A flexible nonlinear modelling framework for nonstationary generalized extreme value analysis in hydroclimatology

Alex J. Cannon*

Meteorological Service of Canada, Environment Canada, Vancouver, British Columbia, Canada V6C 3S5

Abstract:
Parameters in a generalized extreme value (GEV) distribution are specified as a function of covariates using a conditional density network (CDN), which is a probabilistic extension of the multilayer perceptron neural network. If the covariate is time or is dependent on time, then the GEV-CDN model can be used to perform nonlinear, nonstationary GEV analysis of hydrological or climatological time series. Owing to the flexibility of the neural network architecture, the model is capable of representing a wide range of nonstationary relationships. Model parameters are estimated by generalized maximum likelihood, an approach that is tailored to the estimation of GEV parameters from geophysical time series. Model complexity is identified using the Bayesian information criterion and the Akaike information criterion with small sample size correction. Monte Carlo simulations are used to validate GEV-CDN performance on four simple synthetic problems. The model is then demonstrated on precipitation data from southern California, a series that exhibits nonstationarity due to interannual/interdecadal climatic variability.

KEY WORDS extreme value analysis; nonstationary; statistical modelling; neural network; nonlinear hydroclimatology

INTRODUCTION

The distribution of a series of extreme values computed from long sequences of data asymptotically approaches the generalized extreme value (GEV) distribution (Jenkinson, 1955) as the number of samples becomes large. The extreme value theorem, which is the extreme value analogue of the central limit theorem (Coles, 2001), forms the basis for extreme value analysis of meteorological and hydrological series, for example, annual maxima of rainfall or streamflow observations, and, in turn, the estimation of design criteria for engineering structures (Maidment, 1993). The main assumptions are (i) that the series of extremes is suitably long, (ii) that the elements of the series are independent and identically distributed, and (iii) that the series is stationary, meaning that its statistical properties are independent of time. There is ample evidence that the hydroclimatic system is nonstationary on time scales relevant to applied extreme value analysis, whether due to natural climate variability or anthropogenic climate change (Jain and Lall, 2001; Rial et al., 2004; Milly et al., 2008). The assumption of stationarity in extreme value analysis is therefore questionable, and new methods that explicitly allow for nonstationarity in the GEV distribution parameters are required (Coles, 2001).

For example, Kharin and Zwiers (2004) allowed for linear trends in the location and shape and a log-linear trend in the scale parameter of the GEV distribution with time, whereas Wang et al. (2004) considered models with covariate-dependent changes in the location and scale parameters. Parameter estimates were made via the principle of maximum likelihood (ML). El Adlouni et al. (2007) extended this approach by fitting parameters via the generalized maximum likelihood (GML) approach of Martins and Stedinger (2000), which specifies a geophysically realistic prior distribution for the shape parameter within a Bayesian framework. The GML method performed better than the standard ML method for small sample sizes.

In each of these studies, simple parametric models (all linear or log-linear in the parameters) were specified a priori for the nonstationary dependence of the GEV parameters on the covariates. In practice, an assumption of linearity may not be appropriate. For example, Kharin and Zwiers (2004) allowed for nonlinear trends in precipitation and temperature extremes over a 110-year transient global climate model simulation by estimating linear trends in GEV parameters based on a series of overlapping 51-year time windows. In a GEV analysis of winter precipitation in the United States, Schubert et al. (2008) fit separate distributions to La Niña, neutral, and El Niño years to account for a nonlinear relationship between precipitation extremes in the southwest and El Niño-Southern oscillation (ENSO) conditions.

As an alternative, nonparametric and semi-parametric approaches to nonstationary extreme value analysis have been developed to overcome the linearity assumption of the conditional GEV models described above. For example, Koenker and Schorfheide (1994) used quantile
regression splines to assess nonlinear trends in quantiles of a global temperature series, Hall and Tajvidi (2000) fit local-linear trend models to wind storm intensity and temperature series, and Gaetan and Grigoliotto (2004) used dynamic smoothing models to assess trends in a monthly maximum temperature extremes. Such models are, however, generally limited to analysing nonlinear relationships involving a single covariate. Methods based on generalised additive models, for example, those by Chavez-Demoulin and Davison (2005), which analysed relationships between extremes in alpine temperatures and altitude/North Atlantic Oscillation, and Yee and Stephenson (2007), who considered the dependence between annual maximum sea-levels and time/ENSO, allow multiple covariates, but cannot model interactions between covariates without a priori specification of the form of the interactions by the modeller.

In this study, a more flexible nonlinear model for nonstationary extreme value analysis is proposed. Parameters of the GEV distribution are specified as a function of covariates using a conditional density network (CDN) (Neuneier et al., 1994; Bishop, 2006; Cawley et al., 2007; Cannon, 2008), which is a probabilistic extension of the commonly used multilayer perceptron (MLP) neural network (Gardner and Dorling, 1998; Hsieh and Tang, 1998; Dawson and Wilby, 2001). The MLP is a universal function approximator that can model nonlinear relationships, including ones involving unspecified interactions between multiple covariates. If one of the covariates is time (or is dependent on time), then the GEV-CDN model performs nonstationary GEV analysis. Model parameters are estimated following the GML approach of Martins and Stedinger (2000), and model complexity is identified using two model selection criteria. Confidence limits for parameters and estimated quantiles are estimated by the bootstrap. Monte Carlo simulations are used to validate performance on synthetic test problems introduced by El Adlouni et al. (2007). Finally, models are applied to precipitation data from southern California, a time series that exhibits nonstationarity due to the influence of climate variability on interannual/interdecadal time scales.

