**Title:- A novel multi-site simulation methodology for streamflow applications and its applications to the Ohio River Basin.**

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

Abstract:- A novel methodology for the prediction and simulation of streamflow extremes has been proposed. This methodology has been applied to the Ohio River Basin. This method deals with multi-site simulations using dimensionally reduced components of the field. Further the spatial aspects of the region is also accounted in the method. The North American Oscillation (NAO) has been shown to influence the extremes and provides predictability pathways across a multi-year horizons. The Ohio River Basin also reveals several characteristic aspects of large scale hydrology sectors.

**Introduction:-**

Precipitation/Streamflow Importance

Floods account for large fraction of the losses included in natural hazards. As urbanization and land use change increases this risk is likely going to increase in the future.

History of the Ohio River Basin

The Ohio River Basin has a long history of large-scale basin wide floods with major events occurring in 1927, 1945, 1997 and 2011. Unlike other common flooding mechanisms which rely on the land use characteristics, urban landscape and soil, the flooding in this region is dominated by persistent precipitation and/or snow melt events(Farnham, Doss‐Gollin, & Lall, 2018; Nakamura, Lall, Kushnir, Robertson, & Seager, 2013).

Risk Analysis Component

Current state of the art for risk analysis with respect to streamflow and precipitation involves point risk estimation at each site. This method ignores the spatio-temporal aspect of risks consequently leading to underestimation of true risk faced at the site. Here we review the risk at a basin wide scale considering the non-stationarity in the hydroclimatic time series. A proper representation of this risk is important for crucial infrastructure systems to prevent the occurrence of cascading failures and large-scale losses.

A proper representation of risk should be able to capture the underlying non-linear, non-stationary spatio-temporal structure across the entire basin and not just at individual sites. Traditional time series models like the Box-Jenkins which assumes stationarity in the data have failed in this aspect.

Walkthrough the paper

Section 2 elaborates on the data used in this study and the various methods to extract the underlying structure of streamflow extremities on a basin-wide level. The results are presented in Section 3 along with discussion on their meaning. The results section is further subdivided into parts which help evaluate the underlying non-linear structure, the simulating the occurrence of extremes and finally in prediction of these extremes. The conclusion and possibilities of extension of this methodology to various river basins is covered in Section 4.

**Data & Methods:-**

Streamflow Data and Annual Maxima

The streamflow data were restricted to the Ohio River Basin, the region of interest in this study. The data were downloaded from USGS and EPA sources using the ‘dataRetrieval’ (Hirsch & De Cicco, 2015) package in R (R Core Team, 2016). Only stream gauges which had no missing daily data from 01/01/1937 to 12/31/2017 were selected. At each site the time series was transformed from daily to annual, using the annual maximum for each year as a measure of extreme streamflow. Using these criteria, the final number of sites was 30 with each location having 81 years of data from 1937-2017.

Climate Indices

The climate indices used in this study are El-Nino Southern Oscillation (ENSO), Pacific Decadal Oscillation(PDO) and North Atlantic Oscillation(NAO) along with their interactions(Newman, Compo, & Alexander, 2003). The ENSO SST anomaly time series was taken from HadISST Nino 3.4 index. The NAO data was computed as the pressure difference between Gibraltar and Iceland(Jones, Jonsson, & Wheeler, 1997). The PDO was taken from NOAA ERSSTv5. The index included was the mean Feb-May values for each index. The interactions between the three components were computed as the product of indices.

Principal Component Analysis

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

|  |  |  |
| --- | --- | --- |
|  |  | 1 |
|  |  |  |

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 dimensional reduction based on a variance cut-off.

Wavelet Analysis

Hydroclimatic time-series commonly are non-stationary and have low frequency variations and decadal cycles. 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 El-Nino Southern Oscillation and similar climate indices, streamflow and precipitation trends, atmospheric cold fronts and analysis of turbulent flows.

The wavelet analysis was carried out using the code provided by (Torrence & Compo, 1998) and was implemented in R.

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, Lall, & Khalil, 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 in its name regresses itself with the previous p components, and is represented as follows

|  |  |  |
| --- | --- | --- |
|  |  | 1 |

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 .

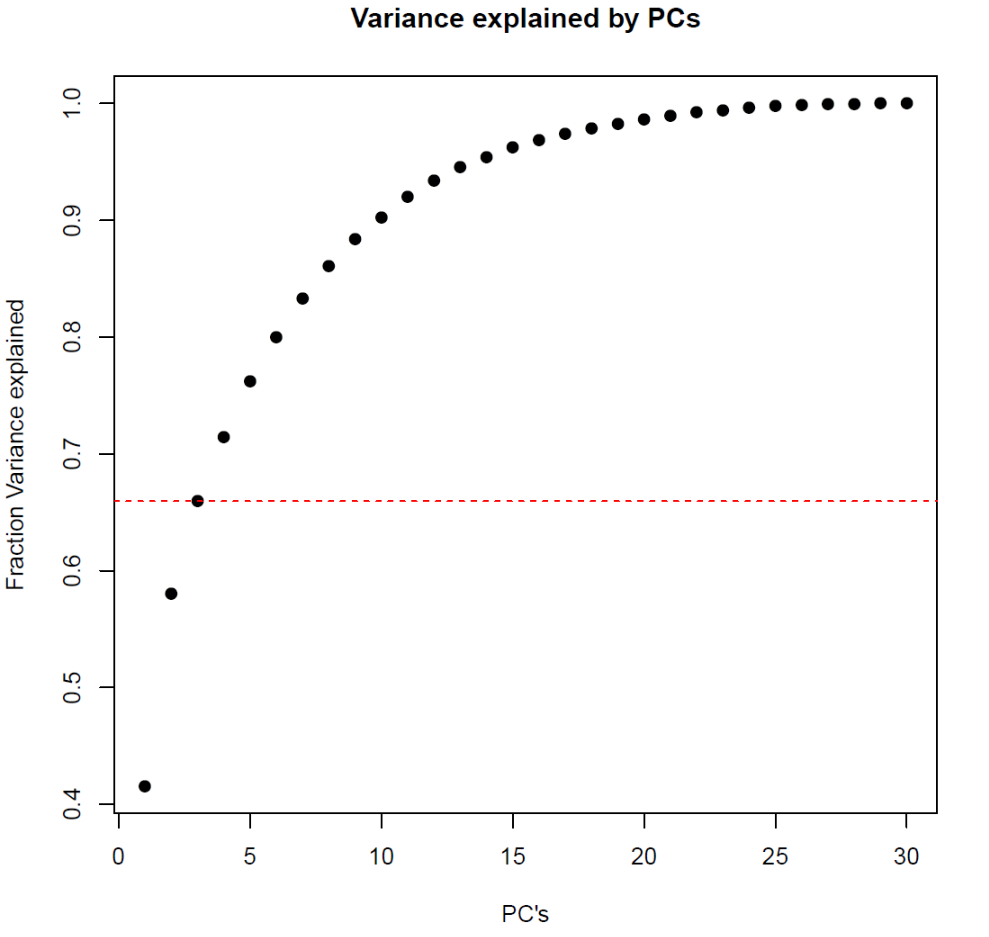
For a SETAR(2,2) model with the 1st 2 denoting the number of threshold regimes and second 2 denoting the AR model associated with each regime, would be written as:

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.

**Results & Discussion: -**

PCA

The Principal Component Analysis (elaborated in Section XX) is applied to the annual streamflow extremes, consisting of 30 sites and 81 years data. The variance explained by the each of the 30 principal components is shown in Figure XX. The first PC explains about 40% of the total variance across of the field with the second and third PC explaining 17% and 9% 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. These 3 PCs are utilized for further analysis.



*Figure XX: Cumulative Variance explained by PCs*

Wavelet Clustering

Why Wavelets – Current state of the art risk assessments focus only on point risk through means similar to a 10-yr flood or a 100-yr flood. Considering a basin wide effect, as is shown in this study, these 10-yr events across regions are not i.i.d and show distinct spatio-temporal structure. The use of PC-Wavelet helps capture the temporal structure by means of the time-frequency methods utilized in the wavelets and encapsulation of the spatial structure by means of eigenvectors of the principal components. If the 10-yr events were truly i.i.d, the total annual occurrence of such events across the region would follow a Poisson process with the expected value being 1/10 times the number of stations. Figure XX shows that this is not observed with periods of heightened activity and long periods of no occurrence of such events across the region.

