**Title: - Diagnosis, simulation and prediction of inter-annual and longer variations of multi-site, annual maximum streamflow in the Ohio River Basin.**

Keywords: - Multi-site Streamflow Simulations, Climate Extremes, Risk Analysis, NAO, Ohio River Basin.

*Abstract*:

Many authors have presented evidence for the modulation of rainfall induced floods by ENSO and other climate modes. How such information should be used for flood risk preparation and mitigation is an open question. This paper presents a systematic approach to diagnose, simulate and predict inter-annual variations in regional flood incidence as a step towards an approach towards adaptation to hydroclimatic variability. Long records of annual maximum streamflow from 30 sites in the Ohio River Basin are analyzed using wavelets, cluster analysis and principal component analysis to identify a dimension reduction that best informs regional spatio-temporal patterns, and their associated hemispheric climate modes. The characteristic time scale of the patterns identified vary from 5-7-year peaks to long term secular trends. Stochastic models for multi-site, multi-year simulation and prediction of annual maximum streamflow at each site are then developed and tested. Model simulations reproduce observed inter-annual and longer frequency domain structure and could be used for new dynamic flood risk mitigation strategies.

**Introduction: -**

Globally floods caused half a million deaths and affected nearly 3 billion people from 1980 to 2009(Melillo, Richmond, & Yohe, 2014). Flood losses are expected to worsen with increasing coastal migration, land-use changes and development in the floodplains(Knowles & Kunreuther, 2014; Thomas & Leichenko, 2011).

On a global scale, El Nino-Southern Oscillation is the most prominent and examined case of interannual climate variability affecting hydrological processes globally by means of teleconnections. (Philip J. Ward et al., 2014) show the existence of global flood risk (computed in terms of soci-economic factors) anomalies during El Nino and La Nina years, which affect nearly half of the global land surface. Further (Philip J. Ward, Beets, Bouwer, Aerts, & Renssen, 2010) show that in the extra-tropics, ENSO has a greater impact on the extremes flows than the mean annual discharge, with the relationship between between floods and ENSO being non-stationary in nature with present of trends (P. J. Ward, Eisner, Flörke, Dettinger, & Kummu, 2014). (Lee, Ward, & Block, 2018) show global-scale long-lead predictions of average streamflow in the peak season using large scale climate patterns of ENSO, PDO, NAO and AMO. ENSO also has been shown to modulate the extreme hydrological events including frequency of extreme flooding events (Jain & Lall, 2000, 2001), with changes in flood occurrence probabilities conditional on ENSO events globally(Khalil, Kwon, Lall, Miranda, & Skees, 2007).

Regional manifestations of similar phenomena like the Pacific Decadal Oscillation (PDO) and the North Atlantic Oscillation (NAO)-associated large scale atmospheric anomalies, exhibiting low frequency modes, which may result in variability in flood risk on a decadal scale (Olsen 1999). Larger basin wide flooding events are shown to be related to large scale atmospheric anomalies (Nakamura 2013, Hirschboeck 1988)

Most past work on flood risk estimation considers regional or at site analyses of the probabilities of extreme events, with limited attention to the spatio-temporal and climate aspect of risks (Kwon et al 2008). The assumption of stationarity is questioned on the grounds of anthropogenic climate change(Milly et al., 2008) and also due to the temporal structure in hydroclimatic variables (Jain & Lall, 2001). This led to the incorporation of considerations of non-stationarity in statistical flood risk models using covariates ((Filho & Lall, 2003; Grantz, Rajagopalan, Clark, & Zagona, 2005; Regonda, Rajagopalan, Clark, & Zagona, 2006; Towler et al., 2010) refs).

Grantz et al developed a technique incorporating climate predictors which modulate seasonal streamflow and further using them as predictors of seasonal streamflow forecasts by means of local regression(Grantz et al., 2005). (Regonda et al., 2006) developed multi-site total spring streamflow ensemble predictions using principal component analysis and locally weighted regression along with large-scale climate features and snow water equivalent as additional covariate. (Filho & Lall, 2003) used ENSO and equatorial Atlantic SST as climate precursors in a semiparametric approach for multi-site streamflow ensemble forecasts in Cearra, Brazil. (Towler et al., 2010) use extreme value statistical theory based generalized extreme value (GEV) distribution to model single site maximum winter monthly streamflow, with the non-stationarity explicitly accounted by means of climate covariates.

However, the temporal clustering of flood risk highlighted in (Jain & Lall, 2001)may or may not be readily addressed by such methods, who showed that the clusters associated with low frequency climate modes can lead to significant under/over design of flood control instruments (structural or financial). Further, given a positive discount rate, near term flood risk mitigation investments contribute more strongly to a benefit cost analysis than those in the distant future. Consequently, even a stochastic prediction of the risk of regional floods over the next several years, may have high economic value relative to the highly uncertain climate change induced flood risk predictions that are 30 to 50 years into the future (Doss-Gollin eta. al., 2018). A proper representation of this dynamic risk would cover the non-stationary spatio-temporal structure across the entire basin and not just at individual sites, to address the potential cascading failure, or compound risk induced by the underlying climate pattern (Cheng, AghaKouchak, Gilleland, & Katz, 2014).

The primary research questions of the work presented were:

1. How can regional space-time-frequency relationships in flood extremes be systematically identified from multi-site data, for the Ohio River Basin? Does spatial dimension reduction followed by temporal/frequency pattern identification, or frequency identification followed by spatial dimension reduction lead to complementary results, or is one approach more effective? How do the identified regional space-time modes for annual maximum floods relate to known, hemispheric climate modes?
2. What is the appropriate form of a statistical model that can simulate the multi-site time series to reproduce the space-time statistics identified on both the basin scale and at stream-gauge level?
3. Can one probabilistically forecast the regional and at site annual maximum floods over the next few years?

