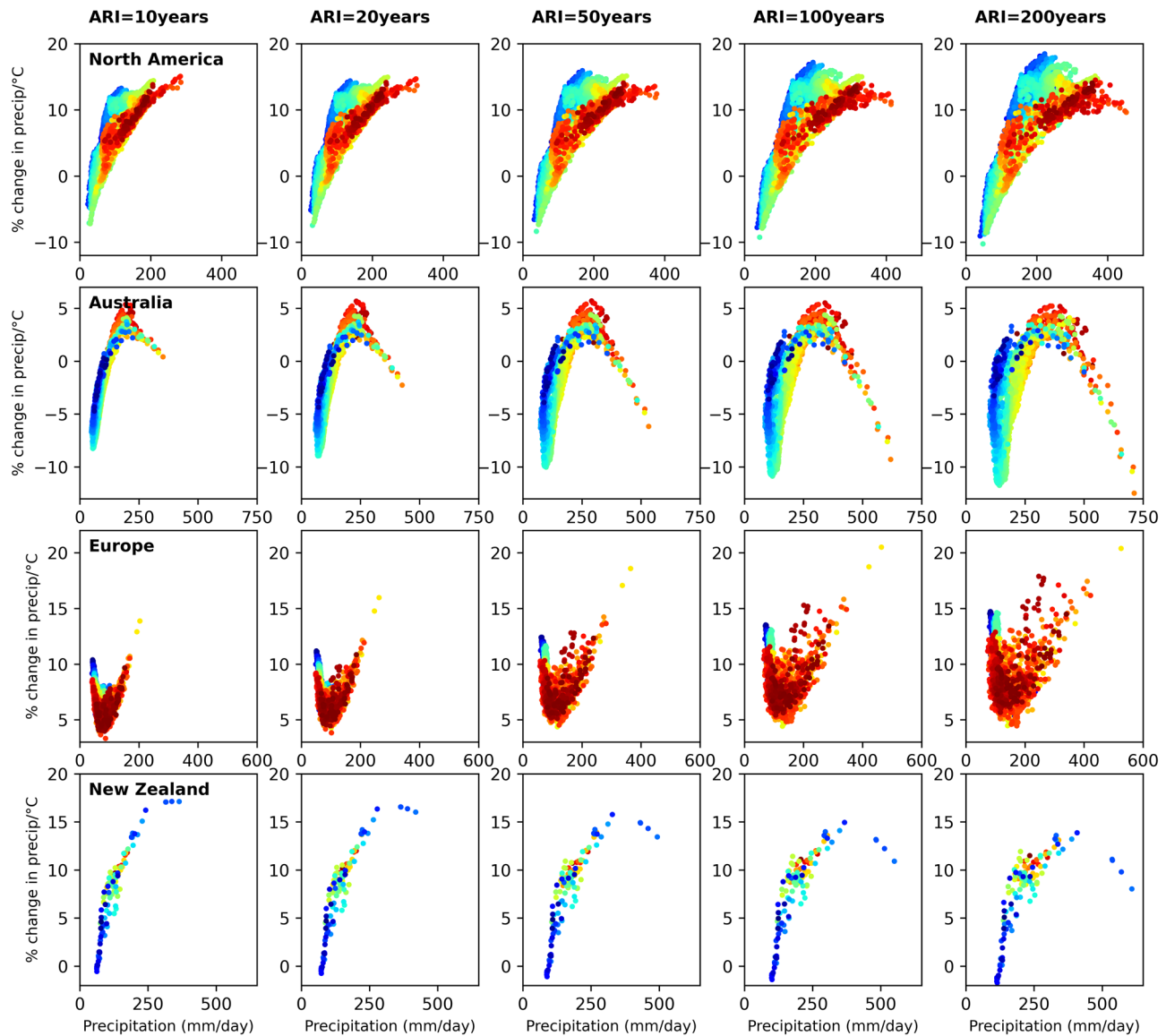


Fig. 6 Sensitivity of extreme precipitation to T'_{Global} vs. precipitation depth. Rows of panels show results for North America, Australia, Europe and New Zealand. Columns show results at ARIs of 10, 20, 50, 100 and 200 years. For clarity, the sensitivities were evaluated at $T'_{\text{Global}} = 1.2^\circ \text{C}$. Results are shown for one out of every 1000 locations in each region. More poleward locations are shown in blue and more equatorward locations in red as for the figures above.

larger sensitivities to climate change. This remains true for different latitudes - noting that the more northern sites generally experience larger extreme precipitation depths (see also Fig. 2 and Supplementary Fig. S2). The behaviour over Australia is quite different. While the relationship between precipitation sensitivity and precipitation depth is compact, it shows strongly non-linear behaviour, with sites with intermediate extreme precipitation depths (e.g., between 100 mm day^{-1} and $350\text{--}500 \text{ mm day}^{-1}$, the latter increasing with ARI) showing positive sensitivities, while sites with very low or very high extreme precipitation show negative sensitivities. This relationship appears to hold true independent of latitude.

For Europe the relationships are neither linear nor particularly compact. There is some sense that, in contrast to Australia, it is the sites with extreme precipitation depths closer to the mean that exhibit lower sensitivities to climate change. Sites with large extreme precipitation depths show the highest sensitivity to T'_{Global} .

For New Zealand the relationships are compact but indicate a divergence from linearity at the highest extreme precipitation values, especially at longer ARIs.

Change in the average recurrence interval for present day 1% AEP events. The sensitivities of the severity of EPEs to climate change were presented above. Here we consider the corollary, i.e., changes in the frequency of EPEs inferred from the CNN. Specifically, we consider present-day ($T'_{\text{Global}} = 1.1^\circ \text{C}$) 1% AEP events (1-in-100 year events) and consider how frequently these events occur (1-in- x year events) at $T'_{\text{Global}} = 2.0^\circ \text{C}$. Maps of the ARI at 2°C for 1-in-100 year events at 1.1°C , for all four regions, are shown in Fig. 7. While the results presented in Fig. 4, practitioners often wish to understand how the ARI for EPEs is expected to change with climate change. For North America, there are regions over the Intermountain West where the ARI extends to beyond 150 years, but over most of the continent the ARI decreases; the interquartile range for the continent is 60 to 93 years. There are regions along the Pacific north-west coast where the frequency of EPEs is projected to double under a 2°C warmer world compared to present day.

