# Modeling Non-stationary Temperature Extremes with ProNEVA: A Case Study of Paris and Detroit

by

**Zehao Zhang**

A Research Paper

presented to the University of Waterloo

in partial fulfillment of the requirements for the degree of

Master of Mathematics

in

Actuarial Science

Waterloo, Ontario, Canada
May 23, 2025

# Table of Contents

# 1 Introduction

In the context of ongoing climate change, extreme temperature events such as heatwaves are becoming increasingly frequent and intense, posing serious risks to human health, infrastructure, and ecosystems (Ghimire et al., 2023; Ghalavand et al., 2025). Understanding and quantifying these rare events is essential for designing resilient urban systems and informed climate adaptation strategies.

Extreme Value Theory (EVT) provides a solid statistical foundation for modeling such extremes (Coles, 2001). Among the main EVT approaches, the block maxima method, based on the Generalized Extreme Value (GEV) distribution, is particularly suitable for analyzing annual maxima derived from temperature records. While classical applications often assume stationarity, the assumption that the statistical properties of extremes remain constant over time is increasingly challenged in light of climate trends (Katz, 2013; Cheng et al., 2014).

To address this, non-stationary extensions of the GEV model allow the distribution parameters, including location, scale, and shape, to evolve with covariates such as time or climate indices. However, estimating these models introduces substantial complexity. The ProNEVA (Process-informed Nonstationary Extreme Value Analysis) framework (Ragno et al., 2019a) offers a Bayesian approach for fitting both stationary and non-stationary GEV models, incorporating covariate effects, flexible model structures, and full uncertainty quantification through advanced Markov chain Monte Carlo techniques.

This study draws on over a century of gridded monthly temperature records with consistent temporal coverage and high spatial resolution. Such long-term observational data are critical for identifying meaningful trends in temperature extremes and for developing robust non-stationary models. The use of block maxima derived from these records ensures comparability across regions and strengthens the reliability of parameter estimation.

From a practical perspective, accurately modeling temperature extremes has important implications for climate adaptation and infrastructure planning. Design standards for buildings, transportation systems, and public services often rely on historical extremes. If

future intensification of extreme events is overlooked by assuming stationarity, this may result in infrastructure underdesign and heightened societal vulnerability. By examining how model selection affects risk estimates under different climatic conditions, this study contributes not only to statistical modeling but also to the evidence base for climate-resilient decision-making.

In this study, we apply the ProNEVA toolbox to evaluate extreme temperature behavior in two contrasting urban environments. Paris, France shows a statistically significant upward trend in annual maximum temperature, while Detroit, United States does not. Through a comparative analysis, we assess the suitability and performance of stationary versus non-stationary GEV models under these different climatic contexts. Model selection is guided by information criteria (AIC and BIC), prediction error metrics (RMSE and NSE), and goodness-of-fit diagnostics such as the Mann–Kendall and White tests (Renard and Lang, 2013; Salas and Obeysekera, 2014).

The next section introduces the theoretical foundations of stationary and non-stationary GEV models and outlines the ProNEVA methodology for parameter estimation and model diagnosis. Section 3 presents detailed case studies for Paris and Detroit, including trend testing, model fitting, and return level analysis. Section 4 concludes the study with key findings and future research directions.

# 2 Theoretical Framework

## 2.1 Stationary GEV model

The GEV distribution is used to model block maxima (Coles, 2001), especially for time series data. Dividing observations into time-based blocks is an effective way to reduce temporal dependence. For example, if we want to analyzing daily precipitation data, consecutive days often exhibit strong dependence due to persistent weather systems, such as storms. To reduce this dependency, a block maxima approach can be applied by dividing the data into annual blocks and extracting the maximum daily rainfall within each year. This method extracts a sequence of approximately independent extreme values, aligning better with the assumptions of classical extreme value theory.

The cumulative distribution function of GEV distribution defined as: (Coles, 2001)

$$G(z) = \begin{cases} \exp\left\{-\left[1+\xi\left(\dfrac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}, & \text{if } \xi \neq 0 \\ \\ \exp\left\{-\exp\left[-\left(\dfrac{z-\mu}{\sigma}\right)\right]\right\}, & \text{if } \xi = 0 \end{cases} \tag{1}$$

where $1+\xi(\frac{z-\mu}{\sigma}) > 0$. The model have 3 parameters: a location parameter, $\mu$, which determines where the distribution is positioned; a scale parameter, $\sigma$, which controls the spread or variability of values around the location; a shape parameter, $\xi$, which determines the heaviness and direction of the distribution's tail. The parameters satisfy $-\infty < \mu < \infty$, $\sigma > 0$, and $-\infty < \xi < \infty$. Depending on the value of the shape parameter $\xi$, the GEV distribution takes one of three forms: Weibull when $\xi < 0$, Gumbel when $\xi = 0$, and Fréchet when $\xi > 0$. The distribution is considered stationary when all parameters are assumed to remain constant across covariates, such as time.

## 2.2 Non-Stationary GEV Models

However, with the changing climate, the assumption of stationarity may no longer hold. In many applications, the parameters of the GEV distribution may vary with time or other

covariates, motivating the use of non-stationary models.

To account for non-stationarity, the location, scale, and shape parameters are modeled as functions of a covariate $x_c$, denoted by $\mu(x_c)$, $\sigma(x_c)$, and $\xi(x_c)$, respectively (Coles, 2001). The covariate $x_c$ is not restricted to a single variable and may have multiple dimensions. In climate-related applications, $x_c$ can represent time (to capture long-term trends), large-scale climate indices such as the El Niño–Southern Oscillation (ENSO), or other environmental factors such as humidity or precipitation. We can revise the equation of stationary GEV distribution, (1), into a non-stationary form as:

$$G(z \mid x_c) = \begin{cases} \exp\left\{-\left[1+\xi(x_c)\left(\dfrac{z-\mu(x_c)}{\sigma(x_c)}\right)\right]^{-1/\xi(x_c)}\right\}, & \text{if } \xi \neq 0 \\ \exp\left\{-\exp\left[-\left(\dfrac{z-\mu(x_c)}{\sigma(x_c)}\right)\right]\right\}, & \text{if } \xi = 0 \end{cases} \tag{2}$$

Typically, non-stationarity is introduced in the location and/or scale parameters, while the shape parameter is often held constant for model stability and interpretability. Compared to the stationary GEV model, non-stationary models involve greater complexity due to the functional relationships between parameters and covariates.

## 2.3 ProNEVA Methodology

To address the complexities of non-stationary extreme value analysis, Ragno et al. (2019a) developed a comprehensive framework known as Process-informed Nonstationary Extreme Value Analysis (ProNEVA). This framework has been implemented as a powerful and user-friendly MATLAB toolbox that facilitates both stationary and non-stationary extreme value modeling.

