Czech Technical University in Prague
Faculty of Civil Engineering
Department of Hydraulic Structures

DIPLOMA THESIS

Modeling of synthetic multivariate hydrological series

Modelování syntetických hydrologických řad v systému stanic

Main thesis supervisor: doc. Dr. Ing. Pavel Fošumpaur

January 2018 Waldemar Grešík

ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE
Fakulta stavební
Thákurova 7, 166 29 Praha 6

ZADÁNÍ DIPLOMOVÉ PRÁCE

I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: Bc. Grešík Jméno: Waldemar Osobní číslo: 409838

Zadávající katedra: Katedra hydrotechniky

Studijní program: Stavební inženýrství

Studijní obor: Vodní hospodářství a vodní stavby

II. ÚDAJE K DIPLOMOVÉ PRÁCI

Název diplomové práce: Modelování syntetických hydrologických řad v systému stanic

Název diplomové práce anglicky: Modeling of synthetic mutlivariable hydrological series

Pokyny pro vypracování:
Diplomová práce se v úvodní části zaměří na rešerši soudobých metodických postupů pro modelování syntetických hydrologických řad v systému stanic pro potřeby stochastického řešení zásobní funkce nádrží a vodohospodářských soustav. Do hodnocení bude zařazena také metoda hlavní komponenty (Principal Component Analysis) a novější metoda nezávislé komponenty (Independent Component Analysis). Na podkladě získaných poznatků bude sestaven matematický model náhodných průtokových řad v měsíčním kroku, který bude verifikován pro vybranou případovou studii. Model umožní zachování shody pravděpodobnostních vlastností reálných a syntetických průtokových řad včetně autokorelačních vazeb a křížových korelací mezi jednotlivými profily.

Seznam doporučené literatury:
Nacházel, K.: Stochastické metody ve vodním hospodářství. Vydavatelství ČVUT, 2000.
Nacházel, K.: Estimation Theory in Hydrology and Water Systems. Academia, Prague, 1993.
Hipel, K.W., McLeod, A.I.: Time Series Modelling of Water Resources and Environmental Systems. Elsevier Science Pub Co, 1994.

Jméno vedoucího diplomové práce: doc. Dr. Ing. Pavel Fošumpaur

Datum zadání diplomové práce: 22.6.2017 Termín odevzdání diplomové práce: 7.1.2018
Údaj uveďte v souladu s datem v časovém plánu příslušného ak. roku

[redacted]
Podpis vedoucího práce

[redacted]
Podpis vedoucího katedry

III. PŘEVZETÍ ZADÁNÍ

Beru na vědomí, že jsem povinen vypracovat diplomovou práci samostatně, bez cizí pomoci, s výjimkou poskytnutých konzultací. Seznam použité literatury, jiných pramenů a jmen konzultantů je nutné uvést v diplomové práci a při citování postupovat v souladu s metodickou příručkou ČVUT „Jak psát vysokoškolské závěrečné práce“ a metodickým pokynem ČVUT „O dodržování etických principů při přípravě vysokoškolských závěrečných prací“.

22. 6. 2017
Datum převzetí zadání

[redacted]
Podpis studenta(ky)

Declaration

I hereby, declare that this thesis is my own work and that, to the best of my knowledge and belief, it contains no material which has been accepted or submitted for the award of and other degree or diploma. I also declare that, to the best of my knowledge and belief, this thesis contains no material previously published or written by any other person except where due reference is made in the text of the thesis.

…………………………………

…………………………..

In Cork, Ireland   Waldemar Grešík

ACKNOWLEDGEMENTS

I would like to thank the academy. And also my supervisor for guidance.

ABSTRACT

Based on its applicability and features, two approaches for multivariate time series modelling were discussed. The first, method based Principal Component Analysis is much more simple and direct method, having the advantage of closed form computational processes and therefore holding much smaller computational burden. Its disadvantage is, that it theoretically destroys part of the mutual information that the multivariate data contain, because it preserves only raw mutual correlations between stations but not higher order dependencies. The basis is that it searches for transformation that has been designed based on the covariance matrix, which is a low order statistical characteristic of data. The second, method based on Independent component analysis, theoretically preserves even those higher order dependencies, because it extracts from the data more mutual information and is therefore able to reapply this information to independent univariate synthetic time series that were generated individually.

The practical part of this thesis involved construction of the PCA method based multivariate model and evaluation of its performance. Regarding the preservation of the correlation structure the model performed arguably quite well, having total error as a performance measure explained in section 7.4 around 3.40% for the autocorrelation structure of lag 1 of the data set and total error of 7.86% for the cross-correlation structure describing mutual relationships of the multivariate data.

In traditional applications of streamflow data the generated time series did not deviate extensively from expected outcomes, making the model’s output usable in some classical water management solutions. However, there were some drawback of the model’s performance especially in water reservoir operation solutions, where the model produced data that underestimated storage capacity requirements for longer time series.

KEY WORDS

Synthetic time series, stochastic models in hydrology, modelling of time series,

CONTENT

1 LAY SUMMARY

Mathematical modeling of synthetic time series is nowadays basic method used during the design process of many water management engineering solutions of different kinds. We use this method in cases, where the available historical time series are not sufficient for proper design solutions of engineering problems. This is usually because the available historical time series are too short and therefore produce unreliable assumptions about future behaviour of the phenomenon they represent. Mathematical modelling enables us to create much longer synthetic time series, based on the historical real data, with the same sets of statistical characteristics. It provides us with more reliable data usable for more complex problems, and data that meet requirements on input materials demanded by the present legislation. [38]

There is fundamental difference between simulation of single time series, called univariate, and simulation of sets of inter-related time series together, called multivariate time series modelling. While the methodical approach for univariate time series modelling has been soundly established over many years of its research and application in the last century, modelling of multivariate time series is still experiencing intensive development and completely new methods have been introduced in recent decades, mainly thanks to advance in computer technologies. The problem of accurately simulating multivariate time series along with their inner dependence and mutual information stored within is non-trivial problem and more complex solution must be employed. [19]

Two main methods broadly used for this problem, Principal Component Analysis and Independent Component Analysis are going being described and compared in this thesis. They both aim to discover and describe mathematically the spatial and temporal dependence within the historical data and based on that to design a model which will maintain the dependence relationships in the new modelled time series. While the first method is older and its computational difficulty is much smaller, it will be shown, that the latter, much younger method of these two, goes further with the analysis of the dependence and therefore preserves the statistical characteristics better. [21]

2 GOALS OF THIS THESIS

The goals of this thesis are to summarize currently used methodological approaches for modelling of synthetic hydrological time series in systems of stations for purposes of solving water storage function of water reservoirs and watercourse systems and to investigate into new methods recently emerged or still emerging in water management engineering. Main focus will be given to methods of Principal Component Analysis and Independent Component Analysis, whereas this thesis aims to evaluate their usability and possible application and to assess the pros and cons in their employment on multivariate hydrological time series.

Secondly, this thesis aims to produce mathematical model for generation of random monthly averages streamflow time series and verify the model for chosen case study. The model will allow preservation of correspondence of probability characteristics of the real and the synthetic streamflow time series including month-wise autocorrelations and cross-correlation structure among individual profiles.

