Modeling of Basin Scale Hydro-meteorological association by Hybrid Wavelet-ARX approach

Mayank Suman & Rajib Maity

Indian Institute of Technology Kharagpur

May 8, 2020

Introduction

- Global Circulation Model (GCM) prediction for climatic/meteorological variable is better than hydrological variable.
- Inter-relation (hydro-meteorological association) is challenging to model due to sptio-temporal variability; however if modeled, it will help in ensuring future water security with changing climate.
- The hydro-meteorological association should be continuously evolving with time.
- Meteorological forcing might not necessarily be affecting all frequency ranges of the hydrological variable.

Hydro-meteorological association has been studied for:

Literature	Conclusion\Purpose
Ahmad and Simonovic, (2005)	Analysis of hydrograph
Lau and Kim, (2012)	Teleconnection of hydrometeorological extremes (2010 Pakistan flood and Russian heat wave)
Ishak *et al.*, (2013)	Decreasing error in wind speed calculation
Shih *et al.*, (2014)	Development of early flood warning system
Durocher *et al.*, (2016)	Predicting mean annual streamflow in Quebec river using North Atlantic Oscillation Index
Adnan *et al.*, (2019)	Streamflow prediction in mountainous basin

Methodology

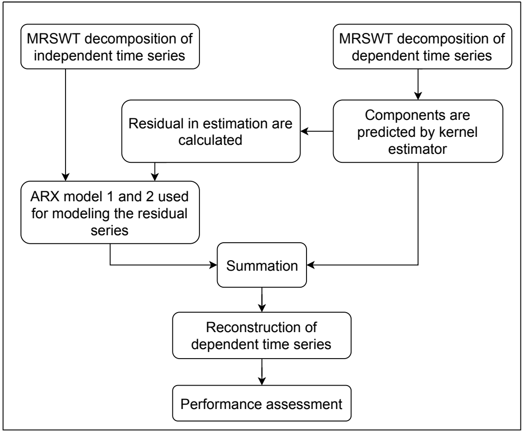


Figure 1: Proposed Wavelet-Auto-Regressive model with eXogenous inputs (ARX) Model: Methodological Overview

Hypothesis

- Due to continuous evolution, the hydro-meteorological association should be more pronounced at constituent wavelet level.
- Wavelet component may have two parts:
  - Memory: modeled using auto-regressive model.
  - Exogenous part: Driven by affecting/exogenous variables.
- After separating memory part, the exogenous part can be modeled as function of exogenous factors.

Methodology

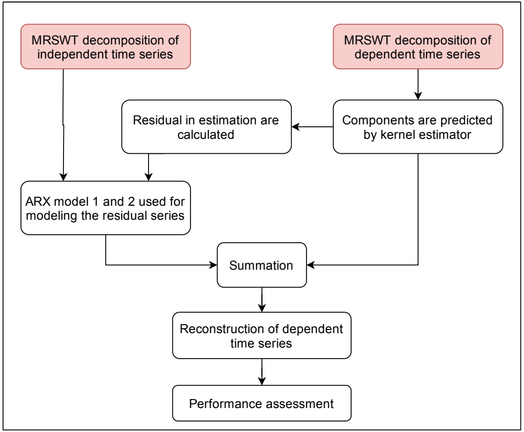
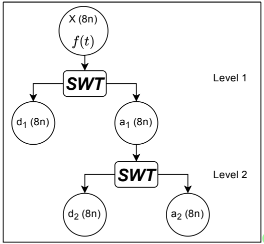
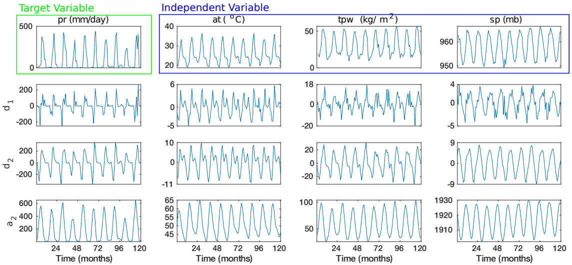


Figure 1: Proposed Wavelet-ARX Model: Methodological Overview



$$f(t) = \sum_{k} a_{n,k}\varphi_{n,k}(t) + \sum_{j=1}^{n}\sum_{k} d_{j,k}\psi_{j,k}(t)$$



Methodology

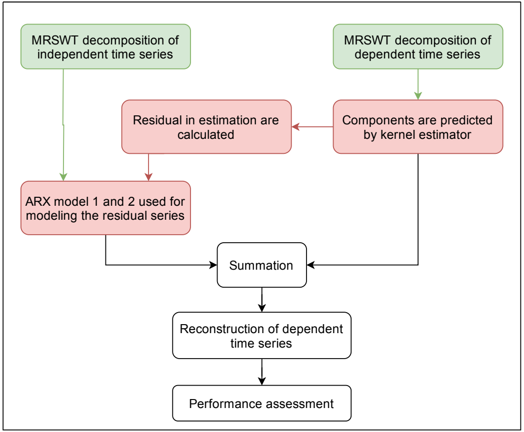


Figure 1: Proposed Wavelet-ARX Model: Methodological Overview

Kernel Estimator for dependent series components:

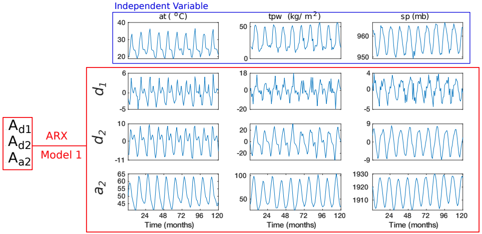
$$\tilde{a}_n(m+1) = \frac{\sum_{t=1}^{m} a_n(t) K\left(\frac{t-(m+1)}{h}\right)}{\sum_{t=1}^{m} K\left(\frac{t-(m+1)}{h}\right)}$$

$$\tilde{d}_j(m+1) = \frac{\sum_{t=1}^{m} d_j(t) K\left(\frac{t-(m+1)}{h}\right)}{\sum_{t=1}^{m} K\left(\frac{t-(m+1)}{h}\right)}$$

ARX models for residuals:

$$A_{c,y}(t) = \sum_{k=1}^{p} S_k A_{c,y}(t-p) + \sum_{l=1}^{c} \sum_{j=0}^{q} T_{l,j} C_{x,l}(t-j) + E$$

The models differ in selection of exogenous components:



Methodology

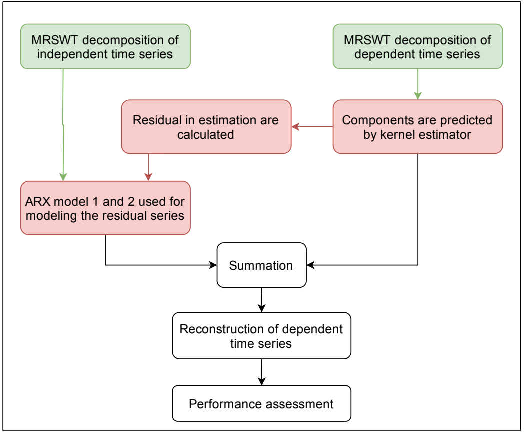


Figure 1: Proposed Wavelet-ARX Model: Methodological Overview

Kernel Estimator for dependent series components:

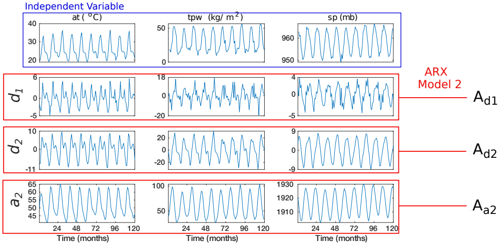
$$\tilde{a}_n(m+1) = \frac{\sum_{t=1}^{m} a_n(t) K\left(\frac{t-(m+1)}{h}\right)}{\sum_{t=1}^{m} K\left(\frac{t-(m+1)}{h}\right)}$$

$$\tilde{d}_j(m+1) = \frac{\sum_{t=1}^{m} d_j(t) K\left(\frac{t-(m+1)}{h}\right)}{\sum_{t=1}^{m} K\left(\frac{t-(m+1)}{h}\right)}$$

ARX models for residuals:

$$A_{c,y}(t) = \sum_{k=1}^{p} S_k A_{c,y}(t-p) + \sum_{l=1}^{c} \sum_{j=0}^{q} T_{l,j} C_{x,l}(t-j) + E$$

The models differ in selection of exogenous components:



Methodology

$$\hat{f}(t) = \sum_k \hat{a}_{n,k}\varphi_{N,k}(t) + \sum_{j=1}^{n}\sum_k \hat{d}_{j,k}\psi_{j,k}(t)$$

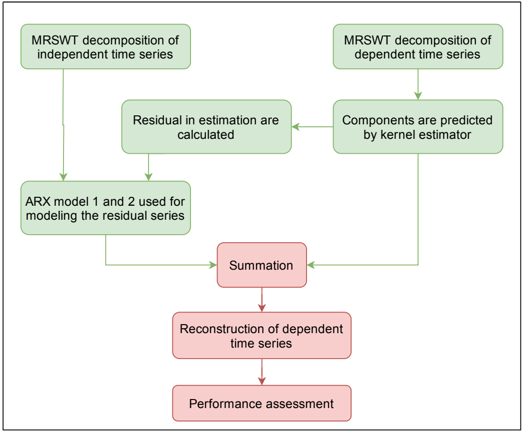


Figure 1: Proposed Wavelet-ARX Model: Methodological Overview

Coefficient of Determination (R^2), Refined Index of Agreement (D_r), Mean Absolute Error (MAE) and unbiased Root Mean Square Error $(uRMSE)$ are used for assessing model performance.