### 2.3.1 Parameter Assumption

In ProNEVA, non-stationarity is introduced by expressing the GEV parameters as functions of one or more covariates. The framework supports several functional forms for the

parameter-covariate relationships, including linear, quadratic, and exponential, as summarized in Table 1.

| **Type** | **Model** | $\mu(X_c)$ | $\ln(\sigma(X_c))$ | $\xi(X_c)$ |
|---|---|---|---|---|
| Linear | $B_l \cdot X_c + A_l$ | ✓ | ✓ | ✓ |
| Quadratic | $C_q \cdot X_c^2 + B_q \cdot X_c + A_q$ | ✓ | ✓ | – |
| Exponential | $A_e \cdot \exp(B_e \cdot X_c)$ | ✓ | – | – |

Table 1: GEV dependence models supported in ProNEVA and their applicability to the location $\mu(X_c)$, log-scale $\ln(\sigma(X_c))$, and shape $\xi(X_c)$ parameters

The functional form selected for one GEV parameter does not necessarily influence the choice of form for the others. To maintain the positivity of the scale parameter $\sigma(X_c)$, it is expressed on the logarithmic scale, following prior recommendations (Coles, 2001; Katz, 2013). Consequently, the exponential form is excluded from modeling $\sigma(X_c)$. In addition, the shape parameter $\xi(X_c)$ is well known for being difficult to estimate accurately, especially when working with limited time series data (Coles, 2001). As such, only the linear form is used for $\xi(X_c)$ in ProNEVA (Ragno et al., 2019a).

### 2.3.2 Parameter Estimation

In the stationary case, the parameters of the GEV distribution are typically estimated using the Maximum Likelihood Estimation (MLE) method, which provides closed-form or easily computable solutions under well-behaved likelihood surfaces. However, in the non-stationary setting, obtaining analytical solutions becomes considerably more difficult due to the increased complexity and dimensionality of the parameter space.

ProNEVA employs a Bayesian framework to estimate the parameters of both stationary and non-stationary extreme value analysis (EVA) models. The Bayesian approach enables a comprehensive quantification of uncertainties stemming from model structure, input data, and estimation procedures (Sadegh et al., 2017, 2018b). According to Bayes' theorem, the posterior distribution is proportional to the product of the prior distribution and the

likelihood function. For independent observations, the posterior density takes the form:

$$p(\boldsymbol{\theta} \mid \tilde{\mathbf{Y}}) \propto p(\boldsymbol{\theta}) \prod_{i=1}^{n} p(\tilde{y}_i \mid \boldsymbol{\theta})$$

where $\boldsymbol{\theta}$ represents the parameter vector, and $\tilde{\mathbf{Y}}$ denotes the observed data.

In ProNEVA, prior distributions for the model parameters can be specified as uniform, normal, or gamma, depending on the analyst's prior beliefs and the nature of the model (Sadegh et al., 2018b). For non-stationary models, the parameter vector includes additional terms associated with covariate effects, leading to a higher-dimensional inference problem.

To sample from the posterior, ProNEVA adopts a hybrid-evolution Markov Chain Monte Carlo (MCMC) method (Ragno et al., 2019a), originally introduced by Sadegh et al. (2017). This method runs multiple chains in parallel, dynamically sharing information between them. It combines global and local search strategies by integrating differential evolution (Storn and Price, 1997), adaptive Metropolis updates (Haario et al., 2001), and snooker update moves (Ter Braak and Vrugt, 2006). Intelligent initialization (Duan et al., 1993) improves convergence speed, and the Gelman–Rubin statistic $\hat{R}$ (Gelman and Rubin, 1992) is used to monitor convergence, with a recommended threshold of 1.2.

### 2.3.3 ProNEVA Analysis Procedure

Having described the parameter assumptions and the Bayesian estimation process, this section outlines the overall analysis procedure of the ProNEVA framework. The process integrates prior information, observed data, and Bayesian inference to generate return level analyses under both stationary and non-stationary assumptions. Figure 1 illustrates the complete modeling workflow.

In the Bayesian framework, we denote the collection of parameters to be estimated as $\boldsymbol{\theta} = (\theta_\mu, \theta_\sigma, \theta_\xi)$, where each component corresponds to the coefficients within the respective functional forms of $\mu(x_c)$, $\sigma(x_c)$, and $\xi(x_c)$. The dimension of each parameter vector depends on the specific form chosen (e.g., linear or quadratic), as described in Table 1. The subscript $N_{x_c}$ in observations is the number of covariates.

