Climate Change Adaptation: Uncertainty Associated with Sequential Finite Period Decisions under Nonstationarity

James Doss-Gollin[1,2], Upmanu Lall[1,2], and David Farnham[1,2]

[1]Columbia Water Center, Columbia University
[2]Department of Earth and Environmental Engineering, Columbia University

January 23, 2018

Abstract

Hydroclimatic systems exhibit organized low-frequency and regime-like variability at multiple time scales, causing the risk associated with climate extremes such as floods and droughts to vary at multiple timescales. Despite broad recognition of this nonstationarity, there has been little theoretical development of ideas for the design and operation of infrastructure considering the regime structure of such changes and their potential predictability. This theory is well-suited to analysis of flood adaptation strategies with fixed project operation period M, but not for analysis of sequential decision strategies with M. In this paper we illustrate some basic considerations of uncertainty and risk analysis associated with the sequential decisions that may need to be made for climate risk adaptation in a nonstationary adaptation. We simulate synthetic data and fit the resulting simulations in a Bayesian context using stationary and non-stationary models to show that `RESULT`. This is a first conceptual step to the development of a full sequential decision analysis approach under nonstationarity.

1 Introduction

Hydroclimatic systems vary on many timescales, including regime-like variability and secular variability (Hodgkins et al., 2017; Hurst, 1951; Matalas, 2012; Merz et al., 2014; Milly et al., 2008; Sveinsson et al., 2005). This multi-scale variability greatly complicates the estimation of future flood sequences, leading to "low confidence in projections of changes in fluvial floods" (IPCC, 2012). Yet the same memory and low-frequency variability that complicate projecting flood sequences far into the future may also impart some predictability at short timescales (Jain and Lall, 2001). The existence of this information gap suggests that approaches for adapting to hydroclimate risks, such as floods, which have short project operation periods, may be preferential in some cases to large and permanent projects with long project operations.

Classical methods for design of infrastructure first specify a return period T (i.e. 100 years), and then design a project that protects against this T-year event. Much of the flood frequency analysis (FFA) literature has tended to focus on the details of flood frequency estimation and on trend detection associated with climate change and land use cover (Kidson and Richards, 2016; Merz et al., 2014; Merz and Blöschl, 2008). The mathematical problem of interest then becomes estimation of the probability distribution $f(p_T)$, where p_T is binary event that the threshold Q_T is exceeded:

$$p_T(t) \equiv \mathbb{P}[Q(t) \geq Q_T] \quad \text{for} \quad t = [t_0+1, \ldots, t_0+M]$$

and where t_0 is the first year forecast and M is the project operation period. Although in a stationary setting an unbiased estimator of $Pr[Q(t) \geq Q_T]$ converges asymtotically to $1 - \frac{1}{T}$, the asymmetric uncertainty distribution of the estimate has also prompted the consideration of the potential for over- or under-design for floods if the conditional mean of the estimated distribution of Q_T is used for design (Stedinger, 1997)[1].

Alternatively, risk-based design methods (RBDM; see Rosner et al., 2014) recognize that the appropriate level of design may itself be a decision variable. Lall (1987) considered the uncertainty in the frequency with which floods may exceed the design level for different values of p (equivalent to different values of T), and also for different sample sizes for estimation N and different project operation periods M. The duration M may reflect the physical life of a proposed infrastructure element, or its economic life. In the context of financial instruments which may be used for risk mitigation, M specifies the length of the contract and Q_T the threshold at which a payout is triggered. In such a setting, the mathematical problem is to estimate the future distribution of $Q(t)$ rather than just the future distribution of the binary outcome p_T.

Under classical assumptions of stationarity, all estimates are assumed independent of time. Yet even under stationarity, Jain and Lall (2001) showed that low-frequency variability can lead to a high probability of surprise – particularly when the length of the record used for estimation N is short. This finding is particularly important to keep in mind given that low-frequency change has been found to be large as compared to detected trends in the observed record (Hodgkins et al., 2017).