3. INTRODUCTION

Mathematical modeling of synthetic time series is nowadays basic method used during the design process of many water management engineering solutions of different kinds. We use this method in cases, where the available historical time series are not sufficient for proper design solutions of engineering problems. This is usually because the available historical time series are too short and therefore produce unreliable assumptions about future behaviour of the phenomenon they represent. Mathematical modelling enables us to create much longer synthetic time series, based on the historical real data, with the same sets of statistical characteristics. It provides us with more reliable data usable for more complex problems, and data that meet requirements on input materials demanded by the present legislation. [38]

There is fundamental difference between simulation of single time series, called univariate, and simulation of sets of inter-related time series together, called multivariate time series modelling. While the methodical approach for univariate time series modelling has been soundly established over many years of its research and application in the last century, modelling of multivariate time series is still experiencing intensive development and completely new methods have been introduced in recent decades, mainly thanks to advance in computer technologies. The problem of accurately simulating multivariate time series along with their inner dependence and mutual information stored within is non-trivial problem and more complex solution must be employed. [19]

Two main methods broadly used for this problem, Principal Component Analysis and Independent Component Analysis are going being described and compared in this thesis. They both aim to discover and describe mathematically the spatial and temporal dependence within the historical data and based on that to design a model which will maintain the dependence relationships in the new modelled time series. While the first method is older and its computational difficulty is much smaller, it will be shown, that the latter, much younger method of these two, goes further with the analysis of the dependence and therefore preserves the statistical characteristics better. [21]

4. THEORETICAL BACKGROUND

4.1 Basic terms

At first it is desirable to define basic terms used in this thesis and the way in which hydrological time series are mathematically approached here. When considering time series in hydrology, they can be viewed as discrete random variable samples. The discharge in a river is a *stochastic process* which we cannot accurately and fully describe mathematically, we can only observe its *realizations*, whereas discharge in a river has always only one realization. All the values of the realization form a *population* and by measuring the discharge we take values from the population and we create a *random sample* - values are drawn from *sample space*, which is set of all possible outcomes of the process (for discharge of a river the sample space is infinite set of real numbers).[2, 3]

We also say that we are recording observations of the realization of the process. The set of observations in chronological order is then referred to as *time series*. One times series accounts for one set of observations of one realization and is represented by one random variable q_i. When conducting time series analysis, we are trying to estimate the characteristics of the population, but because only a sample of the population is available to us, we do that by computing characteristics of the sample. The characteristics of the population are called *statistical parameters*, and we are estimating them with *statistics* or *estimators* such as *sample mean*.[3, 4] Therefore in this thesis, whenever *mean*, *standard deviation* or any statistical characteristics is mentioned, it is always referred to *statistic*, characteristic of the sample, unless it is stated otherwise.

Multivariate time series are sets of observations of more than one variable, such as flow rates at different measuring stations over the same period of time, which is also the case of this thesis. Each station is represented by one time series, mathematically a vector of certain length, but when conducting mathematical operations with the time series, we sometimes view it as a random variable.

The observed time series are referred to as real or historical time series, while realizations of the process generated by a mathematical model are called artificial or synthetic time series. When modeling a synthetic time series, the goal is to find a model which bets fits the real historical time series. When the model is fitted to an existed realization of a stochastic process with reasonable accuracy, it should possess the same statistical properties as the process that generated the existent realization. This procedure is called *time series analysis*.[19] The model can be afterwards used to simulate other possible realizations during the same period of time or generate longer time series for other applications.

4.2 Time series modeling in water management

Time series modeling is an essential discipline in hydrological forecasting. When designing water reservoir or any water structure, future forecast of hydrological events are necessary inputs in water management. These events are random in nature and therefore we must equip ourselves with statistics and probability theory to bring ourselves closer to accurate forecasts of these events.

Sometimes, for the design of a water structure, will suffice to predict the frequency of occurrence of certain hydrological event, using statistical method called frequency analysis. With this method, we can predict for example how often can we expect flood of certain magnitude, or rather to predict the maximum flood magnitude in certain period of time, typically 100, 1000 or 5000 years, generally known as N-year events.[25]

In some other cases, like assessing proper functioning of water reservoir or water systems meeting required goals, it is necessary to take into consideration the actual development of hydrological event over a period of time, looking at the whole sequence, not only the sequence's extremes. Typical example of required data is set of monthly averages of discharge in a catchment over next number of years (again: 100, 1000, 5000, etc.). In this case, not only the magnitude of the values matters, but their order in the sequence matters too. This type of hydrological forecasting can be made by methods called synthetic data generation, also referred to as time series modeling.[39]

While some methods of hydrological forecasting, for example frequency analysis, considers only certain values from historical records, like the extreme discharge magnitudes and the frequency of their occurrence, time series modeling considers in its process the whole continuous sets of observations of a given phenomenon. This makes it more difficult to select reliable records for this method, but it also means the forecasts made by times series modeling are more complex and more plausibly describe the reality. Moreover, when properly modelled, synthetic data series can be used to determine N-year events as well, simply by observation of the time series.[25], [32]

Hydrological forecasting techniques based of streamflow time-series modelling enables engineers and researchers examine possible scenarios for water resources systems with chronological context to it. Synthetic time series can simulate possible behaviour of a water system with realistic succession of events and their context within the continuous development of flow-rate, which is especially important in periods of drought or on the contrary during high-flow periods. This is essential in tasks like designing optimal operation of water systems, irrigation systems, water supply systems, systems upon which hydroelectric power stations rely, flood management planning or risk assessment of reliability of water systems. Variety of generated scenarios and diversity of conditions can largely contribute to creation of more efficient solutions of these tasks and perfected strategies. [40]

4.3 Statistical characteristics of a time series

The properties of a time series are described through statistical moments and other characteristics. These characteristics are needed construct a mathematical model and then to measure its goodness of fit.

This section of the thesis aims to clarify the usage of basic statistical and mathematical terms and expressions, rather than explain to detail their mathematical definitions. It is for the purpose of not confusing some fundamental statistical elements as the terminology is not always uniform in all publications dedicated to this topic. In addition, mathematical symbols and operators used in this thesis are explained to a reasonable extent in the Glossary at the end of the thesis.

4.3.1 Statistical moments

Mean

When it a mean of a variable is being mentioned in this thesis, it is always referred to mean of its sample, *sample mean*, conventionally denoted by $\bar{x}$, not a mean of its population, which is usually expressed as μ_x and is being computed from the whole population of a variable, which is something never available to us in hydrology, and therefore it is never the case here either. Also, it is always referred to *arithmetic mean*, computed as in Equation 4.1, where n is the number of elements in the sample.[30]

$$\bar{x} = \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right) \tag{4.1}$$

Standard deviation and Variance

Both *standard deviation* and *variance* express the spread, or dispersion of values of a variable from its mean. Standard deviation is being expressed here as σ_x, and variance, being its square as σ_x^2. Without the availability of the whole population, standard deviation and variance are being computed by expressions Equation 4.1. Because this standard deviation is computed using the sample mean $\bar{x}$, it is therefore sometimes called *uncorrected sample standard deviation*.[30]

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}};\ \ \sigma_x^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} \tag{4.1}$$

Skewness

The skewness measures the lack of symmetry of a distribution. It is being denoted by κ and expressed through standard deviation σ, as in Equation 4.1.[52] κ

$$\kappa_x = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{\sigma_x^3} \tag{4.1}$$

Skewness can be used to determine normality of a distribution (Normal distribution is symmetric about the mean and therefore has a skewness of zero), wherefore it has a use in this thesis in determining effectiveness of normalization transforms used. For example D'Agostino's K-squared test is a technique commonly used to measure the departure of a distribution from normality through skewness and kurtosis. [11]