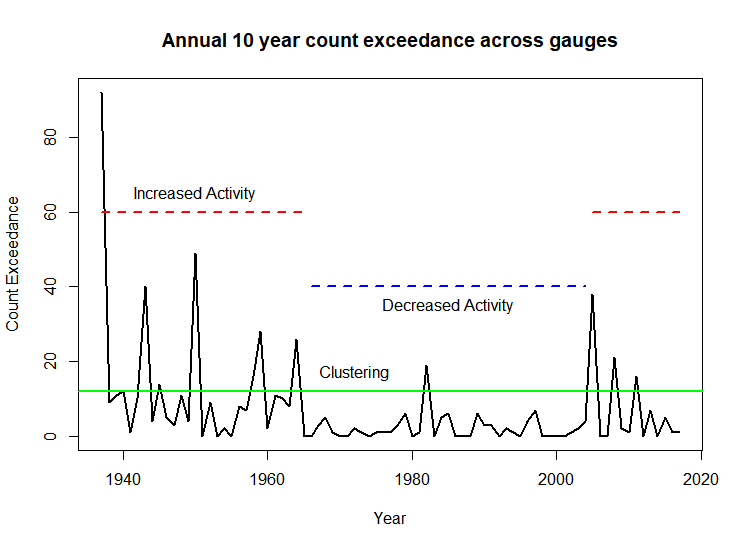
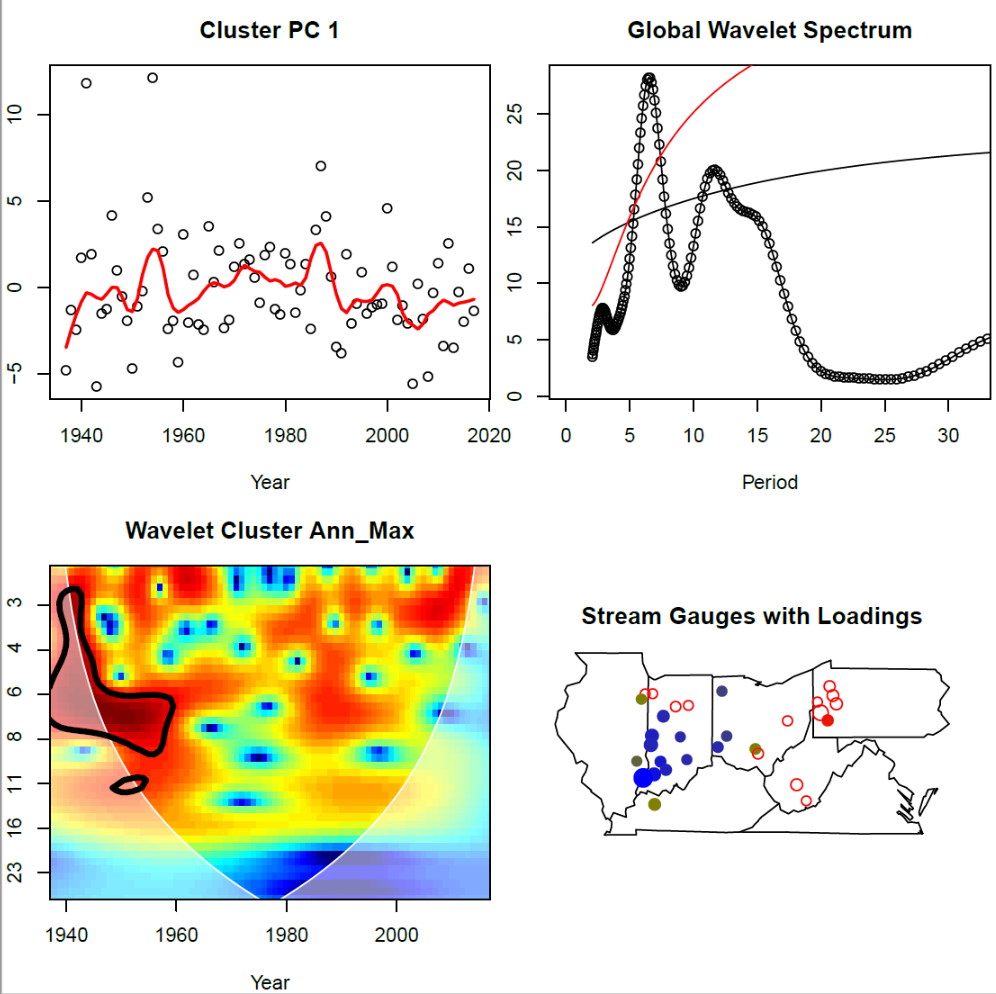
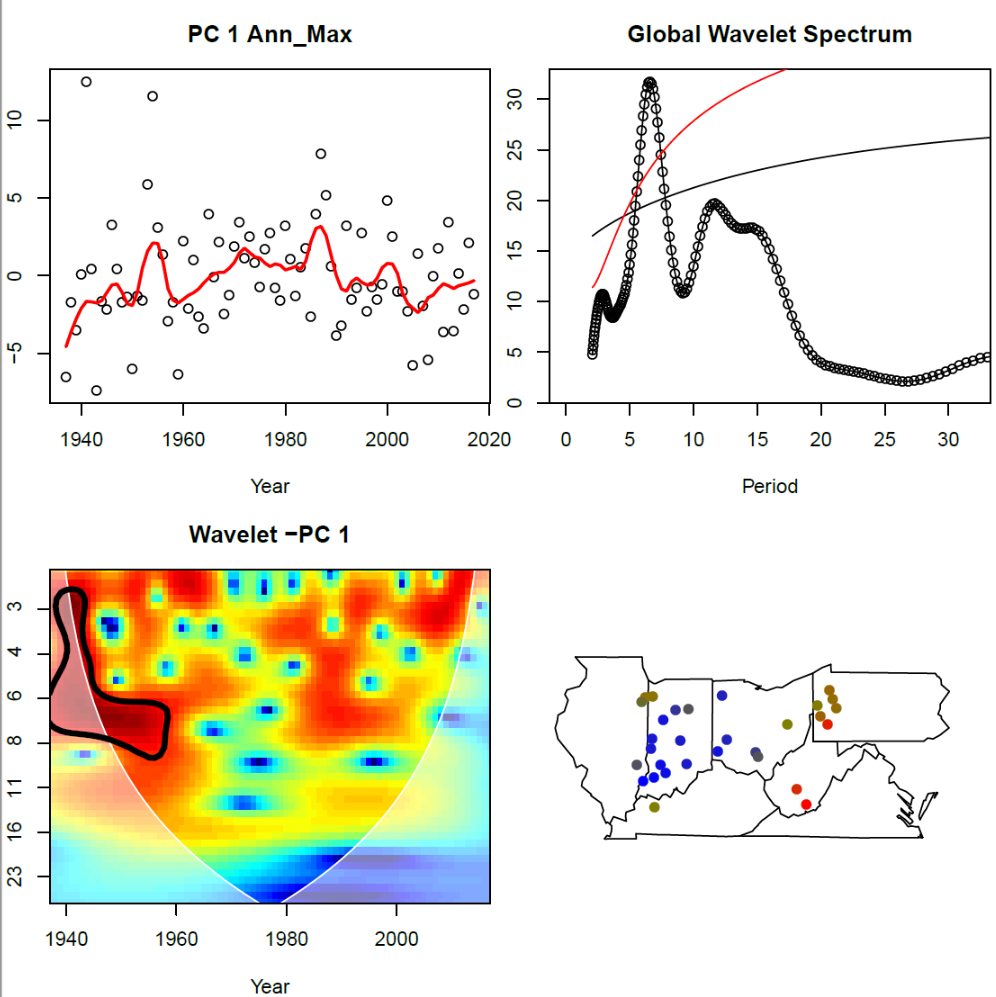