Of course, the hydroclimate systems are not stationary, due to climate, river modification, and land use change (Merz et al., 2014; Milly et al., 2008). However, predicting future hydrologic behavior in a particular location is inherently difficult given limited data and complex, highly nonlinear physics. Physical scaling arguments for an intensified hydrologic cycle under climate change (see Muller et al., 2011; O'Gorman, 2015) do not in general hold over land (Byrne and O'Gorman, 2015; Shaw et al., 2016). Attempts to use a "model chain" approach emcompassing: (i) emission scenario; (ii) general circulation model (GCM); (iii) downscaling; (iv) hydrological catchment model; and (v) flood frequency analysis, typically with bias correction applied at several steps, can lead to results which are difficult to interpret in a probabilistic context (Dankers and Feyen, 2009; Dittes et al., 2017; Merz et al., 2014; Ott et al., 2013). In particular, the downscaling and bias correction methods typically applied assume stationary relationships (i.e. between GCM rainfall and observed rainfall) which have no physical basis, particularly in a changing climate. Recent work has also sought to shorten the above model chain by modeling $Q(X(t))$, where X is a set of climate state variables from a GCM run, and where the relationship between X and Q can more reasonably be assumed stationary (Delgado et al., 2014; Hall et al., 2014; Silva et al., 2016). Alternatively, purely statistical approaches have extended classical flood frequency analysis by incorporating a time trend in the parameters of statistical distributions (Obeysekera and Salas, 2014; Serinaldi and Kilsby, 2015; Strupczewski et al., 2001; Vogel et al., 2011). In this case, however, the analyst can choose not only the distribution but also the time parameterization, if any, for each parameter, leading to many researcher degrees of freedom ("forking paths" or "multiple comparisons"; Gelman and Loken, 2013) and exacerbating the problems caused by lack of theory for model choice even under nonstationarity (Kidson and Richards, 2016).

The wide variety of methods, each with its own advantages and disadvantages, for estimating the future distribution of flooding $p(Q(t))$ illustrates the theoretical need for approaches to incorporate uncertain estimates of future floods into a decision framework. As we have argued, however, low-frequency variability or system memory may lead to less uncertain probabilistic forecasts at short times than long times. There is thus *a critical theoretical need for decision frameworks which can compare projects of different operation period* using an arbitrary choice of model(s) of future flood behavior. Sequential decision models (see Howard, 1960; Russell, 2003) allow decision-makers to optimize not only what action to take, but also when to take it. Sequential decision models have been used in the hydrological

[1] Vogel198x citation

literature for sea wall optimization (Lickley et al., 2014)[2].

In this context, we consider financial risk mitigation instruments in which the price is a function of the amount of coverage C, the expected probability that the insurance is triggered $\mathbb{E}(p_T) \equiv \frac{1}{T}$, and the variance (or more generally full uncertainty distribution) associated with this estimate $\mathbb{V}(p_T)$. Understanding the full uncertainty associated with thresholds of interest is relevant to many other climate adaptation problems, including infrastructure design for flood management, because we are faced with the task of estimating the infrastructure's probability of failure p_T for the next M years, as well as the uncertainty and bias in this estimate. However, to properly analyse it we need detailed information on complex cost and loss structures; we thus focus on the illustrative example of financial instruemnts for clarity.

In the following sections we illustrate some very basic considerations of uncertainty and risk analysis associated with the sequential decisions that may need to be made for climate risk adaptation in a nonstationary environment. This is a first conceptual step to the development of a full sequential decision analysis approach under nonstationarity. We proceed as follows . . ..

2 Methods

We use a synthetic data approach to generate sequences of annual-maximum floods using a known generating mechanism, then fit these sequences in a Bayesian framework to commonly used nonstationary flood models to evaluate their performance under different scenarios and criteria.

2.1 Synthetic Flood Sequence Generation

Extensive discussion in the literature has considered the appropriate choice of distribution for annual-maximum flood sequences. Although three-paremeter models are better able to represent extreme values than two parameter models (IACWD, 1982; Vogel and Wilson, 1996), we assume that flood sequences are *conditionally* lognormal, given a mean and standard deviation which vary in time:

$$\log Q(t) \big| \mu(t), \sigma(t) \sim \mathcal{N}(\mu(t), \sigma(t)). \tag{1}$$

where $\mathcal{N}$ denotes a normal distribution and $\mu(t), \sigma(t)$ are the location and scale parameters, respectively.