**GEV DISTRIBUTION**

The GEV distribution (Jenkinson, 1955) is specified by three parameters: the location \( \mu \), the scale \( \alpha (\alpha > 0) \), and the shape \( \kappa \). The shape parameter \( \kappa \) is the main determinant of the behaviour of the tails of the distribution. Following the hydroclimatological convention, negative values of \( \kappa \) correspond to positive skewness. The cumulative density function (cdf) of a random variable \( y \) drawn from a GEV distribution is given by

\[
F(y; \mu, \alpha, \kappa) = \exp \left[ - \left( 1 - \kappa \frac{y - \mu}{\alpha} \right)^{1/\kappa} \right], \quad \kappa \neq 0, \quad 1 - \kappa \frac{y - \mu}{\alpha} > 0
\]  

(1)

\[
F(y; \mu, \alpha, \kappa) = 0 \quad \text{if } \kappa = 0
\]

and the probability density function (pdf) is given by

\[
f(y; \mu, \alpha, \kappa) = \frac{1}{\alpha} \left( 1 - \kappa \frac{y - \mu}{\alpha} \right)^{(1-\kappa)/\kappa} \exp \left[ - \left( 1 - \kappa \frac{y - \mu}{\alpha} \right)^{1/\kappa} \right], \quad \kappa \neq 0
\]

(3)

\[
f(y; \mu, \alpha) = \frac{1}{\alpha} \exp \left[ - \left( \frac{y - \mu}{\alpha} \right) \right] \exp \left[ - \exp \left( - \frac{y - \mu}{\alpha} \right) \right], \quad \kappa = 0
\]

(4)

Parameters in the GEV distribution can be estimated via the method of moments or \( L \) moments (Hosking, 1990), or by the principle of ML (Coles and Dixon, 1999), where the goal is to identify the most ‘likely’ set of parameters by maximizing the likelihood function

\[
L(\mu, \alpha, \kappa|y) = \prod_{i=1}^{N} f(y(t); \mu, \alpha, \kappa)
\]

(5)

where \( y = \{y(t), \quad t = 1, \ldots, N\} \) is a series of \( N \) independent observations. For convenience, one often minimizes the negative of the log-likelihood function instead. For small samples, ML estimates can be unstable relative to the method of moments (which restricts values of \( \kappa \) to be greater than \(-1/3\)) due to the generation of physically unrealistic \( \kappa \) values (Coles and Dixon, 1999; Martins and Stedinger, 2000). To remedy this problem, Coles and Dixon (1999) and Martins and Stedinger, 2000 modified the ML approach so that \( \kappa \) is forced to take more realistic values. In the penalized ML approach of Coles and Dixon (1999), a penalty term \( \pi(\kappa) \) is added to the likelihood function, leading to

\[
L^{(p)}(\mu, \alpha, \kappa|y) = L(\mu, \alpha, \kappa|y)\pi(\kappa)
\]

(6)

where the recommended penalty is of the form

\[
\pi^{(p)}(\kappa) = \begin{cases} 
1 & \text{if } \kappa \geq 0, \\
\exp \left( 1 - \frac{1}{1 + \kappa} \right) & \text{if } 0 > \kappa > -1, \\
0 & \text{if } \kappa \leq -1.
\end{cases}
\]

(7)

The shape parameter \( \kappa \) is forced to be greater than \(-1\), with values close to \(-1\) penalized more than larger values.

Martins and Stedinger (2000) developed the GML estimator, which applies a more restrictive penalty (or prior distribution) on \( \kappa \)

\[
\pi^{(s)}(\kappa) = \text{Beta}(\kappa + 0.5; \quad c_1, c_2),
\]

(8)

in which the shape parameter is limited to the range \(-0.5 \leq \kappa \leq 0.5\) and where Beta denotes the pdf of a beta distribution with shape parameters \( c_1 \) and \( c_2 \). Martins and Stedinger (2000) recommended that \( c_1 \) and \( c_2 \) be set equal to 6 and 9, respectively, resulting in a pdf
with a mode at $-0.1$ and $\sim 90\%$ of its probability mass concentrated over $\kappa$ values between $-0.3$ and $+0.1$. In this study, $c_1$ and $c_2$ are instead set equal to $2$ and $3.3$, respectively, which results in a broader pdf with a mode of approximately $-0.2$ and approximately $90\%$ of its probability mass concentrated between $-0.4$ and $+0.2$. As shown in Figure 1, this penalty is intermediate between those recommended by Coles and Dixon (1999) and Martins and Stedinger (2000). The result is a prior distribution for $\kappa$ that fits many natural processes well. For example, Kyselý and Picek (2007) analysed daily precipitation extremes at 78 stations in the Czech Republic and found values of $\kappa$ ranging from $-0.37$ to $+0.16$, which corresponds to slightly more than $80\%$ of the probability mass of $\pi(\kappa)$ with $c_1 = 2$ and $c_2 = 3.3$.

**NONSTATIONARY GEV**

El Adlouni et al. (2007) extended the GML estimator to cases where the GEV parameters are functions of covariates. Parameters are estimated by the Markov Chain Monte Carlo method rather than the numerical optimization approach of Martins and Stedinger (2000). If the covariate is time, or is dependent on time, then this amounts to nonstationary GEV estimation within a Bayesian framework. El Adlouni et al. (2007) considered nonstationary relationships in the location $\mu$ and scale $\alpha$ parameters with time of the form

$$
\mu(t) = \beta_1 + \beta_2 t + \beta_3 t^2 \tag{9}
$$

$$
\alpha(t) = \exp(\delta_1 + \delta_2 t) \tag{10}
$$

where $\beta_1$, $\beta_2$, $\beta_3$, $\delta_1$, and $\delta_2$ are parameters that must be estimated from the data. Based on Equations (9) and (10), four models of differing complexity were defined: GEV0, which assumes that all parameters are independent of time ($\beta_2 = 0$, $\beta_3 = 0$, and $\delta_2 = 0$), GEV1, which assumes a linear change in the location parameter ($\beta_1 = 0$ and $\delta_2 = 0$), GEV2, which assumes a quadratic change in the location parameter ($\delta_2 = 0$, and GEV11, which assumes a linear change in the location and a log-linear change in the scale parameter ($\beta_3 = 0$). The shape $\kappa$ is assumed to be constant with time in all models.