4.3.2 Probability density function and Cumulative distribution function

Cumulative distribution function (CDF) and Probability density function (PDF), are two closely related statistical tools, but should not be confused with each other. The first, PDF, describes the probability $\mathbb{P}$ of a variable x taking certain values a – it describes probability of its value distribution. Here, PDF is being expressed as in Equation 4.1. [30]

$$p(x) = \mathbb{P}(x = a) \tag{4.1}$$

The second, CDF, also called probability distribution function or cumulative frequency function, describes probability of a variable taking values less then certain values as in Equation 4.1.[30]

$$F(x) = \mathbb{P}(x \leq a) \tag{4.1}$$

The direct relation between PDF and CDF is, that PDF is CDF's derivative:[30]

$$f(x) = F'(x) \tag{4.1}$$

4.3.3 Covariance and correlation coefficient

Covariance is understood as a measurement of the linear relationship between two variables, for example x and y, and it can be expressed by a single value computed by the Equation 4.1. [52]

$$\sigma_{x,y} = cov(x, y) = E[(x - E\{x\})(y - E\{y\})] \tag{4.1}$$

When mentioning correlation coefficient here, it is always referred to standard Pearson product-moment correlation coefficient, sometimes shortly PCC, which is expressed through covariance of two variables divided by the product of their standard deviation. Therefore they both describe linear relationships between variables, but the difference between PCC and covariance is, that PCC is scaled and takes only values between -1 and 1, while covariance can take any ℝeal values. Positive or negative values of PCC signify positive or negative correlation respectively, when PCC equals zero, there is no correlation between the two examined variables at all. The Equation 4.1 expresses computing PCC for two different variables.[20]

$$\rho_{x,y} = \frac{cov(x, y)}{\sigma_x \sigma_y} \tag{4.1}$$

4.3.4 Partial autocorrelation function

Autocorrelation function, also called serial correlation, describes correlation of a signal with latter itself. It is computed as in Equation 4.1, between observations separated by k time intervals. [19], [53]

$$\rho_k = \rho_{x_t, x_{t+k}} = \frac{E[(x_t - E\{x\})(x_{t+k} - E\{x\})]}{\sigma_{x_t}\sigma_{x_{t+k}}} = \tag{4.4}$$

Partial autocorrelation function (PACF) is then expressing the development of autocorrelation function at different time lags k. It is very important tool for deciding up to which extent the is a time series autocorrelated. It can be then used to determine suitable order of autoregressive model.

4.3.1 Covariance matrix of multivariate and univariate time series

Mutual covariances of n variables of a vector $\mathbf{x}$ are being expressed by the *covariance matrix* of $n \times n$ dimensions, denoted by $\mathbf{\Sigma_x}$ expressed by Equation 4.2, or by more compact formulation in Equation 4.3. This can represent covariance matrix of multivariate time series where every variable x_i stands for one time series or more specifically one streamflow station. [52]

$$\mathbf{\Sigma_x} = cov[\mathbf{x}] = \sigma \begin{Bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{Bmatrix} = \begin{bmatrix} \sigma_{x_1,x_1} & \sigma_{x_1,x_2} & \cdots & \sigma_{x_1,x_n} \\ \sigma_{x_2,x_1} & \sigma_{x_2,x_2} & \cdots & \sigma_{x_2,x_n} \\ \vdots & & & \vdots \\ \sigma_{x_n,x_1} & \sigma_{x_n,x_2} & \cdots & \sigma_{x_n,x_n} \end{bmatrix} \tag{4.2}$$

$$cov[\mathbf{x}] = [(\mathbf{x} - E[\mathbf{x}])(\mathbf{x} - E[\mathbf{x}])^T] = E(\mathbf{x}\mathbf{x}^T) \tag{4.3}$$

While univariate time series is being in mathematical modelling considered as single variable (see section 4.1), it is possible to express the covariances among sequences of the same time series in univariate covariance matrix. In a discrete time series vector with elements equidistantly spaced from each other by time intervals we can define the *covariance* of a variable x_t with latter itself x_{t+k}, k-time intervals apart, by formulation in Equation (4.4). [19]

$$\gamma_k = cov[x_t, x_{t+k}] = E[(x_t - E\{x\})(x_{t+k} - E\{x\})] \tag{4.4}$$

Then the *univariate covariance matrix* can be defined as (4.5). [52]

$$cov[x_t] = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{N-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{N-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots & \gamma_{N-3} \\ \vdots & & & & \vdots \\ \gamma_{N-1} & \gamma_{N-2} & \gamma_{N-3} & \cdots & \gamma_0 \end{bmatrix} \tag{4.5}$$

4.4 Time series analysis

4.5. Univariate time series modelling

Univariate time series modelling has already long history in the world and Czech Republic too. Many models are based on some sort of autoregressive process, describing inner time dependence of the observations. Basic type is Autoregressive model of order p, which determining the value of an observations based on p number of previous observations. However, despite, or maybe thanks to, its simplicity, if employed correctly and smartly, it can yield very satisfactory results.

As the time series modelling discipline developed, additional features were being incorporated into the models with autoregressive basis, such that they take into account non-stationarity or periodicity of a time series or trends found in the raw data.

Worth mentioning is also disaggregation modeling and method of fragments, which has a tradition in Czech Republic and has been widely used here.

4.6 Types of models for univariate time series generation

4.6.1 Autoregressive model

Autoregressive model (hereinafter AR model) is type of so-called Box-Jenkings models, which are linear non-seasonal models, assuming stationarity of the data. Moreover, it does not account for potential periodicity of the phenomenon it is supposed to simulate. Hydrological time series, including streamflow time series, exhibit very strong and obvious periodicity and it is desirable to simulate it in the synthetic data. Therefore, the AR model cannot be used directly on full time series, but first the periods must be recognized and then it is to determine with how many autoregressive coefficients is the series going to be simulated.

If data consists of monthly averages and the aim is to simulate also time series consisted of monthly averages for observations, the suitable approach is to separate raw data according to months, therefore creating twelve vectors of observations, one for each month. Then the correlation coefficients for each month are determined individually.

Generally, the Autoregressive process of order p is described as in Equation 4.A. [19]

$$z_t = \varphi_1 z_{t-1} + \varphi_2 z_{t-2} + \cdots + \varphi_p z_{t-p} + \varepsilon_t \tag{4.A}$$

Construction of the model thus relies on computing the required correlation coefficients φ and generating the white noise. For generating white noise, pseudorandom generators in any computational software can be safely used. The elements comprised of previous observations and corresponding correlation coefficients are adding up to the deterministic component of the model. The white noise along with its coefficient is representing the stochastic (random) component.

The order of the model p determines number of the elements used and number of the correlation coefficients φ used. Here the correlation coefficient of order p for series of variable x with l number of observations is defined as in Equation 4.0. [38]

$$\varphi_p = corr\left(x_1 \ldots x_{l-p}; x_{p+1} \ldots x_l\right) = \rho_{x,y} = \frac{cov(x_{1\ldots l-p}, x_{p+1\ldots l})}{\sigma_{x_{1\ldots l-p}} \sigma_{x_{p+1\ldots l}}} \tag{4.0}$$

The order p is chosen so that it reflects the behaviour of the time series most accurately. There exist tests to determine up to which order there is significant relationships between the observations. For modelling time series with daily observations, it is natural, that the value of the observation, for example flowrate in a stream, might depend on more than one previous observation, so choosing higher order AR model might be appropriate. For monthly data, it is much more likely that the monthly average depends only on the previous month and AR(1) model can be employed in such case. History of research on hydrological time series modelling on Czech Technical University showed, that low order models (with order 1 or 2) yield the best results and higher order models are for this application on monthly streamflow data much less predictable.[13]

Base of Equation 4.A, autoregressive model of order 1 would then look like in Equation 4.B. In this model, another coefficient γ was applied to reduce the contribution of the white noise term, the model then looks like in Equation 4.C. The coefficient is taking values from 0 to 1 and is being computed as in Equation 4.D.[49]

$$z_t = \varphi_1 z_{t-1} + \varepsilon_t \tag{4.B}$$

$$z_t = \varphi_1 z_{t-1} + \gamma \varepsilon_t \tag{4.C}$$

$$\gamma = \sqrt{1 - {\varphi_1}^2} \tag{4.D}$$

4.6 Multivariate time series modelling

4.6.1 Multivariate time series

True challenge in time series modeling in water management comes when there is need to generate multiple time series for multiple stations which are in relative proximity and their waterflows are mutually correlated. These inter-correlated time series are referred to as multivariate time series or sometimes also as multivariable time series. This applies for example to a system of water reservoirs all belonging to the same watershed or to adjacent sites affected by similar conditions, we call them *multiple reservoir systems*.