$$R^2 = (r)^2; \qquad r = \frac{Cov(Y, \hat{Y})}{\sigma_y \sigma_{\hat{y}}}$$

$$Dr = \begin{cases} 1 - D_{r\_frac} & \text{for } D_{r\_frac} \leq 1 \\ \frac{1}{D_{r\_frac}} - 1 & \text{for } D_{r\_frac} > 1 \end{cases}$$

$$D_{r\_frac} = \frac{\sum_{i=1}^{n} \left|Y_i - \hat{Y}_i\right|}{2\sum_{i=1}^{n} \left|Y_i - \overline{Y}\right|}$$

$$MAE = \frac{\sum \left|Y_i - \hat{Y}_i\right|}{n}$$

$$uRMSE = \sqrt{\frac{\sum((Y - \overline{Y}) - (\hat{Y} - \overline{\hat{Y}}))^2}{n}}$$

Relative Importance Analysis

Dominance Analysis

Measure the average association contributed by individual input.

$$Y = M(\boxed{X_1}, X_2, X_3)$$

$$DARIM_1 = \text{Mean}\left\{\begin{array}{ll} R^2(M(X_1, X_2)) & - R^2(M(X_2)) \\ R^2(M(X_1, X_3)) & - R^2(M(X_3)) \\ R^2(M(X_1, X_2, X_3)) & - R^2(M(X_2, X_3)) \end{array}\right\}$$

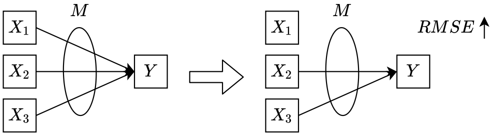
Hence,

$$DARIM_i = \frac{\sum_{j=1}^{2^{N-1}-1} \left(R^2_{(M(\{PO_r, I_i\}))} - R^2_{(M(PO_r))}\right)}{2^{N-1}-1}$$

Birnbaum Importance Measure

Measures the probability of a input being critical for proper functioning of the model.

Assumption:



$RMSE(M(X_1))$	$RMSE(M(X_1, X_2))$	$RMSE(M(X_1, X_2, X_3))$
$RMSE(M(X_2))$	$RMSE(M(X_2, X_3))$	
$RMSE(M(X_3))$	$RMSE(M(X_1, X_3))$	

$u_1 = 1 - \frac{3}{4} = 0.25$ $\quad u_2 = 0.25$ $\quad u_3 = 0.5$

$$BIM_2 = u_1 u_3 = 0.125$$

$$BIM_i = u_1 u_2 u_3 \cdots u_{i-1} u_{i+1} \cdots u_N$$

Prediction of total monthly precipitation in UMB

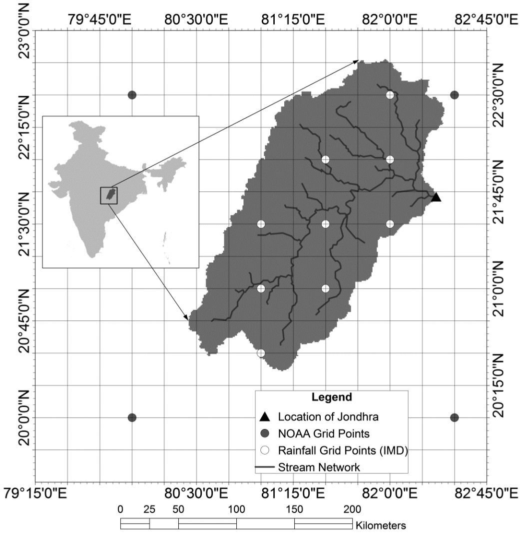


Figure 2: Study Basin – Upper Mahanadi Basin

Table 1: Details of data utilized for predicting monthly rainfall over UMB

Independent Time Series	Source	Temporal Res.	R^2 with rainfall
Avg. Surface Air Temp.	NOAA*	Monthly	0.0015
Precipitable Water Content	NOAA	Monthly	0.7682
Surface Pressure	NOAA	Monthly	0.4576
Avg. Air Temp. (925 mb)	NOAA	Monthly	0.0102
Avg. Air Temp. (700 mb)	NOAA	Monthly	0.2739
Avg. Air Temp. (500 mb)	NOAA	Monthly	0.5996
Avg. Air Temp. (200 mb)	NOAA	Monthly	0.5799
Avg. Sp. Humidity (925 mb)	NOAA	Monthly	0.7265
Avg. Sp. Humidity (850 mb)	NOAA	Monthly	0.7444
Avg. Geop. Height (925 mb)	NOAA	Monthly	0.5039
Avg. Geop. Height (500 mb)	NOAA	Monthly	0.0063
Avg. Geop. Height (200 mb)	NOAA	Monthly	0.5120
Avg. U Wind (925 mb)	NOAA	Monthly	0.4277
Avg. U Wind (200 mb)	NOAA	Monthly	0.6403
Avg. V Wind (925 mb)	NOAA	Monthly	0.0003
Avg. V Wind (200 mb)	NOAA	Monthly	0.2426

* (Kalnay *et al.*, 1996)

Target Series: Basin precipitation recorded by India Meteorological Department (IMD) (0.5°latitude × 0.5°longitude)(Rajeevan *et al.*, 2008)