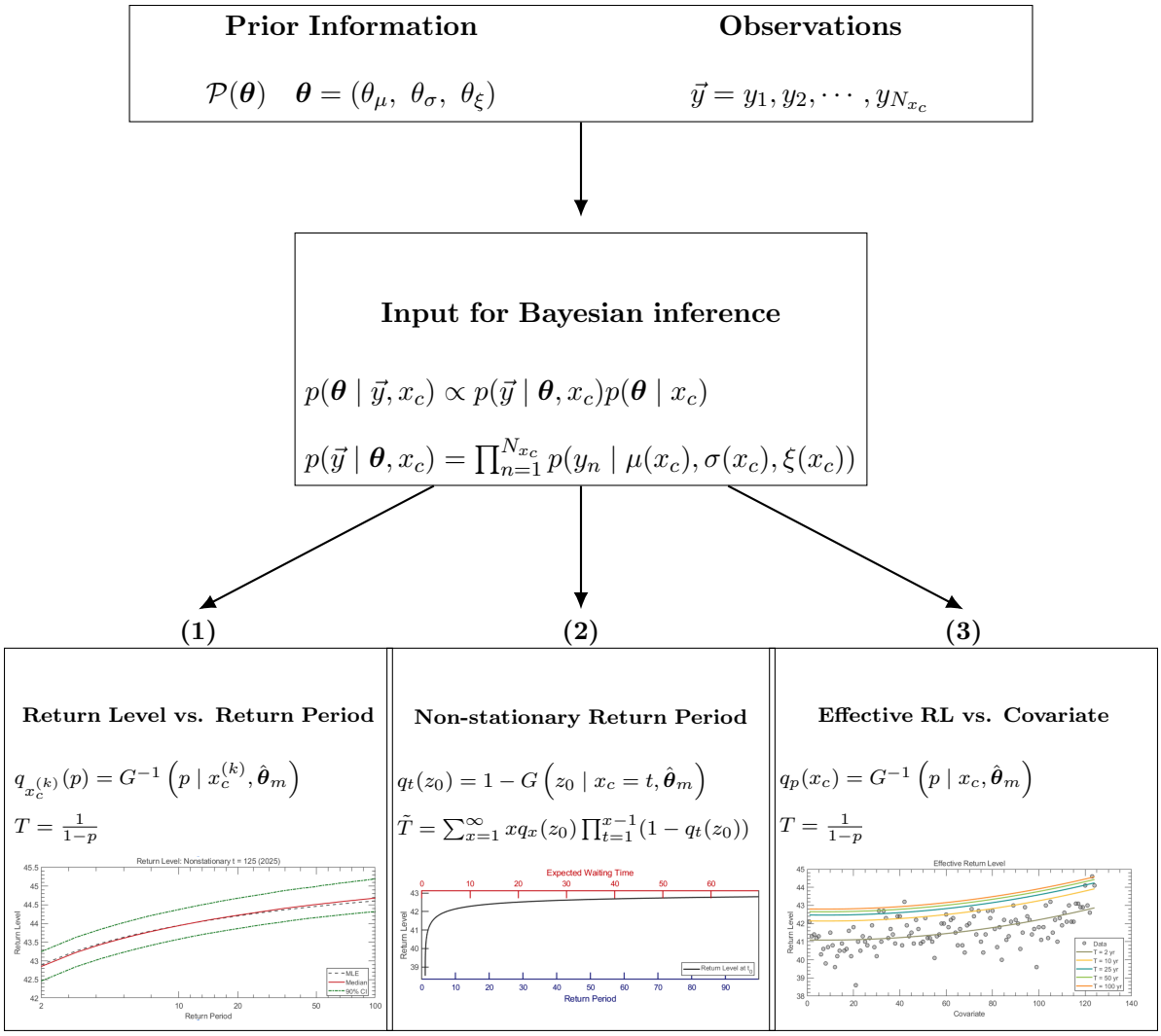


Figure 1: ProNEVA's non-stationary GEV framework for extreme value analysis. Model outputs include: (1) return level as a function of return period under fixed covariate quantiles; (2) non-stationary return period (expected waiting time) for a fixed return level; and (3) effective return levels as functions of covariate under fixed return periods

In graph (1), $q$ represents the return level and $T = 1/(1-p)$ is the corresponding return period, where $p$ is the non-exceedance probability. The covariate value $x_c^{(k)}$ denotes the $k$-th quantile of $x_c$. The choice of $k$ is user-defined and can be adjusted to reflect different levels of risk preference, with higher quantiles corresponding to more extreme covariate conditions. Here, $G^{-1}$ denotes the inverse of the GEV cumulative distribution function, and $\hat{\boldsymbol{\theta}}_m$ represents the maximum a posteriori (MAP; also known as the posterior mode, see

Gelman et al. (2013)) estimate of the parameter vector $\boldsymbol{\theta}$, obtained via hybrid-evolution MCMC sampling.

In graph (2), we analyze the expected waiting time under a non-stationary setting, which is applicable only when the covariate $x_c$ represents time. We first establish a baseline return level $z_0$ corresponding to a given return period at time $t = 0$. Then, using this fixed threshold $z_0$, we compute the time-varying exceedance probabilities $q_t(z_0)$ and derive the expected waiting time $\tilde{T}$ accordingly. Here, $\tilde{T}$ is assumed to follow a geometric distribution under non-stationary exceedance probabilities, following the formulation proposed by Salas and Obeysekera (2014).

In graph (3), we fix the return period $T$ and evaluate how the return level $q_p(x_c)$ varies with the covariate $x_c$. Common return periods include $T = 2$, 10, 25, 50, and 100 years.

### 2.3.4 Model Diagnosis and Selection Metrics

Before fitting the non-stationary Generalized Extreme Value (GEV) models, we first examine whether non-stationarity exists in the data. To this end, we apply the Mann–Kendall (MK) trend test and the White test. The MK test is a non-parametric method commonly used to detect monotonic trends in extreme values over time, without requiring any specific distributional assumptions (Coles, 2001). The White test, on the other hand, is used to detect heteroskedasticity, i.e., whether the variance of the residuals changes with respect to covariate values.

Based on the outcomes of these tests, we decide whether certain model parameters should be treated as non-stationary, i.e., modeled as functions of time or covariates.

Once the models are fitted, we evaluate their adequacy and performance using a set of model selection and diagnostic metrics. To assess how well the fitted distributions align with the observed data, we perform the Kolmogorov–Smirnov (K–S) test (Massey Jr, 1951), and then compute the rejection rate (RJ rate), defined as the proportion of K–S test resamples that reject the null hypothesis (Renard and Lang, 2013). A lower RJ rate indicates better agreement between the model-predicted and empirical distributions.

We also use several quantitative indicators to compare competing models. The Akaike Information Criterion (AIC) (Akaike, 1974) and the Bayesian Information Criterion (BIC) (Schwarz, 1978) measure the trade-off between model fit and complexity. Lower values are preferred as they indicate more parsimonious models with adequate explanatory power (Serinaldi and Kilsby, 2015). The Root Mean Squared Error (RMSE) is used to quantify the average prediction error, with smaller values suggesting more accurate predictions. Additionally, the Nash–Sutcliffe Efficiency (NSE), widely used in hydrologic and climatologic modeling, evaluates the predictive skill of the model relative to the observed mean (Sadegh et al., 2018a). An NSE value closer to 1 indicates higher predictive performance.

Together, these metrics provide a comprehensive framework for assessing the adequacy of each model and facilitate selection of the most suitable model under non-stationary conditions by balancing goodness-of-fit, model simplicity, and predictive accuracy.

# 3 Case Study: Extreme Temperature Analysis

To demonstrate the practical application of the ProNEVA toolbox, we first conducted a global trend analysis using the MK test on gridded annual maximum temperature data from the Climatic Research Unit (CRU). The dataset provides monthly maximum temperature (tmx) observations from 1901 to 2024, at a spatial resolution of 0.5° × 0.5°, as detailed in Harris et al. (2020); Climatic Research Unit (2025). For each grid cell, annual maxima were computed and tested for statistically significant monotonic trends.