Within this kind of system, the water reservoirs located within the same watershed are subject to similar precipitation, evaporation, temperatures and climatic conditions generally. Sometimes they even directly affect each other; some of the reservoirs in the system can be simply downstream from others on the same watercourse. The streamflow times series in their profiles therefore exhibit strong both temporal and spatial dependence with each other. The challenge in time series modeling lies in the task to preserve the mutual spatial and temporal dependence between multiple reservoirs in the synthetic time series, more precisely to authentically simulate the dependences.

The problem gets even more challenging when dealing with not only yearly averages but also with monthly average streamflows. Within a year, seasonal trends can be observed and appropriate model or approach which reflects this seasonality must be employed to accurately simulate all the properties of historical records.

4.6.2 Principal Component Analysis

The Principal Component Analysis (PCA) is a method of linear transformation used for decorrelation of data. It can be used to transform set of process realizations which are correlated into data sets, which are mutually linearly uncorrelated. We call these new data sets principal components and we say we are extracting the components from the historical data.[22] The PCA method in the centre of focus of this thesis and its principle and application is explained in detail in section 5.2.

4.6.3 Independent Component Analysis

Independent Component Analysis (ICA) also decorrelates data within its process, but it goes further with minimization of mutual dependencies and looks for so-called independent components, which are statistically independent, which is concept explained in section 5.3.4. By finding components that hold even less mutual information upon each other than PCA it basically stores more of the information on the dependence of raw data and can reapply this information on synthetic data retaining fully both spatial and temporal dependence of the original time series. ICA is also explained and discussed much more thoroughly, further in the thesis, namely in section 5.3 and 5.4.

4.6.4 Other methods for preserving dependence in multivariate data

Neural networks approach

Another approach is to use artificial neural networks which are systems inspired by the functioning of human brain. These systems have very specific information distribution of information in form of analogue pattern signal and learning abilities which enables them to evolve their own solutions for non-linear problems. [3] Application of neural network in time series modelling was presented for example by Spanish-Colombian authors (Ochoa-Rivera et al., 2002). [40] A hybrid model for generation of multivariate streamflow time series based on a multilayer feedforward neural network, simulating deterministic component with random component represented by normally distributed noise accounting for its stochastic part, was proposed by them and compared with autoregressive model of order 2. Their work is following, among others, research papers on hydrological time series modelling by neural networks of Raman and Sunilkumar (1995) examining this approach for bivariate time series and Anmala et al. (2000) dealing with trivariate data. [42] *and* [4] *respectively, as cited in* [40] According to their results, the composite NN-stochastic model yielded much more satisfactory results than the purely stochastic AR model, especially in terms of simulation of longer persisting events within hydrological series, such as drought periods, which is very essential issue concerning optimal operation of water reservoir schemes design processes.[40]

Wavelet-based method with IAAFT algorithm

Another method presented by working paper by Keylock (2012) [27], is as wavelet-based method based on techniques in non-linear physics from the Fourier domain. This approach for multivariate synthetic time series generation claims to not only preserve the cross-correlative structure of the historical records, as is attempted in PCA, but also to simulate nonlinear properties that may be present. The paper works with assertion that among the most important nonlinear property of the streamflow data is the frequent temporal asymmetry of its hydrographs and uses derivative us skewness to describe and address this property. Author previously presented use of method called gradual wavelet reconstruction for synthetic data generation [28], and adapted in the mentioned work this method to the multivariate case. [27]

Because other two mentioned methods (ICA and PCA) relies on covariance matrices in the search for proper orthogonal transformations, they retain in the process only correlation at zero-lag. Proposed method by Keylock aims to preserve Fourier cross-spectrum between multivariate time series, explaining that it yields identical result as an attempt to preserve cross-correlation function. The iterated amplitude adjusted Fourier transform (IAAFT) algorithm extended to application on multivariate series is in the focus the paper. The research also presents comparison of its introduced method with PCA and ICA, with results speaking for advantage in enhancement of preservation of the full cross-correlation function over both ICA and PCA. [27]

4.7 Input and output data

We usually obtain the flow rates as daily, monthly or yearly averages. It is common to generate synthetic time series of monthly o yearly averages, whereas this thesis deals with the more complex case of modeling monthly averaged series, which also includes the solution for yearly averaged data.

Any input data used engineering solutions should be first critically evaluated in terms of their representativeness. In hydrology, one of the main problems with available source data is their scarcity. When conducting time series analysis on streamflow time series, there is available usually only one set of observations on one realization of the process, often also over relatively short period of time. Results of estimating parameters of the streamflow process, based on rather small sample must be used with caution. It is common to work with time series of only few decades of flow rate measurements, while the task is often to generate synthetic series of several thousands of years.

In hydrology, these synthetic time series are created in the same manner as the historical ones - as chronological sequences where the order of the elements matter. Therefore, we can work with them the same way as with the real historical data and use them as an input without further adjustment of the design process.

4.7.1 Critical assessment of source data

In every case, individual assessment of the researcher is needed to consider all possible cause that may have affected the representativeness of the data. Every problem is somehow specific and human evaluation is irreplaceable.

It is important to take into account all the conditions that accompanied the observations and acquisition of the data, in terms of their homogeneity. The researcher must get acquainted with the location where the observations originated, the landscape circumstances of the watershed and its development during the whole period which is represented by the data sample and in the future too. In case of flowrate measurements in a watercourse, following list presents an example of things that must be considered:

- land-use changes in the watershed during the represented period and future prospects
- anthropogenic interventions in all related watercourses
- method of measurement and its development

If homogeneity in any of the factors listed above has not been maintained during the observed period, influenced sections of the data sample must be ‘purged’ of the effect that compromised the coherence or they must be excluded from the analysis and shorter versions of the observations must be used.

5. METHODICAL APPROACH

5.2 Principal Component Analysis

The Principal Component Analysis (PCA) is a method of linear transformation used for decorrelation of data. It can be used to transform set of process realizations which are correlated into data sets, which are mutually linearly uncorrelated. We call these new data sets principal components and we say we are extracting the components from the historical data.[22]

The method was invented by Karl Pearson and was first introduced by him in 1901 in journal article of Philosphical Magazine, '*On lines and planes of closest fit to systems of points in space'*.[41] *as cited in* [26]

The main focus of PCA is to reduce the dimensionality of a data set, while preserving as much variation as possible. The method deals with transformation of interrelated variables, into a new set of uncorrelated variables, the principal components (PCs). The components are ordered by the amount of preserved variance of the original variables, leaving the last PCs least important. [26], [48] This property allows to reduce dimension from n to p of the original data set, by keeping the first p principal components, which explain substantial portion of the variance of the data set.[54]

PCA uses information that is contained inside the covariance matrix (*section 4.6.2*) to derive the transformation matrix which will decorrelate the data.

The method of principal components represents type of orthogonal transformation, that is generally speaking a linear transformation which preserves the inner product. It is being done by orthogonal matrix – a matrix whose inner product of itself and its inverse matrix equals an identity matrix, in other words, its transpose is equal to its inverse. Moreover, two vectors are orthogonal, if and only if their inner product equals zero. Important property of an orthogonal transformation for this particular use, is that the length of the vectors stays the same. [2, 3]

5.3 Independent Component Analysis

While PCA is will yield uncorrelated variables, they are not truly statistically independent. Any residual dependence left in the processed data, can be lost or destroyed during the synthetic data generation. By Independent Component Analysis (ICA) one can achieve higher order independence within processed data and therefore retain both spatial and temporal dependence of the original time series, which will be stored inside the procedures applied to the data during ICA.

