Emerging Developments in Stochastic Hydrology and Their Importance in Water Resources Planning and Management

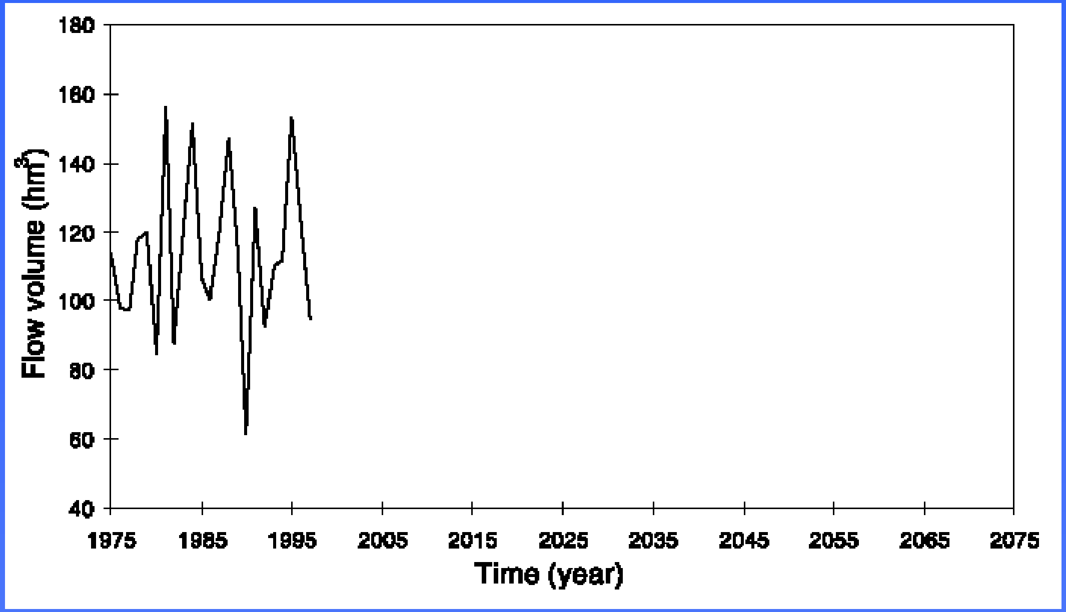
Nesa Ilich, Ph.D., P.Eng.

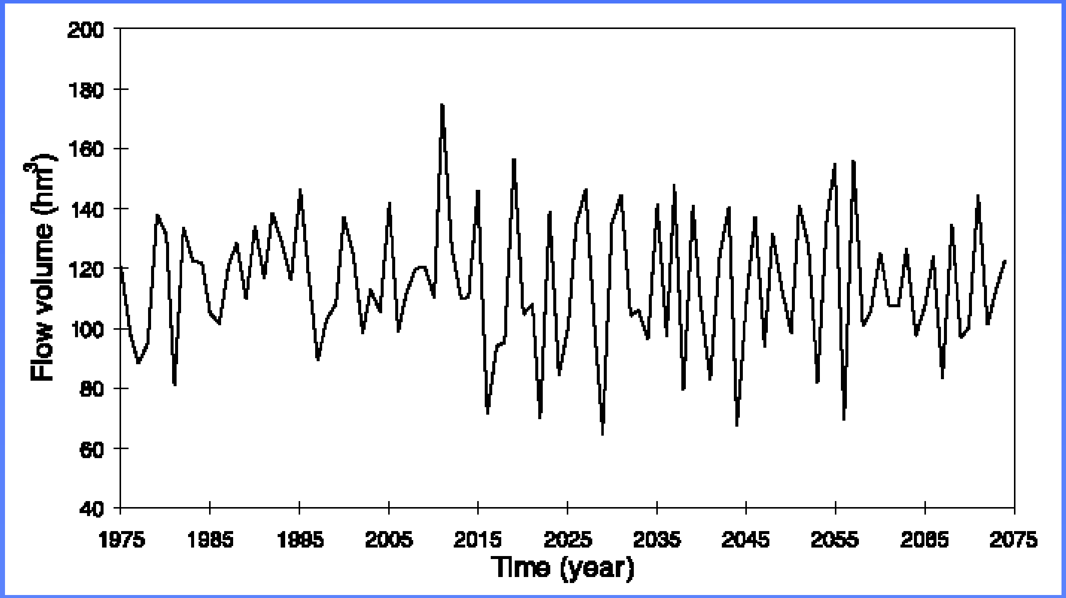
1. Brief Introduction to Stochastic Hydrology
2. Previous Developments
3. A New Approach
4. Case Studies
5. Conclusions and Recommendations

1. What is Stochastic Hydrology ?

1. Search for “Stochastic Hydrology” under Google returns 45,500 hits
2. A peer Reviewed Journal was named “Journal of Stochastic Hydrology and Hydraulics” (Springer publ.)
3. Definition: “Art of Mathematical Modeling of Spatial-Temporal Hydrologic Processes” (V. Yevdjevic, 1989)
4. First published attempt to model lag-1 auto-regressive monthly rainfall by Hannan in 1955
5. The ultimate goal of stochastic hydrology is a computer model that generates long hydrologic time series that are statistically similar to the historic series.

Stochastic Generation of Natural Flows





Continuous vs Single Event Modeling ?

1. Continuous modeling of lengthy time series of inflows is more appropriate for river basin management studies than single event modeling, since it addresses the uncertainty of starting storage levels;
2. Typical input into continuous models are historic natural flow series, which are of limited length.

How are Stochastic Series Used ?

1. As alternative input into river basin modeling studies, whether for water quantity or water quality;
2. In design and sizing of reservoirs in large complex systems;
3. For development of improved operating rules for complex water resources systems

Short History of Previous Developments

1. Thomas and Fiering (1962) developed lag 1 annual and seasonal model for stream flows;
2. Matalas (1967) and Yevdjevic (1972) developed multivariate lag-1 autoregressive models
3. Box and Jenkins (1970) introduced ARMA[p, q] class of models that had a large following in hydrology
4. Valencia and Schaake (1973) proposed disaggregation models to preserve both annual and seasonal statistics; etc.
5. More recent attempts rely on the use of Artificial Intelligence (Neural Networks, SVM, etc.)
6. Numerous books, research centers and hundreds of publications were generated in the last 50 years on this topic.