Results

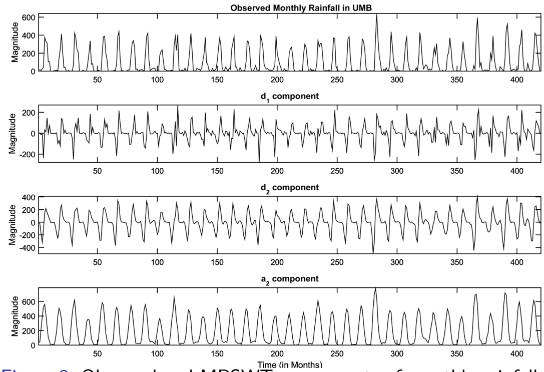


Figure 3: Observed and MRSWT components of monthly rainfall series in the study basin

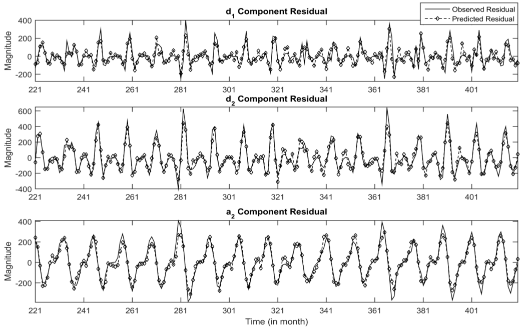


Figure 4: Predicted components residuals from ARX model 2 ($p = 3$ and $q = 2$) as compared to the originally calculated residuals during testing period.

Results

Table 2: Statistics showing model performance for prediction of monthly rainfall over the study basin during the testing period

Model Performance Statistics	No. of Auto-Regressive input (p)	Lag in exogenous time series input (q)											
		ARX (for Comparison)				Model 1				Model 2			
		0	1	2	3	0	1	2	3	0	1	2	3
R^2	1	0.873	0.864	-	-	0.818	0.783	-	-	0.752	0.821	-	-
	2	0.877	0.864	0.862	-	0.901	0.877	0.862	-	0.883	0.897	0.895	-
	3	0.874	0.860	0.859	0.851	0.911	0.885	0.873	0.763	0.887	0.905	0.907	0.895
D_r	1	0.858	0.856	-	-	0.835	0.806	-	-	0.810	0.833	-	-
	2	0.858	0.853	0.853	-	0.871	0.845	0.831	-	0.857	0.864	0.861	-
	3	0.856	0.853	0.851	0.842	0.873	0.847	0.839	0.782	0.855	0.864	0.864	0.855
MAE	1	57.85	59.43	-	-	61.00	67.58	-	-	70.16	61.26	-	-
	2	56.70	59.20	58.32	-	46.15	50.99	54.10	-	50.17	47.05	47.52	-
	3	56.50	59.15	58.84	59.11	44.37	49.46	51.90	71.15	50.01	45.64	45.28	47.89
$uRMSE$	1	56.39	57.60	-	-	60.95	68.82	-	-	71.13	60.63	-	-
	2	55.40	57.55	56.67	-	46.13	50.34	53.23	-	51.04	46.70	47.20	-
	3	55.41	57.67	57.17	58.19	44.35	48.87	51.18	69.90	51.00	45.19	44.83	47.40

Results

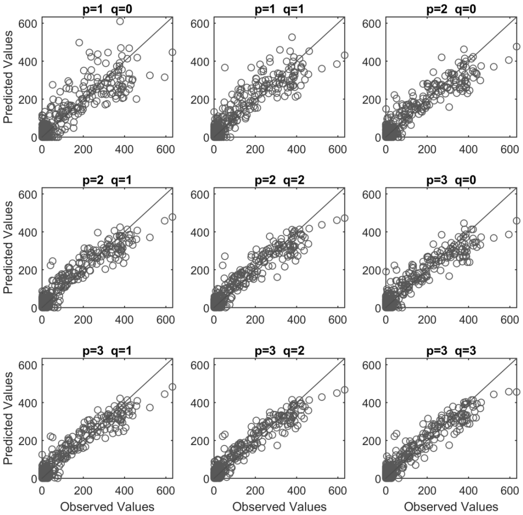


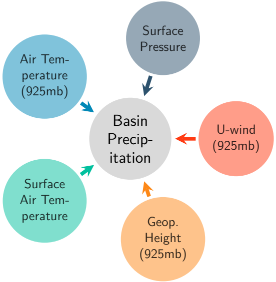
Figure 5: Scatter plot between observed and predicted monthly rainfall over the study basin (in mm/month) for the different combinations of p and q using hybrid Wavelet-ARX model 2.

Results

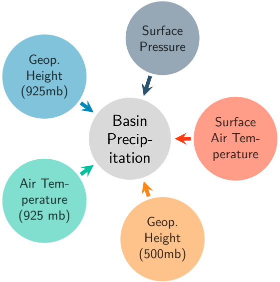
Table 3: Relative Importance Measure for independent variables using Wavelet-ARX model 2 ($p = 3$ and $q = 2$) for monthly rainfall prediction.