**GEV-CDN**

**General framework**

In the GEV-CDN, parameters of the GEV distribution are modelled as a function of covariates using an MLP architecture, which is shown schematically in Figure 2a. The model has $K = 3$ outputs, corresponding to the three GEV parameters ($\mu$, $\alpha$, and $\kappa$). Details on the CDN, which is a probabilistic variant of the standard MLP, are given in Bishop (2006, Ch. 5.6) and Cawley et al. (2007). The reader is referred to Gardner and Dorling (1998), Hsieh and Tang (1998), and Dawson and Wilby (2001) for reviews of the MLP in the context of meteorological, climatological, and hydrological prediction.

Given covariates at time $t$, $x(t) = \{x_i(t), i = 1, \ldots, I\}$, outputs from a GEV-CDN model with $J$ hidden-layer nodes are evaluated as follows. First, the output from the $j$th hidden-layer node $h_j$ is given by applying the hidden-layer activation function $\mu(\cdot)$ to the inner product between the covariates and the input-hidden layer weights $w_{1j}$ plus the bias $b_{1j}$

$$
h_j(t) = m \sum_{i=1}^{I} x_i(t) w_{1ji} + b_{1j} \tag{11}
$$

If the GEV-CDN mapping is to be nonlinear, then $m(\cdot)$ is taken to be a sigmoidal function, e.g. the hyperbolic tangent function $\tanh(\cdot)$. The identity function is adopted if the GEV-CDN mapping is to be strictly linear. The value of the $k$th output from the network $o_k$ is then given by

$$
o_k(t) = \sum_{j=1}^{J} h_j(t) w_{2kj} + b_{2k} \tag{12}
$$

where $w_{2kj}$ are the hidden-output layer weights and $b_{2k}$ are the hidden-output layer biases. Finally, GEV parameters are obtained by applying the output-layer activation functions $g_k(\cdot)$,

$$
\mu(t) = g_1(o_1(t)) = o_1(t) \tag{13}
$$

$$
\alpha(t) = g_2(o_2(t)) = \exp(o_2(t)) \tag{14}
$$

$$
\kappa(t) = g_3(o_3(t)) = \kappa^* \tanh(o_3(t)). \tag{15}
$$

The function $g_2(\cdot)$ forces the scale parameter $\alpha$ to take positive values. The function $g_3(\cdot)$ constrains the shape parameter $\kappa$ to the interval $[-\kappa^*, \kappa^*]$. In this study, GML (Martins and Stedinger, 2000) is adopted, with the value of $\kappa^*$ set to 0.5 to limit the search space.
of $\kappa$ during optimization to the support of the shifted beta distribution prior. The conditional pdf is now given by Equations (3–4) with time-dependent (rather than constant) GEV parameters $\mu(t), \alpha(t),$ and $\kappa(t),$ with the corresponding likelihood,

$$L(\mu, \alpha, \kappa | y) = \prod_{i=1}^{N} f(y(t); \mu(t), \alpha(t), \kappa(t)).$$

A hierarchy of models can be defined by adjusting three aspects of the GEV-CDN model architecture: (i) by specifying either a linear or a nonlinear hidden-layer activation function $m(\cdot); (ii)$ by adjusting the number of hidden-layer nodes $J;$ or (iii) by disconnecting weights leading to output-layer nodes $o_k.$

To illustrate, five GEV-CDN models, shown in order of increasing complexity in Figures 2b–2f, are specified here for the case of a single covariate ($I = 1$), which, in this case, is assumed to be time $x_1(t) = t.$ Following El Adlouni et al. (2007), the shape parameter $\kappa$ is assumed to be constant in all models, although, as shown above, this is not a requirement of the GEV-CDN framework. A description of the hierarchy of GEV-CDN models for this particular case follows.

**CDN-CON**

The CDN-CON model (Figure 2b) assumes that the GEV distribution parameters are constant, i.e. the data are stationary. This can be accomplished in the GEV-CDN framework by disconnecting (i.e. setting to zero) all parameters except for the hidden-output layer biases $b_k.$ From Equations (11)–(15), this reduces to a stationary model in which the GEV parameters are decoupled from the covariates and hidden-layer nodes. The total number of adjustable parameters is thus $P = 3.$ The CDN-CON model is equivalent to the GEV0 model of El Adlouni et al. (2007).

**CDN-LIN**

The CDN-LIN model (Figure 2c) assumes that the network outputs $o_1$ and $o_2$ associated with $\mu(t)$ and $\alpha(t),$ respectively, are a linear function of time, which can be accomplished by setting the hidden-layer activation function $m(\cdot)$ to the identity function and, for simplicity, $J = 1.$ In this example, nonstationary parameters $\mu(t)$ and $\alpha(t)$ are given by Equations (13) and (14). To accommodate constant $\kappa,$ the hidden-output layer weight $w_2$ is disconnected from the model.