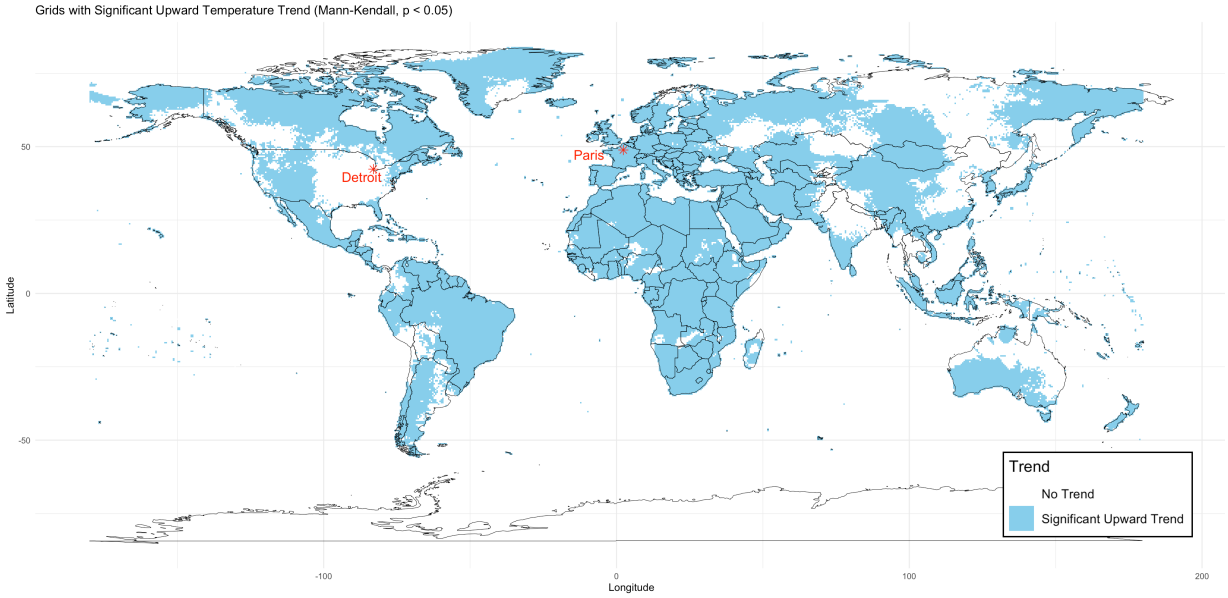


Figure 2: Global map showing locations with significant upward trends (red) in annual maximum temperature based on the MK test at the 5% significance level.

Based on the global trend results (Figure 2), we selected two representative locations for detailed case studies. Paris, France (Latitude: 48.8575° N, Longitude: 2.3514° E), which exhibits a statistically significant trend in annual maximum temperature, was chosen to demonstrate the practical application of the ProNEVA toolbox under non-stationary conditions. In contrast, Detroit, United States (Latitude: 43.4643° N, Longitude: 87.6298° W), which does not show a statistically significant trend, was included to highlight the potential risks of overfitting when applying non-stationary extreme value models in regions without clear temporal changes. These two case studies allow for a comparative evaluation of

model performance and appropriateness under differing climatic trends. Details regarding model fitting procedures, including prior assumptions (Appendix A), MCMC implementation (Appendix B), and code-level adjustments to the ProNEVA toolbox (Appendix C), are provided in the Appendices.

## 3.1 Paris

### 3.1.1 Non-stationary Tests

To assess the non-stationarity of the location and scale parameters in Paris, we conducted both the MK test and the White test. The MK test yielded a statistically significant p-value ($p = 0.0239$), indicating the presence of a potential trend in the location parameter $\mu(t)$. In contrast, the White test returned a non-significant p-value ($p = 0.3621$), suggesting that the scale parameter $\sigma(t)$ can be reasonably assumed to remain constant over time. As there is currently no widely accepted statistical test for detecting time-varying behavior in the shape parameter $\xi(t)$, we opted to fit models both with and without a time-varying $\xi$ for comparison.

### 3.1.2 Model Selection

Model performance was assessed using a suite of statistical metrics, as summarized in Table 2. To identify the most appropriate GEV model for characterizing temperature extremes in Paris, we considered four alternative formulations that differ in how the location parameter $\mu$ varies over time: stationary, linear, quadratic, and exponential trends. For each formulation of $\mu$, we evaluated both constant and linear trend specifications for the shape parameter $\xi$, resulting in a total of eight candidate models.

As shown in Table 2, the best-performing model in terms of AIC and BIC is the one with a quadratic trend in the location parameter $\mu$ and a constant shape parameter $\xi$. This model achieved the lowest AIC (468.78) and BIC (482.89) values, along with the lowest RJ rejection rate (0.55%), indicating strong model adequacy and efficient convergence.

| Model | | RJ rate(%) | AIC | BIC | RMSE | NSE |
|---|---|---|---|---|---|---|
| Stationary | | 0.93 | 480.7766 | 489.2374 | 1.7487 | 0.9846 |
| Linear $\mu$ | constant $\xi$ | 1.23 | 477.5403 | 488.8215 | 1.3470 | 0.9911 |
| | linear $\xi$ | 1.06 | 479.2027 | 493.3041 | 1.4781 | 0.9888 |
| Quad $\mu$ | constant $\xi$ | **0.55** | **468.7836** | **482.8850** | 1.5062 | 0.9884 |
| | linear $\xi$ | 0.86 | 470.5860 | 487.5077 | 1.3484 | 0.9906 |
| Exp $\mu$ | constant $\xi$ | 1.29 | 477.3862 | 488.6673 | **1.1639** | **0.9932** |
| | linear $\xi$ | 1.19 | 479.3009 | 493.4023 | 1.4692 | 0.9894 |

Table 2: Model Comparison - Paris

However, the model with an exponential trend in $\mu$ and constant $\xi$ achieved the best performance in terms of predictive accuracy, yielding the lowest RMSE (1.1639) and highest NSE (0.9932). Despite this, it also exhibited a higher rejection rate (1.29%) and did not outperform the quadratic model significantly in AIC/BIC.

Given these trade-offs, we selected the quadratic $\mu$ with constant $\xi$ model as the final specification for Paris. While the exponential model offered slightly better predictive accuracy, the quadratic model achieved a better balance of parsimony, statistical fit, and stability. Additionally, the minimal improvements in RMSE and NSE did not justify the increased model complexity and higher rejection rate associated with the exponential formulation.

Beyond the quantitative metrics, we also provide a complementary visual evaluation, using quantile-quantile (Q-Q) plots, to illustrate how well the stationary and non-stationary models capture the empirical distribution of extremes. Figure 3 compares the fit of the stationary model (left) and the non-stationary model (right). In both plots, the empirical quantiles (x-axis) are plotted against the theoretical quantiles from the respective fitted models (y-axis), with the 1:1 reference line shown in red.