Section 2 describes the data used. Section 3 summarizes the analysis strategy the methods used for diagnosis, simulation and prediction. The results are presented in Section 4. A discussion section concludes the paper.

**Data:**

The Ohio River Basin has a long history of large-scale basin wide floods, with notable floods in 1937, 1884, 1913, 1845, 1883, 1964, 1907, 1948, 1997, 1933, and 1913. Early Spring, late winter floods in this region tend to be more wide-spread while summer floods are characterized by locally intensive thunderstorms (*National Water Summary 1988-89*, 1991). Unlike other common flooding mechanisms which depend on the land use characteristics, urban landscape and soil, the flooding in this region is dominated by precipitation (persistent or rapid) and/or snow melt events(Farnham, Doss‐Gollin, & Lall, 2018; Nakamura, Lall, Kushnir, Robertson, & Seager, 2013).

Streamflow data were downloaded from USGS and EPA stream gauges using the ‘dataRetrieval’ package (Hirsch & De Cicco, 2015). Stream gauges with no missing daily data from 01/01/1937 to 12/31/2017 and a drainage area greater than 3750 sq. miles were selected. Using these criteria, the final number of sites included were 30 with each location having 81 years of data from 1937-2017. At each site the daily streamflow time series was used to identify the annual maximum for the year as a measure of extreme streamflow.

The Nino 3.4 SST index, used as an indicator of the ENSO phenomena, is computed from HadISST1(Rayner et al., 2003). The monthly index is available at the NOAA Physical Sciences Division website (<https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/Nino34/>). Monthly anomalies of PDO, computed as the first principal component of the Northern Pacific SST(Mantua, Hare, Zhang, Wallace, & Francis, 1997; Zhang, Wallace, & Battisti, 1997), are taken from NOAA ERSSTv5(Information (NCEI), n.d.). Unlike ENSO and PSO, NAO anomaly data was computed as the normalized pressure difference between Gibraltar and Iceland(Jones, Jonsson, & Wheeler, 1997) and is available from the Climate Research Unit (<https://crudata.uea.ac.uk/cru/data/nao/>).The KNMI Climate Explorer (<https://climexp.knmi.nl/start.cgi>) provides a good user interface to extract monthly climate indices from their original sources and was used as the data source, with the original sources listed above.

The indices were converted from monthly to annual scale by computing the mean Feb-March-April values, late winter-early spring period in the ORB, with maximum occurrences(~50%) of extreme streamflow events. The interaction between two indices and was computed as shown in (1).

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

Methods

We consider a dual strategy for the diagnosis and subsequent modeling of the multi-site low frequency variations in annual maximum streamflow across the Ohio River Basin. From a regional perspective, one can anticipate, that the time series at all sites may be modulated at the same frequencies if they are influenced by larger scale climate oscillations that have marked quasi-periodic variability. Such a common large scale driver may also induce a spatial correlation structure in the annual maximum flow series across the sites, even if the annual maximum flows do not occur simultaneously across all the sites in each year. Rather, especially in this region, we note that all the larger annual maximum floods occur in the same season, and often correspond to recurrent synoptic waves of incoming atmospheric moisture every 4-7 days (A. Robertson, Jennifer Nakamura, Kushnir, & Lall, 2015; Nakamura et al., 2013). Thus, some sites may experience the annual maximum event earlier than others, while all sites have elevated flows in such seasons. Consequently, the two complementary approaches we explored were (1) PC-Wavelet: a Principal Component Analysis (PCA) on the spatial correlation matrix of the annual maximum flows for the 30 sites for dimension reduction, followed by a wavelet analysis of the leading Principal Components (PCs) to assess quasi-periodic variability and secular trends; and (2) WaveClust: wavelet analysis of each of the 30 series followed by a hierarchical cluster analysis to select groups of sites with common time-frequency signatures. Often, the first PC represents the spatial mean across the sites – the corresponding eigenvector has roughly similar coefficients. The subsequent PCs identify orthogonal spatial patterns that may have different temporal structure. By contrast, the cluster analysis of the time-frequency structure will directly identify groups of sites that participate in similar climate patterns, but are not necessarily orthogonal or statistically independent. From a diagnostic perspective, the resulting space-time patterns from either method may or may not be similar, and may relate to the larger scale climate modes in different ways. The details of these two approaches follow. An illustration of how the diagnostic analyses can be built on for the simulation and prediction of regional flood variability at inter-annual time scales is provided in Figure 1 below.

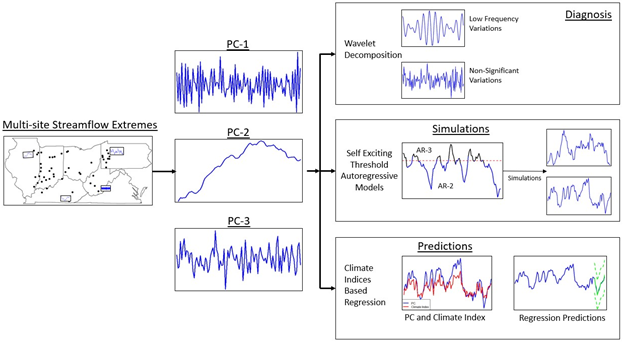


Figure XX: Overall schematic framework of analysis.

Diagnosis of the low frequency hydroclimatic modes

The first step in the analysis of streamflow extremes is checking for presence of low frequency variabilities and inter-annual modes of extremes. Analysis of these modes should include both the basin wide spatial structure of the extremes along with their time-frequency structure.