To show that this leads to a simple linear model, consider the net effect of these choices on Equations (11)–(13) for $\mu(t),$

$$\mu(t) = o_1(t) = w_{11}^{(1)} w_{11}^{(2)} x_1(t) + b_1^{(1)} w_{12}^{(2)} + b_1^{(2)}.$$  

---

The weights and biases can be collapsed down to two effective parameters. A similar equation results for $\alpha(t)$, whereas $\kappa$ is specified by the hidden-output layer bias $b_k^{(2)}$. Discounting redundancies, the total number of adjustable model parameters is thus $P = 5$. The representational capacity of the model does not depend on the number of hidden nodes $J$, as increasing $J$ instead results in additional redundant parameters in the equations rather than a true increase in model flexibility.

As shown in Figure 2c, the CDN-LIN model is equivalent to the GEV11 model of El Adlouni et al. (2007). If weights associated with the scale parameter $\alpha$ are also disconnected from the network, then the CDN-LIN model is equivalent to the GEV1 model of El Adlouni et al. (2007).

CDN-NLIN1, CDN-NLIN2, and CDN-NLIN3

CDN-NLIN models allow for general nonlinear relationships between the covariates and the GEV parameters. The hidden-layer activation function $m(\cdot)$ is set to the hyperbolic tangent rather than the identity function. For a model with $I$ covariates, the total number of parameters, assuming constant $\kappa$, therefore depends on the number of hidden-layer nodes as $P = J(I + 3) + 3$. The MLP is a universal function approximator; the model can approximate any smooth function to an arbitrary degree of accuracy because $J$ increases to infinity (Hornik et al., 1989). However, given the small number of samples available in hydrological and climatological extreme value analyses (usually around 30–100 years of annual extremes), $J$ is limited to values between one and three in this paper. For the CDN-NLIN models specified in Figure 2d, each involving a single covariate, $P$ thus ranges between 7 and 15. Models contain not more than half the number of parameters as the number of years (30) in a standard climate normal period.

The CDN-NLIN1 model takes the same form as CDN-LIN, except, as noted above, $m(\cdot)$ is set to the hyperbolic tangent function. With a single hidden-layer node, $J = 1$, this is the simplest form of nonlinear GEV-CDN model. The hyperbolic tangent can approximate linear $\{\tanh(wx) \simeq wx \text{ as } w \to 0\}$ and step functions (as $w \to \pm \infty$).

The CDN-NLIN2 model (Figure 2e) adds a second hidden-layer node to CDN-NLIN1. The GEV2 model of El Adlouni et al. (2007) would be subsumed by the CDN-NLIN2 model, although CDN-NLIN2 is capable of approximating more complicated functions than a second-order polynomial. A Z-shaped continuous curve can, for example, be described by a neural network with two hidden nodes (Christiansen, 2005).

The CDN-NLIN3 model (Figure 2f) adds a third hidden-layer node to CDN-NLIN2, and, as a result, can describe yet more complicated nonlinear relationships. Further insight is given by Carpenter and Barthelemy (1993), who compared approximations made by polynomials and MLP neural networks with similar numbers of parameters.

MONTE CARLO SIMULATIONS

Simulation procedure

Monte Carlo simulations based on the nonstationary GEV1, GEV2, and GEV11 test cases of El Adlouni et al. (2007) are used to evaluate the performance of the GEV-CDN model. A fourth test case, GEVstep, is added to illustrate model performance when faced with a step change in GEV parameters.

Following El Adlouni et al. (2007), Monte Carlo simulations involve fitting GEV-CDN models to 50 samples of a single covariate, $t = 1, \ldots, 50$, representing time, with time-dependent GEV parameters as specified in Figure 3. One thousand trials are run for each test case and value of $\kappa (\kappa = -0.1, -0.2, -0.3)$. In each trial, samples are randomly generated based on the specified nonstationary GEV distributions, and GEV-CDN models are fit to the random samples. The bias and root mean squared error (rmse) of predicted 0.5, 0.8, 0.9, 0.99, and 0.999 $\tau$-quantiles, which, in a stationary context correspond to return periods of 2, 5, 10, 100, and 1000 years, respectively, are calculated with respect to the true quantiles. For the GEV1, GEV2, and GEVstep test cases, CDN-LIN and CDN-NLIN models are fit with stationary $\alpha$ and $\kappa$; only $\mu$ is dependent on time. For the GEV11 test case, both $\mu$ and $\alpha$ are allowed to be time dependent, whereas $\kappa$ is stationary.

For all GEV-CDN models, neural network weights and biases are estimated by minimizing the GML cost function using a Broyden–Fletcher–Goldfarb–Shanno–quasi-Newton optimization algorithm as implemented by Nash (1990). To avoid convergence to a shallow local minimum of the error surface, the optimization algorithm is run 100 times, each time starting from different initial weights and biases. Parameters associated with the maximum GML over the 100 random restarts are selected as the final parameters.

Model selection

The appropriate GEV-CDN model architecture for a given dataset is selected by fitting increasingly complicated models and choosing the one that minimizes the Akaike information criterion with small sample size correction (AICc) (Akaike, 1974; Hurvich and Tsai, 1989) or the Bayesian information criterion (BIC) (Schwarz, 1978). Both AICc and BIC are cost-complexity model selection criteria that penalize the negative log-likelihood as a function of the number of model parameters $P$. The objective then is to choose the most parsimonious model that is capable of accounting for the true (but unknown) deterministic function responsible for generating the $N$ observations. Overfitting, that is fitting to noise in the finite dataset rather than the underlying signal, is thus avoided. Theoretical justifications for AICc and BIC in the context of model selection can be motivated either by information-theoretic or Bayesian arguments. The reader is referred to Burnham and Anderson (2004) for more details.
AICc and BIC are given, respectively, by

$$\text{AICc} = -2 \log(L) + 2P + \frac{2P(P + 1)}{N - P - 1} \quad (18)$$

and

$$\text{BIC} = -2 \log(L) + P \log(N) \quad (19)$$

where BIC penalizes more complicated models more than does AICc. As an alternative to AICc and BIC, both of which are estimated directly from the training dataset, more computationally intensive split-sample methods, such as cross-validation, can also be used for model selection. Although this paper relies on information criteria for this purpose, the reader is referred to Smyth (2000) for a comparison between models selected via cross-validation and BIC in a ML context. The utility of AICc and BIC for nonstationary model selection is assessed in the following sections.