Overall, both models show reasonable agreement with the observed data. However,

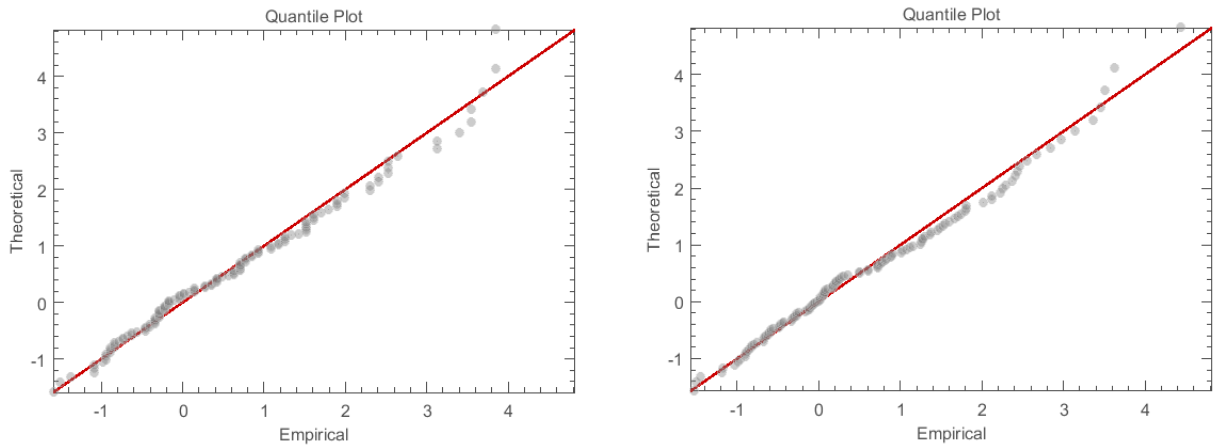


Figure 3: Q-Q plots comparing stationary (left) and non-stationary (right) model fits in 2025

the non-stationary model (right) exhibits a noticeably tighter alignment with the 1:1 line, especially in the upper tail. This indicates an improved fit to extreme values, suggesting that the non-stationary formulation more effectively captures the evolving behavior of temperature extremes. These graphical diagnostics complement the statistical evidence from model comparison metrics (e.g., AIC, BIC, KS test), reinforcing the suitability of the non-stationary framework under a changing climate.

### 3.1.3 Temporal Evolution of Return Levels

To further investigate how climate change affects extreme temperatures, return level plots were generated for two distinct years: 1962 (t = 62) and 2025 (t = 125) using the non-stationary GEV model as shown in Figure 4.

In 1962, the estimated 10-year return level was approximately 26.6°C, with a 90% confidence interval between 26.2°C and 27°C. By 2025, the 10-year return level had increased to approximately 28.8°C, with a 90% CI spanning 29.7°C to 28°C.

This shift implies that events once considered rare are becoming more frequent and more extreme. The growing gap between curves over time underscores the importance of accounting for non-stationarity in climate risk projections.

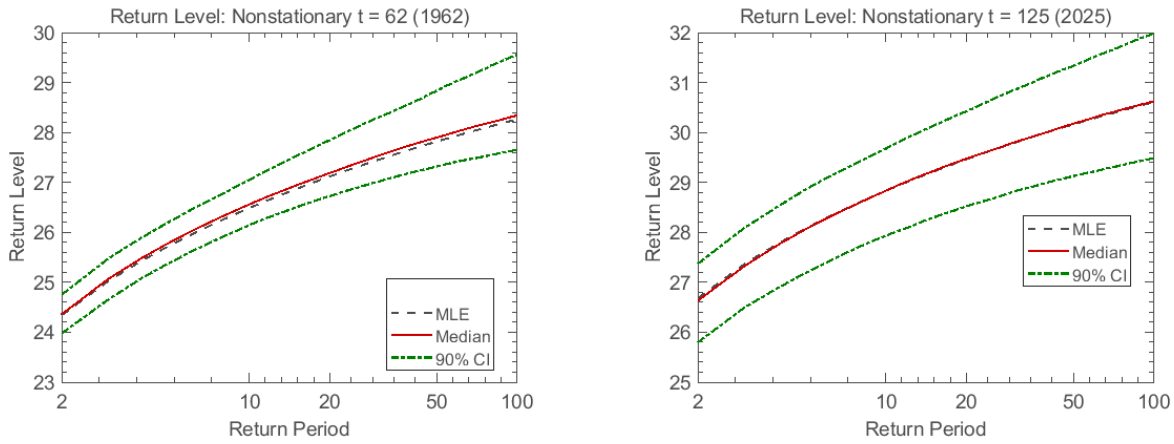


Figure 4: Return level curves in 1962 and 2025 under the non-stationary model

### 3.1.4 Non-stationary Return Period and Expected Waiting Time

While the previous section highlights how return levels shift over time due to a non-stationary climate, it is also important to consider how the recurrence of such extremes should be interpreted under this changing baseline.

Figure 5 compares return period (blue x-axis) and expected waiting time (red x-axis) for the same return levels. Although both metrics describe how often an event may occur, they reflect different assumptions.

The return period is defined with respect to the initial time point $t = 0$, assuming that distribution parameters remain fixed at their baseline values. In contrast, the expected waiting time accounts for time-varying characteristics and reflects the average time until such an event occurs under the non-stationary trajectory.

For example, at $t = 0$, the return level associated with a 20-year return period has an expected waiting time of approximately 14 years, which is clearly much shorter than the nominal 20 years. This aligns with our intuition: under a changing climate, extremes may occur sooner than a stationary return period would suggest.

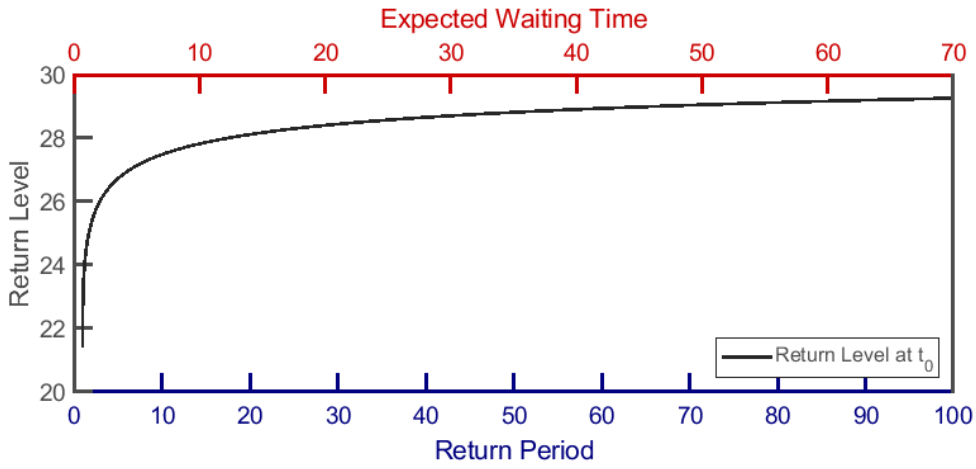


Figure 5: Return level versus return period and expected waiting time

### 3.1.5 Effective Return Level Over Time

Figure 6 displays the effective return levels as a function of time (used here as a covariate), across a range of return periods (2, 10, 25, 50, and 100 years). Each curve exhibits a distinct U-shape, with return levels decreasing initially and then rising as time progresses. This non-linear pattern reflects the influence of a quadratic trend in the location parameter $\mu(t)$, which is specified as a quadratic function of time in our non-stationary model.