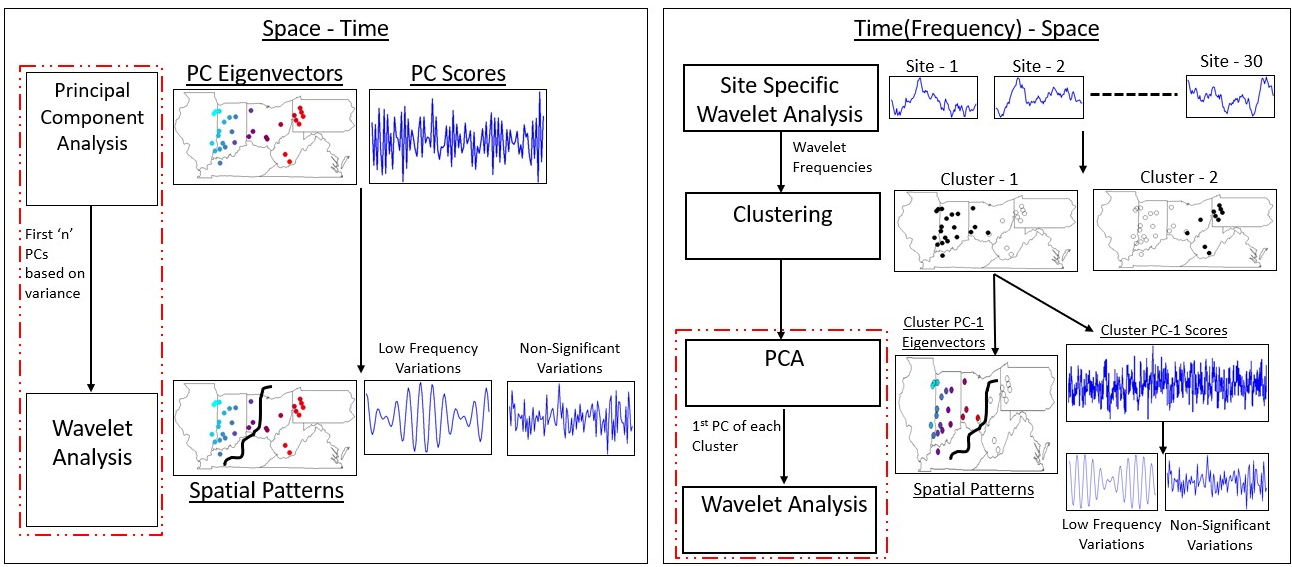


Figure XX: Schematic of the methods (a)PC-Wavelet (b) Wave-Clust for the diagnosis of the inter-annual and low frequency modes of extreme climate variability.

Figure XX above shows the combined use of Principal Component Analysis and Wavelet Analysis to analyze the underlying basin wide spatio-temporal structure of low frequency modes of extremes. The methods are broadly classified by their chronology with which they analyze the spatio-temporal structure of the modes of extremes.

The first method analyses the spatial dimension followed by the time-frequency aspects across the entire basin. This method, PC-Wavelet, utilizes PCA to obtain the leading modes across the region. The PC eigenvectors give the spatial structure of prominent modes of extremes whereas PC scores give the overall temporal structure of that mode, which are then subjected to wavelet analysis to separate the low frequency variation from the non-significant variations and the noise. Here number of leading PCs to be analyzed is decided by the variance explained each of the modes. Figure XX (a) shows this with the maps denoting the spatial features of a hypothetical PC with a clear east-west divide and the PC score, after wavelet analysis, revealing a prominent low-frequency variation.

The second method analyzes the time-frequency domain first followed by the spatial structure. This method, called Wave-Clust, uses the wavelet frequencies (Refer Annex \_\_ for detailed description of the wavelet frequencies and the Wavelet Analysis) to cluster the sites, with separate application of the wavelet analysis at each site. The clustering is computed using hierarchical clustering and the number of clusters defined by its dendrogram. The computed clusters then undergo PCA, where unlike the PC-Wavelet, only the leading mode of variability is extracted for further wavelet analysis of its rotated scores.

Simulation Methods: -

Threshold Auto-Regressive Model introduced by Tong and discussed in detail in (Tong, 1990), a popular class of nonlinear time series modelling are simple to understand yet have been shown to be rich enough to generate complex nonlinear dynamics like limit cycles. TAR models are non-linear regime switching models, linear in each regime and able to switch between regimes based on a threshold model. Self-Exciting Threshold Auto-Regressive (SETAR) models are subset of the more general TAR models, where the threshold variable is a lagged value of the time series itself(Tong, 1990).

Consider a time series of length n, shown as , a generalized two regime SETAR model is denoted by SETAR(2;p,q) is mathematically denoted as:

|  |  |  |
| --- | --- | --- |
|  |  | (2) |

where the are the coefficients of the AR processes with denoting the lag and the regime. is the threshold value which serves as the transition parameter and is the delay parameter denoting the lagged value which is the transition variable. are the residuals for time and regime . The framework in (2) can be easily extended to three regime SETAR model. The best model fit was computed using the ‘tsDyn’ package (Stigler, 2010) in R.

SETAR have been used to model hydrological time series in particular streamflow data. (Komorník, Komorníková, Mesiar, Szökeová, & Szolgay, 2006) used nonlinear regimes including SETAR and proposed model modified TAR to forecast and model the monthly flows of rivers in the Tatry Alpine region in Slovakia. The regime structure in this region arise from the differences in mechanism driving flow – snowmelt in spring and convective precipitation in summer. (Amiri, 2015) used five classes of non-linear time series models to capture the Colorado river discharge dynamics and showed that SETAR models performed better than the other four models.(Tongal & Berndtsson, 2017) used a SETAR, k-NN and ANN to model daily streamflow of three rivers in Sweden and found that SETAR had the best 7-day ahead forecasting. (Tongal & Booij, 2016) used SETAR models to capture the non-linear dynamics of nine rivers in the western U.S. uses a hybrid SETAR-GARCH approach to model streamflow with applications shown for rivers in Iran. To the author’s best knowledge, this is the only study which uses SETAR to model multi-site PC transformed annual maximum streamflow data, with a distinct regime structure unlike daily and monthly flows.