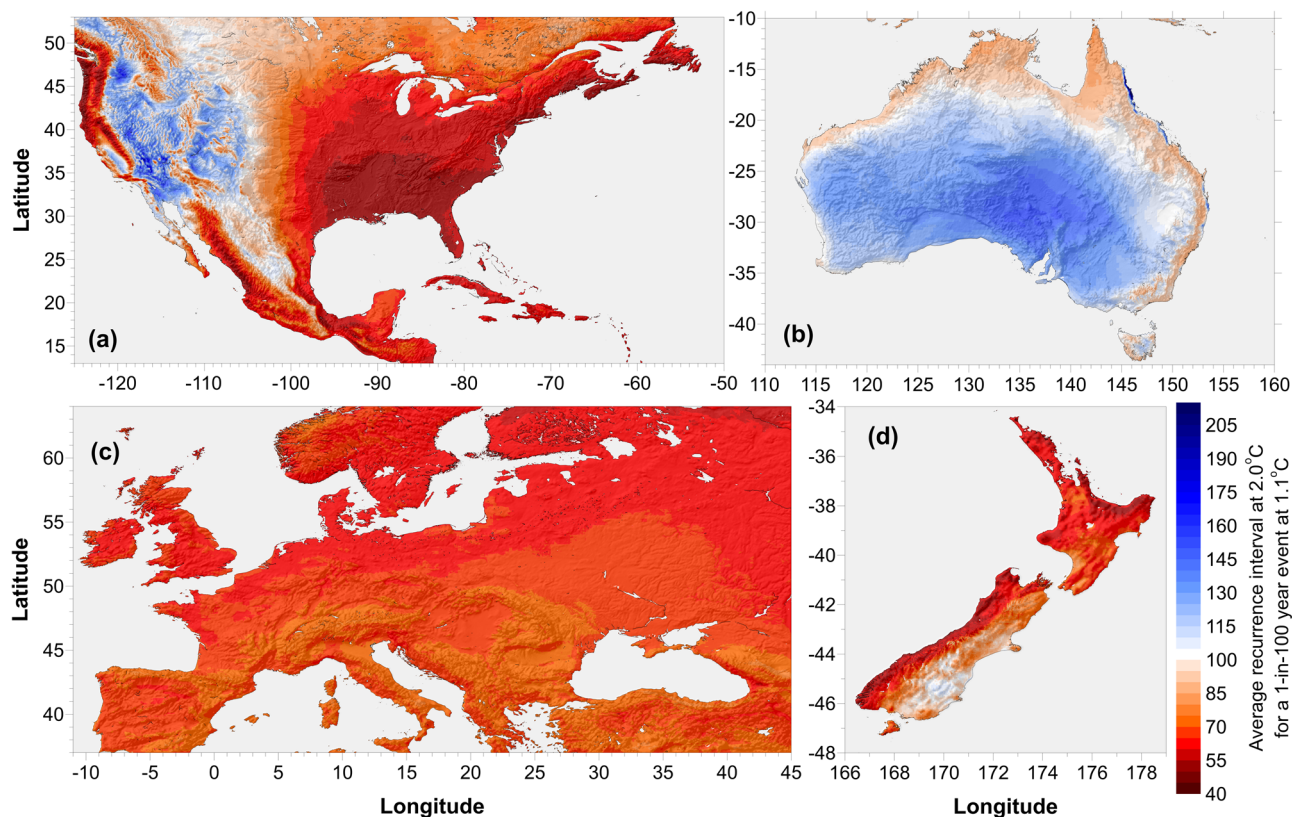


Fig. 7 Change in average recurrence interval. The frequency (1-in- x year) at which 1-in-100 year events under present day conditions ($T'_{\text{Global}} = 1.1^\circ\text{C}$) will occur under $T'_{\text{Global}} = 2.0^\circ\text{C}$ for (a) North America, (b) Australia, (c) Europe and (d) New Zealand.

Over the Cape York Peninsula and much of the coastal regions of eastern and northern Australia, the ARI reduces slightly to 80–90 years, though over most of Australia the ARI increases beyond 100 years; interquartile range of 103 to 131 years.

Over Europe there is an almost uniform reduction in the ARI under a 2°C warmer world with an interquartile range of 64 to 69 years. Similarly, over New Zealand, the ARI reduces over almost all of the country. Over the west coast of the South Island, the far north-east coast, the East Cape and the coastal regions of the Bay of Plenty, the frequency of EPEs under a 2°C world is expected to almost double compared to present day.

Discussion

The CNN detailed here represents, to the best of our knowledge, a new approach to simulating past, present and expected future changes in extreme precipitation. The results demonstrate the ability of a CNN to learn the statistics of EPEs, and, by including a climate covariate in that learning, to infer the sensitivity of extreme precipitation to climate change. The inferred sensitivity often deviates from the CCR expectation of a 7% increase per 1°C of warming. This is primarily because changes in local and large-scale dynamics strongly affect regional precipitation responses to climate change. For example, locally enhanced sea surface temperature warming along coastal regions can lead to stronger moisture intrusions and torrential rainfall in monsoon regions such as Asia^{40,41} as well to heavier coastal rainfall during tropical cyclones, such as in East Asia and the southeast United States^{42,43}. The vertical distribution of temperature change also affects local lapse rates and stability and can thus locally enhance precipitation in regions where instability increases, while regions that become more stable may experience a decrease in heavy precipitation (such as in the sub-tropics)⁵. Finally, large-scale shifts in storm tracks in response to climate change, such as a

poleward shift of the mid-latitude storm tracks, also affects where heavy precipitation occurs^{44,45}; regions receiving more storms may experience precipitation increases that surpass the CCR, while regions receiving fewer storms may experience a decrease in heavy precipitation with climate change (such as in the sub-tropics). Human activity independent of climate change, such as land-use change and aerosol emissions, can also affect local precipitation through changes in soil moisture availability⁴⁶, increased urban heat island-induced moisture convergence⁴⁷, and changes in aerosol-induced cloud condensation, latent heating, and efficiency in precipitation production^{48,49}. By using this large global data set of observed precipitation, which is inherently capturing such processes, to train the CNN, we have shown the CNN is able to learn these spatial heterogeneities without needing to simulate (or know about) the dynamical and thermodynamical processes driving the rainfall.