Current State of the Art (I)

1. Existing stream flow generation models (HEC-4, LAST, SPIGOT) typically work only with monthly time steps with assumed autoregressive lag of 1;
2. Models are not user friendly and usually not well documented;
3. They require intimate expert knowledge to setup and calibrate all parameters;
4. Some models may require manual intervention by the user to correct time series to fit the desired statistics or comply with the model constraints

$$Y_t = C_1 Y_{t-1} + C_2 + \sigma N[0, 1]$$

Current State of the Art (II)

5. A model that has gained wide acceptance among practitioners does not yet exist;
6. Some of the reasons are:

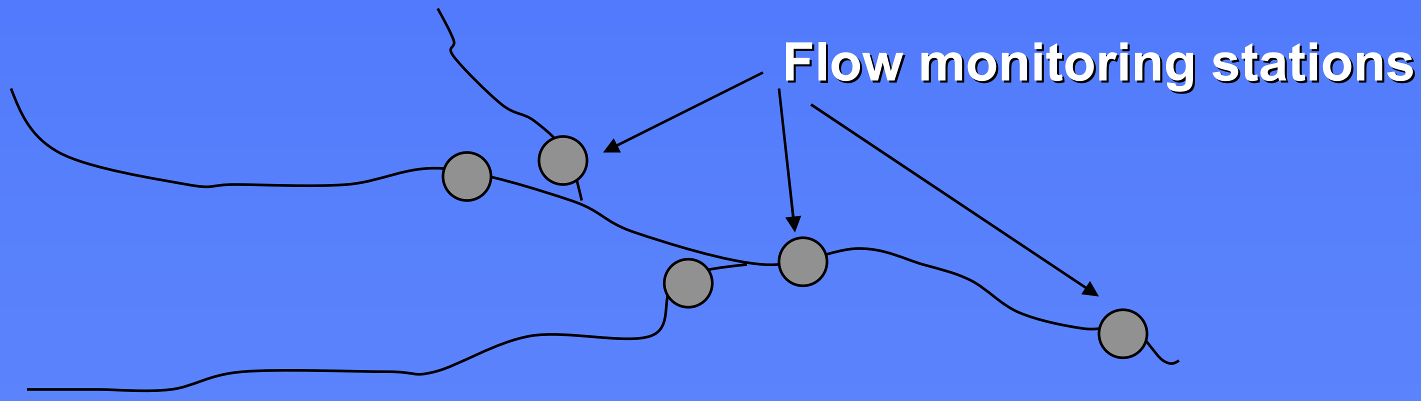
   - too much focus on the Box – Jenkins models, in spite of clear evidence that they cannot be implemented with smaller time steps, lags greater than 2 and on intermittent series

   - Difficult nature of hydrologic series, with
     - periodic and seasonal nature
     - intermittency and long memory
     - temporal and spatial statistical dependence

Spatial and Temporal Statistical Dependence

Cross-correlation indicates statistical dependence between various series. It can be expressed for various lags k (k = 0, n)

Auto-correlation indicates statistical dependence between flows in different time intervals for a single site. It varies seasonally, and depends on the length of time step.



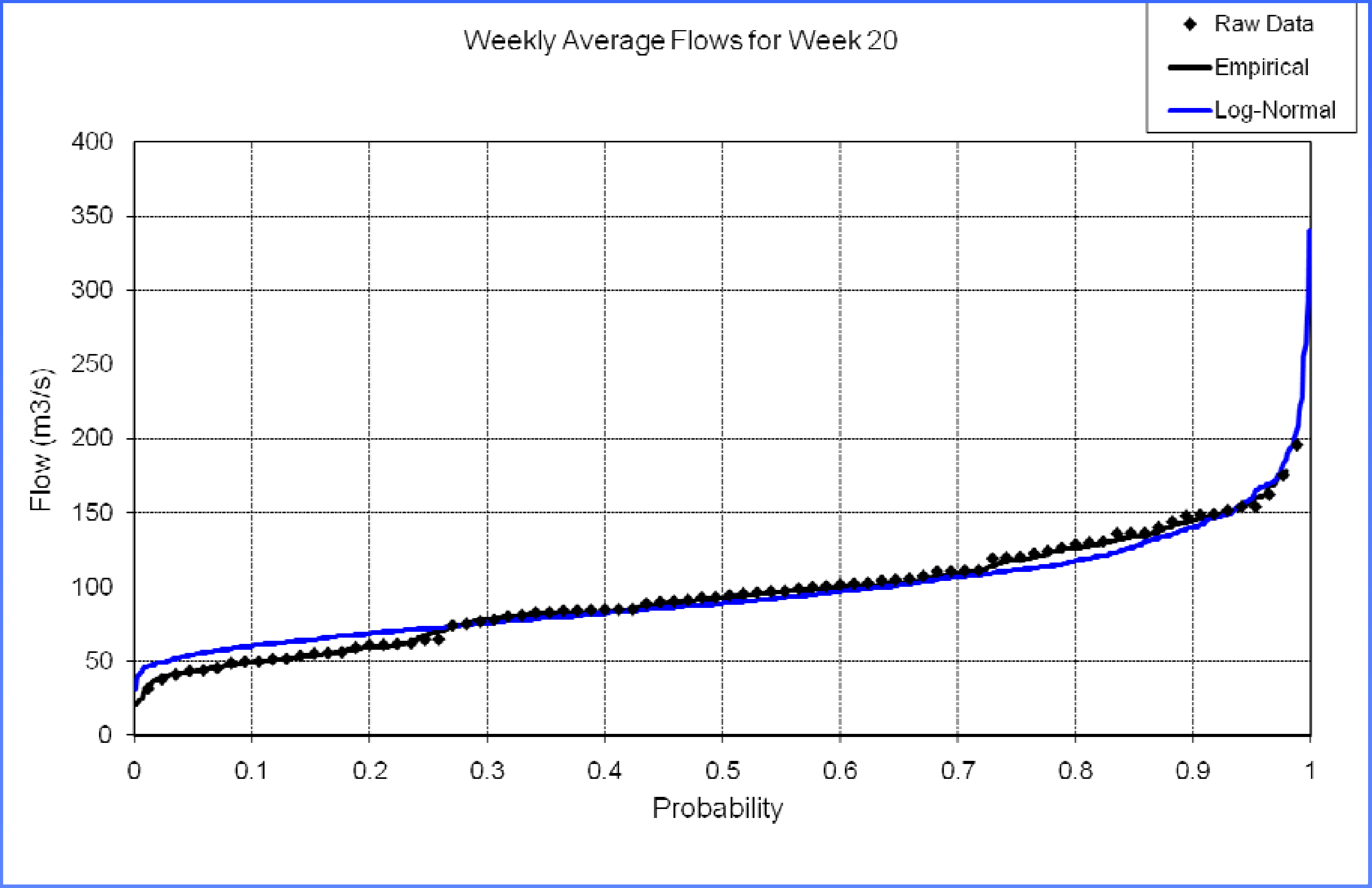
Stochastic Generation of Natural Flows

The relevant *weekly* statistics to be preserved are:

- **Annual mean;**
- **Annual standard deviation;**
- **Annual auto correlation;**
- **Annual cross-correlation between various stations;**
- **Weekly mean;**
- **Weekly standard deviation;**
- **Weekly auto correlation;**
- **Weekly cross-correlation between various stations; and,**
- **Weekly probability distribution function.**

Proposed Methodology

Step 1: Generate 1000 years of data for each week using an Empirical Kernel-type distribution



Stochastic Generation of Natural Flows

Step 2 consists of re-ordering position of the elements of columns 2, 3, etc. in a systematic way until the desired lag correlations and cross correlations are preserved.

Year	STATION 1					STATION 2					STATION 3				
	weeks					weeks					weeks				
	1	2	.	.	52	1	2	.	.	52	1	2	.	.	52
1932	$X_{1,1}$	$X_{1,2}$			$X_{1,52}$	$Y_{1,1}$	$Y_{1,2}$			$Y_{1,52}$	$Z_{1,1}$	$Z_{1,2}$			$Z_{1,52}$
1934	$X_{2,1}$	$X_{2,2}$			$X_{2,52}$	$Y_{2,1}$	$Y_{2,2}$			$Y_{2,52}$	$Z_{2,1}$	$Z_{2,2}$			$Z_{2,52}$
1935	$X_{3,1}$	$X_{3,2}$			$X_{3,52}$	$Y_{3,1}$	$Y_{3,2}$			$Y_{3,52}$	$Z_{3,1}$	$Z_{3,2}$			$Z_{3,52}$
1936	$X_{4,1}$	$X_{4,2}$			$X_{4,52}$	$Y_{4,1}$	$Y_{4,2}$			$Y_{4,52}$	$Z_{4,1}$	$Z_{4,2}$			$Z_{4,52}$
1937	$X_{5,1}$	$X_{5,2}$			$X_{5,52}$	$Y_{5,1}$	$Y_{5,2}$			$Y_{5,52}$	$Z_{5,1}$	$Z_{5,2}$			$Z_{5,52}$
1938	$X_{6,1}$	$X_{6,2}$			$X_{6,52}$	$Y_{6,1}$	$Y_{6,2}$			$Y_{6,52}$	$Z_{6,1}$	$Z_{6,2}$			$Z_{6,52}$
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
1999	$X_{10,1}$	$X_{10,2}$			$X_{10,52}$	$Y_{10,1}$	$Y_{10,2}$			$Y_{10,52}$	$Z_{10,1}$	$Z_{10,2}$			$Z_{10,52}$
2000	$X_{11,1}$	$X_{11,2}$			$X_{11,52}$	$Y_{11,1}$	$Y_{11,2}$			$Y_{11,52}$	$Z_{11,1}$	$Z_{11,2}$			$Z_{11,52}$
2001	$X_{12,1}$	$X_{12,2}$			$X_{12,52}$	$Y_{12,1}$	$Y_{12,2}$			$Y_{12,52}$	$Z_{12,1}$	$Z_{12,2}$			$Z_{12,52}$
2002	$X_{13,1}$	$X_{13,2}$			$X_{13,52}$	$Y_{13,1}$	$Y_{13,2}$			$Y_{13,52}$	$Z_{13,1}$	$Z_{13,2}$			$Z_{13,52}$

Stochastic Generation of Natural Flows

Step 2 ensures the following statistics are preserved:

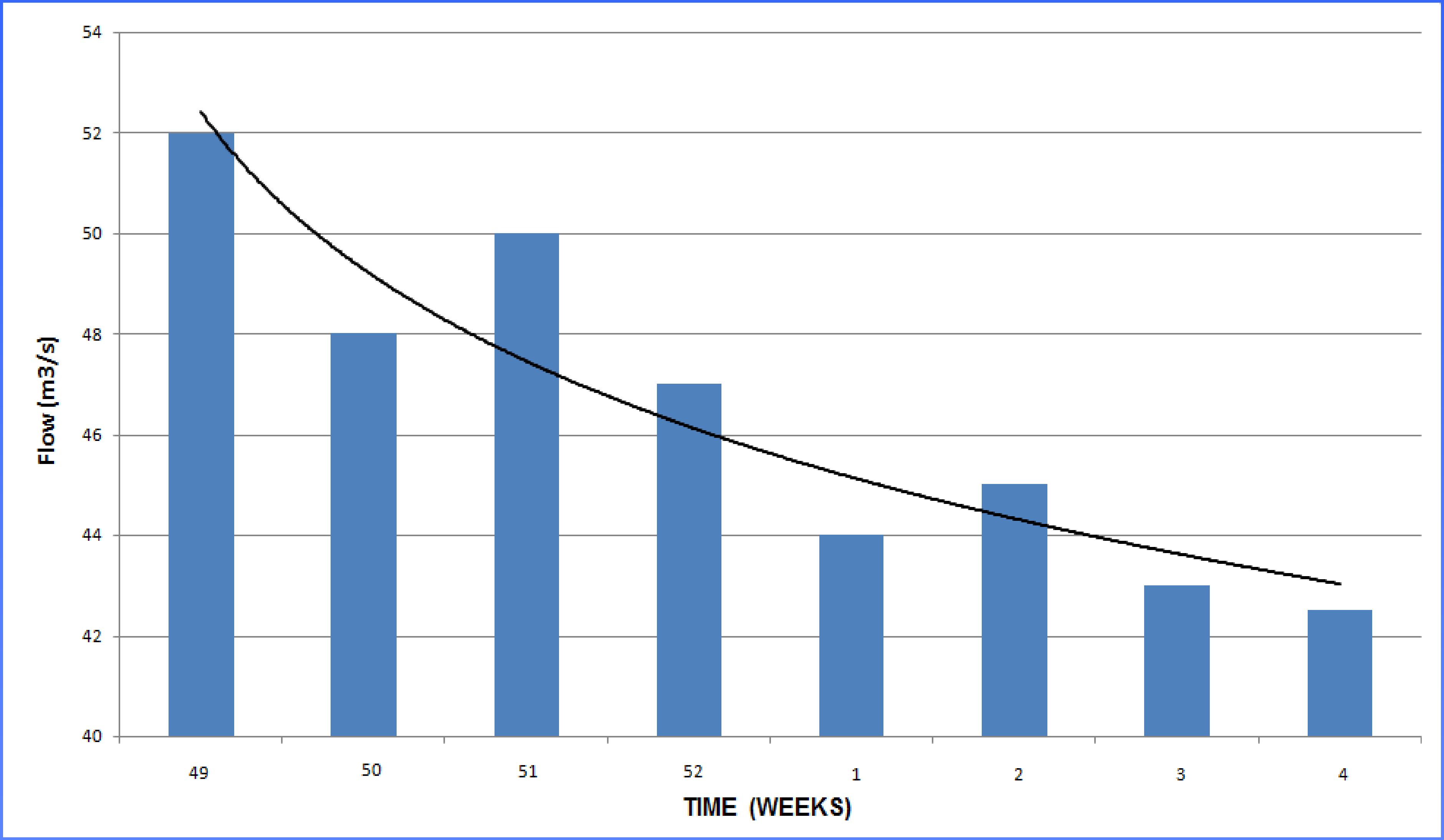
- **Annual mean;**
- **Annual standard deviation;**
- **Annual auto correlation;**
- **Annual cross-correlation between various stations;**
- **Weekly mean;**
- **Weekly standard deviation;**
- **Weekly auto correlation;**
- **Weekly cross-correlation between various stations; and,**
- **Weekly probability distribution functions.**