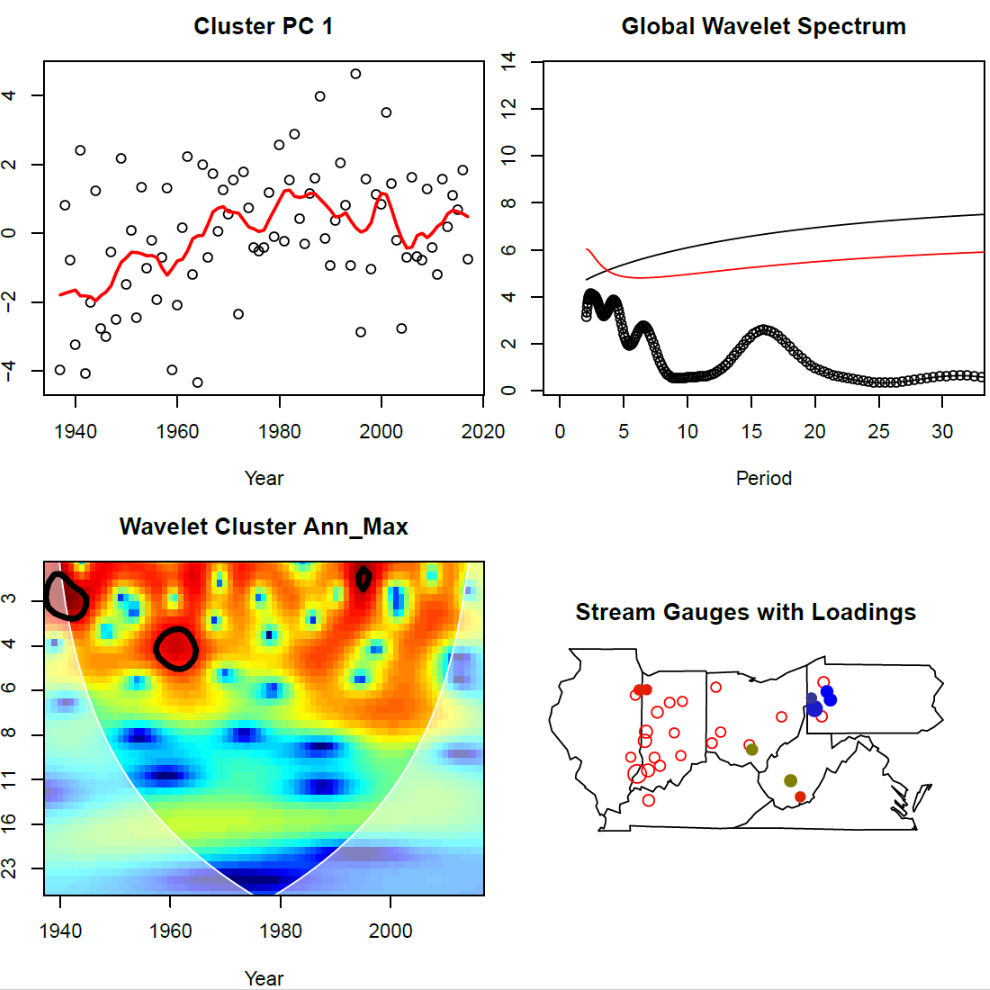
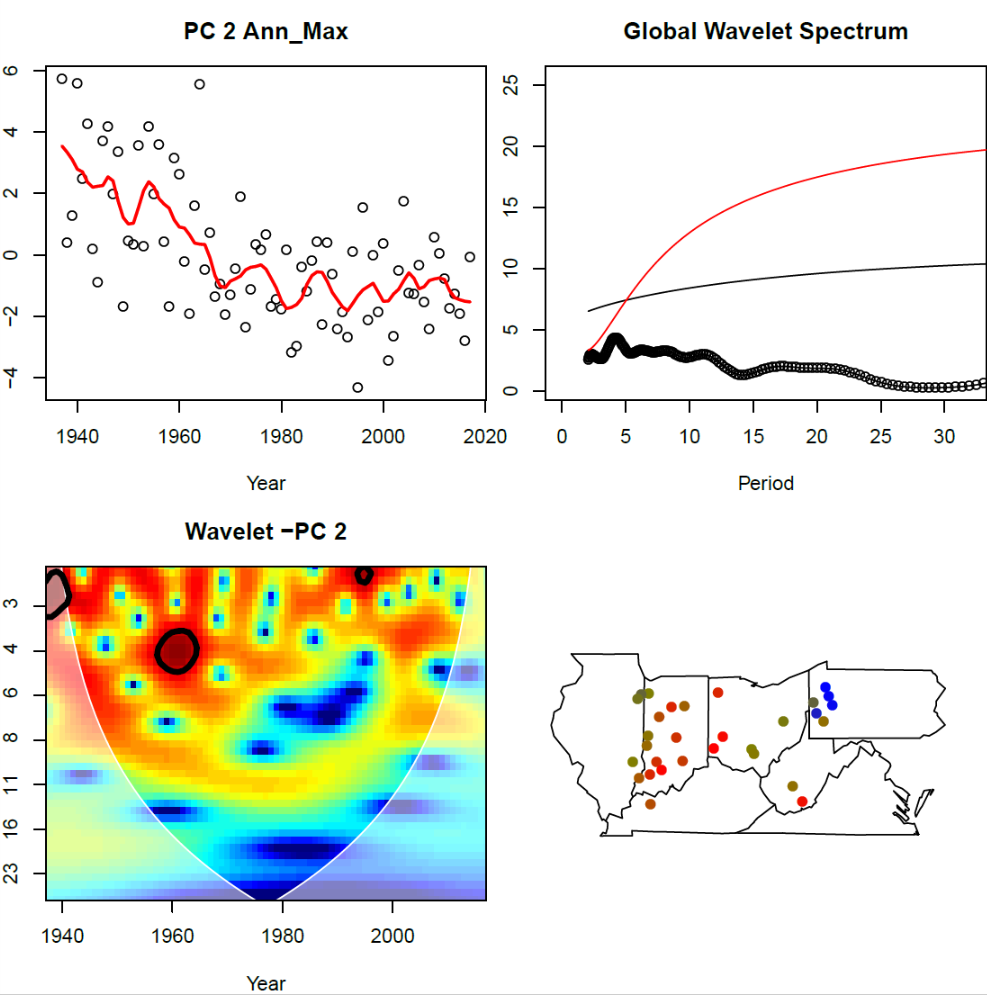
Figure XX: Annual count occurrence of 10-yr events across the 10 sites.

Explain the Figure - The Figure XX below shows the results from the PC-Wavelet and Wavelet-Clustering methodologies. For the implicit clustering-wavelet methodology, the 3 PC selected above further undergo wavelet analysis. In case of the explicit clustering-wavelet method, for sake for comparison, the number of selected clusters is 3, with each cluster’s 1st PC undergoing a wavelet analysis. (Refer Annex \_\_\_ to see the visual hierarchical clustering and location of the three clusters.)