This work has shown the value of precipitation measurements from the GHCN. Recent declines in the number of reporting stations and/or the volume of data reported, compromised the application of this method in many regions. We encourage National Meteorological and Hydrometeorological Services to maintain and expand surface climate observing sites as these observational data are essential to conducting the type of analysis presented here that can then inform adaptation actions to EPEs.

There are several advantages to this CNN approach:

1. When being trained, the CNN can use station records of any length. It is not necessary to screen for sites with long homogeneous precipitation measurement series. Every block maximum from every site counts towards the training of the CNN. This greatly increases the number of sites that can be used for such analyses.

2. It is independent of global or regional climate models which are costly to run and have shortcomings in simulating precipitation extremes.
3. Once trained, the CNN can simulate expected changes in extreme precipitation under a wide range of greenhouse gas emissions scenarios as represented by different time series of T'_{Global} ; other climate covariates could be employed in future work for more detailed regional studies.

There are, however, some important caveats. The first is that the CNN can only derive precipitation sensitivities to climate change from historical observations and the sensitivities derived here from observations from 1960 to 2019 may not necessarily apply in the future, i.e., those sensitivities may themselves change. For example, over some region, past negative trends in the severity of EPEs may switch to positive trends in the future as a result of non-linearities in the climate system making the sensitivities derived from historical observations inapplicable. There is a large body of work [e.g., ref. 50–52] that highlights the challenges of determining climate trends in the presence of internal climate variability whose timescales may be similar to the lengths of the observational records. This CNN, lacking any intrinsic representation of the physics of the climate system and potential triggers for such non-linear behaviour, has no ability to learn of these non-linearities and incorporate them into its learning of precipitation sensitivities. A focus for future work will be to explore the possibility of using observational data as the basis for the CNN to learn the stationary statistics of extreme precipitation, and a preliminary learning of the non-stationary statistics, but then to augment the observational data with transient model simulations from which the CNN, using transfer learning, updates its representation of the non-stationarity of extreme precipitation.

Useful projections of expected changes in extreme precipitation require multiple lines of evidence. The CNN-based approach presented here to infer precipitation sensitivities that can then be used as a basis to project expected changes in extreme precipitation constitutes an additional line of evidence. This work highlights the value that artificial intelligence can bring to climate applications, in this case capitalising on the ability of a CNN to learn the regional and topographical variability of extreme precipitation statistics and, importantly, how those statistics change with global mean surface temperature, allowing the CNN to simulate EPE statistics continuously across space and at very high spatial resolution. As these artificial intelligence-based methods evolve, and as we learn how to imbue artificial intelligence-based methods with intrinsic knowledge of the physics underlying the system (so-called physics informed CNNs; ref. 53,54), these methods will become more useful in combining observations of the past with our best understanding of the physics of the system to provide robust projections of expected future changes.

Methods

Data. Daily precipitation data were obtained from the Global Historical Climatology Network (GHCN). The number of sites, and the volume of data provided for each site through the GHCN, has declined considerably in the past few decades. As such, for this analysis which requires long precipitation records over a dense network, only GHCN data for North America, Europe and Australia were found to be sufficient; for New Zealand the GHCN data needed to be supplemented with data from the New Zealand National Climate Database (CLIDB). Therefore, only data from sites within these four target regions of North America, Europe, Australia and New Zealand were considered for this analysis.

Data from 1950 to 2019 from the combined GHCN and CLIDB database (hereafter GHCN+CLIDB) were used in this study, with the first decade being withheld from the CNN training to be used for evaluation. Due to a lack of measurements in New Zealand during the 1950s, data from 1960 to 2019 were used instead, with the evaluation period covering the first five years (1960–1964). Data flagged within the GHCN database as poor quality were removed. Annual maximum daily precipitation values (so-called block maxima; $Rx1day$) were calculated for each site in each year for which 360 or more daily precipitation values were available.

Site selection started by identifying eight test sites for which long time series of daily precipitation data were available and which represented large population centres over a diverse range of climates. These were Los Angeles, Miami, New York, London, Berlin, Darwin, Sydney and Auckland. First, the site furthest from all of these evaluation sites was added to the list of sites to be used for CNN training. Then the site furthest from the eight evaluation sites and the site on the training list was identified and added to the list of sites for training. This process was repeated until 10,000 GHCN sites were on the list of sites to be used for training the CNN. As a consequence, the closest two sites to be used for CNN training were 28 km apart. Data from the eight evaluation sites were not used in the CNN training, nor data recorded during the evaluation time period. To ensure each site covered a reasonable time period, only stations with at least ten years of block maxima were selected.

As a climate covariate to allow the CNN to learn of non-stationarity in the GEV probability density functions, annual mean global mean near surface temperature data were obtained from HadCRUT.5.0.1.0⁵⁵. Anomalies were then calculated with respect to the period 1850 to 1900. To remove year-to-year variability from, e.g., El Niño events, a Savitzky–Golay filter was applied to the anomaly time series with a window length of 21 years and a polynomial order of 3.