**GEV1 test case**

For the GEV1 test case, where the nonstationarity is linear, a CDN-LIN model with constant $\alpha$ and $\kappa$ is functionally equivalent to the linear model evaluated by El Adlouni et al. (2007). Performance intercomparisons between the neural network and the corresponding quantiles from the fitted GEV1 model in terms of bias and rmse are therefore not instructive. In these cases, the Monte Carlo simulations will instead focus on issues related to model selection, namely, (i) the ability of AICc and BIC to select the appropriate level of model complexity and (ii) the cost of model misspecification.

Average values of BIC, AICc, bias, and rmse are given in Table I for each of the five GEV-CDN models (CDN-CON, CDN-LIN, CDN-NLIN1, CDN-NLIN2, and CDN-NLIN3) and three values of $\kappa$. In all cases, both BIC and AICc correctly recommend the CDN-LIN model. As expected given the linear nonstationarity of the GEV1 test case, values of bias and rmse are typically best for the CDN-LIN model. Quantile estimation by the CDN-LIN model is effectively unbiased for all but the highest $\tau$-quantile, which is consistent with the results found by El Adlouni et al. (2007) for the GML estimator. Similarly, values of rmse are relatively low. Although results for the standard ML estimator were also computed during the study, they are not reported because they support the findings of El Adlouni et al. (2007). Namely, GML quantile estimates are less biased and have lower rmse relative to ML estimates.

The cost of model misspecification is generally high. For example, choosing a stationary CDN-CON model

![Figure 3](https://example.com/figure3.png)

Figure 3. Each panel shows, from bottom to top, time series of 0.5, 0.8, 0.9, 0.99, and 0.999 $\tau$-quantiles for (a) GEV1, (b) GEV11, (c) GEV2, and (d) GEVstep test cases (with $\kappa = -0.2$) defined by Equations (9) and (10). Equation parameters are given in the plot titles.
Table I. Model performance of CDN-CON, CDN-LIN, CDN-NL1, CDN-NL1, CDN-NL2, and CDN-NL3 models on the Monte Carlo simulation for the GEV1 test case. Given the linearity of the GEV1 model, the \textit{a priori} expectation is for CDN-LIN to yield the best performance; the column header associated with the CDN-LIN model is shown in bold to reflect this fact. For each combination of performance statistic and shape parameter \(\kappa\), the value corresponding to the model that actually performed best (e.g., the minimum value for BIC, AICc, and rmse, and the value closest to zero for bias) is shown in bold italics.

<table>
<thead>
<tr>
<th>(\kappa)</th>
<th>(-0.1)</th>
<th>(-0.2)</th>
<th>(-0.3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistic</td>
<td>CON</td>
<td>LIN</td>
<td>NL1</td>
</tr>
<tr>
<td>BIC</td>
<td>269.4</td>
<td>175.0</td>
<td>181.5</td>
</tr>
<tr>
<td>AICc</td>
<td>264.2</td>
<td>168.2</td>
<td>172.0</td>
</tr>
<tr>
<td>bias</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\tau = 0.5)</td>
<td>0.05</td>
<td>(-0.01)</td>
<td>(-0.02)</td>
</tr>
<tr>
<td>(0.8)</td>
<td>1.69</td>
<td>(-0.03)</td>
<td>(-0.06)</td>
</tr>
<tr>
<td>(0.9)</td>
<td>2.40</td>
<td>(-0.04)</td>
<td>(-0.05)</td>
</tr>
<tr>
<td>(0.99)</td>
<td>3.09</td>
<td>(0.19)</td>
<td>0.35</td>
</tr>
<tr>
<td>(0.999)</td>
<td>1.75</td>
<td>(1.13)</td>
<td>1.83</td>
</tr>
<tr>
<td>rmse</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\tau = 0.5)</td>
<td>2.90</td>
<td>(0.21)</td>
<td>0.27</td>
</tr>
<tr>
<td>(0.8)</td>
<td>3.35</td>
<td>(0.29)</td>
<td>0.35</td>
</tr>
<tr>
<td>(0.9)</td>
<td>3.78</td>
<td>(0.38)</td>
<td>0.44</td>
</tr>
<tr>
<td>(0.99)</td>
<td>4.34</td>
<td>(1.24)</td>
<td>1.42</td>
</tr>
<tr>
<td>(0.999)</td>
<td>4.02</td>
<td>(3.41)</td>
<td>4.06</td>
</tr>
</tbody>
</table>
increases the absolute bias and rmse by a factor of 2 or more (and, in most cases, by an order of magnitude) for almost all combinations of $\kappa$ and $\tau < 0.999$. Also, with $\tau < 0.999$ the increase in rmse and (absolute) bias associated with selection of a nonlinear model (CDN-NLIN1, CDN-NLIN2, or CDN-NLIN3) is typically lower than the increase associated with selection of the stationary model (CDN-CON), which means that overfitting has less effect on model performance than does improperly by assuming stationarity. With $\tau = 0.999$, however, performance of the nonlinear models can be slightly worse than the stationary model, in particular for $\kappa > -0.3$.

**GEV11 test case**

In the GEV11 test case, the location parameter is a function of time, and the scale parameter is a log-linear function of time. Similar to the previous test case, a CDN-LIN model with constant $\kappa$ is thus functionally equivalent to the corresponding GEV11 model. The *a priori* expectation then is for the CDN-LIN model to be selected by the BIC and AICc cost-complexity model selection criteria. As shown in Table II, this is indeed the case for all three values of $\kappa$.