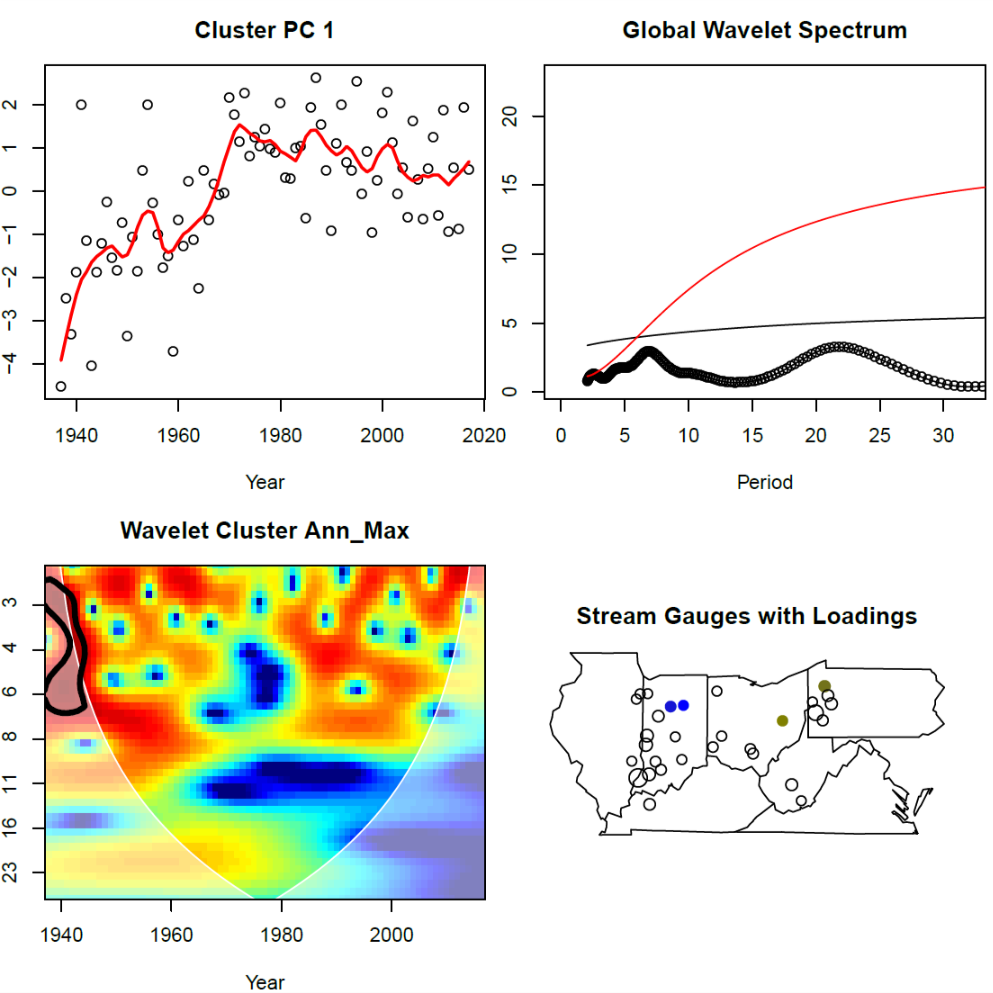
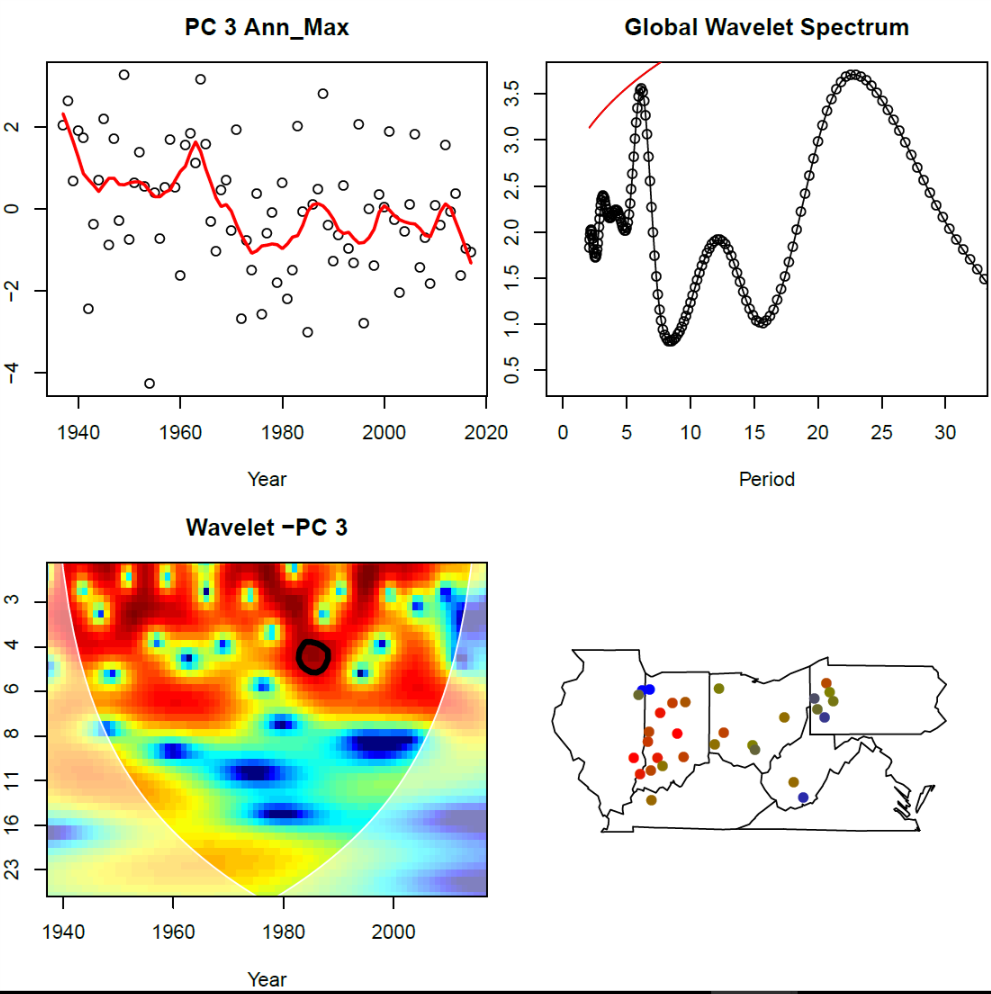
1. (b)



(c) (d)



(e) (f)



*Figure XX: - Implicit Clustering-Wavelet Analysis(a,c,e) and Explicit Clustering-Wavelet Analysis (b,d,f) for the 1st the PCs. For each subplot: - Top-left – The PCs 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) for each site.*

As seen from Figure XX, the output from the two methods for the first two rows are nearly identical. This is especially true for the first row which is (a) PC-1 of the entire field i.e. implicit method and (b) PC-1 of the explicit method. The global wavelet spectrum in both have peak in the 5-7 yrs. period which cross the corresponding 90% confidence intervals over red and white noise. The wavelet power spectrum for both are nearly identical with similar regions in frequency-time space crossing the confidence threshold. Further the distribution of the absolute values of the PC loadings across the two field is similar. This leads to the inference that the region of Indiana, eastern Illinois and western Ohio has dynamics which dominantly drives the upper Ohio River Basin. These dynamical features then could be possibly tied to a global mode of variability (explained further in section XX)

Like the first row, there exist a lot of similarities between the (c) PC-2 of the entire field and (d) first PC of the 2nd cluster. The differences between the (e) 3rd PC of the entire field and (f) first PC of the 3rd cluster are much more pronounced.

The similarities across the results of the two methodologies provide validation for both methods. It also provides regions have a similar spectral signature possibly pointing to regions have similar dynamics or forcing by similar global external climatic variability modes.

AR Simulations

Using the process of iterative model building with each step corresponding with increased model complexity, model which was fit first to the data was the Auto-Regressive (AR) Model. Based on the AIC values a suitable AR order was decided and the model coefficients are computed. The AR models fit to the PCs were of order 4,4 and 0 correspondingly. The AR Models were then simulated for 1000 times and the mean, standard deviation, minimum and maximum; spectral signature and pdf and cdf values of the simulated series and the original PC were compared. These models were successful in simulating the PC mean, whereas it has some decent success in replicating the standard deviation, the minimum and maximum values.

This is followed by reconstructing the field data from the simulated PCs since the loadings(eigenvectors) were computed. The similar comparison process is repeated but in this case each simulation is transformed to the actual 30 stream-gauges. Using three PC’s the simulation skill improves for the site-specific standard deviation yet there is a consistent bias in simulations with the simulations over-estimating it. The simulation for the minimum and maximum values does not improve much. Thus, the AR Models described above have shown a degree of ability to simulate a multi-site flow but there is area for improvements here. (Refer annex B for further details). Note:- We strongly suggest to fit a PC-AR followed by a PC-SETAR model to check for model improvements as a trade-off for the more complex model i.e are the results justified.

SETAR Simulations

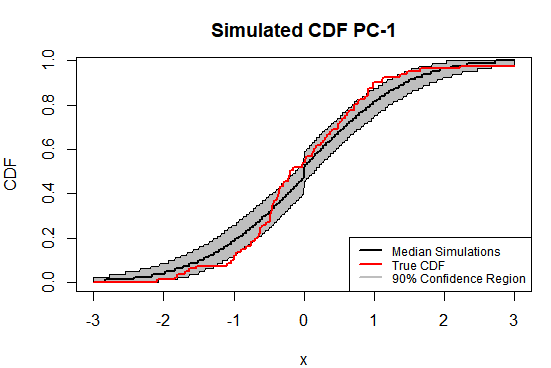
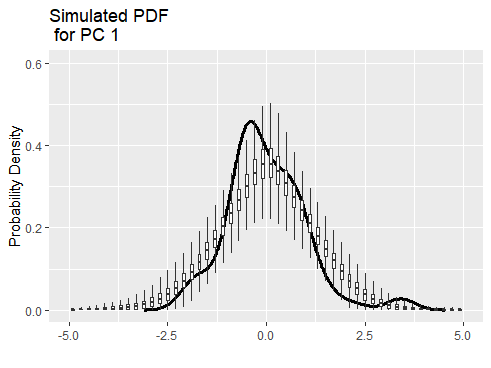
Better than AR - As was seen in the section above, the PC-AR Model was not able to capture the variability in the PC space and correspondingly wasn’t able to project that variability onto the multisite flow. To bridge this shortcoming a SETAR model was fit to the PCs. The SETAR models are known to model and simulate much more complex dynamical processes including limit cycles. As was seen in the section above, based on the cumulative variance criteria, the first three PCs were selected.