Independent Time Series	BIM ($\times 10^{-6}$)	$DARIM$ ($\times 10^{-3}$)
Avg. Surface Air Temp.	6.48	4.10
Precipitable Water Content	4.83	3.94
Surface Pressure	8.13	6.39
Avg. Air Temp. (925 mb)	7.14	4.76
Avg. Air Temp. (700 mb)	4.77	2.00
Avg. Air Temp. (500 mb)	4.20	-1.38
Avg. Air Temp. (200 mb)	5.03	1.04
Avg. Sp. Humidity (925 mb)	5.64	3.88
Avg. Sp. Humidity (850 mb)	4.12	1.06
Avg. Geop. Height (925 mb)	6.10	5.47
Avg. Geop. Height (500 mb)	5.64	4.53
Avg. Geop. Height (200 mb)	5.47	1.83
Avg. U Wind (925 mb)	6.06	3.29
Avg. U Wind (200 mb)	5.48	2.47
Avg. V Wind (925 mb)	5.22	0.82
Avg. V Wind (200 mb)	3.62	-1.83

BIM



DARIM



All Variables (Table 2)

	BIM	DARIM	All Variables (Table 2)
R^2	0.907	0.910	0.907
D_r	0.867	0.869	0.864
MAE	44.72	44.25	45.28
$uRMSE$	44.53	44.01	44.83

Take Home

- The association between meteorological and hydrological variables are more prominent on wavelet component level, resulting in Wavelet-ARX model outperforming ARX model in most of the cases.
- WT-ARX model 2 outperforms model 1 when exogenous input components are considered, hence, the WT components of input/target variables at same level (same frequency bins) are more associated.
- Additionally, the relative importance analysis of meteorological variables helps in identifying the forcings having stronger hydro-climatic association with hydrological variables.
- For instance, the meteorological variable having highest association with total monthly precipitation in UMB are surface pressure, average geo-potential height at 925mb, average air temperature at 925mb, average geo-potential height at 500mb and average surface air temperature.

Further Reading

Suman, M. and Maity, R., 2019. Hybrid Wavelet-ARX approach for modeling association between rainfall and meteorological forcings at river basin scale. Journal of Hydrology, 577, p.123918.

Journal of Hydrology 577 (2019) 123918

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

ELSEVIER

JOURNAL OF HYDROLOGY

Research papers

Hybrid Wavelet-ARX approach for modeling association between rainfall and meteorological forcings at river basin scale

Mayank Suman[a], Rajib Maity[b,*]

[a] School of Water Resources, Indian Institute of Technology Kharagpur, Kharagpur 721302, West Bengal, India

[b] Department of Civil Engineering, Indian Institute of Technology Kharagpur, Kharagpur 721302, West Bengal, India

A R T I C L E I N F O

This manuscript was handled by A. Bardossy, Editor-in-Chief

Keywords:
Hydroclimatic association
Rainfall prediction/simulation
Climate change
Hybrid Wavelet-ARX model
Relative importance analysis

A B S T R A C T

Interaction between meteorological and hydrologic processes is challenging to model owing to their high spatio-temporal variability. The understanding of their associations can help to ensure future fresh water security with changing climate. In this study, due to continuously evolving nature of these interactions, the hydrological and meteorological variables are studied on wavelet component level. Multi-Resolution Stationary Wavelet Transformation (MRSWT) is used to transform the independent (climatic variable) and dependent (hydrological variable) time series into their components. The components of the dependent time series are modeled using a kernel-based auto-regressive (AR) model for separating their memory part. The residuals are hypothesized to be the effect of interaction of predictor variables and thus, are modeled using the MRSWT components of meteorological variables in an auto-regressive model with exogenous inputs (ARX). Finally, the predicted residuals (effect of climatic variables) are added to the component estimated by kernel-based AR estimator (memory of dependent series components) to obtain the predicted components of the dependent hydrologic variable, which are then inverse-transformed to obtain the predicted dependent hydrologic variable. The developed hybrid Wavelet-ARX is found to capture the information about relationship between synthetically generated data better than a simple ARX model. The model is then applied to predict total monthly rainfall over Upper Mahanadi Basin and is found to effectively extract the information from the poorly associated hydro meteorological variables. While the potential of Wavelet-ARX is found to be impressive for hydro meteorological applications, additionally, discarding some climatic inputs on the basis of their relative importance may lead to better prediction by the developed model. The developed model is suitable for extracting climatic forcings and is desirable in a changing climate.

1. Introduction

association as far as rainfall is concerned in the basin.

It is hypothesized that the inter-relationship between meteor-

References

Adnan, Rana Muhammad, Zhongmin Liang, Salim Heddam, Mohammad Zounemat-Kermani, Ozgur Kisi, and Binquan Li (2019). "Least square support vector machine and multivariate adaptive regression splines for streamflow prediction in mountainous basin using hydro-meteorological data as inputs". In: *Journal of Hydrology*, p. 124371.