Stochastic Generation of Natural Flows

Step 3 consists of re-ordering of the entire rows in a systematic way until the desired annual lag correlations and the lag correlations between ending weeks of year *i-1* and starting weeks of year *i* are preserved.

Year	STATION 1					STATION 2					STATION 3				
	weeks					weeks					weeks				
	1	2	.	.	52	1	2	.	.	52	1	2	.	.	52
1932	$X_{1,1}$	$X_{1,2}$			$X_{1,52}$	$Y_{1,1}$	$Y_{1,2}$			$Y_{1,52}$	$Z_{1,1}$	$Z_{1,2}$			$Z_{1,52}$
1934	$X_{2,1}$	$X_{2,2}$			$X_{2,52}$	$Y_{2,1}$	$Y_{2,2}$			$Y_{2,52}$	$Z_{2,1}$	$Z_{2,2}$			$Z_{2,52}$
1935	$X_{3,1}$	$X_{3,2}$			$X_{3,52}$	$Y_{3,1}$	$Y_{3,2}$			$Y_{3,52}$	$Z_{3,1}$	$Z_{3,2}$			$Z_{3,52}$
1936	$X_{4,1}$	$X_{4,2}$			$X_{4,52}$	$Y_{4,1}$	$Y_{4,2}$			$Y_{4,52}$	$Z_{4,1}$	$Z_{4,2}$			$Z_{4,52}$
1937	$X_{5,1}$	$X_{5,2}$			$X_{5,52}$	$Y_{5,1}$	$Y_{5,2}$			$Y_{5,52}$	$Z_{5,1}$	$Z_{5,2}$			$Z_{5,52}$
1938	$X_{6,1}$	$X_{6,2}$			$X_{6,52}$	$Y_{6,1}$	$Y_{6,2}$			$Y_{6,52}$	$Z_{6,1}$	$Z_{6,2}$			$Z_{6,52}$
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
1999	$X_{10,1}$	$X_{10,2}$			$X_{10,52}$	$Y_{10,1}$	$Y_{10,2}$			$Y_{10,52}$	$Z_{10,1}$	$Z_{10,2}$			$Z_{10,52}$
2000	$X_{11,1}$	$X_{11,2}$			$X_{11,52}$	$Y_{11,1}$	$Y_{11,2}$			$Y_{11,52}$	$Z_{11,1}$	$Z_{11,2}$			$Z_{11,52}$
2001	$X_{12,1}$	$X_{12,2}$			$X_{12,52}$	$Y_{12,1}$	$Y_{12,2}$			$Y_{12,52}$	$Z_{12,1}$	$Z_{12,2}$			$Z_{12,52}$
2002	$X_{13,1}$	$X_{13,2}$			$X_{13,52}$	$Y_{13,1}$	$Y_{13,2}$			$Y_{13,52}$	$Z_{13,1}$	$Z_{13,2}$			$Z_{13,52}$