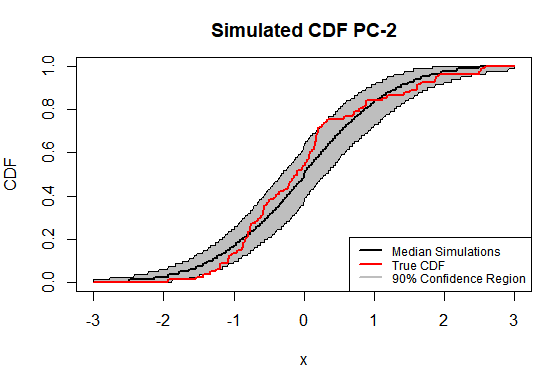
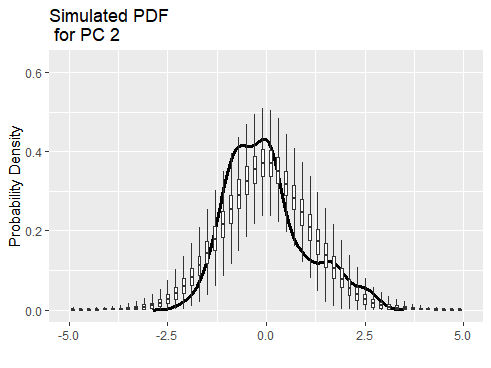
How were the models fit – A separate SETAR model was fit for each PC with AIC being the fitting criteria given an embedding dimension, like an AR model. Given a embedding dimension, based on the AIC the upper and lower auto-regressive order dynamics, threshold lag and threshold value are computed. We have assumed that the dynamics was can captured best with a two regime process instead of three or four with length of our time series( 81 years) being the chief consideration behind this assumption. The fit between embedding dimensions was to difficult to compare and the model was evaluated on a best visual fit based on the simulations. The R package tsDyn was utilized to fitting SETAR models.

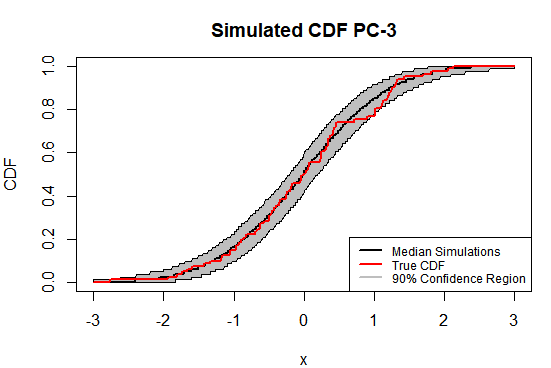
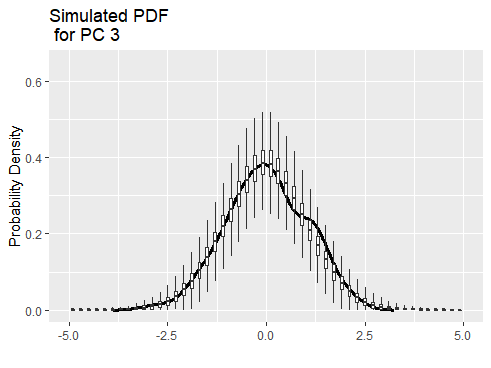
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Name | Embedding Dimension | Lower Regime Dynamics | Upper Regime Dynamics | Threshold Lag | Threshold Value |
| PC1 | 4 | 4 | 4 | 3 | 0.4894 |
| PC2 | 2 | 1 | 2 | 1 | -0.432 |
| PC3 | 1 | 1 | 1 | 1 | -0.245 |

Table XX:- SETAR Model Specifications for each PC.

Model Simulations and Transformations – Table XX shows the SETAR model parameters for the three different PCs. Each SETAR Model was simulated 1000 times, for a run-length of 81 which corresponds to the time series length. This provides a clear benchmark to check for reproducibility and comparisons of key features of the PCs. The key parameters of interest to assess the model skill are the probability density function, cumulative density function, the spectral signature and the spatial variation across the river basin.

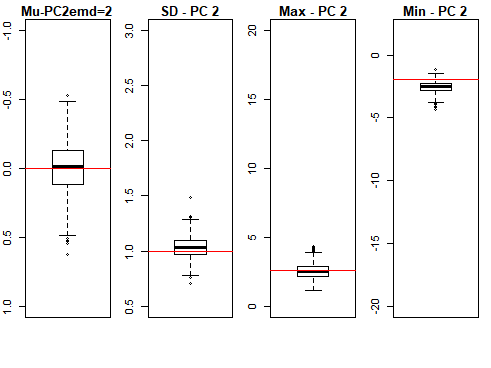
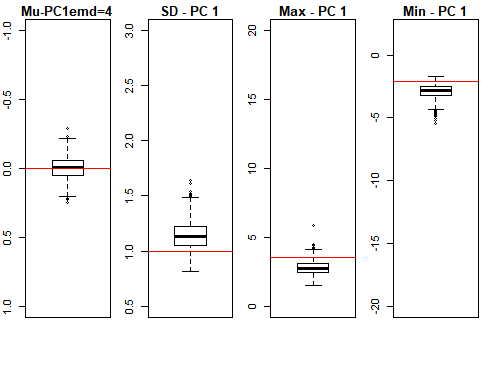


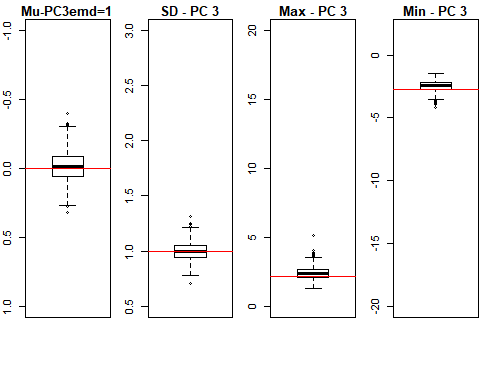




*Figure XX:- Probability Densities(right) and Cumulative Densities(left) for each of the PCs for the 1000 simulations. The black line is the density of the original PC. Density Estimation done by means of kernels.*

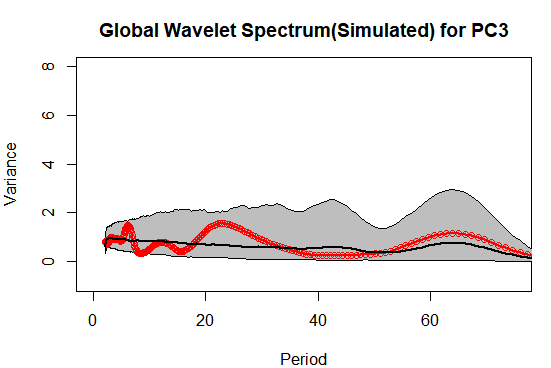
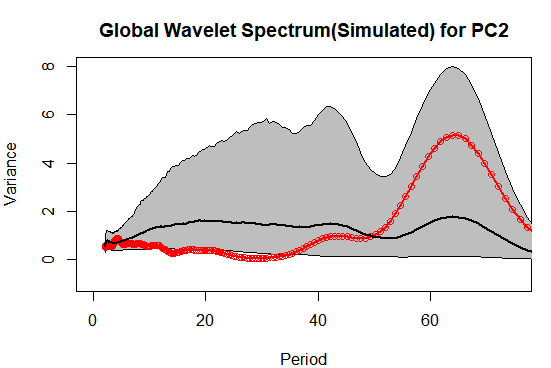
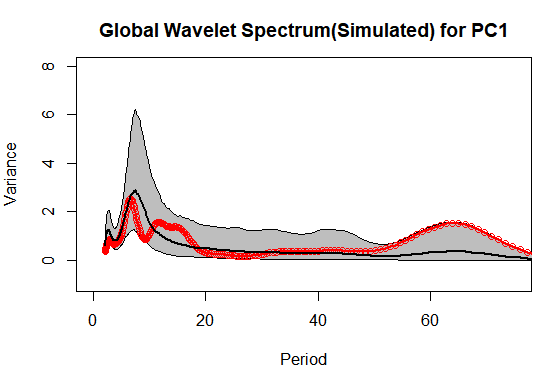
CDF, PDF and Moments – Figure XX shows the simulated CDF and PDF for the PCs and the Figure XX shows the moments for the PC along with the true values. SETAR simulations can capture the internal dynamics of the PC and consequently the multi-site flow much better with an ability to generate minimum and maximum values outside the time series values, which has a positive effect on planning in water management scenarios as unseen phenomena which is plausible under the model dynamics is accounted in the planning.. The PDF simulations are gaussian, leading to some differences with the true values but comparisons with the CDF shows that the SETAR model is able to capture the distribution with the tails being captured accurately. Checking the CDF, PDF and the moments for the true data along with the simulations lends confidence to the SETAR models and its ability to capture the dynamics over the river basin.