Ahmad, Sajjad and Slobodan P Simonovic (2005). "An artificial neural network model for generating hydrograph from hydro-meteorological parameters". In: *Journal of Hydrology* 315.1-4, pp. 236–251.

Bosq, Denis (2012). *Nonparametric statistics for stochastic processes: estimation and prediction*. Vol. 110. Springer Science & Business Media.

Burrus, CS, RA Gopinath, and H Guo (1998). *Introduction to Wavelets and Wavelet Transforms: A Primer*.

Cao, Liangyue, Yiguang Hong, Haiping Fang, and Guowei He (1995). "Predicting chaotic time series with wavelet networks". In: *Physica D: Nonlinear Phenomena* 85.1-2, pp. 225–238.

Daubechies, Ingrid (1992). *Ten lectures on wavelets*. Vol. 61. Siam.

Durocher, Martin, Tae Sam Lee, Taha BMJ Ouarda, and Fateh Chebana (2016). "Hybrid signal detection approach for hydro-meteorological variables combining EMD and cross-wavelet analysis". In: *International Journal of Climatology* 36.4, pp. 1600–1613.

Ishak, Asnor Muizan, Renji Remesan, Prashant K Srivastava, Tanvir Islam, and Dawei Han (2013). "Error correction modelling of wind speed through hydro-meteorological parameters and mesoscale model: a hybrid approach". In: *Water resources management* 27.1, pp. 1–23.

References

Kalnay, Eugenia *et al.* (1996). “The NCEP/NCAR 40-year reanalysis project”. In: *Bulletin of the American meteorological Society* 77.3, pp. 437–472.

Lau, William KM and K M Kim (2012). “The 2010 Pakistan flood and Russian heat wave: Teleconnection of hydrometeorological extremes”. In: *Journal of Hydrometeorology* 13.1, pp. 392–403.

Maity, Rajib and D Nagesh Kumar (2008). “Probabilistic prediction of hydroclimatic variables with nonparametric quantification of uncertainty”. In: *Journal of Geophysical Research: Atmospheres* 113.D14.

Rajeevan, M, J Bhate, and Ashok K Jaswal (2008). “Analysis of variability and trends of extreme rainfall events over India using 104 years of gridded daily rainfall data”. In: *Geophysical Research Letters* 35.18.

Shih, Dong-Sin, Cheng-Hsin Chen, and Gour-Tsyh Yeh (2014). “Improving our understanding of flood forecasting using earlier hydro-meteorological intelligence”. In: *Journal of hydrology* 512, pp. 470–481.

Thank You

Additional Slides

Wavelet Transform

- Need of Transform – The usual representation (Amplitude vs Time) is not always the best for analysis.
- Help in separating slow moving and fast moving components of time series.
- Wavelet Transform (WT) : transforms signals into coefficients of scaled and shifted version of wavelet function.
- Due to finite domain, scaling and shifting of wavelet function enable it to catch most of intermittent disturbances of different durations.
- WT represents the signal in terms of frequency and time domain.
- Unlike Fourier transform, WT provide the temporal information along with frequency information.
- Well suited for the study of multi-scale, non-stationary signals (Daubechies, 1992; Burrus *et al.*, 1998).
- Multi-Resolution Wavelet Transform (MRWT): WT can again be applied on approximate component to produce components at even lower level or frequency ranges (Burrus *et al.*, 1998).
- WT are of many types based different shifting and scaling schemes
  - Continuous Wavelet Transform (CWT)
  - Discrete Wavelet Transform (DWT)
  - Stationary Wavelet Transform (SWT)

Wavelet Function

- Wavelet functions are finite domain square-integrable disturbance of unit energy and zero mean amplitude. Ex.- Haar, Daubechies, Morlet etc.

$$\int \psi(t)dt = 0 \qquad \qquad \iint |\psi(t)|^2 \, dt = 1$$

- Wavelet functions can separate only one half (higher) of frequency range, hence, they have another function called scaling function to separate the other half frequency range from signal.
- Haar wavelet:

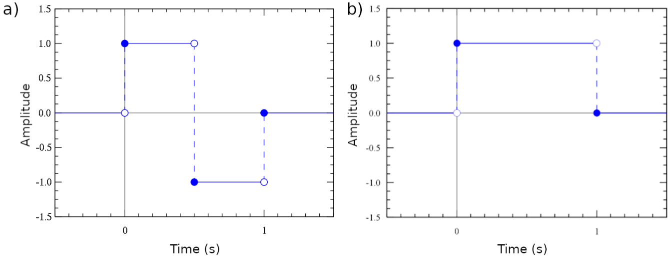


Figure 6: a) Haar wavelet function b) Corresponding scaling fuction

Continuous Wavelet Transform

If $\psi(t)$ is wavelet function then its shifted and scaled functions $(\psi_{a,b}(t))$ are obtained as:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \tag{1}$$

where $a \in \mathbb{R}^+$, and $b, t \in \mathbb{R}$. a and b are scaling and shifting parameters respectively. Wavelet transform is given by:

$$W_f(a,b) = \frac{1}{\sqrt{C_\psi}} \int f(t)\psi^*_{a,b}(t)dt \tag{2}$$

where the * denotes complex conjugate, $C_\psi = 2\pi \int \left|\widehat{\psi}(\omega)\right|^2 /\omega d\omega$. The $\widehat{}$ denotes the Fourier transform $\widehat{\psi}(\omega) = \int e^{-i\omega t}\psi(t)dt/\sqrt{2\pi}$.