The curvature becomes more pronounced for longer return periods, indicating that the magnitude of extreme events is increasing more sharply in recent years. This behavior suggests that not only are extreme heat events becoming more frequent, but their severity is also accelerating with time. As such, the figure provides strong visual evidence for a non-linear intensification of extremes, consistent with the hypothesized quadratic temporal evolution.

## 3.2 Detroit

Following the analysis of Paris, we applied the same set of statistical tests to the Detroit series to evaluate potential non-stationarity. In contrast to Paris, however, the results

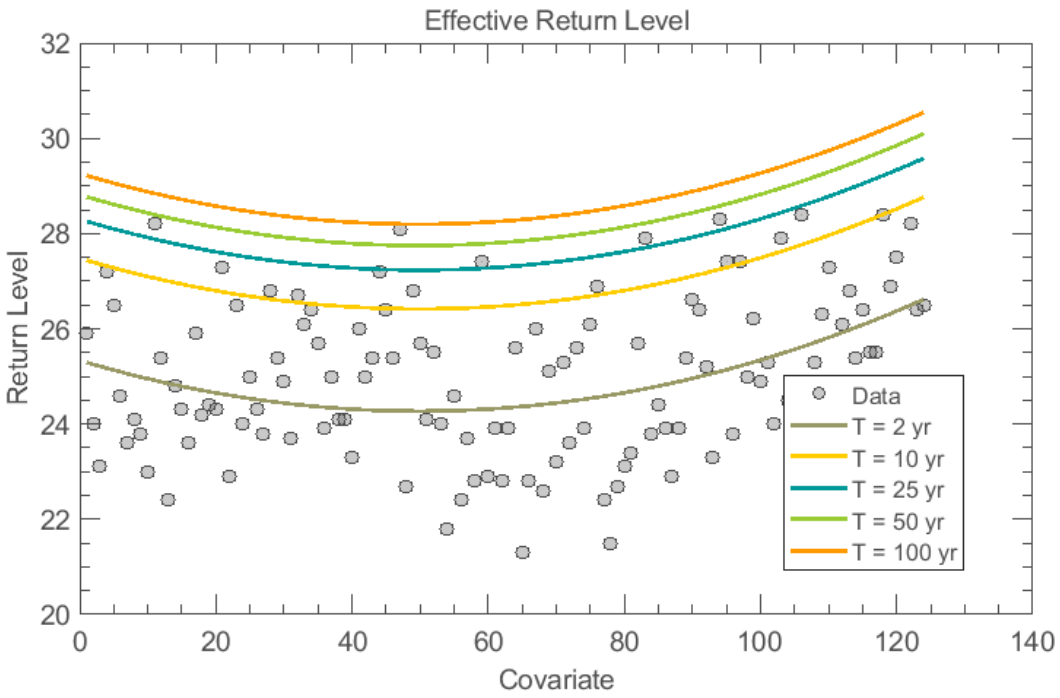


Figure 6: Effective return levels over time for multiple return periods

showed no significant trends in either the location or scale parameters. The MK trend test yielded a non-significant p-value ($p = 0.5715$), indicating no evidence of a potential trend in the location parameter $\mu(t)$. Similarly, the White test returned a non-significant p-value ($p = 0.5461$), suggesting that the scale parameter $\sigma(t)$ can be reasonably assumed to remain constant over time. These results provide no statistical support for incorporating time-varying components into the model.

Nevertheless, to illustrate the potential pitfalls of overfitting and the limitations of indiscriminately applying non-stationary models, we proceeded to fit a full suite of non-stationary specifications for demonstration purposes. Specifically, we assumed that the location parameter $\mu(t)$ follows a linear, quadratic, or exponential trend, while holding the scale parameter $\sigma$ constant. As in the Paris analysis, we considered both constant and linearly time-varying forms for the shape parameter $\xi(t)$, resulting in seven non-stationary candidate models in addition to the stationary baseline.

### 3.2.1 Model Selection

Model comparison results for Detroit are summarized in Table 3. Consistent with expectations, the stationary model, despite its simplicity, achieved the best performance in terms of AIC (427.30) and BIC (435.76), and had the lowest RJ rejection rate (0.22%), indicating strong adequacy and stable convergence.

| **Model** | | **RJ rate**(%) | **AIC** | **BIC** | **RMSE** | **NSE** |
|---|---|---|---|---|---|---|
| Stationary | | **0.22** | **427.3011** | **435.7620** | 1.7748 | 0.9847 |
| Linear $\mu$ | constant $\xi$ | 0.42 | 428.7951 | 440.0762 | 1.6725 | 0.9857 |
| | linear $\xi$ | 0.27 | 430.6551 | 444.7565 | **1.4397** | **0.9894** |
| Quad $\mu$ | constant $\xi$ | 0.32 | 430.9495 | 445.0509 | 1.7299 | 0.9854 |
| | linear $\xi$ | 0.42 | 432.6488 | 449.5705 | 1.5231 | 0.9883 |
| Exp $\mu$ | constant $\xi$ | 0.26 | 428.7252 | 440.0063 | 1.7233 | 0.9850 |
| | linear $\xi$ | 0.29 | 430.8067 | 444.9081 | 1.5155 | 0.9887 |

Table 3: Model Comparison - Detroit

Although the linear $\mu$ with linear $\xi$ model yielded the best predictive accuracy, with the lowest RMSE (1.4397) and highest NSE (0.9894), these improvements over the stationary model were marginal. Furthermore, the non-stationary specification incurred increased model complexity, a higher rejection rate (0.27%), and substantially worse AIC/BIC scores.

Other non-stationary variants displayed similar trade-offs. In several cases, they performed worse overall in terms of information criteria and convergence reliability, as reflected by higher AIC/BIC values and elevated RJ rejection rates.

Overall, these results show that non-stationary models, though more flexible, are not necessarily better. Without significant trends, their added complexity can lead to overfitting and reduced interpretability. For Detroit, the stationary model remains the most defensible choice both empirically and conceptually.

# 4 Conclusion

This study employed the ProNEVA framework to examine non-stationary behavior in extreme temperature data at two contrasting urban locations: Paris, France, and Detroit, United States. By applying both statistical trend tests and Bayesian model fitting, we evaluated the adequacy of stationary versus non-stationary GEV models under varying climatic conditions.

For Paris, where a statistically significant upward trend in the location parameter was detected, non-stationary models outperformed the stationary counterpart. The selected quadratic trend for $\mu(t)$ provided a balanced trade-off between statistical fit, predictive performance, and model stability. Visual diagnostics such as Q–Q plots and return level trajectories further reinforced the conclusion that extremes in Paris are intensifying over time, consistent with the evolving climate.

In contrast, Detroit exhibited no significant trends in either location or scale parameters. While non-stationary models offered slight improvements in RMSE and NSE, these gains were marginal and came at the cost of higher model complexity, rejection rates, and worse AIC/BIC scores. The stationary model, by contrast, demonstrated strong convergence and parsimony, underscoring the risks of overfitting when applying non-stationary models indiscriminately.