To model serial correlation and memory in the system, we use a two-state Markov chain model representing a "wet" and a "dry" state to describe the time evolution of the paramaters $\mu(t), \sigma(t)$. We also allow a trend term to model climate change or other secular change. For simplicity we assume a constant coefficient of variation α so that $\sigma(t) = \alpha\mu(t)$. In this Markov model, the distribution of μ (and correspondingly σ) at time t depends only on time and on the state of the system at time t, S_t

$$\mu(t) = \begin{cases} \mu_1 + \beta_1(t - t_0) & \text{if} \quad S_t = 1 \quad (\text{"wet" state}) \\ \mu_2 + \beta_2(t - t_0) & \text{if} \quad S_t = 2 \quad (\text{"dry" state}) \end{cases} \tag{2}$$

In order for state 1 to be the "wet" state, we enforce $\mu_1 > \mu_2$.

The transition from one state to the next is modeled by stochastic matrix $P_{2\times2}$. We consider here a symmetric system, in which π represents the probability of persisting in a particular state:

$$P = \begin{bmatrix} \pi & (1-\pi) \\ (1-\pi) & \pi \end{bmatrix}. \tag{3}$$

[2]need some citations for sequential decisions

We also consider the role of low-frequency variability by allowing the probability of persistence to vary sinusoidally in time. To constrain the probabilities to lie in $[0,1]$ we take the inverse logit transformation of the sinusoidal term:

$$\pi = \pi(t) = \text{logit}^{-1}\left[\pi_0 + A_\pi \sin\left(\frac{2\pi(t-t_0)}{\tau} + \phi\right)\right], \tag{4}$$

where τ represents the period of the low-variability oscillation, π_0 the expected value of π, A_π the amplitude of the sinusoidal oscillation, and ϕ the phase of the oscillation taken to be a random variable in $[0, 2\pi)$.

If we define $\mathbf{t} = [t_0 - N + 1, \ldots, t_0 + M]$, the general procedure for generating a flood sequence of $(N + M)$ years is:

1. Select parameters $\pi_0, A_\pi, \tau, \beta, \mu_1, \mu_2, \alpha_1, \alpha_2$
2. Randomly choose $\phi \in [0, 2\pi)$
3. Calculate the persistance probability $\pi(t)$ for $t \in \mathbf{t}$
4. Simulate a sequence of states $S(t)$
5. For each $t \in \mathbf{t}$, draw $Q(t)$ using eq. (1) using the values of μ, σ corresponding to the state simulated for that year.

2.2 Parameter Estimation

Because we do not, in general, know the true distribution of annual-maximum floods, we compare a stationary and non-stationary parameterizations of the generalized extreme value (GEV) statistical distribution (eq. 5) for estimating the distribution of $Q(t)$ for the future M-year period.

$$f(x|\mu,\sigma,\xi) = \frac{1}{\sigma}\begin{cases}\left[1+\xi\frac{x-\mu}{\sigma}\right]^{(-1/\xi)-1}\exp\left[-\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-1/\xi}\right], & \xi \neq 0\\ \exp\left[-\frac{x-\mu}{\sigma}\right]\exp\left[-\exp\left(-\frac{x-\mu}{\sigma}\right)\right], & \xi = 0\end{cases} \tag{5}$$

Where μ is the location parameter, σ is the shape parameter, and ξ is the scale parameter of the GEV distribution. For each simulation, the first N years ($t = [t_0 - N + 1, \ldots, t_0]$) are treated as observations. Then, once the appropriate time parameterization is selected (table 1), the parameters are fit in a Bayesian framework to fully represent the posterior uncertainty. Computation is carried out in the `python` environment, with simulation using the `numpy` package (van der Walt et al., 2011) and fitting using the No-U-Turn Sampler (NUTS) (Hoffman and Gelman, 2014) as implemented in the `stan` probabilistic programming language (Carpenter et al., 2016). Weak (uninformative) priors are chosen based on recommendations from Martins and Stedinger (2000); the exact parameterizations are described in following sections which describe specific experiments.

We also use a Hidden Markov Model (HMM) to fit to the data.[3]

2.3 Flood Estimation and Uncertainty

Finally, we consider a contract in which an insurer agrees to pay the entire coverage cost C every time that the trigger threshold Q_T is exceeded. Although Q_T is in general a decision variable, we treat

[3] Need to implement! Check `https://github.com/jmschrei/pomegranate/blob/master/tutorials/Tutorial_3_Hidden_Markov_Models.ipynb` for a nice Bayesian implementation

it as fixed. Then, using the posterior distribution of the parameters of the chosen statistical model (from table 1) we estimate the probability that $Q(t) \geq Q_T$ for $t \in \mathbf{t}$, which is equivalent to $p_T(t)$.