5.3.1 History

The Independent component analysis was introduced for the first time by French authors, namely Jeanny Hérault and Bernard Ans, in 1984 in a journal article for *Comptes rendus de l'Académie des Sciences*, with Christian Jutten joining them in 1985 for conference proceedings in Paris and Nice. [17], [5], [18] *as cited in* [21]

Many concepts of ICA, including connection between negentropy and mutual information, or estimation of the components by minimization of the mutual information were most likely introduced in work of Pierre Comon (1994).[9] *as cited in* [21]

Probably the most extensive work on ICA have been done the Finnish team Aapo Hyvärinen, Juha Karhunen, and Erkki Oja from University of Helsinky, with the main reference book published in 2001.[21] Work of the Finnish authors is the most significant source for this thesis.

Contribution to the topic with summary of previous work was also done by James V. Stone, under MIT in 2004.[44] Research about ICA with regard to hydrology, which is of particular interest in this thesis, have been done by Westra, Brown, Lall and Sharma in 2007 and is being continued.[54]

5.3.2 Introduction

Independent component analysis is closely related to the problem called *Blind Source Separation* (BSS). That is an extraction of source signals from mixed observations without any or very little additional information to the observed data, hence the name ‘blind’.[7] It comes as an solution to the *cocktail party problem*. That is a well known mathematical problem, popularize in Cherry (1953), where there is number of signals observed, for example from microphones placed on different locations in a room where number of source signals, e.g. people speaking in the room, are present.[23] *as cited in* [36] In this problem, each microphone records different combination of the speech signals, each signal having different strength in each of those combinations. The goal is to estimate the source signals – speeches, as they were recorded directly, uninfluenced and unpolluted by other signals. [1] The only assumption on the source signals, is that they are independent in their origin. That is also one of the basic assumptions of the ICA.

Tu put it in simplified way, there can be two basic cases where ICA can be used. One is to find good linear representation of multivariate data. By transforming the variables, it aims to discover some hidden information on the data set that describes the underlying structure. The idea behind ICA is that the independent components correspond to some real physical parts of the process that generated the observations. [21]

5.3.3 Application

Even though ICA is relatively young method, it is widely used across all scientific fields and professions. Because of generality of ICA, its use can be found in medicine, financial markets or audio-visual sciences and several more, including natural sciences, where working with hydrological time series belongs. This section presents other uses and applications than the one of interest in this thesis, which is streamflow synthetic data generation. It is also fair to mention that at first it has found its application in signal processing, where ICA was used to separate the source signals from recorded mixtures of signals as it was explained in section 4.5.2 for 'blind source separation'.

Medicine

Many diagnostic medical devices are using ICA to analyse their outputs, which usually are dimensionless electric signals, to find underlying factors, corresponding to some activities of the body. For example *magnetoencephalography* - it is a functional neuroimaging technique for mapping brain activity. [33] Magnetoencephalography device records the signals emitted by brain. However, the signals are being mixed up in the sensors of the device placed on the head of the patient and ICA is used to re-extract them.[22]

Another health care diagnostic method called *optical coherence tomography* (OCT) used for biomedical imaging, with applications in ophthalmology or dermatology, can utilize ICA techniques to process output images, reducing speckle noise present in them and thus helping the process of interpretation of the images. [6]

Macroeconomics

Utilization of ICA analysis can be find in econometry. Many macroeconomic indicators are not strictly deterministic and are influenced by some hidden factors, which cannot be observed. Finding the background structure or driving mechanism of parallel time series such as currency exchange rates or stock prices may be useful both for econometry application and forecasting on financial markets.[21]

Some other applications

In audio-visual sciences there is also broad area of application of ICA, such as separating and purifying specific sources on sound recordings, removing noise from images, extraction features from images, or finding filters for natural images. [24]

Considering ICA deals with signals, other obvious area of employment for ICA is telecommunications. where it can be used for example for separation of signals interfering with other signals in mobile communications. [22]

Another use can be found in abundance quantification for hyperspectral imagery, more specifically for endmember extraction, a field where PCA was commonly applied in history, but not always yielded satisfactory results, as frequent scarcity of endmembers can result in very small influence on data variance, which is something that PCA reflects poorly, while ICA can be implemented more effectively. [50]

5.4 Mathematical background of ICA

5.4.1 The Central Limit Theorem

The Central Limit Theorem is a key concept in probability theory, allowing for application of many statistical methods. The theorem establishes, that the sum of independent random samples with any type of distribution, tends toward normal distribution, or at least is more gaussian than any original sample. With more data samples, the approximation of normal distribution gets better.[53], [22] The logic of the theorem, or more accurately its reverse, is adapted in ICA, which searches for transformation that will result in as non-Gaussian variables as possible. Its search is based only on the observed mixtures, which are by the central limit theorem assumptions much more Gaussian than their underlying signals.[54]

5.4.2 Data representation

When we denote the n mixtures x_i and the m source signals as c_j we can express the observed singals as linear combinations in Equation 4.-2 or expanded form in Equation 4.-1. The coefficients a_{ij} are real numbers representing the 'strengths' of the original signals in different combinations, the source signals c_j are called the *independent components*.

$$x_i = \sum_j a_{ij}\, c_j, \text{where } i = 1 \dots n, j = 1 \dots m \tag{4.-2}$$

$$\begin{array}{ccccccccc} x_1 & = & a_{11}c_1 & + & a_{12}c_2 & + & \dots & + & a_{1m}c_m \\ x_2 & = & a_{21}c_1 & + & a_{22}c_2 & + & \dots & + & a_{2m}c_m \\ & \vdots & & & & & & & \\ x_n & = & a_{n1}c_1 & + & a_{n2}c_2 & + & \dots & + & a_{nm}c_m \end{array} \tag{4.-1}$$

Using vector and matrix notations to the above equations, they can be written as in Equations 4.x and 4.x, where **A** is the $n \times m$ matrix of linear real coefficients a_{ij}.We call the matrix **A** the *mixing matrix*, as it explains how the components got mixed and resulted in the observed mixtures.

$$\begin{Bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{Bmatrix} = \mathbf{A} \begin{Bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{Bmatrix} \tag{4.-2}$$

$$\mathbf{x} = \mathbf{Ac} \tag{4.-2}$$

The mixing matrix **A** is unknown to us, and so are the components c_j , all we have are the observed mixtures. If we knew the mixing matrix, we could use its inverse to find the components, which is basically the approach for finding the solution. The goal is to find the inverse of the mixing matrix such that the resulting components have the desired properties or more specifically stastistical characteristics. Most important restriction is that they should be as independent as possible.

5.4.3 Objective of ICA - Statistical independence

When applying ICA to a data set, we are trying find as independent components as possible, so it might be useful to explain, how the statistical independence is defined.

In mathematical terms, two random variables x and y are *statistically independent* if their joint probability density functions can be expressed as a product of their individual (marginal) probability density functions (explained in section 4.3.2) as expressed in Equation 4.0. This definition also applies to their cumulative density functions, as in Equation 4.0. Both their PDFs and CDFs must be factorizable. [30], [52], [21] In layman's terms, variables are independent if changes in one do not affect the other, and they hold no information about each other.

$$p(x, y) = p(x)p(y) \tag{4.0}$$

If two variables are *uncorrelated*, their covariance is 0, and if they are not constants, which we can safely assume they are not when considering natural random variables, their Pearson correlation coefficient is also zero – Equation 4.0.[30]

$$\rho_{x,y} = \frac{cov(x, y)}{\sigma_x \sigma_y} = 0 \tag{4.1}$$

But that means only that they are partly independent and they can still contain significant amount of *mutual information*. Simply: when two variables are independent it implies they are also uncorrelated, but not vice versa. Uncorrelated variables are also statistically independent if and only if their joint probability distribution $F(x, y)$ is *normal*, which is inferred from the assumptions of central limit theorem.[53] Therefore, when searching for independence, the independent variables must not have normal distribution – they must be non-Gaussian.