CNN architecture. A schematic of the CNN architecture is displayed in Supplementary Fig. S5. This architecture is based on high-performing similar architectures with long heritages in machine vision^{56–58}. For each weather station in GHCN+CLIDB, the surrounding 64×64 cells at four different grid resolutions ($1/256^\circ$, $1/64^\circ$, $1/16^\circ$, and $1/4^\circ$), selected for their representation of different spatial scales over which precipitation might be homogeneous, were extracted from a land-sea mask and an elevation data set to provide 2D features (referred to as channels in machine vision) which can be used by the CNN to learn the dependence of the precipitation on the spatial morphology of the regions around each site at these different spatial scales. The 64×64 cell size was selected as a compromise between memory restrictions and the need to represent the spatial morphology of the precipitation around each site (i.e. out to scales of $16^\circ \times 16^\circ$) - there is unlikely to be relevant information beyond these scales. These 64×64 grids can be considered as images with two features (viz., land-sea mask and topography/elevation). An additional three features, calculated at each resolution, were used to encode positional information for each grid cell relative to the location of the weather station or location of interest. These three variables were the north-south distance, the east-west distance, and the Euclidean distance, between the grid cell or pixel and the centre pixel or grid cell.

At the highest resolution, $1/256^\circ$, an additional four features are added, namely: (i) the elevation of the station above sea level, (ii) the latitude of the station, and (iii) the sine and (iv) the cosine of the longitude of the station. These additional point-based features were repeated across 64×64 pixels so that they could be

concatenated with the other five features (see top left of Supplementary Fig. S5). All features were first normalised by subtracting the mean and dividing by the standard deviation.

The CNN architecture comprises four parallel branches, for each input feature scale, that are initially processed separately, but then later concatenated deeper in the network. The architecture includes a series of ResNet blocks⁵⁶ as well as many of the design features detailed in⁵⁷. Each resnet block consists of 3×3 convolutions, Mish⁵⁹ activation functions, and batch normalisations. ResNet blocks, also known as residual blocks, are a key component of the ResNet architecture for deep CNNs used in image classification and other computer vision tasks⁵⁶. ResNet blocks enable the training of very deep neural networks by solving the vanishing gradient problem. In a standard CNN, each layer transforms the input data into a feature representation which is then passed to the next layer. However, as the number of layers in a network increases, it becomes increasingly difficult to train the network due to the gradients becoming vanishingly small, preventing the network from learning. To solve this problem, ResNet blocks introduce skip connections that allow information to bypass one or more layers in the network. ResNet blocks take the input feature map and passes it through a sequence of convolutional layers and non-linear activation functions (in our case the Mish activation function), before adding the original input feature map to the output of the block. This creates a shortcut connection that allows the gradient to be directly propagated back to the earlier layers of the network, thereby mitigating the problem of vanishingly small gradients.

A version of the CNN, with a single branch and where all input scales were concatenated from the start, was initially developed. However, the four-branch version of the architecture shown in Supplementary Fig. S5 leads to better results. The reasoning behind this modification is that we seek to add a strong prior to the CNN such that each spatial resolution should initially be treated independently, only later combining the information after important features were extracted at each scale separately.

Pseudocode for the ResNet block is shown in the bottom right of Supplementary Fig. S5. The spatial resolution progressively halves each time by setting the stride of the first 3×3 convolution within a ResNet block to two. The number of intermediate features is progressively increased as the spatial resolution decreases. The final layer consists of an adaptive average pool, averaging the $2048 \times 2 \times 2$ output to $2048 \times 1 \times 1$, which is then flattened and fed through a final linear layer, with a dropout⁶⁰ probability of 50%. This linear layer outputs six coefficients (μ , σ and ξ and the coefficients describing their dependence on T'_{Global}).

Training the CNN. The loss function used to train the CNN is:

$$f(x) = -\log \epsilon + \begin{cases} \frac{1}{\sigma} \left[1 + \xi \left(\frac{x-\mu}{\sigma} \right) \right]^{-\frac{1}{\xi}-1} \exp \left[- \left(1 + \xi \left(\frac{x-\mu}{\sigma} \right) \right)^{-\frac{1}{\xi}} \right], & \text{if } \|\xi\| > \epsilon, \\ \frac{1}{\sigma} \exp \left[- \frac{x-\mu}{\sigma} - \exp \left(- \frac{x-\mu}{\sigma} \right) \right], & \text{if } \|\xi\| \leq \epsilon, \end{cases} \quad (3)$$

where x is a single observed block maximum, μ is the location parameter, $\sigma > 0$ is the scale parameter, and ξ is the shape parameter. An extra parameter ϵ , which we set to 10^{-5} , is added for numerical stability.

As alluded to above, the block maxima from the 10,000 training sites were divided into a training data set (1960 to 2019 for North America, Europe and Australia; 1965 to 2019 for New Zealand) and an evaluation data set (1950 to 1959 for North America, Europe and Australia; 1960 to 1964 for New Zealand). The later period was used for training since this period experienced a larger change in climate, allowing the CNN to

learn more robustly the non-stationarity in the GEV parameters. The CNN was trained to minimise the average negative log-likelihood of the expanded GEV distribution (see equation (3)) of a single randomly sampled batch of 64 block maxima. The RAdam optimizer⁶¹, with a weight decay of 0.01, was used as the inner optimizer within Lookahead⁶² which was configured with a slow weight step of 0.5 and synchronization period of 6 steps. Training was performed over four epochs with the learning rate fixed at 10^{-3} for epochs 1 and 2 and then decreasing to 10^{-7} by the end of epoch 4 using a cosine decay schedule. Early stopping was employed in selecting the epoch with the lowest validation loss. Training was first performed exclusively on the subset of training sites for North America. The CNN was then trained separately for Australia, New Zealand and Europe, in each case taking the pre-trained weights from North America and freezing all weights apart from those in the final output layer. Training was then performed for one epoch with a learning rate of 10^{-3} . All weights were then unfrozen and training was performed for another four epochs using the same training method as North America. The code was implemented using PyTorch⁶³ and fastai⁶⁴.