In terms of model performance, the pattern of results is similar to that reported in the previous section for the GEV1 test case. As expected, the CDN-LIN model is relatively unbiased and gives the lowest values rmse for all $\kappa$ and $\tau$. The cost of model misspecification is, again, quite high. Specifying either stationary or nonlinear models can lead to large drops in model performance.

**GEV2 test case**

For the GEV2 test case, where the location parameter has a nonlinear dependence on time, the main objective is to compare the performance of the CDN-NLIN models against the GEV2 model of El Adlouni et al. (2007). In contrast to the GEV1 and GEV11 test cases, where CDN-LIN models are functionally equivalent to the corresponding GEV1 or GEV11 models, the CDN-NLIN model approximates the quadratic signal specified by the GEV2 model via the nonlinearity introduced by the sigmoidal activation functions in the neural network’s hidden layer. As indicated by Christiansen (2005), accurate approximation of a quadratic function should be possible using an MLP with two hidden nodes, i.e. the CDN-NLIN2 model.

Results from the GEV2 model and the GEV-CDN models are shown in Table III. Among the GEV-CDN models, the CDN-NLIN2 model, as expected, minimizes both BIC and AICc for all values of $\kappa$. Differences in BIC and AICc between the CDN-NLIN2 model and the GEV2 model are small, meaning that the information loss resulting from the use of the CDN-NLIN2 model instead of the GEV2 model is likely to be minimal (Burnham and Anderson, 2004). Comparing bias and rmse values between the two models supports this conclusion. Bias and rmse values are comparable for $\tau < 0.999$, whereas performance statistics for $\tau = 0.999$ tend to favour the
Table III. Model performance of CDN-NLIN1, CDN-NLIN2, and CDN-NLIN3 models on the Monte Carlo simulation for the GEV2 test case. For comparison, results from GEV2 model fits to Equations (9) and (10) are also shown. Among the GEV-CDN models, the a priori expectation is for CDN-NLIN2 to yield the best performance; the column header associated with this model is shown in bold to reflect this fact. For each combination of performance statistic and shape parameter \( \kappa \), the value corresponding to the GEV-CDN model that actually performed best (e.g. the minimum value for BIC, AICc, and rmse, and the value closest to zero for bias) is shown in bold italics.

<table>
<thead>
<tr>
<th>( \kappa )</th>
<th>-0.1</th>
<th>-0.2</th>
<th>-0.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>GEV2</td>
<td>NLIN1</td>
<td>NLIN2</td>
<td>NLIN3</td>
</tr>
<tr>
<td>Statistic</td>
<td>181.3</td>
<td>192.6</td>
<td>187.0</td>
</tr>
<tr>
<td>BIC</td>
<td>173.1</td>
<td>183.1</td>
<td>174.3</td>
</tr>
<tr>
<td>Bias</td>
<td>0.10</td>
<td>0.07</td>
<td>0.04</td>
</tr>
<tr>
<td>GEVstep test case</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AICc</td>
<td>0.8</td>
<td>0.11</td>
<td>0.14</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.9</td>
<td>0.09</td>
<td>0.24</td>
</tr>
<tr>
<td>NLIN1</td>
<td>0.99</td>
<td>0.50</td>
<td>0.54</td>
</tr>
<tr>
<td>NLIN2</td>
<td>0.99</td>
<td>0.72</td>
<td>3.23</td>
</tr>
<tr>
<td>NLIN3</td>
<td>0.80</td>
<td>0.59</td>
<td>0.65</td>
</tr>
</tbody>
</table>

GEV2 model for \( \kappa = -0.1 \) and \( \kappa = -0.2 \) but the CDN-NLIN2 model for \( \kappa = -0.3 \).

Confidence intervals

Confidence intervals for GEV distribution parameters and associated quantiles can be estimated by bootstrap-based methods. For nonstationary GEV models, Kharin and Zwiers (2004) and Khaliq et al. (2006) recommend the residual bootstrap. Following Khaliq et al. (2006), the method proceeds by (i) fitting a nonstationary GEV model to the observed data; (ii) transforming residuals from the fitted model to be identically distributed

\[
\varepsilon(t) = \left[ 1 - \kappa(t) \frac{(\gamma(t) - \mu(t))}{\alpha(t)} \right]^{1/\kappa}
\]

where \( \varepsilon(t) \) is the \( r \)th transformed residual, and \( \mu(t), \alpha(t), \) and \( \kappa(t) \) are GEV parameters from the fitted model; (iii) resampling the transformed residuals with replacement to form a bootstrapped set of residuals \( \{\varepsilon^{(b)}(t), t = 1, \ldots, N\} \);

(v) fitting a nonstationary GEV model to the bootstrapped samples; (vi) estimating parameters and quantiles from the fitted model; and (vii) repeating steps (i) to (vi) for a large number of times.

In a comparison of bootstrap-based methods for estimating confidence intervals of stationary extreme value models, Kyselý (2008) founds the parametric bootstrap to outperform a resampling-based bootstrap. In the parametric bootstrap, bootstrap samples [steps (ii) to (iv) in previous paragraph] are generated by randomly sampling directly from the fitted distribution, rather than by resampling and rescaling model residuals. In either case, confidence intervals can be formed by calculating percentiles (e.g. the 5th and 95th percentiles for the 90% confidence interval) of the bootstrapped parameter/quantile estimates. Bias-corrected alternatives to the percentile method are reviewed by Kyselý (2008), but their application is beyond the scope of this study.