If the basis wavelet or mother wavelet $\psi(t)$ is orthogonal (Daubechies, 1992), then the inverse of wavelet transformation is given by:

$$f(t) = \frac{1}{\sqrt{C_\psi}} \iint \frac{W_f(a,b)\psi_{a,b}(t)}{a^2} da\ db \tag{3}$$

Dyadic Discrete Wavelet Transform

DWT is formed when shifting and scaling parameters are taken discrete. The discrete wavelet if sampled over dyadic space, time grid, then they are called dyadic discrete wavelets (Cao *et al.*, 1995).

$$\psi_{a,b}(t) = \frac{1}{\sqrt{2^j}}\psi\left(\frac{t}{2^j} - k\right) \tag{4}$$

where $j, k \in \mathbb{Z}$. The DWT is similar to CWT but it is applied in discrete terms.

As per Nyquist– Shannon sampling theorem, subband coding (downsampling by avoiding every second sample) is carried out to avoid redundancy in component series as suggested by following relation:

$$N_o B_c = N_c B_o \tag{5}$$

where N and B shown length and bandwidth of signal. Subscript o and c mean parent and component series. Hence, $N_c = {}^{N_o}/_2$

Stationary Wavelet Transform

- SWT is specially designed to avoid the transition-invariance of DWT.
- SWT achieves transition-variance by avoiding down-sampling of components as per Nyquist–Shannon sampling theorem and up-sampling the filter coefficient.
- Despite redundancies in components, SWT reduces the complexity of signal analysis as both input signal and its components have equal length.
- In this study, the SWT is used as preferred WT.

Filter theory basis of DWT and SWT

- DWT or SWT can be applied as pair of high pass and low pass filters.
- High pass filter is obtained from wavelet function and low pass filter is obtained from corresponding scaling function.
- Component from high pass filter is called detailed component and other is called approximate component.
- The approximate and detailed components show trend and local disturbance respectively in the parent signal.

Wavelet components at time t is projection of $f(t)$ over daughter wavelet functions.

$$W_f(a,b) = \frac{1}{\sqrt{C_\psi}} \int f(t)\psi^*_{a,b}(t)dt \tag{6}$$

Given constant j, the equation is convolution of $f(t)$ with dilated, reflected and normalized mother wavelet $h(t) = \frac{1}{2^j}\psi\left(\frac{-t}{2^j}\right)$. For reference convolution between $p(t)$ and $q(t)$ is given by

$$(p*q)(t) = \int_{-\infty}^{\infty} p(\tau)g(t-\tau)d\tau \tag{7}$$

Hence, for Haar wavelet, the dilated reflected and normalized filters are $h = \frac{1}{\sqrt{2}}[1,-1]$ and $g[n] = \frac{1}{\sqrt{2}}[1,1]$.

Multi-Resolution Wavelet Transform

- Multi-Resolution Wavelet Transform (MRWT) provides the detailed and approximate components at even lower levels by using low pass filter component from higher level as input to wavelet transform at each subsequent level.
- MRWT enhances the accuracy of prediction.
- The MRWT is named on the basis of the wavelet transform algorithm being used repeatedly like Multi-Resolution Discrete Wavelet Transform (MRDWT) or Multi-Resolution Stationary Wavelet Transform (MRSWT).
- Irrespective of WT used, after application of MRWT a function can be represented as

$$f(t) = \sum_{k} a_{n,k}\varphi_{n,k}(t) + \sum_{j=1}^{n}\sum_{k} d_{j,k}\psi_{j,k}(t) \tag{8}$$

where $a_{n,k}$ is called the coarse or approximate coefficient and $d_{j,k}$ is called the detailed component or wavelet coefficient at level j. The n is the maximum level of decomposition of MRWT, k denotes the shift parameter of wavelet functions.

Scaling parameter

- Scale parameter (a) is inversely related to the wave frequency.
- High scale or low frequency corresponds to non-detailed global view (of the signal), and low scale or high frequency corresponds to detailed view.
- Dilate $(a > 1)$ or compress $(a < 1)$ mother wavelet function.