For a particular year, the fair price for the insurance premium is $\mathbb{E}(p_T)C$. However, there is typically also a risk premium associated with the policy that prices the uncertainty in this estimate (since the trigger Q_T is specified in the contract and thus not uncertain). This risk premium is proportional to the variance of p_T, or more generally to its uncertainty distribution. The insurance premium for a particular year is thus

$$Z(t) = C[\mathbb{E}(p_T(t)) + R\mathbb{V}(p_T(t))] \tag{6}$$

where $R \geq 0$ is a coefficient describing the cost of tue uncertainty in the estimate of $p_T(t)$.

In this paper we consider the price of a contract of M years to be simply the summed price of the contract for each individual year. We assume that the discount rate is equal to inflation (i.e. the inflation-adjusted discount rate is zero) for simplicity. A more comprehensive pricing mechanism for the financial instrument could incorporate more or different information about the uncertainty in the estimate, and should consider the price of the M years as a portfolio (i.e. should consider correlation between the different years). Such an approach, however, is not necessary for this conceptual study.

Model Name	Description
stationary	$Q \sim \text{GEV}\,(\mu, \sigma, \xi)$
location trend	$Q \sim \text{GEV}\,\big(\mu_0 + \beta_\mu(t - t_0), \sigma, \xi\big)$
full trend	$Q \sim \text{GEV}\,\big(\mu_0 + \beta_\mu(t - t_0), \sigma_0 + \beta_\sigma(t - t_0), \xi_0 + \beta_\xi(t - t_0)\big)$
HMM	log-normal Hidden Markov Model with # states chosen by BIC

Table 1: Summary of models used for fitting

3 Stationary, Static Risk

Consider static risk and a stationary process, and partiuclarly implications of M and N. Much of this is already laid out in Lall (1987).

4 Stationary, Static Risk

Consider static risk and a non-stationary process. Estimate using drift or nonstationary models using full N of the most recent N (i.e. what Vogel is proposing). Illustrate via simulation for different signal to noise ratio of nonstationary terms and M and N.

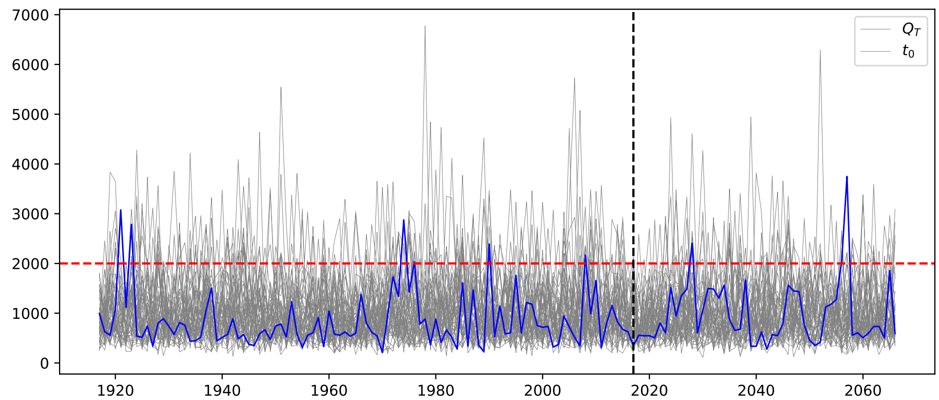

Figure 1: 100 synthetic stationary streamflow sequences with $N = 100, M = 50$. The blue line highlights a single randomly chosen sequence for clarity. The sequences were generated with parameters `list here`.