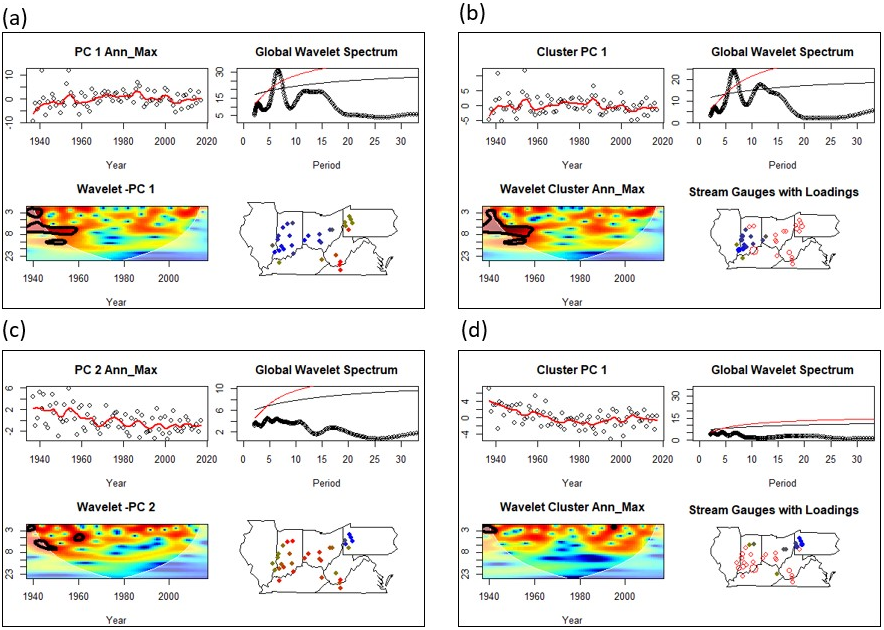
Along with the PC-SETAR approach, a simpler PC-AR and PC-WARM(Kwon, Lall, & Khalil, 2007) were used to capture the dynamics but the SETAR models performed the best. Refer Annex **XX** for additional details regarding SETAR models and the results from the PC-AR and PC-WARM.

**Results & Discussion**: **-**

Diagnosis

The diagnosis portion of the analysis is mainly concerned with the understanding of modes of streamflow extremes through their spatio-temporal structure. The results from the PC-Wavelet and Wave-Clust methods are shown in Figure XX.

Explain the Figure’s Results – The 1st and 2nd principal components of extreme streamflow explain about 40% and 14% of the total variance and were used for the subsequent analysis shown in Figure 2. (Annex \_\_ includes the same plot for PC3~9% variance)



*Figure 2: - PCWavelet Analysis (a, c) and WaveClust Analysis (b, d) for the 1st PC(a, b) and 2nd PC(c,d). For each subplot: - Top-left – The PC with the loess line in red. Top-right – Global Wavelet Spectrum. The red and black lines correspond to the 90% significance red and white noise respectively. Bottom-Left – Wavelet Coherence for each PC. Bottom-Right – The PCs loadings (absolute values from blue to red) for each site.*

As seen from Figure 2, the output from PC-Wavelet and Wave-Clust for PC 1-2 and correspondingly cluster 1-2 are nearly identical. Focusing first on the temporal dimension, the wavelet spectrum for PC-1 of the entire field i.e. PC-Wavelet(2a) shows a sharp peak around 7 years. This is nearly replicated in the 1st PC of Cluster-1 as computed by WaveClust(2b), which also shows a periodicity around the 12-yr cycle. Further the 7-yr. peak in spectrum in both cases crosses the 90% confidence interval over red and white noise respectively, with significance computed against mean background spectrums(Torrence & Compo, 1998). Unlike the global wavelet spectrum(top-right), the power spectrum(bottom-left) gives a time-frequency dimension. Here again, both show similar regions in time (up to 1960s) where the near 7 yr. periodicity is dominant, with the strength of the lower frequencies carrying forward until the early 1970’s.

The PC-2(2c) and Cluster-2 PC-1(2d) both show strong trends(top-left). 2-c is monotonically decreasing throughout as seen by the loess fit and is contrasted against 2-d which decreases up to the 80’s and then stabilizes. The global wavelet spectrum and power spectrum do not help uncover the time-frequency structure due to the underlying trend.

For both methods, the colors scheme for the gauges (bottom-right) corresponds to the absolute value of the eigenvectors with bluer shades denoting higher values and color fill denoting cluster membership for WaveClust. Focusing on the spatial field, PC-1 and Cluster-1 PC-1 have the same pattern with the region of Indiana, eastern Illinois and western Ohio being the dominant mode across the Ohio River Basin. This region corresponds to the mainly Wabash river basin and some parts in the lower Ohio river basin(Refer Annex \_\_ to map of the river basin breakup). The PC-2 and Cluster-2 PC-1 is physically located in western Pennsylvania and eastern Ohio. These are the sites in the Allegheny, Monongahela, Muskingum and upper Ohio river basins.

The similarities between these approaches break down on reaching PC-3 and Cluster-3 PC-1(plot shown in the annex). This is not surprising as the variance explained by PC-3 is less compared to PC 1-2 and cluster-3 does not have a lot of members for both to have strong signals over noise.PC-3 and cluster-3 PC-1(not shown) are situated in the central region mainly Green, Kentucky and middle Ohio river basins. (None of the sites in this dataset lie in the Cumberland and Tennessee river basins).

The leading mode in both methods shows dependence in the western portion of the Ohio River Basin with a key 7-year periodicity. The second leading mode shows dependence on the eastern portion of the basin with no characteristic low frequency variations but presence of a trend. Another way of looking at this is that the leading mode captures the dynamics of the western part (Wabash) along with the aggregated part(lower Ohio) basin. In part the leading mode contains filter information about entire basin preceding it. The second mode is separated from the leading mode by Ohio and captures the dynamics driving the upscale and smaller basin, whose aggregated information is captured in the leading mode.