*Figure XX:- Simulation of the key characteristics of the PC – Mean, Standard Deviation, Minimum and Maximum values. The red line denotes the actual data values.*

Spectral Signature – Hydroclimatic time series are characterized by low frequency and decadal oscillations which are difficult to capture by traditional statistical models. This makes accurately capturing the spectral signature over lower frequencies crucial to understanding and characterization of risks faced across the region. Using the wavelet analysis, the global wavelet spectrums of the PC and the simulated time series are computed, giving a characterization time-frequency domain. Figure XX shows SETAR model can simulate and thus capture the spectral signature of the individual three PCs. For the 1st PC the 5-7-year cycle is simulated prominently. The SETAR model for the 2nd PC can incorporate the long-term secular trend structure in the data as is seen for periods greater than 75 years. This combined with the replications of the PC moments – mean, standard deviation, minimum and maximum values show that SETAR is able to incorporate the decadal and larger variations in streamflow variability.



*Figure XX:- 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 – Another crucial step in assessing the method skill is transformation of the simulations from low dimensional PC space to the higher dimension original space comprising of the 30 sites. Using the 1000 simulations for the 3 PC’s the simulations were retransformed using equation \_\_\_ 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. The sites 4,8,9 and 14 which have the greatest deviation from their true CDF have the lowest eigenvectors for the 1st PC which contributes the largest variance.

Conclusion – Thus reducing the dimensional to a higher dimension system, thereby focusing on the main signals and modelling and simulating them has shown promise to model higher order systems with greater accuracy. This method could be easily extended to other regions or river basin throughout the world given enough data availability.

Streamflow extremes predictability

Question of Predictability - Since the SETAR models have shown an ability to capture the dynamics in the system and simulate the system, the next obvious question arises is: - Can the SETAR model or any other model predict the streamflow extremes in the near future? The previous section uses the inherent dynamics of the field by means of the principal components to model the system.

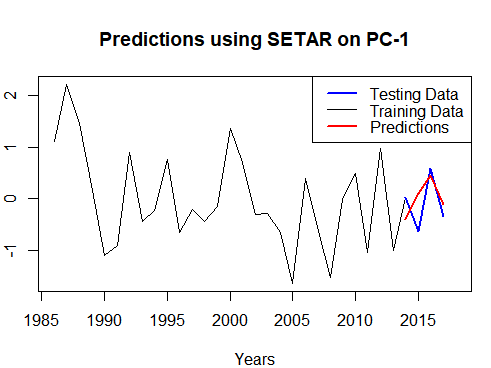
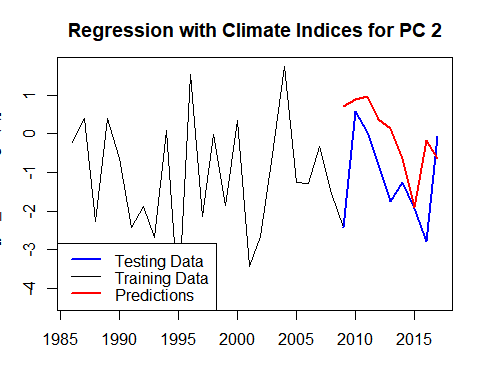
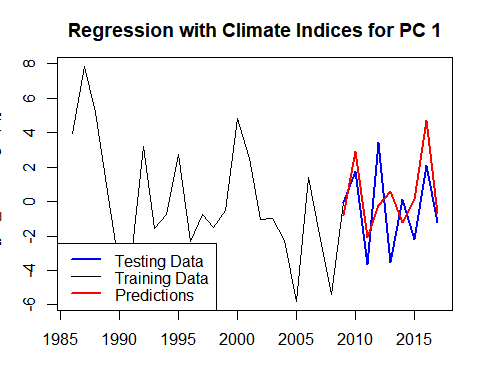


Figure XX:- Predictions based on the SETAR model applied to the 1st PC.

SETAR Baseline – As a baseline model to compete against the models using climate indices, predictions were made on the SETAR model using 77 years (1937-2013) as training data and 4 years (2014-2017) as testing data to validate the predictions. The figure XX shows that the SETAR model applied to the 1st PC, which explains about 40% of the total variance, having decent skill. We now transform this PC to the original streamflow field of 30 sites and make predictions for 4 years using the SETAR model(Results not shown). Unlike the NAO based regression model elaborated below, the SETAR model exhibits skill over a smaller prediction horizon. Extending to larger time horizons leads to significant loss of prediction skill.

Regression of Climate Indices

Climate Indices Intro - The dominant modes of climate variability are known to affect distant regional climate over varying temporal scales. These major modes are characterized as climate indices with the chief among them being the El Nino Southern Oscillation (ENSO), North Atlantic Oscillation (NAO), Pacific Decadal Oscillation (PDO) and the Atlantic Multidecadal oscillation (AMO). The Ohio River Basin located near in North American Continent near the Great Lakes, is affected by dominant global modes of variability by means of teleconnections. The climate indices used were ENSO, PDO and NAO along with their interactions. The mean Feb-May climatology was used to predict the streamflow maxima, using ordinary linear regression. These indices are used to make predictions in the PC-space and are then transformed to predictions in the river basin domain space. A simple linear regression was utilized to fit the model, with comparisons between different models evaluated by AIC criteria.



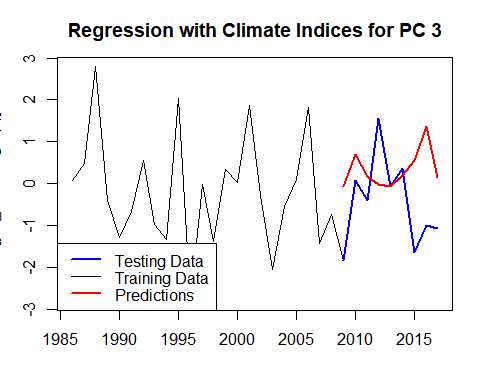
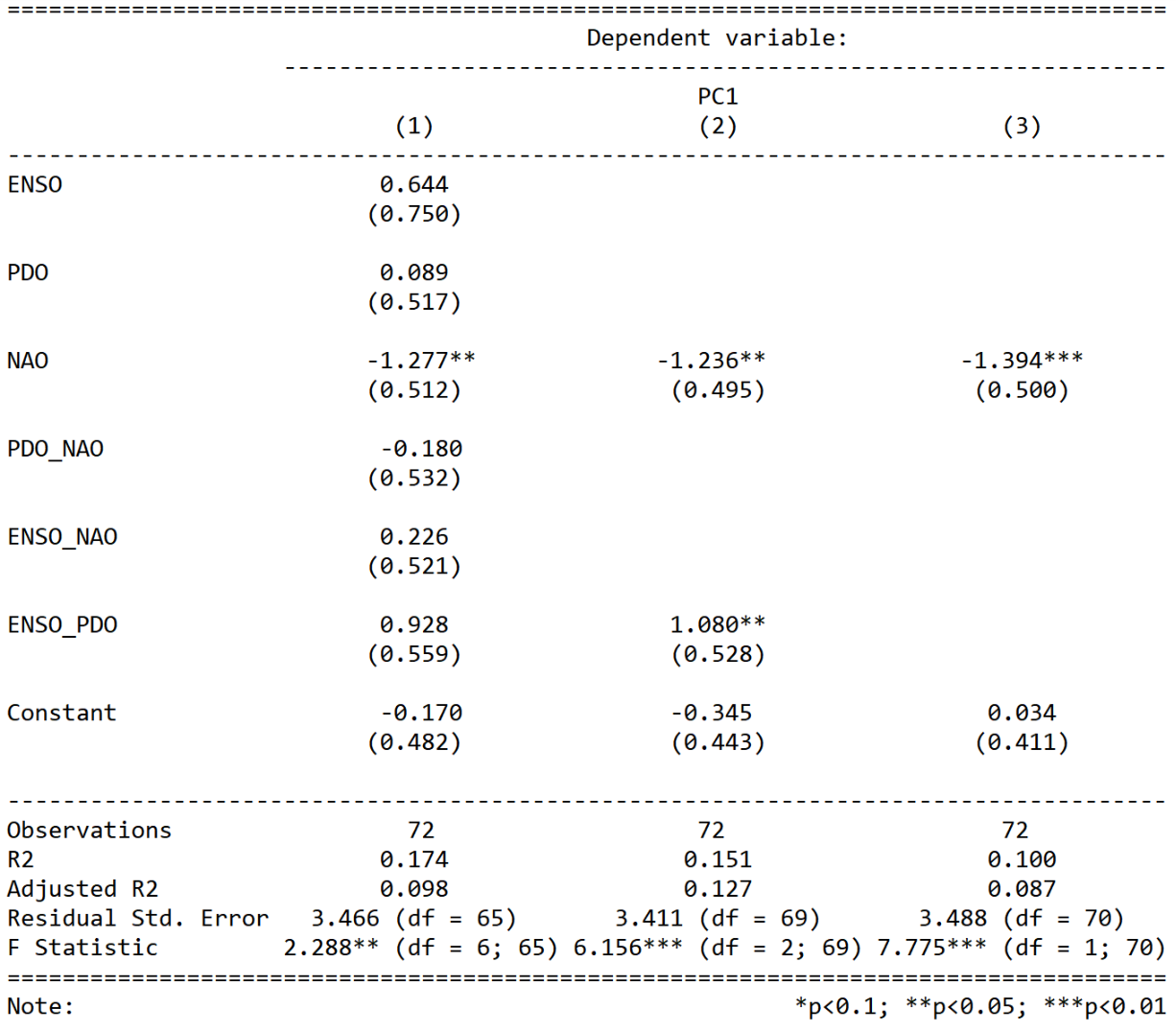