5.4.4 Non-Gaussianity

Many variables encountered in statistical theory tend to have normal distributions.[22] Some authors also argue that this also applies to many physical quantities appearing in natural environment. [15] The fact that ICA searches for non-Gaussian variables, precludes the usage of ICA in many applications, where the wanted independent variables are assumed to have normal distribution. Independent components can be mathematically found, but they probably will not correspond to real variables that were searched for.

The same fact also complicates application in hydrological time series modelling, although not for the same reason. We are not searching for particular existing variables, only independent underlying factors representing the behaviour of our data, and distributions of these components are unknown to us. They need not to be normally distributed in terms of the representation. However, for further modelling procedures the components to be as normal as possible, because we estimate important characteristics, like auto-regression parameters, from them, used in univariate models, which often use the assumption of normal variables, as they are designed also to model normally distributed synthetic data. The assumption concerns not only mean and variance, but also skewness and behaviour of tails. [8], [39] Fitting of the autoregressive model requires that the probability distribution of the modelled variable matches the distribution of the transformed sample from which the statistics are estimated. The most desirable and easiest way to approach this restriction is to model the synthetic data Gaussian with mean 0 and variance 1.[38]

A solution offers itself here: to apply normalization procedure only after the ICA transformation directly on the estimated components. Normalizing will not affect their independence **after it was found.** This is however non-trivial task as many widely used normalization procedures, with effects that are desirable for hydrological modelling (such as not magnifying the variance dispersion), work on data that are positive only, which data after ICA transformation along with its pre-processing procedures never are. Even when applying the fastICA algorithm alone, the algorithm, when maximizing the non-Gaussianity, will reach for negative values.

Among possible solutions of this ICA property are to use normalization transformation that can work also with negative values and yields desirable results, modify fastICA algorithm so it ouputs positive data only, or to use different method of synthetic series generation such as bootstrapping, or other resampling methods, that will accurately simulate even non-Gaussian distributions.

5.4.5 Measures of non-Gaussianity

First, a measure of non-Gaussianity must be established, in order to mathematically define the search for the independent components in ICA. Among known measures are Kurtosis and Negenthropy.

Kurtosis is probably more classical measure. Being the scaled version of fourth moment, it tells us something about the behaviour of the peak and the tails of the distribution.[53] It can be defined by general formula in Equation 5.1.

$$\zeta_x = E[x^4] - 3(E[x^2])^2 \tag{5.1}$$

Although for standardized variable we can write kurtosis as in Equation 5.2.[53]

$$\zeta_x = E\left[\left(\frac{x - \bar{x}}{\sigma_x}\right)^4\right] - 3 \tag{5.1}$$

Same as for skewness, the kurtosis will be zero for Gaussian variables because the following relationship will apply.[21]

$$E[x^4] = 3(E[x^2])^2 \tag{5.1}$$

Kurtosis can take both positive or negative values (from -2 to infinity), which signifies so-called super-Gaussian or sub-Gaussian distribution respectively.[53] Super-Gaussian distribution's PDF has spiky peak and heavy tails, and is much more common when it comes to hydrological data, such as flowrate.[56] However, absolute value of kurtosis is being usually used to measure non-Gaussianity. The problem with kurtosis is, that it is very sensitive to ouliers and these are very common in streamflow data.[16] It is however much simpler both computationally and theoretically than the latter method.

Negentropy

Negentropy is also being used as a measure of non-Gaussianity and is defined through differential entropy, which is a concept from information theory. Entropy is sometime being called measure of randomness – the more unpredictable a variable is, the higher its entropy is. Mathematically it is defined as in Equation 5.3, where $H(x)$ is entropy of variable x. For vectors, modified definition can be used, called differential entropy, expressed in Equation 5.4.

$$H(x) = -\sum_i \mathbb{P}(\mathrm{x} = a_i) \cdot \ln[\mathbb{P}(\mathrm{x} = a_i)] \tag{5.1}$$

$$H(\mathbf{x}) = -\int p(\mathbf{x}) \cdot \ln[p(\mathbf{x})] \cdot d\mathbf{x} \tag{5.1}$$

The most important feature of entropy for herein application is, that it is largest for Gaussian variables. [10] *as cited in* [21] To obtain measure of non-Gaussianity, the definition of differential entropy must be further modified into definition of Negentropy which contrarily to entropy smallest for Gaussian variables – it is always non-negative and it goes to 0 for Gaussian variables. The definition of Negentropy $J(\mathbf{x})$ of vector $\mathbf{x}$ is as follows, in Equation 5.3.

$$J(\mathbf{x}) = H(\mathbf{x}_{gaussian}) - H(\mathbf{x}) \tag{5.1}$$

The vector of random variables $\mathbf{x}_{gaussian}$ is vector of Gaussian variables with the same covariance matrix as **x**.

5.4.6 Ambiguities of ICA

Among ambiguities of ICA belongs unknown variances and signs of the components as well as their order. Because of the nature of the transform it is impossible to determine the variances of the independent components. When we are searching for the mixing matrix, which is unknown, there is no unequivocal sole solution, because with different matrix with columns multiplied by unknown scalar, or with matrix that is a linear combination of other possible mixing matrices, we can accomplish similar results in terms of independence. Because of the same reason, there is also no way to determine the signs the components. Because of this ambiguity, the variances of all the components are sometimes being scaled to 1, by diving the combinations with standard deviation. [21], [44]

However, in many applications, this ambiguity is not significant. [21] It is also not important for multivariate hydrological time series modelling, where the goal is not to find components representing some particular existing latent variables as that would be impossible, it is sufficient to find components that are independent and represent the overall variance of the multivariate data.

5.5 Application of ICA to multivariate streamflow series

Before using Independent Component Analysis, we must first prepare the data. Some of these techniques are not necessary but they can be very useful and can make the process of estimation of ICAs much easier. It includes for example centering and whitening the input data.

The inconsistency of synthetic data is a result from either bias (systematic error), lying within the chosen techniques and procedures the model is based on, or from a random error.

The random error should not be significant for many utilizations of the modelled data, especially when longer time series are modelled. When 10 000 years long time series are being generated, it of course give more space for any random error, but they also lose significance at least in terms of sample statistics and long term behaviour parameters of the system. However random error can cause for example generation of unrealistically high flowrate at certain point of the time series, which might affect some tasks the series is being used to solve, e.g. frequency analysis and related applications.

The bias on the other hand, is of more concern. It can be stored within applied transformations or methods used to achieve the deterministic component of the model.

5.5.1 Standardization

By standardization it is here meant making the data zero mean with standard deviation equal to 1. We can achieve centering of the input data simply by substracting the mean from the time series (4.1). This will make the data zero mean, which is done solely to simplify the upcoming process. When we also divide each time series or its part by standard deviation, we then achieve in having data sets that are scaled to the same variance (and standard deviation) of 1, which also simplifies next processeces. Standardized variable is denoted as x_0, and computed as in equation 4.1.

$$x_0 = \frac{x - E\{x\}}{\sigma_x} \tag{4.1}$$

5.5.2 Whitening

Whitening is another process that can help prepare the data for further analysis. It is a linear transformation which transforms elements of a vector into new vector whose elements are uncorrelated. Covariance matrix of the original vector is used for this transformation, hence it must be known.[21]

The whitening transformation will convert the input vector into *white noise* vector whose covariance matrix is an identity matrix (the elements have no correlation with each other and they all have variance equal to one), as expressed in Equation 4.5, where $\hat{\mathbf{x}}$ is the whitened vector.

$$E(\hat{\mathbf{x}}\hat{\mathbf{x}}^T) = I \tag{4.5}$$

One of the ways how to perform whitening transformation is to use Eigen Value Decomposition, denoted as EVD. The EVD can be applied to covariance matrix of original vectors and it will yield corresponding matrix of eigenvectors $\mathbf{E}$ and diagonal matrix of eigenvalues $\mathbf{D}$. Eigenvectors are important mathematical elements with convenient properties when using linear transformation matrices. Eigenvector e of linear transformation matrix $\mathbf{A}$ does not change direction when this transformation is applied to it, it only changes scale by λ, which is its corresponding eigenvalue.[45] This relationship is represented by Equation 4.6.