Reordering of synthetic years in Step 3



Case Study

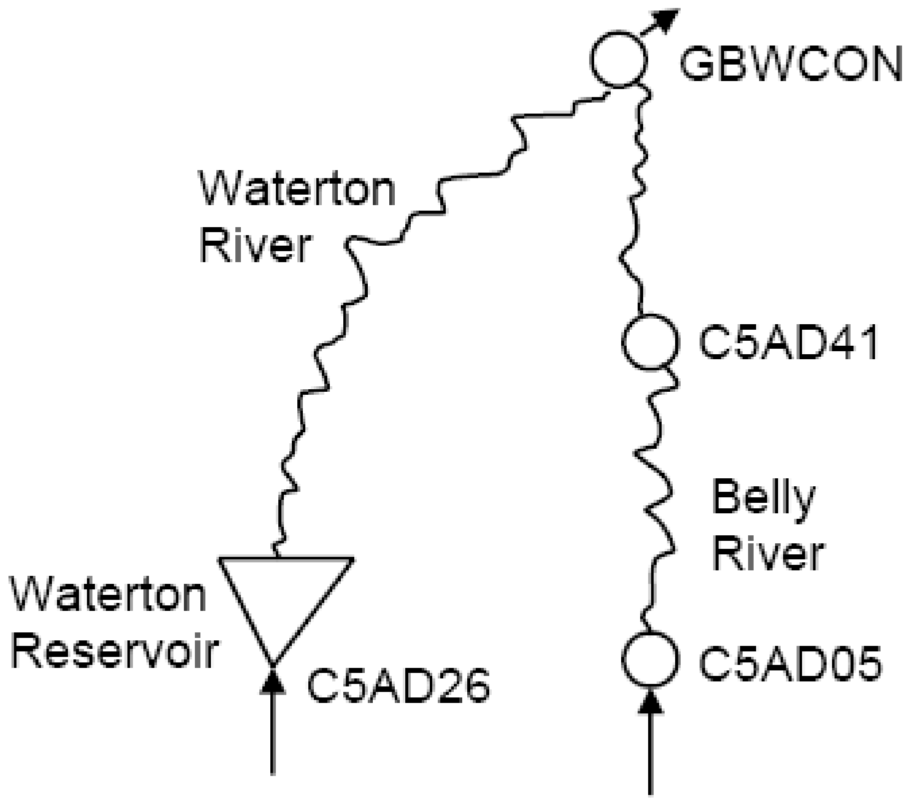


Figure 1. Locations of Selected Stations Relative to Each Other

Ilich, N. and J. Despotovic. 2008. A Simple Method for Effective Multi–Site Generation of Stochastic Hydrologic Time Series. Journal or Stochastic Environmental Research and Risk Assessment, Vol. 22 (2), p. 265-279

Case Study Results

	Summary of Simulated Annual Statistics			
Mean	34.342	9.511	9.069	22.246
Stdev	9.682	2.302	2.018	5.987
Skew	0.969	0.890	0.832	0.905

	Summary of Historic Annual Statistics			
Mean	34.336	9.510	9.069	22.247
StDev	9.796	2.305	1.995	6.030
Skew	1.092	0.693	0.369	0.838

Cross-Correlation Matrix of Simulated Annual Flows

	Station 1	Station 2	Station 3	Station 4
Station 1	1.000	0.962	0.936	0.985
Station 2		1.000	0.977	0.943
Station 3			1.000	0.944
Station 4				1.000

Cross-Correlation Matrix of Historic Annual Flows

	Station 1	Station 2	Station 3	Station 4
Station 1	1.000	0.966	0.942	0.988
Station 2		1.000	0.980	0.946
Station 3			1.000	0.945
Station 4				1.000

Case Study 2: Bow River Basin

Historic Series								
	Stn 1	Stn 2	Stn 3	Stn 4	Stn 5	Stn 6	Stn 7	Stn 8
Min	3.8	12.5	0.0	0.0	1.6	1.8	17.5	0.0
Avg	38.3	90.3	1.9	0.6	19.4	9.2	124.3	1.1
Max	254.9	512.7	49.2	31.9	500.4	128.1	1066.1	44.3
Stdev	44.7	89.7	3.8	1.5	38.4	9.5	127.7	2.5

Generated Series								
	Stn 1	Stn 2	Stn 3	Stn 4	Stn 5	Stn 6	Stn 7	Stn 8
Min	2.2	7.3	0.0	0.0	0.9	1.0	10.2	0.0
Avg	38.3	90.2	1.9	0.6	19.5	9.2	124.4	1.1
Max	454.1	950.3	75.8	44.6	990.5	186.9	1616.9	63.2
Stdev	46.3	92.5	4.0	1.6	41.7	10.1	133.8	2.7

Historic Series							
	Stn 9	Stn 10	Stn 11	Stn 12	Stn 13	Stn 14	Stn 15
Min	11.3	0.5	1.1	12.6	13.0	12.6	1.8
Avg	82.7	6.6	7.3	92.6	93.3	92.9	9.0
Max	472.6	201.9	40.1	543.3	545.0	544.4	122.2
Stdev	84.7	14.3	5.8	92.3	92.8	92.6	9.3

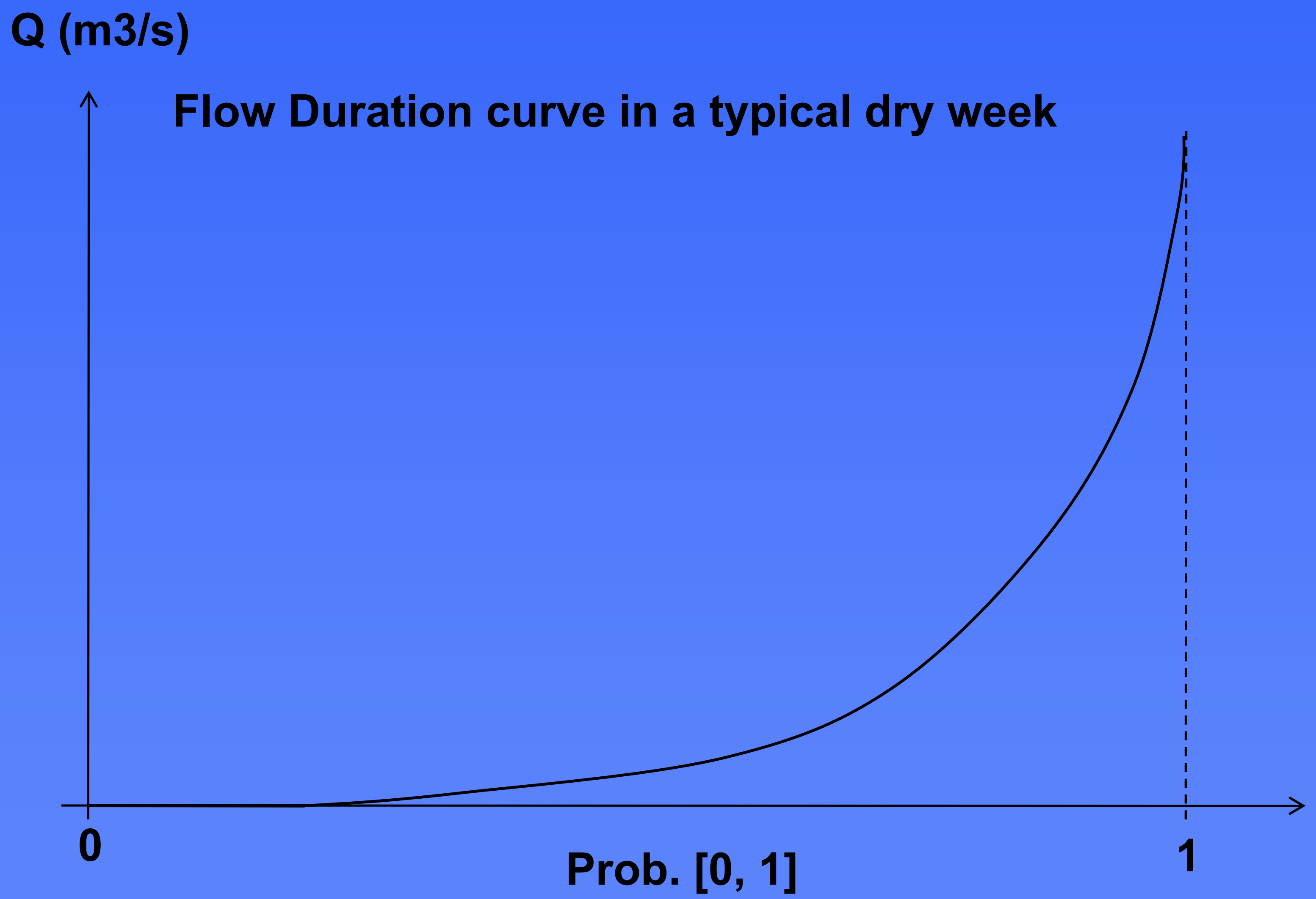
Generated Series							
	Stn 9	Stn 10	Stn 11	Stn 12	Stn 13	Stn 14	Stn 15
Min	6.5	0.3	0.7	7.4	7.5	7.4	1.0
Avg	82.6	6.7	7.3	92.5	93.2	92.8	9.0
Max	859.9	372.3	68.4	970.9	971.8	971.2	181.2
Stdev	87.3	15.5	6.2	95.5	96.0	95.8	10.0