Monte Carlo simulations are used to evaluate whether empirical coverage probabilities of residual and parametric bootstrapped confidence intervals for the GEV-CDN model parameters and quantiles match those expected. GEV-CDN models, selected based on minimum AICc and BIC statistics from Tables I–IV, are fit to 50 samples with time-dependent GEV parameters as specified in Figure 3. One thousand trials are run for each test case. For each Monte Carlo trial, confidence intervals are
constructed by fitting GEV-CDN models to 500 bootstrapped datasets and calculating percentiles of the resulting parameter/quantile distributions. Empirical coverage proportions for 90\% confidence intervals of the 0.5, 0.8, 0.9, 0.99, and 0.999 τ-quantiles, along with μ, σ, and κ parameters, are calculated and compared with the nominal value of 0.9. Results are shown in Table V for GEV1, GEV11, GEV2, and GEVstep test cases.

Bootstrapped confidence intervals for the location parameter μ and scale parameter α tend to be too narrow for both the residual and parametric bootstrap. Coverage proportions for α, in particular, are much lower than 0.9 for all test cases. Conversely, the parametric bootstrap leads to confidence intervals for the shape parameter κ that are too broad, with coverage proportions exceeding 0.9 in 11 of 12 test cases. In terms of confidence intervals for quantiles, coverage proportions for low to intermediate quantiles (τ = 0.5, 0.8, and 0.9) are, again, narrower than expected for both the residual and parametric bootstrap. Results are mixed when moving to higher quantiles (τ = 0.99 and 0.999). The residual bootstrap continues to generate confidence intervals that are too narrow, whereas the parametric bootstrap generates broader than expected confidence intervals. Aggregated over all test cases and values of τ, the parametric bootstrap performs better than the residual bootstrap; coverage proportions are within 0.05 of the nominal value in 34 out of 60 combinations of test case and τ for the parametric bootstrap, but just 23 of 60 combinations for the residual bootstrap.

### PRECIPITATION DATA

Following El Adlouni et al. (2007), nonstationary GEV-CDN models are demonstrated on annual precipitation data recorded at Randsburg, California (station number 047253, latitude 35.22, longitude 117.39, elevation 36 m). In the analysis by El Adlouni et al. (2007), stationary (GEV0) and nonstationary (GEV1 and GEV2) models were fit to the precipitation series. The southern oscillation index (SOI) was used as a covariate describing ENSO conditions in the two nonstationary models. The GEV2 model, representing a quadratic dependence between the GEV location parameter μ and SOI, was found to best describe the relationship between annual precipitation and ENSO at the station of interest. The analysis is replicated here using the GEV-CDN modelling framework. In addition to the SOI, the Pacific decadal oscillation (PDO) index is included as a second covariate, as the PDO, which varies on an interdecadal rather than interannual time scale, has been found to modulate the effect of ENSO teleconnections on precipitation in the western United States (Gershunov and Barnett, 1998; McCabe and Dettinger, 1999; Gutzler et al., 2002).

Precipitation data from 1937 to 2007 are obtained from the Western Regional Climate Centre of the National Oceanographic and Atmospheric Administration. Concurrent PDO and SOI indices for the spring season are
Table V. Empirical coverage proportions for residual and parametric bootstrapped 90% confidence intervals of GEV parameters and 1 quantiles, from GEV-CDN models applied to GEV1, GEV2, and GEVstep test cases. Coverage proportions that differ by more than 0.05 from the nominal value of 0.90 are marked in bold.

<table>
<thead>
<tr>
<th>Test case</th>
<th>Parameter</th>
<th>Coverage Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>GEV1</td>
<td>μ</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>α</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>κ</td>
<td>0.86</td>
</tr>
<tr>
<td>GEV1/3</td>
<td>μ</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>α</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>κ</td>
<td>0.90</td>
</tr>
<tr>
<td>GEVstep</td>
<td>μ</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>α</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>κ</td>
<td>0.92</td>
</tr>
</tbody>
</table>

DISCUSSION AND CONCLUSION

Parameters in a GEV distribution are specified as a function of covariates using a probabilistic variant of the MLP neural network model. If the covariate is time, or is dependent on time, then the resulting GEV-CDN model can be used to perform nonstationary GEV analysis. The use of a neural network architecture means that the model is capable of representing a wide range of linear and nonlinear relationships among covariates and the obtained from the Joint Institute for the Study of the Atmosphere and Ocean at the University of Washington. For reference, annual time series of SOI, PDO, and precipitation data are shown in Figure 4. The SOI is negatively correlated with precipitation, whereas the PDO is positively correlated with precipitation. Negative values of the SOI (e.g. El Niño years) and positive values of the PDO (e.g. warm PDO phase years) tend to coincide with wetter years; the opposite tendency occurs during La Niña and cool PDO phase years. Combinations of PDO and SOI indices are entered as covariates in CDN-LIN, CDN-LIN1, CDN-LIN2, and CDN-LIN3 models. For each combination, models with (i) nonstationary μ, α, and κ parameters; (ii) models with nonstationary μ and α parameters; and (iii) models with a nonstationary μ parameter are fit to precipitation data separately.