$$\mathbf{A}e = \lambda e \tag{4.6}$$

After finding the matrices $\mathbf{E}$ and $\mathbf{D}$, they can be used to create new matrix $\mathbf{V}$ which will perform the required transformation process. There are more ways how to create the whitening transform matrix $\mathbf{V}$, one is represented in Equation 4.7.[31], [54] Transformation of vector $\mathbf{x}$ into white noise vector $\mathbf{z}$ is then expressed in Equation 4.8.

$$\mathbf{V} = \mathbf{D}^{-1/2}\mathbf{E}^T \tag{4.7}$$

$$\mathbf{z} = \mathbf{V}\mathbf{x} \tag{4.8}$$

5.5.3 Normalization transformation

Real streamflow data is never truly normal. By common sense, it is understandable that the probability distribution or the PDF of streamflow data goes to zero very quickly on the left tail from the mean as it cannot exceed to negative values of the horizontal axis. On the right tail on the other hand, there are extremes, even thought they might be scarce, very far from the mean, making the PDF to converge to zero much more slowly. This phenomenon is called heavy right tail, and we speak of heavy-tailed distributions, which is, by definition, every distribution whose tail is heavier than the tail of an exponential distribution. [8]

Box-cox power transform

Box cox is type of power transformation where simple objective function is employed to find optimal coefficient λ, that is subsequently used as a power in following Equation 5.3. Its advantage is, that it deals very well with non-normally behaving tails. Unfortunately for our application, it also has wo major disadvantages. One is that being a very strong transform, it can very sensitive to outliers, unnaturally boosting variance in its reapplication after synthetic data generation. The second disadvantage is, that in its original form, it cannot be reapplied to negative data, which synthetic data generated by AR model always contain. Result of the reapplication on negative data would result in complex numbers and modified version of the transformation needs to be used, like the one in Equation 5.4, if one wishes to employ Box cox power transform.

$$y = \frac{x^{\lambda} - 1}{\lambda} \tag{5.3}$$

$$y = x^{\lambda} \tag{5.4}$$

3-Parameter Log-Normal transform

One of the ways to normalize is through defining so-called 3-Parameter Log-Normal distribution.[43] To normalize variable x, it is searched for variable y, defined in Equation 4.0.

$$y = \ln|x - x_0| \tag{4.8}$$

If this variable y has normal distribution, which is what is needed, the variable x has log-normal distribution, defined by three parameters μ_y, σ_y and x_0, expressed in Equations 4.0, 4.0 and 4.0, where c expresses coefficient of variation and for the 3-parameter log-normal distribution is defined through relation with skewness of x described in Equation 4.0. The three parameters express the mean and standard deviation of the sample space of y and a shift of the distribution, respectively. The shift parameter is present to smooth the asymmetry of the distribution of x. Its PDF can be written as follows, in Equation 4.0.[43]

$$p(x) = \frac{1}{|x - x_0|\sigma_y\sqrt{2\pi}} \cdot \mathrm{e}^{-\frac{(\ln|x-x_0|-\mu_y)^2}{2\sigma_y^2}} \tag{4.8}$$

$$\mu_y = \ln\sigma_x - \ln|c| - 0.5\ln(1 + c^2) = \ln\left(\frac{\sigma_x}{|c| \cdot \sqrt{1 + c^2}}\right) \tag{4.8}$$

$$\sigma_y^2 = \ln(1 + c^2) \tag{4.8}$$

$$x_0 = \bar{x} - \frac{1}{c}\sigma_x \tag{4.8}$$

$$c^3 + 3c - \kappa_x = 0 \tag{4.?}$$

The way to find y is to find the shift parameter x_0, such that the expression Equation 4.? is true. The skewness κ_x of variable x is within this setup defined in Equation 4.8, but in order to find c and subsequently x_0, the κ_x is computed in standard way, as in Equation 4.1.[43]

$$\kappa_x = \left(e^{\sigma_y^2} + 2\right)\sqrt{e^{\sigma_y^2} - 1} \tag{4.8}$$

5.5.3 Independent components estimation

As it was explained, we are looking for components that are as independent as possible. As it was described in section 5.3.4 the components ic will be independent when their PDFs are factorizable as in equation 4.9.[53]

$$p(ic_1, ic_2, \dots, ic_n) = p(ic_1)p(ic_2)\dots p(ic_n) \tag{4.9}$$

The independence can be searched for by looking for maximum non-Gaussianity. That is true, because of the logic of central limit theorem. If the components were Gaussian, their PDFs could be factorizable even if they are not truly independent by origin, but their “independence” could still be proved. With components that are not non-Gaussian, the estimation is not possible. There is an exception allowing for at most one of the independent components being Gaussian.[22]

5.6 Comparison of PCA and ICA

Main purpose of PCA, for which was this technique originally intended, is to maximize variance. In water management, we are using also its orthogonality and uncorrelatedness properties, although these are part of the method mainly to ensure that each of PCs are expressing separate factors. For ICA the separation of components is its main objectives and has more demanding requirements for unrelatedness of the components, as it aims for them being not only uncorrelated but also statistically independent and that not only for Gaussian variables.[53] [26]

Uncorrelated variables with Gaussian distribution are also independent if their joint probability distribution is also Gaussian. For that particular case, PCA can be viewed as an equivalent to ICA and ICA can be considered as generalization of PCA to non-Gaussian data. In scientific community it is sometimes considered that PCA assumes normality of data, but that is not by any means necessary, unless one wants to find principal components with statistical independence (in mathematical definition-wise sense) and does not want to, for any reason possible, use ICA.[2] *as cited in* [26]

Generally, PCA can be used to reduce a data set's dimension before ICA algorithm is applied to it. When dealing with high-dimensional entries it can significantly reduce computational difficulty of the iterative algorithm. However sometime reducing dimension of your source data might not be desirable and it that case the one might want to use PCA mechanisms just to pre-process data for ICA by making them uncorrelated and standardized.[2] Eigenvalue or singular value decompositions are commonly being used for this purpose.[54], [31], [22]

6. MATHEMATICAL MODEL

6.1 Introduction and goals of the mathematical model

6.2 Structure of the model

Whole model was coded in MATLAB R2015a, MathWorks software as a function with variable inputs.

6.2.1 Input data

To simplify the work with source data it is convenient to put all the flow rate time series as input vectors in one matrix **Q**, where each vector represents one station. They are being ordered vertically, so columns correspond to stations and rows correspond to observations at the same time, as in Equation 6.1. We can call the matrix **Q** the input matrix. Obviously, all the vectors must be of the same length l and the resulting matrix is then of dimensions $l \times n$, where n is the number of time series (number of stations).

$$\mathbf{Q} = \begin{bmatrix} q_{11} & Q_{21} & \cdots & q_{n1} \\ q_{12} & q_{22} & \cdots & q_{n2} \\ \vdots & & & \vdots \\ q_{1l} & q_{2l} & \cdots & q_{nl} \end{bmatrix} \qquad (6.1)$$

But technically we will view the source data set as vector **q** of random variables q_i where $i = 1 \ldots l$, Equation 6.2.

$$\mathbf{q} = \begin{Bmatrix} q_1 \\ q_2 \\ \vdots \\ q_l \end{Bmatrix} \qquad (6.2)$$

The matrix **Q** can contain any number of time series of any length, the model performs the procedures independently on the dimension of the input matrix. While this does not affect its functionality, it may significantly affect the computing time and there is no information whether it might affect the efficiency of PCA, although there is no reason to believe so.

6.2.2 Pre-processing

Normalization transformation

Although it is listed here as pre-processing transformation, because it does not have a significant impact on PCA or ICA in terms of looking for uncorrelated or independent components, it is very important part of the process of time series modelling, and the choice of normalization method has fundamental implications.