Case Study 2: Bow River Basin

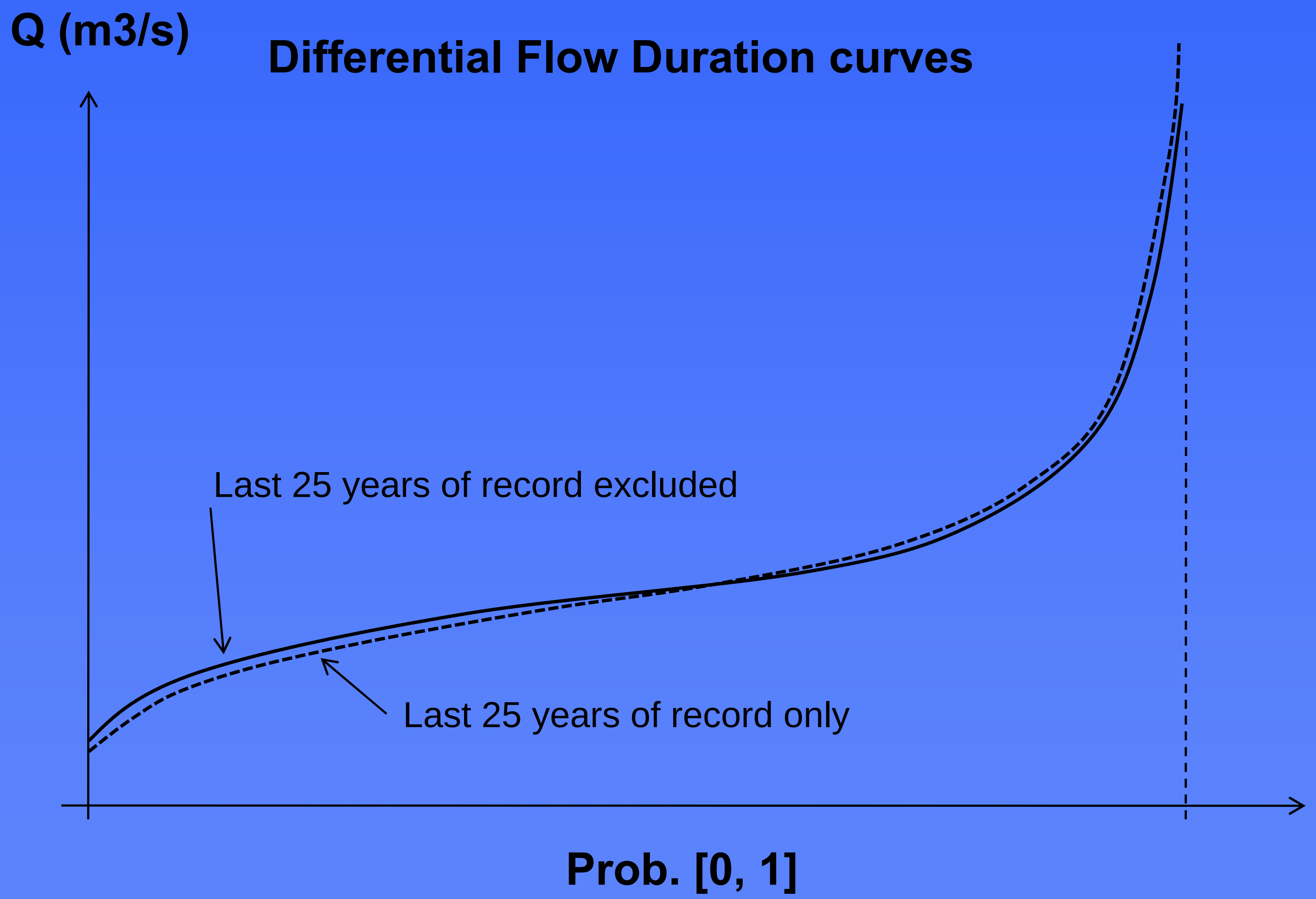
Historic Pearson Cross-Correlations															
	Stn 1	Stn 2	Stn 3	Stn 4	Stn 5	Stn 6	Stn 7	Stn 8	Stn 9	Stn 10	Stn 11	Stn 12	Stn 13	Stn 14	Stn 15
Stn 1	1.000	0.949	0.477	0.185	0.369	0.577	0.848	0.231	0.954	0.370	0.710	0.942	0.940	0.941	0.575
Stn 2		1.000	0.619	0.276	0.479	0.732	0.939	0.315	0.999	0.471	0.833	0.999	0.999	0.999	0.731
Stn 3			1.000	0.507	0.607	0.891	0.746	0.585	0.605	0.545	0.694	0.649	0.651	0.651	0.892
Stn 4				1.000	0.466	0.521	0.436	0.530	0.262	0.448	0.360	0.300	0.311	0.307	0.512
Stn 5					1.000	0.664	0.741	0.676	0.477	0.972	0.402	0.496	0.502	0.499	0.658
Stn 6						1.000	0.851	0.625	0.718	0.614	0.765	0.754	0.759	0.757	1.000
Stn 7							1.000	0.534	0.936	0.721	0.806	0.948	0.950	0.949	0.848
Stn 8								1.000	0.307	0.636	0.338	0.338	0.345	0.342	0.618
Stn 9									1.000	0.470	0.809	0.997	0.997	0.997	0.717
Stn 10										1.000	0.381	0.486	0.492	0.489	0.606
Stn 11											1.000	0.841	0.842	0.842	0.767
Stn 12												1.000	1.000	1.000	0.753
Stn 13													1.000	1.000	0.757
Stn 14														1.000	0.756
Stn 15															1.000