Taken together, these findings emphasize that non-stationary modeling should be guided by empirical evidence rather than assumed by default. While flexible, non-stationary models are not inherently superior and may degrade interpretability and robustness in the absence of meaningful trends.

Future work could explore the inclusion of additional covariates beyond time, such as land use, humidity, or teleconnection indices (e.g., ENSO), to capture complex climate dynamics. Moreover, modeling frameworks that account for potential dependencies among GEV parameters, whether through joint priors or copula based structures, could further improve uncertainty quantification and realism, particularly for risk projections involving rare events.

# References


Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723.

Cheng, L., AghaKouchak, A., Gilleland, E., and Katz, R. W. (2014). Non-stationary extreme value analysis in a changing climate. *Climatic Change*, 127(2):353–369.

Climatic Research Unit (2025). Cru ts v4.09: Gridded time-series dataset of monthly climate observations. https://crudata.uea.ac.uk/cru/data/hrg/. Version released March 19, 2025. Covers 1901–2024. Accessed April 2025.

Coles, S. G. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer.

Duan, Q., Sorooshian, S., and Gupta, V. (1993). Shuffled complex evolution approach for effective and efficient global minimization. *Journal of Optimization Theory and Applications*, 76(3):501–521.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis*. Chapman and Hall/CRC, Boca Raton, FL, 3rd edition.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472.

Ghalavand, M. R., Farajzadeh, M., and Ghavidel Rahimi, Y. (2025). Modeling return levels of non-stationary rainfall extremes due to climate change. *Atmosphere*, 16(2):136.

Ghimire, B., Kharel, G., Gebremichael, E., and Cheng, L. (2023). Evaluating non-stationarity in precipitation intensity-duration-frequency curves for the dallas–fort worth metroplex, texas, usa. *Hydrology*, 10(12):229.

Haario, H., Saksman, E., and Tamminen, J. (2001). An adaptive metropolis algorithm. *Bernoulli*, 7(2):223–242.

Harris, I., Osborn, T. J., Jones, P., and Lister, D. (2020). Version 4 of the cru ts monthly high-resolution gridded multivariate climate dataset. *Scientific Data*, 7:109.

Katz, R. W. (2013). Statistical methods for nonstationary extremes. In AghaKouchak, A., Easterling, D., Hsu, K., Schubert, S., and Sorooshian, S., editors, *Extremes in a Changing Climate: Detection, Analysis and Uncertainty*, pages 15–37. Springer Netherlands, Dordrecht.

Massey Jr, F. J. (1951). The kolmogorov-smirnov test for goodness of fit. *Journal of the American Statistical Association*, 46(253):68–78.

Ragno, E., AghaKouchak, A., Cheng, L., and Sadegh, M. (2019a). A generalized framework for process-informed nonstationary extreme value analysis. *Advances in Water Resources*, 130:270–282.

Ragno, E., AghaKouchak, A., Cheng, L., and Sadegh, M. (2019b). Supplementary material for: A generalized framework for process-informed nonstationary extreme value analysis. Available at: https://ars.els-cdn.com/content/image/1-s2.0-S0309170818309540-mmc1.pdf (Supplementary Data S1). Supplementary information to the article published in *Advances in Water Resources*, Volume 130, 2019, Pages 270–282.

Renard, B. and Lang, M. (2013). Bayesian methods for modeling nonstationary extremes: A review. *Water Resources Research*, 49(1):51–63.

Sadegh, M., Ragno, E., and AghaKouchak, A. (2018a). A statistical framework to analyze hydrological extremes under climate change. *Hydrology and Earth System Sciences*, 22:593–608.

Sadegh, M., Vrugt, J., and Gupta, H. (2018b). A probabilistic approach for estimating the uncertainties of streamflow hydrographs. *Water Resources Research*, 54(5):3487–3510.

Sadegh, M., Vrugt, J., and Tolson, B. (2017). Bayesian modeling of nonstationary precipitation extremes using synthetic design hydrographs. *Water Resources Research*, 53(1):478–500.

Salas, J. D. and Obeysekera, J. (2014). Revisiting the concepts of return period and

risk for nonstationary hydrologic extreme events. *Journal of Hydrologic Engineering*, 19(3):554–568.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464.

Serinaldi, F. and Kilsby, C. G. (2015). Stationarity is undead: Uncertainty dominates the distribution of extremes. *Advances in Water Resources*, 79:17–36.

Storn, R. and Price, K. (1997). Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11(4):341–359.

Ter Braak, C. J. and Vrugt, J. A. (2006). A markov chain monte carlo version of the genetic algorithm differential evolution: easy bayesian computing for real parameter spaces. *Statistics and Computing*, 16(3):239–249.

# APPENDICES

# A Prior Assumptions in ProNEVA GEV Configuration

## A.1 Using the Graphical User Interface

Figure 7 shows the graphical user interface of ProNEVA used to configure prior assumptions for the Generalized Extreme Value (GEV) distribution. The prior distributions and their respective parameters were defined for the three GEV parameters: location ($\mu$), scale ($\sigma$), and shape ($\xi$).

## A.2 Prior Distribution Settings

The Table 4 summarizes the prior configurations for each parameter, followed by a discussion of the assumptions and motivations behind each choice.

| **Parameter** | **Prior Distribution** | **Parameters** |
|---|---|---|
| Location ($\mu$) | Normal | $\mu = 0, \quad \sigma = 100$ |
| Scale ($\sigma$) | Normal | $\mu = 0, \quad \sigma = 10$ |
| Shape ($\xi$) | Normal | $\mu = 0, \quad \sigma = 0.2$ |

Table 4: Prior Distributions and Parameters

The prior distributions for the MCMC parameter estimation were defined according to the default recommendations in the ProNEVA toolbox (Ragno et al., 2019a). Specifically, a weakly-informative normal prior $\mathcal{N}(0, 100)$ was chosen for the location parameter $\mu$ to accommodate a wide range of plausible values without introducing strong bias. For the scale parameter $\sigma$, a prior $\mathcal{N}(0, 10)$ was used, which balances flexibility and numerical stability. The shape parameter $\xi$ was assigned a more concentrated prior $\mathcal{N}(0, 0.2)$, reflecting the