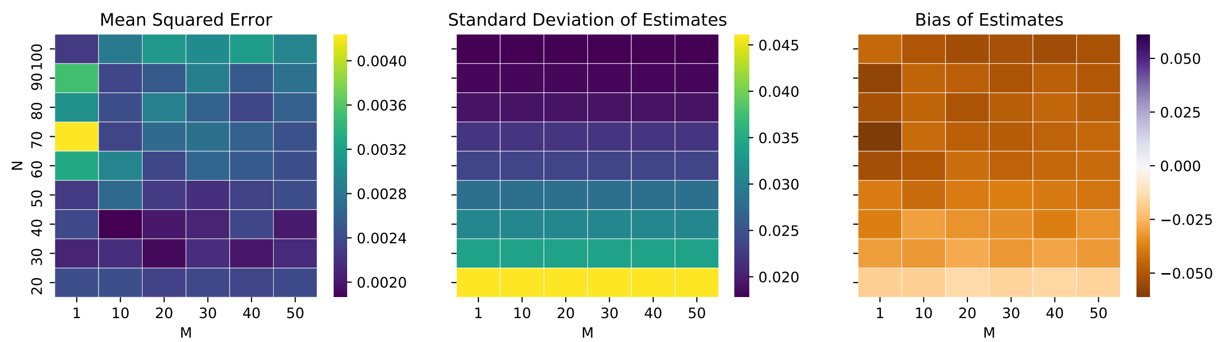

Figure 2: MSE, Bias, and Variance of results when stationary data (fig. 1) fit with statonary GEV model as a function of M, N.

5 Dynamic Risk

Consider dynamic risk, i.e. updating and sequential decisions, a nonstationary generative process, and M and N considerations

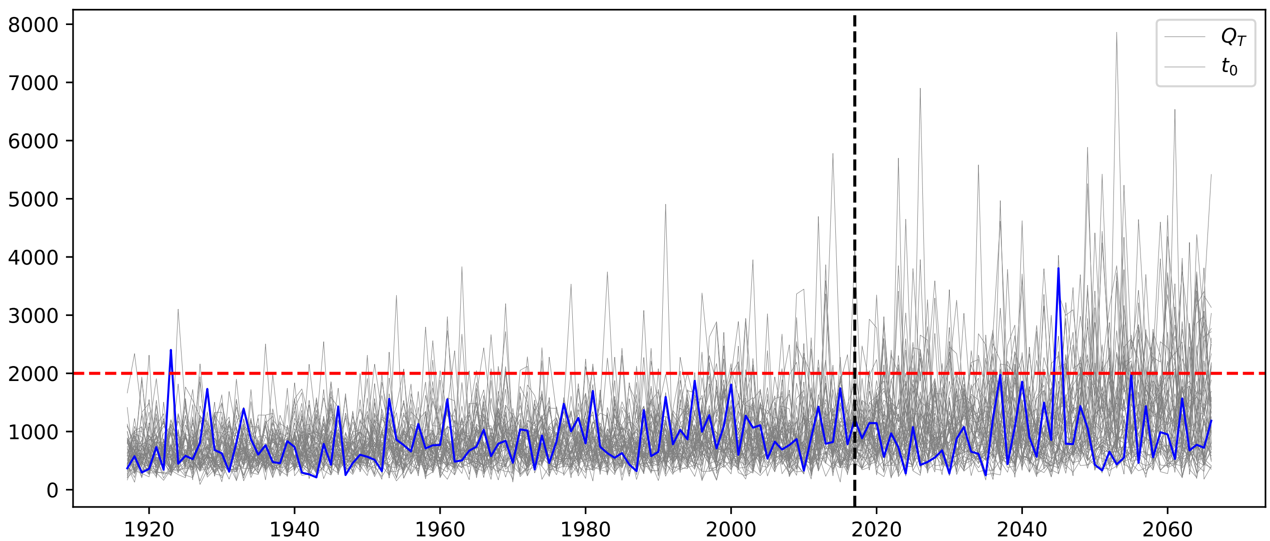

Figure 3: 100 synthetic streamflow sequences with $N = 100, M = 50$. Trend included only on "wet" state.