Generated Pearson Cross-Correlations															
	Stn 1	Stn 2	Stn 3	Stn 4	Stn 5	Stn 6	Stn 7	Stn 8	Stn 9	Stn 10	Stn 11	Stn 12	Stn 13	Stn 14	Stn 15
Stn 1	1.000	0.934	0.466	0.170	0.334	0.554	0.830	0.206	0.942	0.333	0.683	0.928	0.926	0.927	0.553
Stn 2		1.000	0.603	0.246	0.432	0.705	0.928	0.298	0.995	0.419	0.813	0.996	0.995	0.996	0.704
Stn 3			1.000	0.463	0.561	0.841	0.718	0.538	0.590	0.498	0.673	0.628	0.628	0.628	0.845
Stn 4				1.000	0.428	0.448	0.374	0.485	0.237	0.426	0.293	0.260	0.266	0.263	0.440
Stn 5					1.000	0.634	0.647	0.636	0.431	0.925	0.383	0.448	0.451	0.450	0.629
Stn 6						1.000	0.821	0.611	0.692	0.578	0.729	0.724	0.726	0.726	0.995
Stn 7							1.000	0.487	0.924	0.620	0.788	0.938	0.940	0.939	0.819
Stn 8								1.000	0.290	0.601	0.340	0.317	0.320	0.319	0.603
Stn 9									1.000	0.419	0.788	0.993	0.992	0.992	0.691
Stn 10										1.000	0.355	0.433	0.436	0.435	0.573
Stn 11											1.000	0.821	0.822	0.822	0.732
Stn 12												1.000	0.998	0.999	0.723
Stn 13													1.000	0.999	0.725
Stn 14														1.000	0.725
Stn 15															1.000

Possible Extensions:
1. Intermittent Flows



2. Incorporating Climate Change Impacts



3. Additional Extensions

- The algorithm can model non-linear correlations using the standard transform function ln(Q+1)
- The algorithm could incorporate the estimates of annual flow series based on the tree rings data to generate weekly flow series that match both the 90 year weekly as well as 1000 year annual statistics
- With minor modifications the algorithm could be used to provide in-filling of missing data within existing series where there is strong multiple cross correlation with nearby stations that have data

Create and Verify a Basin Operational Model

Develop Input Data based on historic records

↓

Replace the historic data series by longer stochastic series

↓

Solve the basin allocation problem in multiple time step mode

↓

Train AI (pattern recognition) model

↓

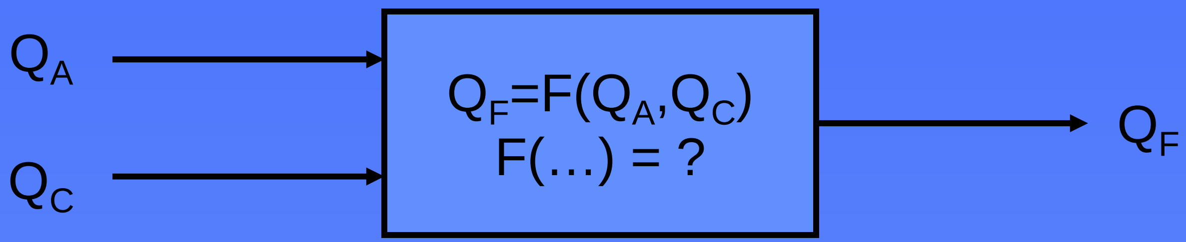
Verify AI model (sequence of single time step runs)

↓

Compare the output from the last step with the historic data to quantify the potential benefits

3. AI (Pattern Recognition Algorithms)

- Used widely in various industries
- Available as DLL libraries
- Require knowledge on the nature of the algorithm (i.e. required input and possible output) and the knowledge of the problem domain