Considering the spatial organization of the modes and their distinct temporal structure leads to the conclusion of presence of large-scale temporal structure in the Ohio River basin especially in the aggregated flow. It also shows that large regions have similar spectral signatures possibly pointing to regions having similar dynamics or forcing by global external climatic variability. The ability to make predictions on these modes, which explain more than half the region’s variability would help characterize the regional risk and provide an easy mechanism to transfer the estimated regional risk to individual sites by means of eigenvectors. Simulation models for these modes, which can include it’s non-stationary regime structure would help in quantification of risk regionally and at specific sites.

modes.

Regional Modes and Climate Modes.

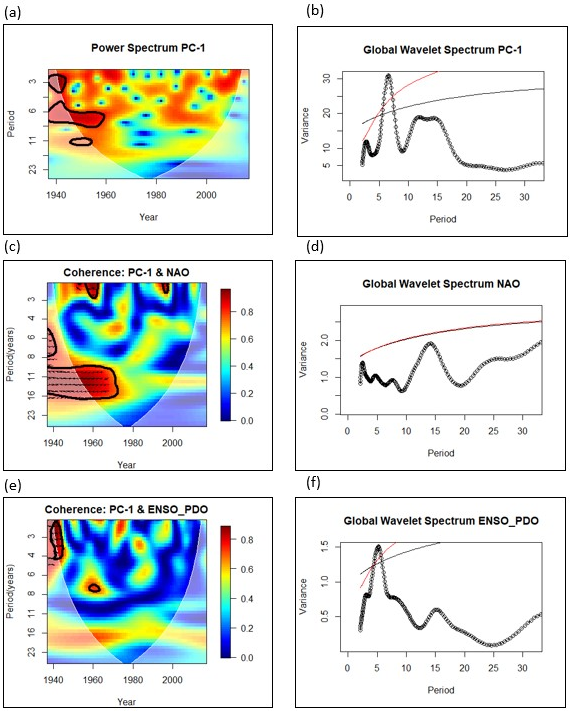


Figure XX:- (a) Wavelet Power Spectrum for PC-1. (b) Wavelet Coherence Power Spectrum between PC-1 and NAO. (c) Global Wavelet Spectrum for PC-1. (d) Global Wavelet Spectrum for NAO.

PC-1, with a total explained variance of 40% is the dominant mode of regional variability in the region and shows a significant negative correlation (-0.31) with NAO and positive with ENSO(0.23) and ENSO-PDO(0.27). (Refer Annex for the complete correlation matrix). Apart from the strong peak in strength of the near 7-year cycle, PC-1 also has a weaker lower frequency 12-16-year cycle, which is not statistically significant, as seen through its power and global wavelet spectrum. This low frequency cycle, like the 7-year cycle is also most active with power up to the 1970s. The wavelet coherence plot between PC-1 and NAO shows that the 11-16-year period is in phase with NAO and is also statistically significant during the same time period. The lower cycle of 11-16 years and the correlation tie the modulating effect of NAO to PC-1 in the mid-1950s.

Thus, the regional extreme streamflow and flood potential as captured by PC-1 is likely modulated to a great extent by NAO, with the teleconnections being particularly strong in the earlier duration of the data.

Simulation

Observations from Diagnosis: - The section above shows presence of non-stationarity in the streamflow extremes along with spatio-temporal structure of the low frequency modes of streamflow extremes. While the AR uses a common dynamic to model the entire process, SETAR assumes presence of threshold and modes, better mirroring the findings seen in the earlier section. The SETAR models are known to model and simulate much more complex dynamical processes including limit cycles (Tong, 1991). For the sake of iterative modelling the AR model was first fit to the PC, followed by SETAR.

PC-AR Brief Results - The AR models fit to the 1-3 PCs were of order 4,4 and 0 respectively, with AIC being the fitting criteria. Each AR Model was simulated 1000 times and the mean, standard deviation, minimum and maximum; spectral signature, probability distribution and cumulative density of the simulated series and the original PC were compared. The AR models were successful in simulating the PC mean, whereas it has some decent success in replicating the standard deviation, the minimum and maximum values. Overall, the PC-AR was not able to capture the variability in the PC space and wasn’t able to project that variability onto the multisite flow.

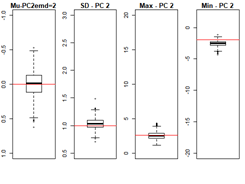
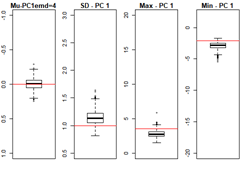
How were the models fit – A separate SETAR model was fit for each PC with AIC being the fitting criteria given an embedding dimension, like an AR model. The R package tsDyn was utilized for fitting SETAR models.

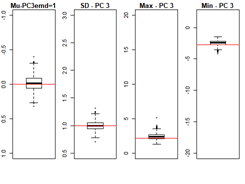
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Name | Embedding Dimension | Lower Regime Dynamics | Upper Regime Dynamics | Threshold Lag | Threshold Value |
| PC1 | 4 | 4 | 4 | 3 | 0.4894 |
| PC2 | 2 | 1 | 2 | 1 | -0.432 |
| PC3 | 1 | 1 | 1 | 1 | -0.245 |

Table 1: - SETAR Model Specifications for each PC.

Brief Results – Table 1 shows the SETAR model parameters for the three different PCs (Refer Annex \_\_ for complete results with lag coefficient values). The model fit for PC-1 was similar to AR process fit for PC-1 with both the upper and lower regime dynamics modelled by AR-4 processes and regime threshold on The similarities between SETAR and AR vanish thereafter with PC-2 being modelled by , but the individual regimes being modelled by distinct AR-Models.