Evaluating the ability of the CNN to capture non-stationary in the GEV functions. To evaluate the ability of the CNN to capture the non-stationarity in the GEV distributions, the negative log-likelihood was calculated for all sites with observations available over the evaluation period, averaged across each region to avoid single site outliers resulting from natural variability in precipitation extremes (Fig. 8). Because minimising the negative log-likelihood is equivalent to maximising the likelihood, the minima in the blue traces in Fig. 8 show the value of T'_{Global} that maximises the probability of observing the precipitation block maxima. For North America and New Zealand the T'_{Global} estimated from the negative log-likelihood traces almost exactly match the expected average T'_{Global} for the corresponding evaluation period. The matches for Europe and Australia are close, indicating that the CNN was able to accurately generalise the non-stationary components of the GEV model beyond the range of the training period.

Evaluating the ability of the CNN to spatially generalise the GEV functions. To further verify the CNN's ability to generalise spatially, and echoing the approach that was followed in "1% annual exceedance probability precipitation levels" and "Change in the magnitude of 1% annual exceedance probability per degree warming", we now compare more quantitatively the CNN approach with an approach where GEV fits are performed individually for single sites. For this purpose, an additional 200 test sites that were not included in the training of the CNN or prior validation, were selected for each region. Each selected site was required to have a minimum of 35 block maxima during the fitting period (1960 to 2019 for North America, Australia and Europe, and 1965 to 2019 for New Zealand). Other than New Zealand which only had 159 additional evaluation sites, the other three regions each had at least 200 sites meeting these criteria. For the regions with more than 200 sites, sites with the most block maxima were selected. Nonetheless, this still resulted in a distinctive lack of suitable sites for some countries in Europe.

Unlike in "1% annual exceedance probability precipitation levels" and "Change in the magnitude of 1% annual exceedance probability per degree warming", to generate the most robust GEV fits possible, these previously unseen sites were allowed to source block maxima back to 1901; each site was required to have at least one block maximum in the period prior

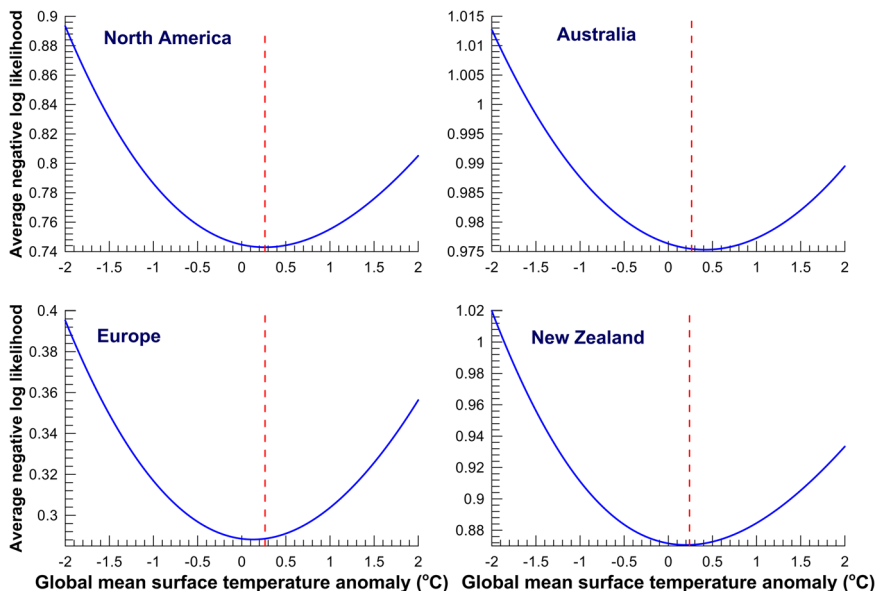


Fig. 8 CNN evaluation of non-stationarity. The negative log-likelihood across all 10,000 sites that had observations available during the evaluation period, averaged for each region at different values of T'_{Global} from -2°C to 2°C (blue lines). The vertical red dashed line in each panel shows the average T'_{Global} over the evaluation period, 0.240°C for New Zealand from 1960 to 1964, and 0.264°C over 1950 to 1959 for the other locations. If the linearly expanded coefficients accurately generalise over the evaluation period, the minimum in the negative log-likelihood curve is expected to coincide with the vertical dashed line.

to 1960/1965 to be used in the evaluation (see below). Furthermore, a stationary GEV was fitted to the block maxima at each site. While this prevents the GEV fit from being used to determine the sensitivity of extreme precipitation to climate warming, it results in a more robust stationary fit.

To provide comparison statistics from the CNN at these new evaluation sites, GEV coefficients were extracted from the cell nearest to the evaluation site from the 1.5 km gridded CNN output and evaluated at a T'_{Global} of 0.2°C . Slightly better CNN results may have been generated by evaluating the CNN at the exact location of the site, especially in mountainous regions, and at an appropriately weighted value of T'_{Global} . As such, the comparison statistics presented below are conservative.

As we are comparing the accuracy of the CNN to a direct GEV fit at a single site, we replicated the same validation split as before - any data before 1960 (or 1965 for New Zealand) were used for evaluation, and any data thereafter were used for the fit (at least 35 years, but typically more). This ensured that any single extreme precipitation event that the CNN was trained on would not impact the comparison. The mean negative log likelihood for all observations during the evaluation period (i.e., before 1960/1965) was then calculated both for the individual site GEV fits and using the GEV coefficients extracted from the CNN; each was evaluated against the same set of block maxima unseen in their training.

The results are summarised in Table 1. Overall, the CNN performs markedly better than a direct fit in North America and Europe, where 74% and 63% of sites, respectively, showed a better fit. For Australia and New Zealand, the results are mixed with the CNN providing a better fit to the observations about half the time. There is some ambiguity in these results however. A site with block maxima closely following a typical GEV function, but with each block maximum inflated by a factor of 10, would generate a very small negative log likelihood (good GEV fit) but would generate highly erroneous 1% AEP precipitation depths. The CNN, evaluated at the same site, would generate a very large negative log likelihood (poor GEV fit) because the CNN generalises spatially and is not influenced by the 10x biased observations at that site.