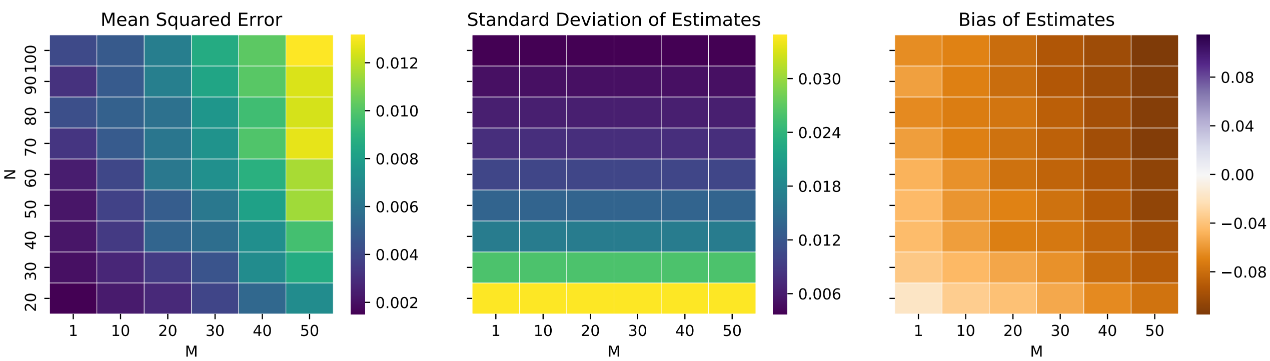

Figure 4: MSE, Bias, and Variance of results when trend data (fig. 3) fit with statonary GEV model as a function of M, N.

6 Discussion

Hopefully something to discuss?

7 Summary

Some interesting conclusions

References

Byrne, Michael P. and Paul A. O'Gorman (2015). "The Response of Precipitation Minus Evapotranspiration to Climate Warming: Why the "Wet-Get-Wetter, Dry-Get-Drier" Scaling Does Not Hold over Land". *Journal of Climate* 28.20. DOI: `10.1175/JCLI-D-15-0369.1`.

Carpenter, Bob et al. (2016). "Stan: A Probabilistic Programming Language". *Journal Of Statistical Software*.

Dankers, Rutger and Luc Feyen (2009). "Flood Hazard in Europe in an Ensemble of Regional Climate Scenarios". *Journal of Geophysical Research* 114.D16.

Delgado, J. M., B. Merz, and H. Apel (2014). "Projecting Flood Hazard under Climate Change: An Alternative Approach to Model Chains". English. *Natural Hazards and Earth System Science* 14.6. DOI: `10.5194/nhess-14-1579-2014`.

Dittes, B. et al. (2017). "Climate Uncertainty in Flood Protection Planning". *Hydrol. Earth Syst. Sci. Discuss.* 2017. DOI: `10.5194/hess-2017-576`.

Gelman, Andrew and Eric Loken (2013). "The Garden of Forking Paths: Why Multiple Comparisons Can Be a Problem, Even When There Is No "Fishing Expedition" or "p-Hacking" and the Research Hypothesis Was Posited Ahead of Time". *Department of Statistics, Columbia University*.

Hall, J. et al. (2014). "Understanding Flood Regime Changes in Europe: A State-of-the-Art Assessment". *Hydrology and Earth System Sciences* 18.7. DOI: `10.5194/hess-18-2735-2014`.

Hodgkins, Glenn A et al. (2017). "Climate-Driven Variability in the Occurrence of Major Floods across North America and Europe". *Journal of Hydrology*.

Hoffman, M D and A Gelman (2014). "The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo." *Journal of Machine Learning Research*.

Howard, Ronald A. (1960). *Dynamic Programming and Markov Processes.* Dynamic programming and Markov processes. Oxford, England: John Wiley.

Hurst, H E (1951). "Long-Term Storage Capacity of Reservoirs". *Transactions of the American Society of Civil Engineers* 116.

IACWD (1982). *Guidelines for Determining Flood Flow Frequency*. Tech. rep. Bulletin 17B. Reston, VA: Interagency Committee on Water Data, Office of Water Data Coordination, USGS.

IPCC (2012). *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*.

Jain, Shaleen and Upmanu Lall (2001). "Floods in a Changing Climate: Does the Past Represent the Future?" *Water Resources Research* 37.12. DOI: `10.1029/2001WR000495`.

Kidson, R and K S Richards (2016). "Flood Frequency Analysis: Assumptions and Alternatives". *Progress in Physical Geography* 29.3.

Lall, Upmanu (1987). "Project Risk Considering Samplng Uncertainties and a Finite Project Operation Period". en. *Application of Frequency and Risk in Water Resources*. Ed. by Vijay P. Singh. Springer Netherlands. DOI: `10.1007/978-94-009-3955-4_23`.

Lickley, Megan J, Ning Lin, and Henry D Jacoby (2014). "Analysis of Coastal Protection under Rising Flood Risk". *Climate Risk Management* 6.