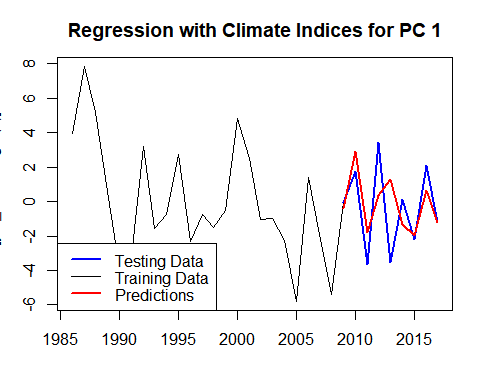
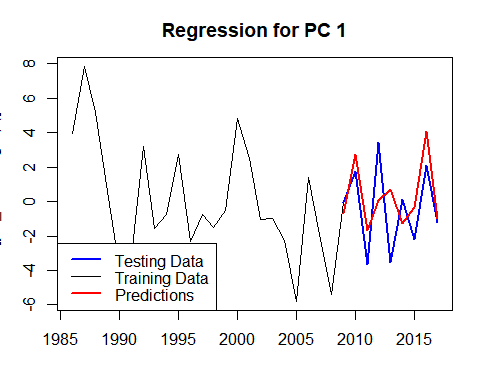
Figure XX:- Predictions for a 9 year horizon using all the above specified climate indices.

All Indices - The results from the predictions of an OLS regression to each PC are presented in Fig XX. OLS using the climate indices can provide predictability in streamflow extremes but only for PC-1 rather than PC-2 and PC-3 with the model prediction skill falling going from the 1st to the 3rd PC. Since PC-1 was the only variable showing predictability and explaining the most variance, the other two were dropped from further analysis in this section.



*Table XX:- Model summary for the linear regressions for all three models. Note:- The variables with the two indices names are the interactions between the two.*

Simpler Models – Based on the significance and the standard errors associated with each of the variables in the OLS model described above, it is clear that certain indices and interactions do not contribute to predictability. Reducing the model complexity, the model was simplified by keeping only NAO and ENSO-PDO interaction term and another model was followed by keeping just NAO. Based on the AIC values, the model with NAO and ENSO-PDO interaction was the best fit. (386 - 388 and 395). The detailed results for each of the three cases is presented in Table XX. Figure XX shows the predictions from the OLS models containing NAO and PDO-ENSO interaction; and just NAO as predictors. NAO and the interaction between PDO and ENSO provide a great predictability pathway, up to 9 years, to predict extremes for the next few years. An obvious caveat to this is that we do not know the corresponding year’s Feb-May NAO index, but if the skill of the GCMs were to improve this would be valuable tool.



*Figure XX:- Predictions based on the linear model using the predictors only (a) ENSO\*PDO and NAO (b) NAO.*

Retransformation - The model specified above was able to predict PC-1, explaining about 40% of the total field variance, for a near decadal period. Checking the utility of this prediction in PC space, by its ability to predict extremes at the site-specific stream gauges was the next step. This was accomplished by fitting the model which includes the covariates NAO and ENSO-PDO interaction and computing the 9-year predictions (This would be the predictions seen in Figure XX (a)). These predictions were then translated to the original stream-gauge data using equation \_\_\_\_\_. The prediction skill was benchmarked against the site-specific Long-Term Mean (LTM) and predictions made by fitting a OLS regression using NAO and ENSO-PDO interaction as covariates at each site. Across the 9-year horizon we see remarkable skill with it being 61% against LTM and 57% against site-specific regression. Over short 4-year horizon this value increases to 75% against the LTM and 81% against the site-specific regression as shown in Figure XX.

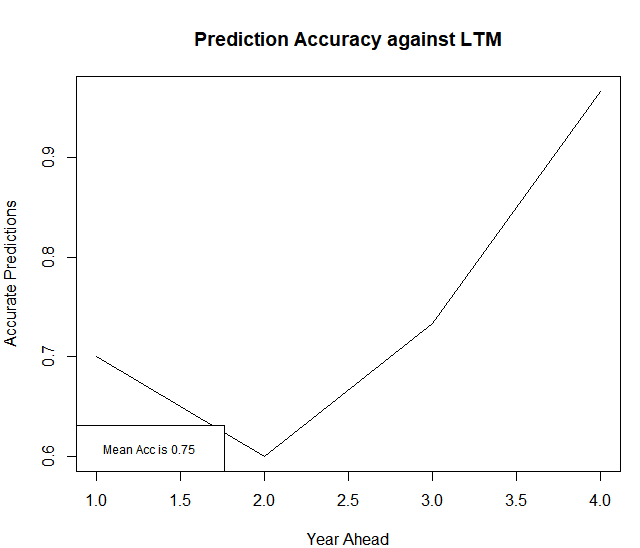
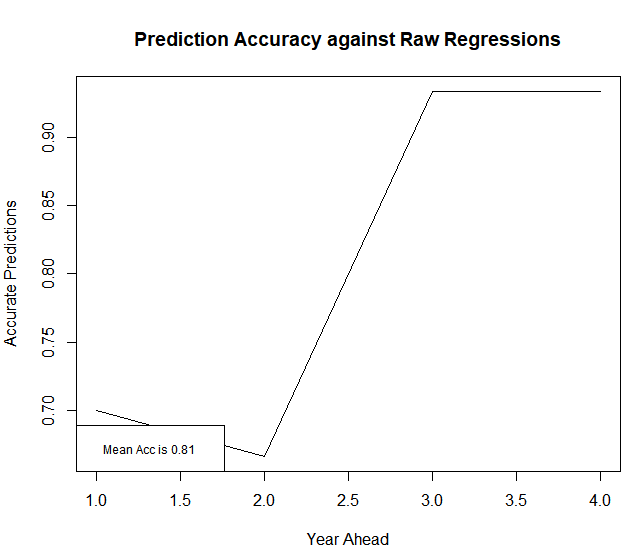
 

Figure XX:- Prediction Skill across the 4 year-horizon against the benchmarks (a) Long Term Mean - LTM (b) Site specific NA and ENSO\*PDO regressions.

The 1st PC provides a clear predictability pathway using well established climate indices over a multi-year horizon. This helps in planning over a multi-year period and provides a clear link to the regional teleconnections to global climatic modes of variability.

**Conclusion**

A new methodology to simulate multi-site river basin streamflow extremes using a combination of dimension reduction by means of a Principal Component Analysis and modelling by a Self-Exciting Threshold Autoregressive Models is presented here. This methodology can accurately simulate the traditional components of the dimensionally reduced components and individual sites - like the mean, standard deviation, minimum and maximum. This simulation method also passes another critical test by reproducing the accurate spectral signature along with the pdf and the cdf of the individual components along with the sites.

The Ohio River Basin with its long history of regional scale precipitation and flooding events accompanied with detailed literature and extensive period of data availability presented an ideal region to study the 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.

Using both explicit and implicit methods of clustering followed by wavelet analysis, led to similar results pointing to different dynamics in different regions of the basin. The regions of Illinois, Indiana and Western Ohio represented by the 1st PC seem to be dynamically uncoupled from Pennsylvania, eastern Ohio and Tennessee.

The North Atlantic Oscillation along with the interaction between the El-Nino Southern Oscillation and Pacific Decadal Oscillation were covariates which showed predictability on a near decadal scale. The fact that the 1st dimensionally reduced component was only used for this purpose is even more promising and this has significant implications for water and disaster management strategies in the Ohio River Basin.

All the methods presented in this paper can be easily extended to other river basins over the globe contingent on there being enough data to justify these analyses. A package \_\_\_ has been developed in \_\_\_\_ for swift implementation of the PC-SETAR methodology.