Table 1 The mean and median negative log likelihoods for both the CNN and the direct GEV fits to each site.					
Location	CNN Mean	GEV Mean	CNN Median	GEV Median	CNN Better
NA	0.733	0.786	0.493	0.541	74%
EUR	0.316	0.326	-0.006	-0.031	63%
AUS	0.801	0.795	0.468	0.482	51%
NZ	0.916	0.890	0.568	0.624	49%

Lower numbers represent better fits to the block maxima. CNN better shows the percentage of times the CNN produced better results than a direct fit at a given site.

The spatial patterns of the differences in the negative log likelihoods between the individual site GEV fits and the CNN-derived GEV fits are shown in Supplementary Fig. S6. Other than potentially in Norway, which contains a region where the CNN appears to perform poorly (potentially due to complex topography), there are no spatially coherent biases that can be discerned. In every region there are nearby stations that show opposing results. There are two explanations for this viz., (1) due to random variations and small sample sizes, and (2) due to anomalous block maxima at a single site as was seen in Supplementary Fig. S4. In the second case, while a direct GEV fit is influenced by this anomalous block maximum, the CNN, having generalised spatially, will perform comparatively poorly. While most sites contained sufficient block maxima for this evaluation, for New Zealand there were far fewer block maxima available during the evaluation period (less than 5 on average). As a result, for New Zealand there is considerable random variation such that no spatially consistent results can be discerned.

Another way of evaluating the benefits of the CNN compared to direct GEV fits is to evaluate the number of very rare events inferred from the fit; essentially a test of the robustness of the inferred long tail of the distribution. In this case a very rare event was considered to be an event with an AEP of less than 0.1%, i.e. a 1-in-1000 year event. The number of such events, in the context

Table 2 The number of rare events derived from both the CNN GEV coefficients and from the direct fit GEV coefficients.

Location	CNN rare events	GEV rare events	Avg obs per Site	Total obs
NA	7	41	51.6	10316
EUR	19	63	47.3	9457
AUS	8	23	55.9	11188
NZ	1	2	4.4	695

Average observations per site is the average number of observations used at each site for evaluation, for both the CNN and direct GEV fits. The total number of 1-year block maxima available through the evaluation period in each region is shown in the rightmost column.

of the total number of 1-year block maxima available, is shown in Table 2. Rare events were estimated to be far more frequent when evaluated from the individual site GEV fits compared to when evaluated from the CNN-derived GEV coefficients. Over all four regions, the CNN derived frequency of rare events is not substantively different from the total number of 1-year block maxima available, i.e., 0.1% of the total number of observations. This suggests that a direct GEV fit is much more likely to underestimate the bounds, or extremes, of a GEV distribution. Given that the CNN saw much more data in its training, including very rare events which it had to account for, it is unsurprising that it models these events with greater fidelity.

This evaluation shows that while the CNN generalises well to unseen locations, it is not always better than a stationary direct fit (noting, however, the ambiguity in such an evaluation as discussed above). If the primary goal is to derive estimates of stationary precipitation return periods, then repeating the CNN training process with no, or fewer, terms expanded in T'_{Global} (see equation (2)), would be beneficial, especially for a region such as New Zealand with limited volumes of data.

Data availability

All data sets used in this study are publicly available. The Global Historical Climate Network (GHCN) data are available from ref. ⁶⁵ while the New Zealand National Climate Database data are available from ref. ⁶⁶. The data required to generate all of the figures in the paper are available through a Zenodo database⁶⁷.

Code availability

All code used to produce the analysis is publicly available from⁶⁸.

Received: 21 November 2022; Accepted: 23 November 2023;