The SETAR Models were simulated 1000 times, for a run-length of 81 which corresponds to the time series length. The simulations capture the mean accurately across all three PCs. Figure XX shows the mean, standard deviation, minimum and maximum values along with the simulated values. In the case of standard deviation there is no bias and the skill in simulation improves over AR (Refer Annex for AR results). Further SETAR is also able to simulate and capture the PC minimum and maximum values, with an ability to generate minimum and maximum values outside the time series values, a needed skill in planning in water management scenarios as unseen phenomena which is plausible under the model dynamics is accounted for in the planning.

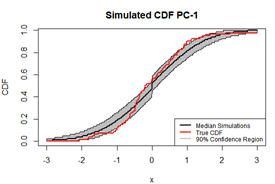
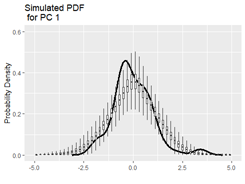


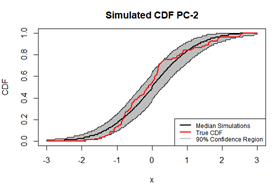
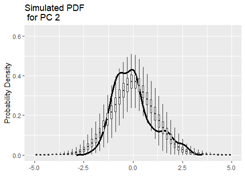


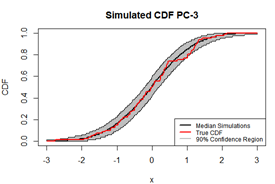
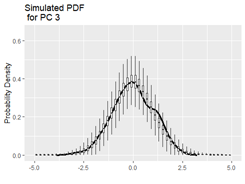
*Figure 4: - SETAR Simulations of the key characteristics of the PC – Mean, Standard Deviation, Minimum and Maximum values. The red line denotes the true PC values.*

Implications of getting the CDF, PDF and Spectrum correct – Further as measure of the skill of the simulation, the true PDF and CDF of the PC’s were compared with simulated PDFs and CDFs. Figure 5 shows the simulated CDF and PDF for the PCs, with simulations showing great skill in capturing the internal dynamics of the PCs.

The PDF simulations are gaussian with the true PCs exhibiting certain non-gaussian characteristics, leading to some differences with the true profile but comparisons with the CDF show that the SETAR model is able to capture the distribution along with the tails accurately. The CFD simulation skill, though much better than the skill seen in the PDF’s shows a similar nature with biases in the PC-1,2 and better skill with PC-3.

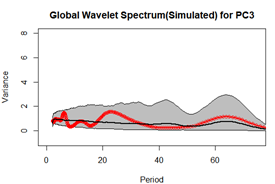
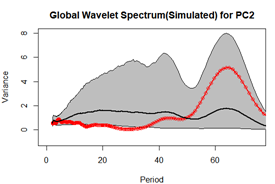
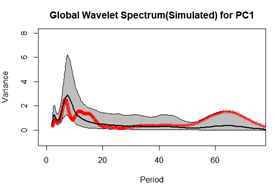






*Figure 5: - Probability Densities(right) and Cumulative Densities(left) for each of the PCs for the 1000 simulations. The black line is the density of the original PC. Density Estimation done by means of kernels.*

Spectral Signature – Using the wavelet analysis, the global wavelet spectrum for PCs and for each simulated time series are computed, giving a characterization of the time-frequency domain. Figure 6 shows SETAR models can simulate and capture the spectral signature of the individual three PCs. For the 1st PC the 5-7-year cycle is simulated prominently, both the simulated signal and the original signal crossing the 90% significance threshold over red and white noise. The SETAR model for the 2nd PC accounts for the long-term secular trend structure in the data.PC-3 which does not show any trend or low frequency variation has a similar nature of the simulated series, lending confidence that the selected SETAR model is flexible to model phenomena with and without low frequency variations. This combined with the replications of the PC moments – mean, standard deviation, minimum and maximum values; PDF and CDF show that SETAR models are able to incorporate the decadal and larger variations in streamflow variability.



*Figure 6: - Simulated Global Wavelet Spectrum for the 1st three PCs. The red line with dots is the actual data, the black thick line is the mean of the 1000 simulations and the thin black lines are the 95th and 5th percentiles of the simulations at each period.*

Transformation & Influence of Eigenvectors – A crucial step in assessing the method skill is transformation of the simulations from PC space to the original space comprising of the 30 sites, Using the 1000 simulations for the 3 PC’s, the simulations were retransformed using eigenvectors to get 1000 simulations at each of the 30 sites, providing a measure to check the simulation skill at each site in the river basin. (Refer Annex C for the transformed data simulations for each site for the moments and the CDFs and moments). These transformed simulations captured with fair success rate the characteristics as measured by mean, standard deviation, minimum and maximum values; PDF and CDF at each site.

The sites 4(rank 2),8(rank 1),9(rank 3) and 14(rank 4) which have the greatest deviation from their true CDF have the lowest absolute value eigenvectors for the 1st PC which contributes the largest variance(~40%). Figure XX shows the site-specific simulation skill depends critically on the PC eigenvectors with CDF simulations for two sites one with low absolute rank and one with high absolute rank, serving as a confirmation of the benefits of transferring to the PC-space and to the overall spatio-temporal structure of each PC.

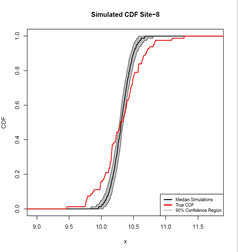
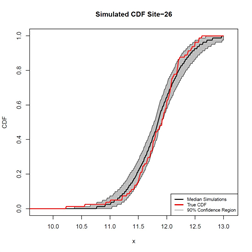


Figure 7: CDF Simulation plots for Site-26 and Site-8. Site 26 has a higher eigenvector associated with PC-1 whereas Site-8 has the lowest eigenvector associated with PC-1.