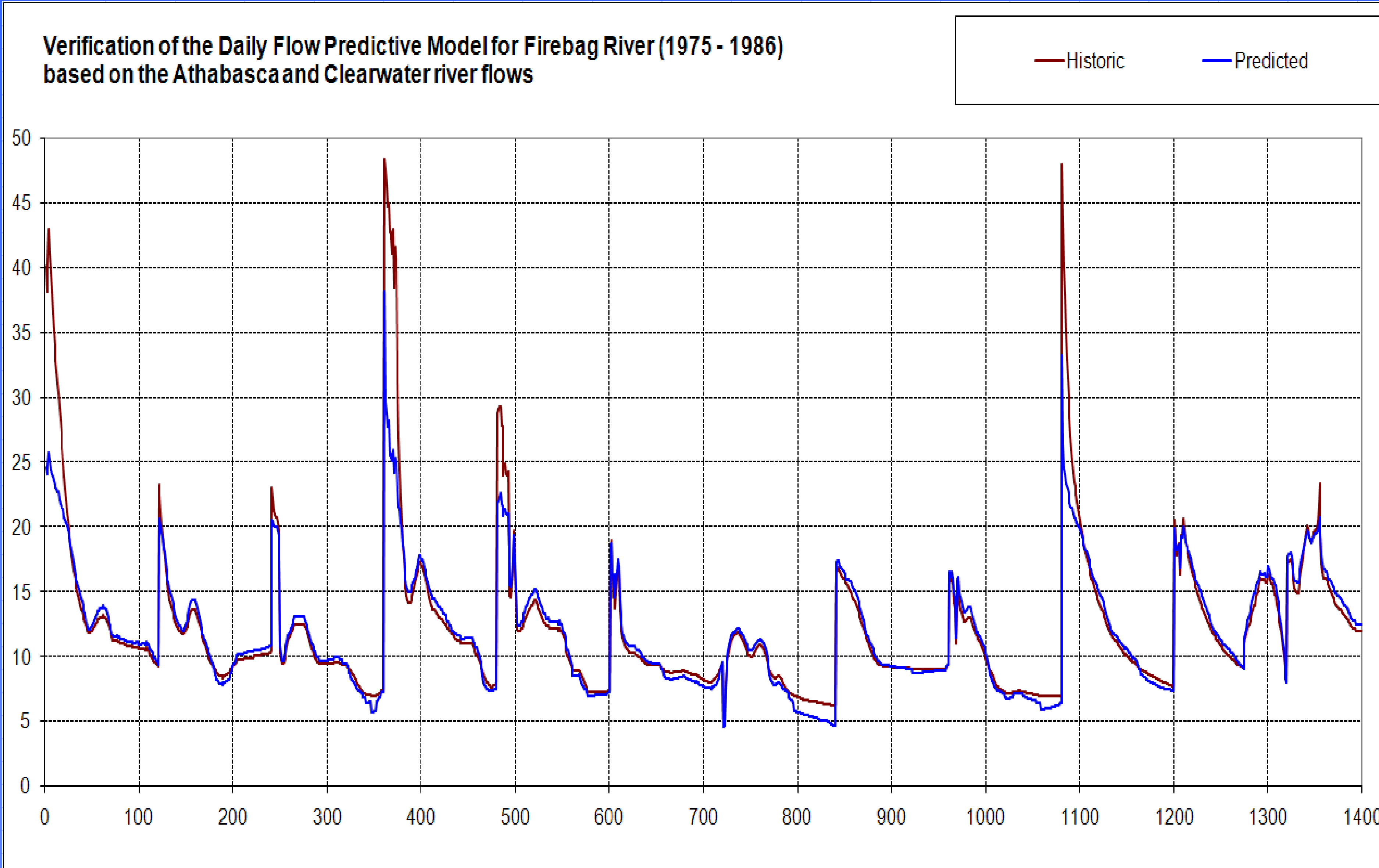


Q_A – Athabasca River at Fort McMurray (Q_{avg}=628 m^3/s)

Q_C – Clearwater River at Draper (Q_{avg}=119 m^3/s)

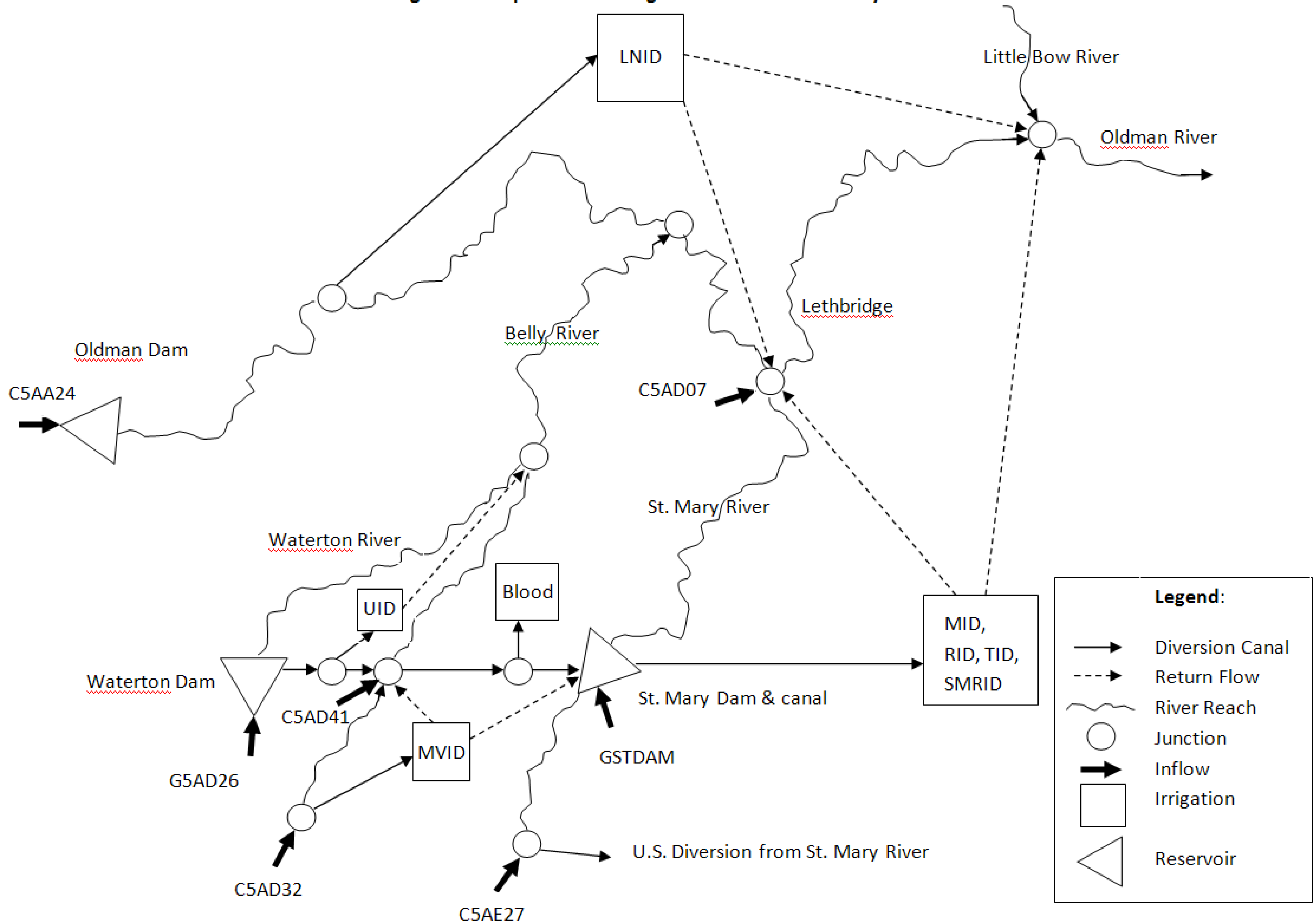
Q_F – Firebag River near the mouth (Q_{avg}=26 m^3/s)

In-filling of Missing Data using AI algorithms



R & D funded by AWRI: Basin Operational Model

Figure 1. Proposed Modeling Schematic of the Test System



The End