The top five covariate/model combinations recommended by AICc and BIC are listed in Table VI. All are nonlinear, with the top four involving both SOI and PDO as covariates. The CDN-LIN2 model with SOI and PDO as predictors of the location μ and scale κ parameters is recommended as the best model. Contour plots of CDN-LIN2 modelled relationships between the covariates and the GEV parameters are shown in Figure 5, along with a sample plot of the τ = 0.90 quantile (10-year return period). Threshold behaviour is present, resulting in two main regions of SOI/PDO space, each with near constant values of μ, κ, and the precipitation quantile. The primary influence on precipitation appears to be SOI: largest values tend to occur when SOI anomalies are more than one standard deviation below the mean (e.g. El Niño years), whereas lowest values occur when the SOI index is near zero or positive (e.g. neutral or La Niña years). The PDO modifies the El Niño signal. When the PDO index exceeds ~1.5 standard deviations from the mean during El Niño events, the GEV location parameter μ increases while the scale parameter κ decreases. As noted previously, neural network models are capable of modelling interactions between covariates without a priori specification of the form of the interaction. In this example, the modification of ENSO impacts by PDO phase would be difficult to capture with a nonstationary linear model. One would, for example, need to explicitly include an interaction term in the regression equation. Conversely, the CDN-LIN2 model detects the nonlinearity automatically from the training data.
location, scale, and shape parameters of the GEV distribution. For example, the neural network can exactly replicate the GEV0, GEV1, and GEV11 models of El Adlouni et al. (2007), and can approximate the GEV2 model with good accuracy. Other forms of nonlinearity, for example, step changes, interactions between covariates, and higher-order polynomial relationships, can also be modelled with the same architecture.

Two model selection criteria, AICc and BIC, correctly identified the generating model for four weakly nonstationary synthetic test datasets. Although this paper focused on results from the single ‘best’ model, model averaging, which involves taking a weighted average of multiple models, has been recommended as a means of improving estimation performance Burnham and Anderson, 2004. This approach has been applied successfully in the context of CDN models by Carney et al. (2005) and is worth exploring for GEV-CDN models.

Confidence intervals calculated via the residual and parametric bootstrap led to overly optimistic estimates of uncertainty for low to moderate \( \tau \)-quantiles. This pattern continued for high \( \tau \)-quantiles with the residual bootstrap, whereas the parametric bootstrap led to broader than expected confidence intervals. Aggregated over all test cases, the parametric bootstrap outperformed the residual bootstrap. In general, results are similar to those found by Kharin and Zwiers (2004) for nonstationary models and Kyselý (2008) for stationary models. It is possible that alternative bootstrap approaches, for example, the bias-adjusted percentile estimators evaluated by Kysely (2008), might yield better calibrated confidence intervals, although improvements were modest for stationary GEV models.

Application of the GEV-CDN models to precipitation data from southern California identified a nonlinear relationship among PDO, SOI, and parameters of the GEV distribution. Results are consistent with other work linking interannual/interdecadal modes of climate variability

---

**Table VI. Top five models in terms of AICc and BIC for Randsburg precipitation series**

<table>
<thead>
<tr>
<th>Model</th>
<th>Nonstationary</th>
<th>Covariate(s)</th>
<th>AICc</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDN-NLIN2</td>
<td>( \mu, \alpha )</td>
<td>SOL, PDO</td>
<td>453.7</td>
<td>476.5</td>
</tr>
<tr>
<td>CDN-NLIN3</td>
<td>( \mu, \alpha )</td>
<td>SOL, PDO</td>
<td>454.8</td>
<td>481.9</td>
</tr>
<tr>
<td>CDN-NLIN2</td>
<td>( \mu, \alpha, \kappa )</td>
<td>SOL, PDO</td>
<td>457.3</td>
<td>482.1</td>
</tr>
<tr>
<td>CDN-NLIN3</td>
<td>( \mu, \alpha, \kappa )</td>
<td>SOL, PDO</td>
<td>465.5</td>
<td>493.4</td>
</tr>
<tr>
<td>CDN-NLIN2</td>
<td>( \mu, \alpha )</td>
<td>SOI</td>
<td>479.5</td>
<td>499.7</td>
</tr>
</tbody>
</table>

---

![Figure 4. Time series of annual precipitation at Randsburg (vertical bars), along with standardized SOI (dashed line) and PDO (dotted line) anomalies.](image1)

**Figure 4.** Time series of annual precipitation at Randsburg (vertical bars), along with standardized SOI (dashed line) and PDO (dotted line) anomalies. Values of \( r \) are Spearman rank correlation coefficients between SOI/PDO and precipitation.

![Figure 5. Contour plots of CDN-NLIN2 relationships among SOI, PDO, and (a) GEV location parameter \( \mu \), (b) GEV scale parameter \( \alpha \), and (c) \( \tau = 0.90 \) precipitation quantile. The GEV shape parameter \( \kappa = -0.02 \). Black dots indicate observed SOI and PDO index values.](image2)

**Figure 5.** Contour plots of CDN-NLIN2 relationships among SOI, PDO, and (a) GEV location parameter \( \mu \), (b) GEV scale parameter \( \alpha \), and (c) \( \tau = 0.90 \) precipitation quantile. The GEV shape parameter \( \kappa = -0.02 \). Black dots indicate observed SOI and PDO index values.
to local hydroclimatological datasets, for example, work by Gershunov and Barnett (1998) among others. Caution is required in the interpretation of the GEV-CDN model due to the small number of samples in some regions of the SOI/PDO phase space. The result does, however, demonstrate the ability of the GEV-CDN model to account for interactions between covariates without their a priori specification.

One perceived limitation of MLP neural networks and, by extension CDN models, is that they are ‘black boxes’ in which modelled relationships are difficult to interpret. It bears noting that sensitivity analysis methods, for example, the one used by Cannon and McKendry (2002), are applicable to CDN models and could be used to identify the form of nonlinear relationships between covariates and GEV distribution parameters or quantiles.

ACKNOWLEDGEMENTS

The author would like to thank three anonymous reviewers for their helpful and constructive comments. Portions of this work were conducted while visiting the Climate Prediction Group in the Department of Earth and Ocean Sciences at The University of British Columbia.

REFERENCES


Copyright © 2009 Her Majesty the Queen in right of Canada. Published by John Wiley & Sons, Ltd Hydrol. Process. 24, 673–685 (2010)