Conclusion – The PC-SETAR methodology has shown the ability to capture the main system dynamics and translate that uncertainty and structure to site specific extreme streamflow. Site-Specific transformations have been successful in replicating site-specific metrics, with skill depending on the site’s contribution to the PC as measured by the eigenvector. Overall the PC-SETAR methodology shows better skills in modelling and simulating the basin wide processes compared to PC-AR, a large part owing to the non-stationary nature.

**Conclusion**

Region of focus/data - The Ohio River Basin with its long history of regional scale precipitation and flooding events and extensive period of data availability presented an ideal region to study low frequency variability and modes of streamflow extremes. Using the annual streamflow maximum as criteria for extremes, 30 stream gauges with data from 1937 to 2018 were selected for this study.

Diagnosis The diagnosis of the low frequency modes of streamflow extremes along their spatio-temporal structure was carried out using the combined PC-Wavelet methods. The chronologically varied PC-Wavelet and Wave-Clust methods yielded similar results pointing to the presence of characteristic spatio-temporal structures and low-frequency modes of extreme streamflow variability. The most dominant mode had a 5-7 yr. recurrence period along with a distinct east-west spatial divide.

Simulations: The novel PC-SETAR methodology was successful in simulating the streamflow extreme dynamics at the basin level and site specificities, capturing the risk on a basin-wide interval and transferring that on a stream-gauge level. This simulation method also passed another critical test by capturing the inter-annual and low frequency variations of streamflow extremes.

Predictions:- The climate based model, using the indices was successful in near-term predictions, with the North Atlantic Oscillation and the interaction between El-Nino Southern Oscillation and Pacific Decadal Oscillation being the two predictors selected during model selection. The model, using only the most dominant mode of variability PC-1, with strong near-term to decadal predictions of annual maximum streamflow, points towards teleconnections and forcings in the Ohio River Basin.

Basin Scale Processes - All the methods and analyses presented in this paper can be easily extended to other river basins, allowing for a structural and global understanding of the role of low-frequency variations in annual streamflow maximums on a basin wide level and manifestation of that risk at a stream-gauge level. A package \_\_\_ has been developed in \_\_\_\_ for swift implementation of the overall analysis methodology.

**References**

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*(4), 433–459. https://doi.org/10.1002/wics.101

Adamowski, J., & Chan, H. F. (2011). A wavelet neural network conjunction model for groundwater level forecasting. *Journal of Hydrology*, *407*(1), 28–40. https://doi.org/10.1016/j.jhydrol.2011.06.013

Bayazit, M. (2015). Nonstationarity of Hydrological Records and Recent Trends in Trend Analysis: A State-of-the-art Review. *Environmental Processes*, *2*(3), 527–542. https://doi.org/10.1007/s40710-015-0081-7

Cheng, L., AghaKouchak, A., Gilleland, E., & Katz, R. W. (2014). Non-stationary extreme value analysis in a changing climate. *Climatic Change*, *127*(2), 353–369. https://doi.org/10.1007/s10584-014-1254-5

Farnham, D. J., Doss‐Gollin, J., & Lall, U. (2018). Regional Extreme Precipitation Events: Robust Inference From Credibly Simulated GCM Variables. *Water Resources Research*, *54*(6), 3809–3824. https://doi.org/10.1002/2017WR021318

Hirsch, R. M., & De Cicco, L. A. (2015). *User guide to Exploration and Graphics for RivEr Trends (EGRET) and dataRetrieval: R packages for hydrologic data* (USGS Numbered Series No. 4-A10; p. 104). Retrieved from U.S. Geological Survey website: http://pubs.er.usgs.gov/publication/tm4A10

Jones, P. D., Jonsson, T., & Wheeler, D. (1997). Extension to the North Atlantic oscillation using early instrumental pressure observations from Gibraltar and south-west Iceland. *International Journal of Climatology*, *17*(13), 1433–1450. https://doi.org/10.1002/(SICI)1097-0088(19971115)17:13<1433::AID-JOC203>3.0.CO;2-P

Kim Tae-Woong, & Valdés Juan B. (2003). Nonlinear Model for Drought Forecasting Based on a Conjunction of Wavelet Transforms and Neural Networks. *Journal of Hydrologic Engineering*, *8*(6), 319–328. https://doi.org/10.1061/(ASCE)1084-0699(2003)8:6(319)

Knowles, S. G., & Kunreuther, H. C. (2014). Troubled Waters: The National Flood Insurance Program in Historical Perspective. *Journal of Policy History*, *26*(3), 327–353. https://doi.org/10.1017/S0898030614000153

Kwon, H.-H., Lall, U., & Khalil, A. F. (2007). Stochastic simulation model for nonstationary time series using an autoregressive wavelet decomposition: Applications to rainfall and temperature. *Water Resources Research*, *43*(5). https://doi.org/10.1029/2006WR005258

Melillo, J. M., Richmond, T. (T. C. ), & Yohe, G. W. (2014). *Climate Change Impacts in the United States: The Third National Climate Assessment*. https://doi.org/10.7930/J0Z31WJ2

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., & Stouffer, R. J. (2008). Stationarity Is Dead: Whither Water Management? *Science*, *319*(5863), 573–574. https://doi.org/10.1126/science.1151915

Nakamura, J., Lall, U., Kushnir, Y., Robertson, A. W., & Seager, R. (2013). Dynamical Structure of Extreme Floods in the U.S. Midwest and the United Kingdom. *Journal of Hydrometeorology*, *14*(2), 485–504. https://doi.org/10.1175/JHM-D-12-059.1

Nalley, D., Adamowski, J., Biswas, A., Gharabaghi, B., & Hu, W. (2019). A multiscale and multivariate analysis of precipitation and streamflow variability in relation to ENSO, NAO and PDO. *Journal of Hydrology*, *574*, 288–307. https://doi.org/10.1016/j.jhydrol.2019.04.024

*National Water Summary 1988-89: Hydrologic Events and Floods and Droughts*. (1991). U.S. Government Printing Office.