Published online: 12 December 2023

References

- van der Wiel, K. & Bintanja, R. Contribution of climatic changes in mean and variability to monthly temperature and precipitation extremes. *Commun. Earth Environ.* **2**, 1 (2021).
- Sebastian, A., Gori, A., Blessing, R. B., van der Wiel, K. & Bass, B. Disentangling the impacts of human and environmental change on catchment response during Hurricane Harvey. *Environ. Res. Lett.* **14**, 124023 (2019).
- Herring, S. C., Hoerling, M. P., Kossin, J. P., Peterson, T. C. & Stott, P. A. Explaining extreme events of 2014 from a climate perspective. *Bullet. Am. Meteorol. Soc.* **96**, S1–S172 (2015).
- Wang, G. et al. The peak structure and future changes of the relationships between extreme precipitation and temperature. *Nat. Clim. Change* **7**, 268–275 (2017).
- Pfahl, S., O’Gorman, P. & Fischer, E. Understanding the regional pattern of projected future changes in extreme precipitation. *Nat. Clim. Change* **7**, 423–427 (2017).
- Martinkova, M. & Kysely, J. Overview of observed clausius-clapeyron scaling of extreme precipitation in midlatitudes. *Atmosphere* **11**, 1–16 (2020).
- Coles, S. *An Introduction to Statistical Modeling of Extreme Values* (Springer-Verlag London, 2004).
- Sippel, S. et al. Combining large model ensembles with extreme value statistics to improve attribution statements of rare events. *Weather Clim. Extrem.* **9**, 25–35 (2015).
- Vose, R. Reference station networks for monitoring climatic change in the conterminous United States. *J. Clim.* **18**, 5390–5395 (2005).
- Huang, W. K., Stein, M. L., McInerney, D. J., Sun, S. & Moyer, E. J. Estimating changes in temperature extremes from millennial-scale climate simulations using generalized extreme value (GEV) distributions. *Adv. Statist. Climatol. Meteorol. Oceanogr.* **2**, 79–103 (2016).
- Li, C., Sinha, E., Horton, D. E., Diffenbaugh, N. S. & Michalak, A. M. Joint bias correction of temperature and precipitation in climate model simulations. *J. Geophys. Res. Atmos.* **119**, 13–153 (2014).
- Ehret, U., Zehe, E., Wulfmeyer, V., Warrach-Sagi, K. & Liebert, J. HESS Opinions “Should we apply bias correction to global and regional climate model data?”. *Hydrol. Earth Syst. Sci.* **16**, 3391–3404 (2012).
- Haerter, J. O., Eggert, B., Moseley, C., Piani, C. & Berg, P. Statistical precipitation bias correction of gridded model data using point measurements. *Geophys. Res. Lett.* **42**, 1919–1929 (2015).
- Posada-Marín, J. A., Rendón, A. M., Salazar, J. F., Mejía, J. F. & Villegas, J. C. WRF downscaling improves ERA-interim representation of precipitation around a tropical Andean valley during El Niño: implications for GCM-scale simulation of precipitation over complex terrain. *Clim. Dyn.* **52**, 3609–3629 (2019).
- Pham, H. X., Shamseldin, A. Y. & Melville, B. W. Projection of future extreme precipitation: a robust assessment of downscaled daily precipitation. *Nat. Hazards* **107**, 311–329 (2021).
- Eden, J. M., Widmann, M., Grawe, D. & Rast, S. Skill, correction, and downscaling of GCM-simulated precipitation. *J. Clim.* **25**, 3970–3984 (2012).
- Venema, V. K. et al. Benchmarking homogenization algorithms for monthly data. *Clim. Past* **8**, 89–115 (2012).
- Toreti, A., Kuglitsch, F. G., Xoplaki, E. & Luterbacher, J. A novel approach for the detection of inhomogeneities affecting climate time series. *J. Appl. Meteorol. Climatol.* **51**, 317–326 (2012).
- Peterson, T. C. et al. Homogeneity adjustments of in situ atmospheric climate data: a review. *Int. J. Climatol. J. Royal Meteorol. Soc.* **18**, 1493–1517 (1998).
- Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inform. Process. Syst.* **25**, 854–906 (2012).
- LeCun, Y., Bengio, Y. & Hinton, G. Deep Learning. *Nature* **521**, 436–444 (2015).
- Wan, H., Zhang, X., Zwiers, F. & Min, S.-K. Attributing northern high-latitude precipitation change over the period 1966–2005 to human influence. *Clim. Dyn.* **45**, 1713–1726 (2014).
- Westra, S., Alexander, L. V. & Zwiers, F. W. Global increasing trends in annual maximum daily precipitation. *J. Clim.* **26**, 3904–3918 (2013).
- Fix, M. J., Cooley, D., Sain, S. R. & Tebaldi, C. A comparison of U.S. precipitation extremes under RCP8.5 and RCP4.5 with an application of pattern scaling. *Clim. Change* **146**, 335–347 (2016).
- Zittis, G., Bruggeman, A. & Lelieveld, J. Revisiting future extreme precipitation trends in the Mediterranean. *Weather Clim. Extrem.* **34**, e00218 (2021).
- Sarojini, B., Stott, P. & Black, E. Detection and attribution of human influence on regional precipitation. *Nat. Clim. Change* **6**, 669–675 (2016).
- Sun, Q., Zhang, X., Zwiers, F., Westra, S. & Alexander, L. V. A global, continental, and regional analysis of changes in extreme precipitation. *J. Clim.* **34**, 243–258 (2021).
- Donat, M. G. et al. Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: The HadEX2 dataset. *J. Geophys. Res.* **118**, 2098–2118 (2013).
- Easterling, D. et al. Precipitation change in the United States. In: climate science special report: fourth national climate assessment, volume i. [Wuebbles, D., Fahey, D., Hibbard, K., Dokken, D., Stewart, B., Maycock, T. (eds.)]. US Global Change Research Program, Washington, 207–230 (2017).
- Kirchmeier-Young, M. C. & Zhang, X. Human influence has intensified extreme precipitation in North America. *Proc. Natl Acad. Sci. USA.* **117**, 13308–13313 (2020).
- Patricola, C. M. & Wehner, M. F. Anthropogenic influences on major tropical cyclone events. *Nature* **563**, 339–346 (2018).
- Dey, R., Lewis, S. C. & Abram, N. J. Investigating observed northwest Australian rainfall trends in coupled model intercomparison project phase 5 detection and attribution experiments. *Int. J. Climatol.* **39**, 112–127 (2019).
- Risbey, J. S., Pook, M. J. & McIntosh, P. C. Spatial trends in synoptic rainfall in southern Australia. *Geophys. Res. Lett.* **40**, 3781–3785 (2013).
- Delworth, T. & Zeng, F. Regional rainfall decline in Australia attributed to anthropogenic greenhouse gases and ozone levels. *Nat. Geosci.* **7**, 583–587 (2013).