**References**

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

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

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

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/

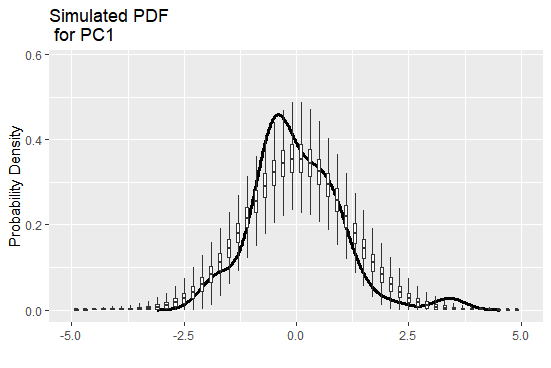
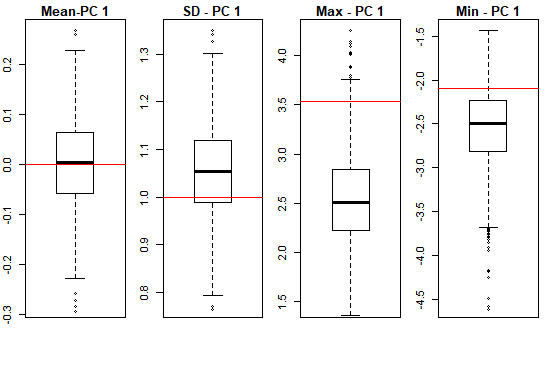
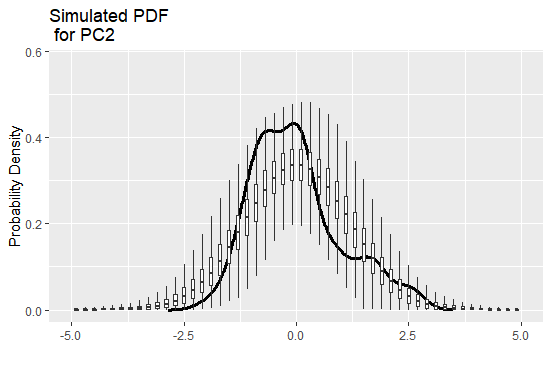
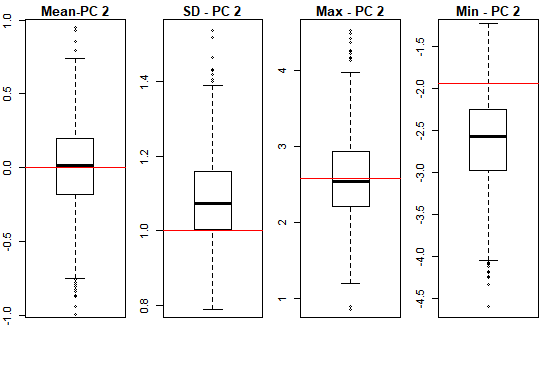
Tong, H. (1990). *Non-linear time series: A dynamical system approach*. Retrieved from http://eprints.lse.ac.uk/6212/

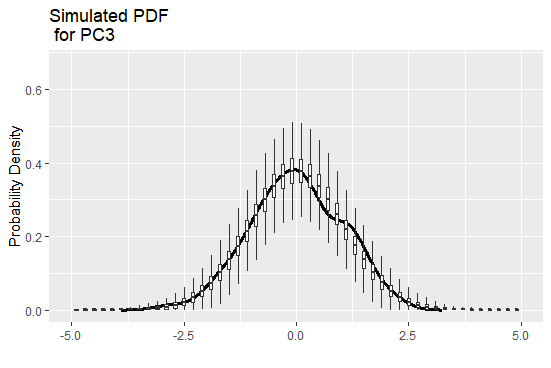
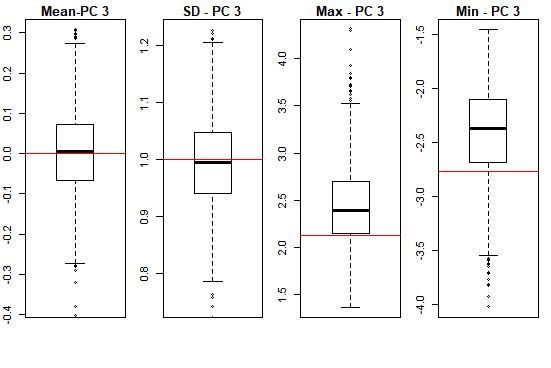
Torrence, C., & Compo, G. P. (1998). A Practical Guide to Wavelet Analysis. *Bulletin of the American Meteorological Society*, *79*(1), 61–78. https://doi.org/10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2

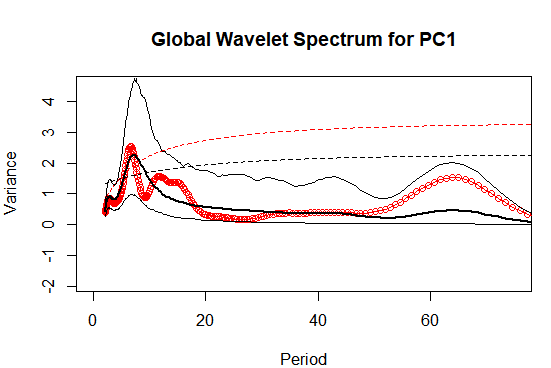
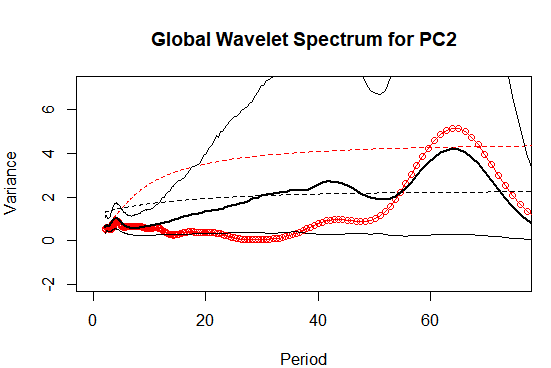
**Supplementary Information**

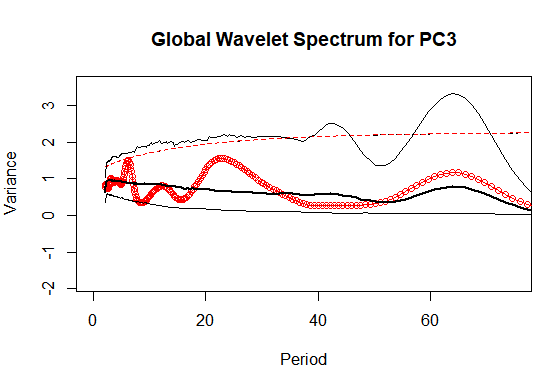
Annex A – Hierarchical Clustering

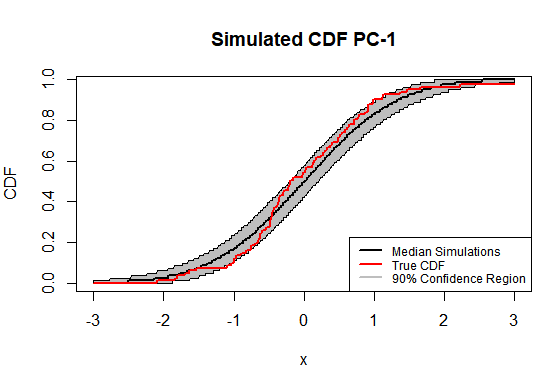
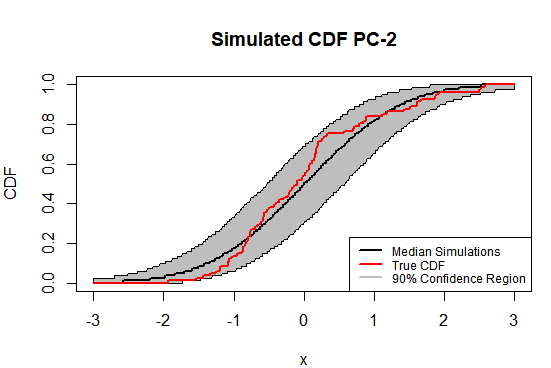
Annex B – Streamflow Replications for AR Modelling.

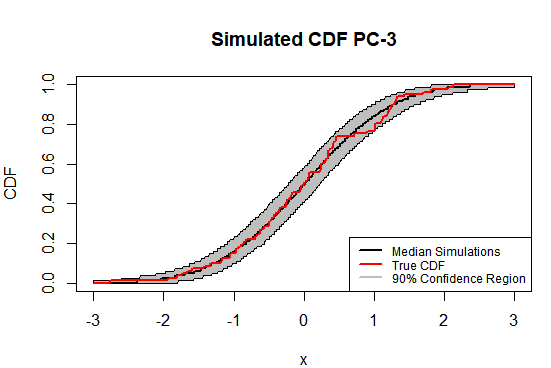
   





The AR model fits as a good first step to model this dynamical process.

Annex C – CDF and Moment Replications for the Individual Sites