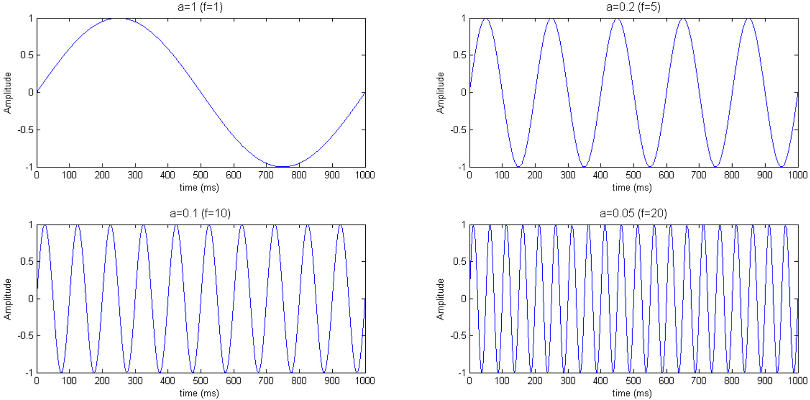
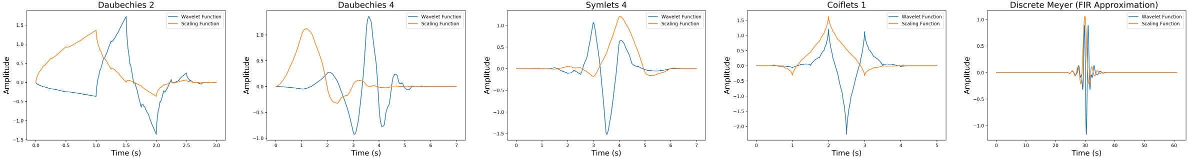


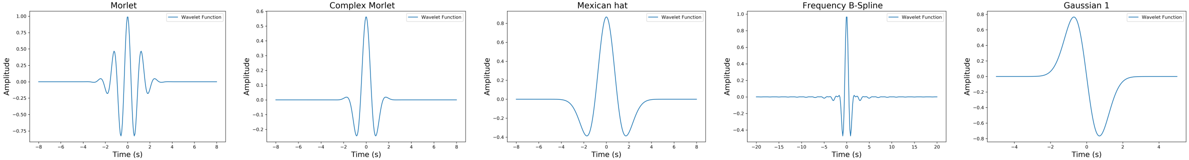
Figure 7: Graphical representation of sine signal of different scales.

Common wavelet functions

Discrete Wavelet Functions



Continuous Wavelet Functions



Kernel Estimator

- Non-parametric estimators for estimating probability distribution of a given data set.
- The components of hydrological time series are estimated as

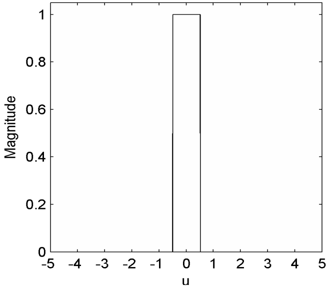
$$\tilde{a}_n(m+1) = \frac{\sum_{t=1}^{m} a_n(t) K\left(\frac{t-(m+1)}{h}\right)}{\sum_{t=1}^{m} K\left(\frac{t-(m+1)}{h}\right)} \tag{9a}$$

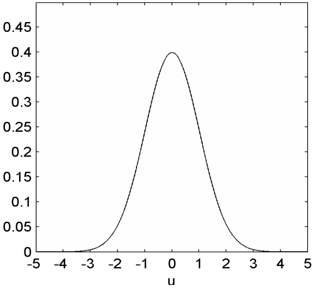
$$\tilde{d}_j(m+1) = \frac{\sum_{t=1}^{m} d_j(t) K\left(\frac{t-(m+1)}{h}\right)}{\sum_{t=1}^{m} K\left(\frac{t-(m+1)}{h}\right)} \tag{9b}$$

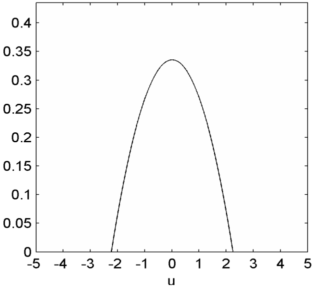
where K represents kernel function, and parameter h is the window of the kernel function. Naïve kernel function (Maity and Nagesh Kumar, 2008; Bosq, 2012) is used in this study.

Different Kernel Functions

Naïve	Normal	Epanechnikov
$K(u) = 1$ for $-0.5 \leq u \leq 0.5$	$K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-u^2}{2}\right)$ for $-\infty \leq u \leq \infty$	$K(u) = \frac{3}{4\sqrt{5}}\left(1 - \frac{u^2}{5}\right)$ for $-\sqrt{5} \leq u \leq \sqrt{5}$







ARX Model

$$A_{c,y}(t) = \sum_{k=1}^{p} S_k A_{c,y}(t-p) + \sum_{l=1}^{c}\sum_{j=0}^{q} T_{l,j} C_{x,l}(t-j) + E \tag{10}$$

where the number of auto-regressive terms and exogenous inputs are represented by p and q respectively.

$A_{c,y}(t)$ is t^{th} time step residual in any one component of the dependent variable Y,
C_x represents the set of selected individual components of the independent variable set X, and
c represent the number of members in set C_x or cardinal number of C_x.
$C_{x,l}$ represents l^{th} member of C_x.
S and T represent the set of regression parameters estimated during model calibration period.