Martins, Eduardo S. and Jery R. Stedinger (2000). "Generalized Maximum-Likelihood Generalized Extreme-Value Quantile Estimators for Hydrologic Data". en. *Water Resources Research* 36.3. DOI: `10.1029/1999WR900330`.

Matalas, Nicholas C (2012). "Comment on the Announced Death of Stationarity". *Journal of Water Resources Planning and Management*.

Merz, B. et al. (2014). "Floods and Climate: Emerging Perspectives for Flood Risk Assessment and Management". *Natural Hazards and Earth System Science* 14.7. DOI: `10.5194/nhess-14-1921-2014`.

Merz, Ralf and Günter Blöschl (2008). "Flood Frequency Hydrology: 2. Combining Data Evidence". *Water Resources Research* 44.8.

Milly, P C D et al. (2008). “Stationarity Is Dead: Whither Water Management?” *Science* 319.5863. DOI: 10.1126/science.1151915.

Muller, Caroline J, Paul A O’Gorman, and Larissa E Back (2011). “Intensification of Precipitation Extremes with Warming in a Cloud-Resolving Model”. *Journal of Climate* 24.11. DOI: 10.1175/2011JCLI3876.1.

Obeysekera, Jayantha and Jose D. Salas (2014). “Quantifying the Uncertainty of Design Floods under Nonstationary Conditions”. *Journal of Hydrologic Engineering* 19.7. DOI: 10.1061/(ASCE)HE.1943-5584.0000931.

O’Gorman, Paul A (2015). “Precipitation Extremes Under Climate Change.” *Current climate change reports* 1.2. DOI: 10.1007/s40641-015-0009-3.

Ott, Irena et al. (2013). “High-Resolution Climate Change Impact Analysis on Medium-Sized River Catchments in Germany: An Ensemble Assessment”. *Journal of Hydrometeorology* 14.4.

Rosner, Ana, Richard M. Vogel, and Paul H. Kirshen (2014). “A Risk-Based Approach to Flood Management Decisions in a Nonstationary World”. en. *Water Resources Research* 50.3. DOI: 10.1002/2013WR014561.

Russell, Stuart J. (2003). *Artificial Intelligence : A Modern Approach*. Ed. by Peter Norvig. 2nd ed. Upper Saddle River, N.J. : Prentice Hall,

Serinaldi, Francesco and Chris G. Kilsby (2015). “Stationarity Is Undead: Uncertainty Dominates the Distribution of Extremes”. *Advances in Water Resources* 77. DOI: 10.1016/j.advwatres.2014.12.013.

Shaw, T. A. et al. (2016). “Storm Track Processes and the Opposing Influences of Climate Change”. en. *Nature Geoscience* 9.9. DOI: 10.1038/ngeo2783.

Silva, Artur Tiago, Mauro Naghettini, and Maria Manuela Portela (2016). “On Some Aspects of Peaks-over-Threshold Modeling of Floods under Nonstationarity Using Climate Covariates”. *Stochastic Environmental Research and Risk Assessment* 30.1.

Stedinger, Jery R. (1997). “Expected Probability and Annual Damage Estimators”. *Journal of Water Resources Planning and Management* 123.2. DOI: 10.1061/(ASCE)0733-9496(1997)123:2(125).

Strupczewski, W G, V P Singh, and W Feluch (2001). “Non-Stationary Approach to at-Site Flood Frequency Modelling I. Maximum Likelihood Estimation”. *Journal of Hydrology* 248.1-4.

Sveinsson, O G, Jose D Salas, and D C Boes (2005). “Prediction of Extreme Events in Hydrologic Processes That Exhibit Abrupt Shifting Patterns — Journal of Hydrologic Engineering — Vol 10, No 4”. *Journal of Hydrologic Engineering*.

van der Walt, Stéfan, S Chris Colbert, and Gaël Varoquaux (2011). “The NumPy Array: A Structure for Efficient Numerical Computation”. *Computing in Science & Engineering* 13.2.

Vogel, Richard M and Ian Wilson (1996). “Probability Distribution of Annual Maximum, Mean, and Minimum Streamflows in the United States”. *Journal of Hydrologic Engineering* 1.2.

Vogel, Richard M, Chad Yaindl, and Meghan Walter (2011). “Nonstationarity: Flood Magnification and Recurrence Reduction Factors in the United States”. *Journal of the American Water Resources Association* 47.3.