Newman, M., Compo, G. P., & Alexander, M. A. (2003). ENSO-Forced Variability of the Pacific Decadal Oscillation. *Journal of Climate*, *16*(23), 3853–3857. https://doi.org/10.1175/1520-0442(2003)016<3853:EVOTPD>2.0.CO;2

R Core Team. (2016). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.* Retrieved from https://www.R-project.org/

SALAS, J. D. (1993). Analysis and modelling of hydrological time series. *Handbook of Hydrology*, *19*. Retrieved from https://ci.nii.ac.jp/naid/10013176141/

Thomas, A., & Leichenko, R. (2011). Adaptation through insurance: Lessons from the NFIP. *International Journal of Climate Change Strategies and Management*. https://doi.org/10.1108/17568691111153401

Tiwari, M. K., & Chatterjee, C. (2010). Development of an accurate and reliable hourly flood forecasting model using wavelet–bootstrap–ANN (WBANN) hybrid approach. *Journal of Hydrology*, *394*(3), 458–470. https://doi.org/10.1016/j.jhydrol.2010.10.001

Tong, H. (1990). *Non-linear time series: A dynamical system approach*. Retrieved from http://eprints.lse.ac.uk/6212/

Torrence, C., & Compo, G. P. (1998). A Practical Guide to Wavelet Analysis. *Bulletin of the American Meteorological Society*, *79*(1), 61–78. https://doi.org/10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2

**Supplementary Information**

Annex A – Individual Methods

Individual Methods

Principal Component Analysis - Principal Component Analysis is a commonly used linear dimension reduction and component extraction methodology for multivariate and multidimensional datasets(Abdi & Williams, 2010). Let x be the m-dimensional observation set. The principal components (PC) of x are computed as follows.

Here is the i’th principal component and are the eigenvectors of the covariance matrix. The eigenvectors are ordered based on their corresponding eigenvalues, giving the successive PC’s the property of explaining lower residual variance. The PC’s also have the property of being mutually uncorrelated.

This makes the computation of the PC’s a problem involving the computation of eigenvectors and the corresponding eigenvalues of the original m-dimensional dataset. Further, the descending ordering of the eigenvectors allows for dimensionality reduction based on a variance cut-off.

Wavelet Analysis - Hydroclimatic time-series commonly are non-stationary and have low frequency variations and decadal cycles (Milly et al., 2008). Wavelet analysis is used as a frequency domain tool to capture and analyze the localized power variations in the time-series. The analysis helps transform a unidimensional time series to a bi-dimensional time-frequency spectrum.

The wavelet analysis has a long history of use in the hydrology and climatological applications like flood and drought forecasting(Kim Tae-Woong & Valdés Juan B., 2003; Tiwari & Chatterjee, 2010), groundwater level(Adamowski & Chan, 2011), and El-Nino Southern Oscillation and similar climate indices and streamflow and precipitation variability(Nalley, Adamowski, Biswas, Gharabaghi, & Hu, 2019).

The wavelet analysis was carried out using the code provided by (Torrence & Compo, 1998) and was implemented in R. Refer (Torrence & Compo, 1998) for a complete description of the mathematical treatment involved in wavelets.

Autoregressive Models - A commonly used model for time series analysis is the Auto-regressive Moving Average(ARMA) models and has a long history of wide hydro-climatic applications(Kwon et al., 2007; SALAS, 1993).The ARMA model can be further characterized by Auto-Regressive(AR) model and the Moving Average(MA) Model. An AR model of order p written as AR(p), and indicative of its name regresses itself with the previous p components, and is represented as follows

An inherent assumption of the Auto-Regressive model is that the errors are gaussian and independent of all . The AR process is modelled as a stationary process if all the roots of the characteristic equation lie outside the unit circle.

Self-Exciting Threshold Autoregressive Models - Self-Exciting Threshold Autoregressive (SETAR) Models, a subset Threshold Auto-regressive (TAR) models, using the concept of piecewise linearity to model much more complex behavior compared to the traditional auto-regressive model(Tong, 1990).

Based on the threshold variable and threshold value the dynamics of the system is different. For SETAR models, this threshold variable is one of the lag variables, hence the name self-exciting.

Here, is the threshold regime, is the embedding dimension, is the lag, are the coefficients associated with the various lag parameters, is the error associated with the kth regime. Thus, for each threshold regime k the dynamics is modelled using an auto-regressive model with different values of and .

For a SETAR (2,2) model with the 1st 2 denoting the number of threshold regimes and second 2 denoting the AR model associated with each regime, would be written as:

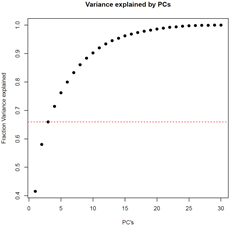
(Equation to be added)

The above example has the threshold on for this example and the threshold can be on any one of the lag parameters. Furthermore, there is no constraint that the order of the AR models should be the same for all the variables.

Annex B: -

Principal Component Analysis

Brief Results - The Principal Component Analysis is applied to the annual streamflow extremes, consisting of 30 sites and 81 years data. The data were log transformed beforehand. The cumulative variance explained by each of the 30 principal components is shown in Figure XX. The first PC explains about 40% of the total variance across the field with the second and third PC explaining 17% and 8% of the variance respectively. Based on the cumulative variance explained by PCs, the first three PCs (marked by the hashed red line in figure XX) explain about 66% of the total variance. Throughout the rest of the analysis, the step of each method involves transformation and subsequent analysis in the PC-space.



*Figure 1: Cumulative Variance explained by PCs. The red dotted line shows the variance explained by the PC 1-3, about 66%.*