35. Solman, S. & Orlanski, I. Poleward shift and change of frontal activity in the southern hemisphere over the last 40 years. *Am. Meteorol. Soc.* **71**, 539–551 (2014).
36. Thompson, D. & Solomon, S. Interpretation of recent southern hemisphere climate change. *Science* **296**, 895–899 (2002).
37. Zhang, X., Wan, H., Zwiers, F. W., Hegerl, G. C. & Min, S.-K. Attributing intensification of precipitation extremes to human influence. *Geophys. Res. Lett.* **40**, 5252–5257 (2013).
38. Min, S.-K., Zhang, X., Zwiers, F. W. & Hegerl, G. C. Human contribution to more-intense precipitation extremes. *Nature* **470**, 378–381 (2011).
39. Glade, T. Establishing the frequency and magnitude of landslide-triggering rainstorm events in New Zealand. *Environ. Geol.* **35**, 160–174 (1998).
40. Krishnan, R. et al. Deciphering the desiccation trend of the south asian monsoon hydroclimate in a warming world. *Clim. Dyn.* **47**, 1007–1027 (2016).
41. Manda, A. et al. Impacts of a warming marginal sea on torrential rainfall organized under the asian summer monsoon. *Sci. Rep.* **4**, 5741 (2014).
42. Mei, W. & Xie, S.-P. Intensification of landfalling typhoons over the northwest pacific since the late 1970s. *Nat. Geosci.* **9**, 753–757 (2016).
43. Trenberth, K. E., Cheng, L., Jacobs, P., Zhang, Y. & Fasullo, J. Hurricane harvey links to ocean heat content and climate change adaptation. *Earth's Future* **6**, 730–744 (2018).
44. Shepherd, T. Atmospheric circulation as a source of uncertainty in climate change projections. *Nat. Geosci.* **7**, 703–708 (2014).
45. Shaw, T. et al. Storm track processes and the opposing influences of climate change. *Nat. Geosci.* **9**, 656–664 (2016).
46. Odoulami, R. C., Abiodun, B. J. & Ajayi, A. E. Modelling the potential impacts of afforestation on extreme precipitation over west africa. *Clim. Dyn.* **52**, 2185–2198 (2019).
47. Shastri, H., Paul, S., Ghosh, S. & Karmakar, S. Impacts of urbanization on indian summer monsoon rainfall extremes. *J. Geophys. Res. Atmos.* **120**, 496–516 (2015).
48. Schmid, P. E. & Niyogi, D. Modeling urban precipitation modification by spatially heterogeneous aerosols. *J. Appl. Meteorol. Climatol.* **56**, 2141–2153 (2017).
49. Sui, C.-H., Satoh, M. & Suzuki, K. Precipitation efficiency and its role in cloud-radiative feedbacks to climate variability. *J. Meteorol. Soc. Japan. Ser. II* **98**, 261–282 (2020).
50. McKinnon, K., Poppick, A., Dunn-Sigouin, E. & Deser, C. An “observational large ensemble” to compare observed and modeled temperature trend uncertainty due to internal variability. *J. Clim.* **30**, 7585–7598 (2017).
51. Marotzke, J. & Forster, P. Forcing, feedback and internal variability in global temperature trends. *Nature* **517**, 565–570 (2015).
52. Mann, M. E., Steinman, B. A. & Miller, S. K. On forced temperature changes, internal variability, and the AMO. *Geophys. Res. Lett.* **41**, 3211–3219 (2014).
53. Zhao, X., Gong, Z., Zhang, Y., Yao, W. & Chen, X. Physics-informed convolutional neural networks for temperature field prediction of heat source layout without labeled data. *arXiv* <https://doi.org/10.48550/arXiv.2109.12482> (2021).
54. Cuomo, S. et al. Scientific machine learning through physics-informed neural networks: Where we are and what's next. *arXiv* <https://doi.org/10.48550/arXiv.2201.05624> (2022).
55. Morice, C. P. et al. An updated assessment of near-surface temperature change from 1850: the HadCRUT5 data set. *J. Geophys. Res. Atmos.* **126**, e2019JD032361 (2021).
56. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 770–778 (2016).
57. He, T. et al. Bag of tricks for image classification with convolutional neural networks. *arXiv* <https://doi.org/10.48550/arXiv.1812.01187> (2018).
58. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inform. Process. Syst.* **25**, 1097–1105 (2012).
59. Misra, D. Mish: A self regularized non-monotonic neural activation function. *arXiv* **4**, 10–48550 (2019).
60. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).
61. Liu, L. et al. On the variance of the adaptive learning rate and beyond. *arXiv* <https://doi.org/10.48550/arXiv.1908.03265> (2019).
62. Zhang, M., Lucas, J., Ba, J. & Hinton, G. E. Lookahead optimizer: k steps forward, 1 step back. *Adv. Neural Inform. Process. Syst.* **32**, 8030–8040 (2019).
63. Paszke, A. et al. PyTorch: An imperative style, high-performance deep learning library. In Wallach, H. et al. *Adv. Neural Inform. Process. Syst.* **32**, 8024–8035 (2019).
64. Howard, J. & Gugger, S. Fastai: a layered API for deep learning. *Information* **11**, 108 (2020).
65. *Global Historical Climate Network (GHCN) Daily*. <https://www.ncei.noaa.gov/products/land-based-station/global-historical-climatology-network-daily> (2023).
66. *New Zealand National Climate Database (CLIDB)*. <https://cliflo.niwa.co.nz/> (2023).
67. Bird, L., Bodeker, G. & Clem, K. Data files for figures in “Sensitivity of extreme precipitation to climate change inferred using artificial intelligence shows high spatial variability” by Bird et al. *Zenodo* <https://doi.org/10.5281/zenodo.10038809> (2023).
68. Bird, L. Code for “Sensitivity of extreme precipitation to climate change inferred using artificial intelligence shows high spatial variability” by Bird et al. *Zenodo* <https://doi.org/10.5281/zenodo.6917840> (2023).

Acknowledgements

This research was funded by the New Zealand Ministry of Business, Innovation and Employment under the Extreme Events and the Emergence of Climate Change (Whakahura) programme, administered through Victoria University of Wellington (contract number RTVU1906). The New Zealand daily rainfall data were kindly provided by Errol Lewthwaite, National Institute of Water and Atmospheric Research. We would also like to thank the National Institute of Water and Atmospheric Research for making the High Intensity Rainfall Design System analyses available online. We would also like to acknowledge the GHCN team for the provision of the GHCN data used in this analysis and the Hadley Centre/Climatic Research Unit Temperature for provision of the annual mean global mean temperature time series.

Author contributions

L.J.B. and G.E.B. developed the idea, analysed the data and wrote the manuscript with substantial input from K.R.C. All authors contributed to the writing of the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s43247-023-01142-4>.

Correspondence and requests for materials should be addressed to Gregory E. Bodeker.

Peer review information *Communications Earth & Environment* thanks Peter Gibson and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Akintomide Akinsanola, Heike Langenberg. A peer review file is available.

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.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023