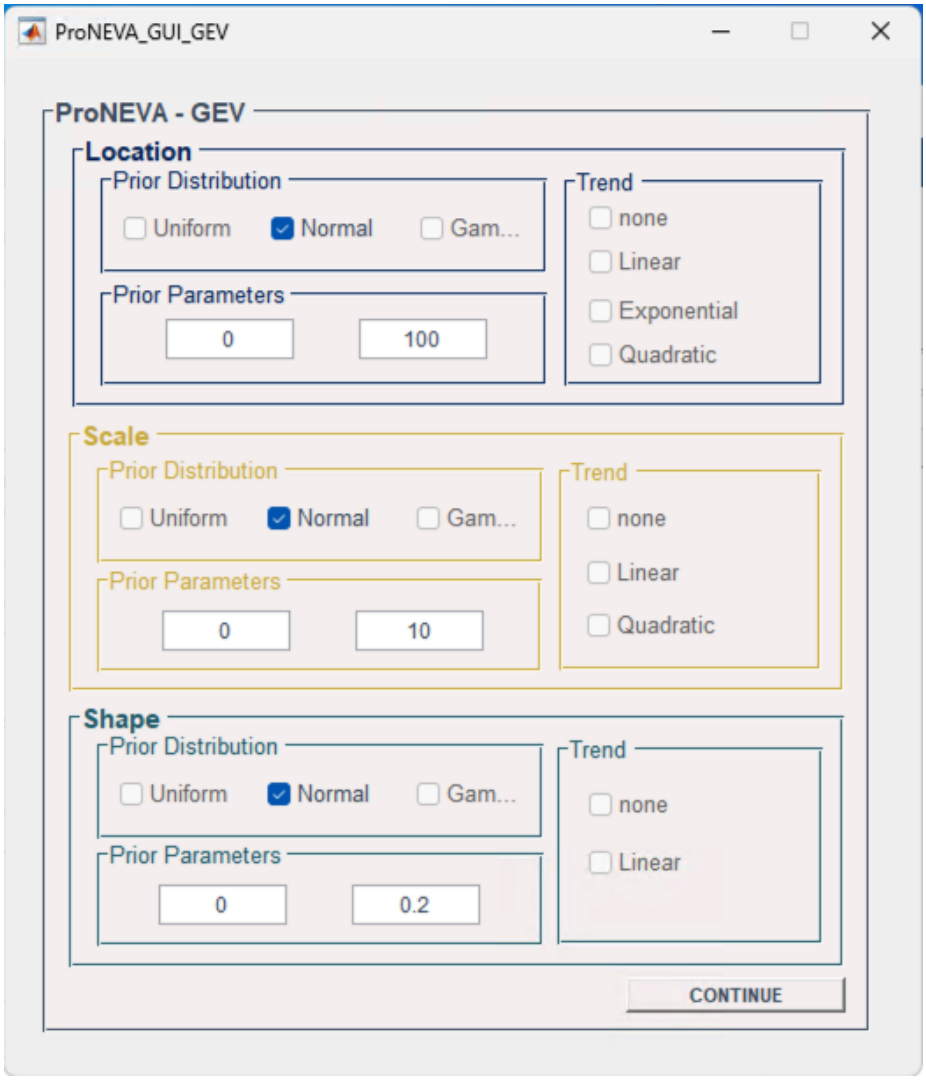


Figure 7: Prior distribution settings for the GEV parameters in ProNEVA GUI

difficulty in estimating this parameter and its strong influence on tail behavior. This prior setup ensures robust and stable inference while maintaining consistency with established practices in nonstationary extreme value analysis.

# B MCMC Assumptions in ProNEVA GEV Inference

## B.1 Using the Graphical Interface

Figure 8 shows the MCMC configuration panel in the ProNEVA GUI. This interface is used to specify key sampling parameters, return periods for analysis, and options for result saving and plotting.

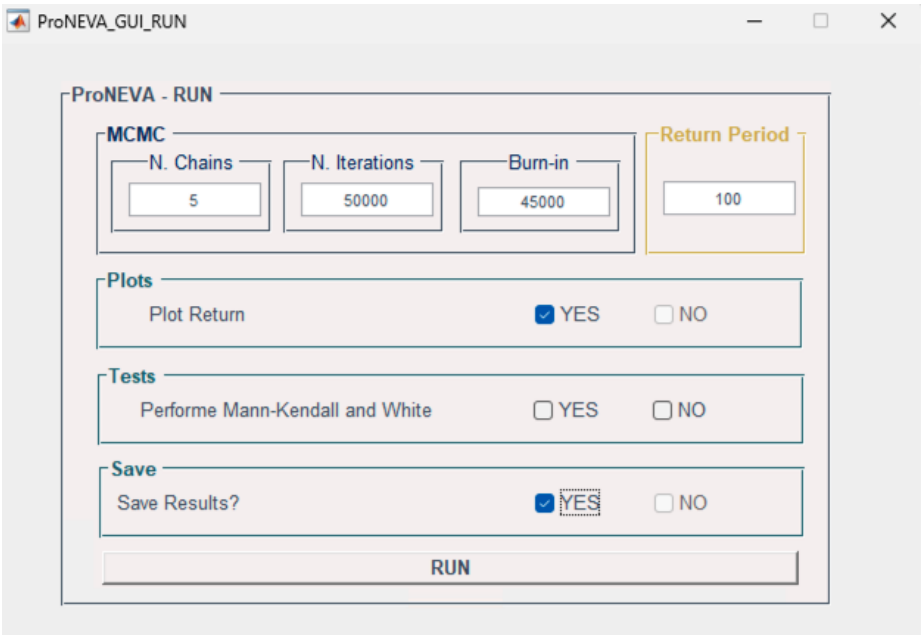


Figure 8: MCMC sampling and analysis options in the ProNEVA GUI

## B.2 MCMC Configuration

The Markov Chain Monte Carlo (MCMC) sampling was configured to explore the posterior distribution of the GEV parameters. The baseline setup for the majority of prior combinations used the following settings:

- **Number of Chains:** 5
- **Number of Iterations:** 50,000
- **Burn-in:** 45,000

- **Return Period:** 100 years

The MCMC configuration used throughout this study included 5 chains, 50,000 iterations, and a burn-in period of 45,000, with a return period set at 100 years. These settings were consistently applied across all model runs to ensure comparability and efficient sampling.

# C ProNEVA Code Adjustment

## Code Adjustments for Reproducibility and Forecast Horizon

To ensure full reproducibility of the results, a random seed was explicitly set in the file `/ProNEVA/RUN_GUI_for_ProNEVA.m` using the command `rng(20885604)`. This modification guarantees that all MCMC chains and derived quantities are reproducible across different runs.

In addition, a minor change was made to the MATLAB plot generating programm: `/ProNEVA/ProNEVApackage/PLOTS.m`. By default, when the covariate type is set to `Time`, the code appends an additional covariate value `VC3 = length(t) + median(t)` to extend the return level plot beyond the observational period. While this setup may be appropriate for datasets with moderate sample sizes, in our case, with $t = 124$, it resulted in a projection over 60 years into the future, which we considered overly speculative. To address this, we modified the code to use `VC3 = length(t) + 1`, thereby limiting the extended forecast to just one year ahead. This change maintains a more interpretable and policy-relevant forecast horizon.

The project code is included in this GitHub repository.