To check if the normalization transformation succeeded, the Kolmogorov–Smirnov one sample test (KS test), named after Andrey Kolmogorov and Nikolai Smirnov, was used. It is a nonparametric test used to compare the probability distribution of tested sample, with reference probability distribution, such as normal distribution. The transformed time series is taken, standardized and compared with standard normal distribution.[46] In MATLAB the corresponding function to execute this test is *[h,p] = kstest(x,y)*. The null hypothesis is that the tested sample comes from a standard normal distribution, which is either accepted resulting in h is 0, or rejected at significance level alpha (implicitly 5%), resulting in h is 1.

$$y = \frac{x^{\lambda} - 1}{\lambda} \tag{1}$$

$$y = x^{\lambda} \tag{2}$$

Standardization

Very simple standardization transformation is used in this model. The aim is to make the data zero-mean with standard deviation equal to 1, as was explained in section [5.6.1]. To do that, the normalized time series N is taken, from which mean is subtracted for every column in every month matrix and then it is divided by its standard deviation, as expressed in Equation 6.1, resulting in standardized times series S.

$$S_{m,n} = \frac{N_{m,n} - \bar{N}_{m,n}}{\sigma_{N_{m,n}}} \tag{6.1}$$

Although some methods of normalization, as for example the MATLAB built-in Box-Cox transformation, already make the data zero mean, it won't affect the data to subtract the mean in every case again in standardization process and it might prove helpful in case the normalization method is changed, so we do not need to change the standardization and de-standardization procedures already implemented within the model.

6.2.3 Employment of Principal Component Analysis

6.2.4 Autoregressive model

6.2.6 Reverse transformations and interpretation of data

6.3 Verification of the model

6.3.1 System of stations for verification of the model

The constructed model was verified on system of stations in Moravian-Silesian region in Czech Republic (hereinafter CR). All stations belong to the same watershed of river Odra. It is a 850 km long river, out of which 112 km is in Czech Republic, where it starts. Its spring can be found in hill formation Oderské Vrchy in region of Olomouc, from where it goes north-east to city of Ostrava and then crosses the border to Poland. Table 6.1 contains names of the stations, the streams they are located on and their average long-term annual flowrate Q_a. From now on, the stations are referred to only by their corresponding numbers in Table 6.1.

Details of these station can be found on webpage of national organization Povodí Ohře. [57] Approximate locations of the stations within the region are shown in map of the region in Appendix 1.

Table 6.1

SYSTEM OF STATIONS in ODRA WATERSHED			
Number	**Station**	**Stream**	**Qa [m3/s]**
1	KS Svinov	Odra	12.135
2	VD Kružberk	Moravice	5.784
3	Děhylov	Opava	14.788
4	VD Šance	Ostravice	3.148
5	VD Morávka	Morávka	1.7
6	Vyšní Lhoty	Morávka	3.355
7	VD Olešná	Olešná	0.392
8	VD Žermanice	Lučina	0.552
9	Slezská Ostrava	Ostravice	13.523
10	Český Těšín	Olše	7.329
11	VD Těrlicko	Stonávka	1.226
12	Věřňovice	Olše	15.521
13	Řeka	Ropičanka	0.304

This thesis is focused on methodology of construction of the model itself and uses arbitrary historical data only to evaluate the model's performance, without consideration of possible water management implications to examined stations. Data that were chosen for this purpose were already used in similar research and their homogeneity and consistency were already evaluated. Critical assessment of the input data as it was described in section 4.7.1 is therefore not included here.

Table 6.2 [58]

STATIONS with WATER RESERVOIR					
Number	**Reservoir Name**	**Storage Capacity** [mil. m^3]	**Watershed area** [km^2]	**Dam Height** [m]	**Flooded Area** [ha]
2	**VD Kružberk**	24.6	567.0	34.5	280.0
4	**VD Šance**	43.1	146.4	65.0	337.0
5	**VD Morávka**	4.9	63.3	39.0	79.5
7	**VD Olešná**	3.5	33.6	18.0	88.0
8	**VD Žermanice**	18.5	45.5	32.0	248.0
11	**VD Těrlicko**	22.0	82.0	25.0	267.6

During the 36 years long period, two outstanding flood events can be seen. One is from July of 1997, and second from May of 2010. Both event affected mainly eastern part of Czech Republic and Silesia, if we restrict our interest on Czech country only. The 2010 event was considerably smaller in scale than the one from 1997, but both represent exceptional incidents within such short period of time.

In terms of statistics, in May 2010 the flows were estimated to have return periods 20 to 50 years in most of the watercourses in Odra and Morava watersheds, in watershed of river Olše (part of Odra watershed) the flows exceeded 100 years values. In July of 1997 the flows had estimated return periods at least 50 years in most watercourses, many exceeded 100 years and at least one river (Opava) registered discharge with return period significantly higher than 100 years. Table 6.3 summarizes the estimates on return periods as a result of research conducted on the records of the two events. [59]

Table 6.3; [57, *p. 48*]

FLOOD EVENTS of 1997 and 2010 in ODRA WATERSHED					
Stream	**Station**	**Culmination flow**		**Return perion**	
		July 1997	*May 2010*	*July 1997*	*May 2010*
		[m^3/s]	[m^3/s]	[years]	[years]
Odra	**Ostrava-Svinov**	688.0	404.0	> 100	20-50
Opava	**Opava**	647.0	76.9	>> 100	2-5
Ostravice	**Ostrava**	898.0	780.0	50	20-50
Odra	**Bohumín**	2160.0	1070.0	> 100	10-20
Olše	**Věřňovice**	673.0	1030.0	20-50	> 100

Table 6.4 [57]

RETURN PERIODS for SYSTEM OF STATIONS in ODRA WATERSHED									
I.	**Station**	**Stream**	**Q1**	**Q2**	**Q5**	**Q10**	**Q20**	**Q50**	**Q100**
1	KS Svinov	Odra	128	180	258	322	392	491	571
2	VD Kružberk	Moravice	52.6	75.5	111	140	173	219	258
3	Děhylov	Opava	101	150	228	296	371	482	576
4	VD Šance	Ostravice	52.8	84.9	132	170	211	267	313
5	VD Morávka	Morávka	21.8	39.4	67	90.7	117	155	187
6	Vyšní Lhoty	Morávka	35.4	57.7	96.5	133	175	241	300
7	VD Olešná	Olešná	9.6	15.8	26.9	37.4	49.9	69.6	87
8	VD Žermanice	Lučina	16.2	23.3	34.1	43.3	53.3	67.7	79.5
9	Slezská Ostrava	Ostravice	186	280	431	565	714	936	1120
10	Český Těšín	Olše	110	164	249	323	405	525	626
11	VD Těrlicko	Stonávka	27.8	40.8	61.2	78.8	98.1	127	150
12	Věřňovice	Olše	182	267	399	512	637	819	970
13	KS Smilovice	Ropičanka	7.32	12.3	20.2	27.2	35	46.4	56.2

7. RESULTS AND DISCUSSION

7.1 Testing the synthetic time series

This section of thesis aims to test the synthetic times series generated with the constructed model with water management problems that are commonly being solved based on streamflow time series. These problems involve following solutions.[34]

a) Requirement for supply storage capacity in a water reservoir and the reliability with which this capacity ensures unimpaired water supply at required rate, along with function expressing volume of water present in the reservoir in progress.
b) Operational function expressing relation between required for supply storage capacity and required improved outflow at given level of temporal reliability.

7.1.1 Water Supply Reliability function – WSR – ad a)

This function is dealing with basic water management solution. Requirement for supply storage capacity is computed, such that the required reliability of the reservoir is satisfied. It also produces plot of function expressing volume of water present in the reservoir in progress. The way WSR works is summarized in computational diagram, in Picture 7.1.

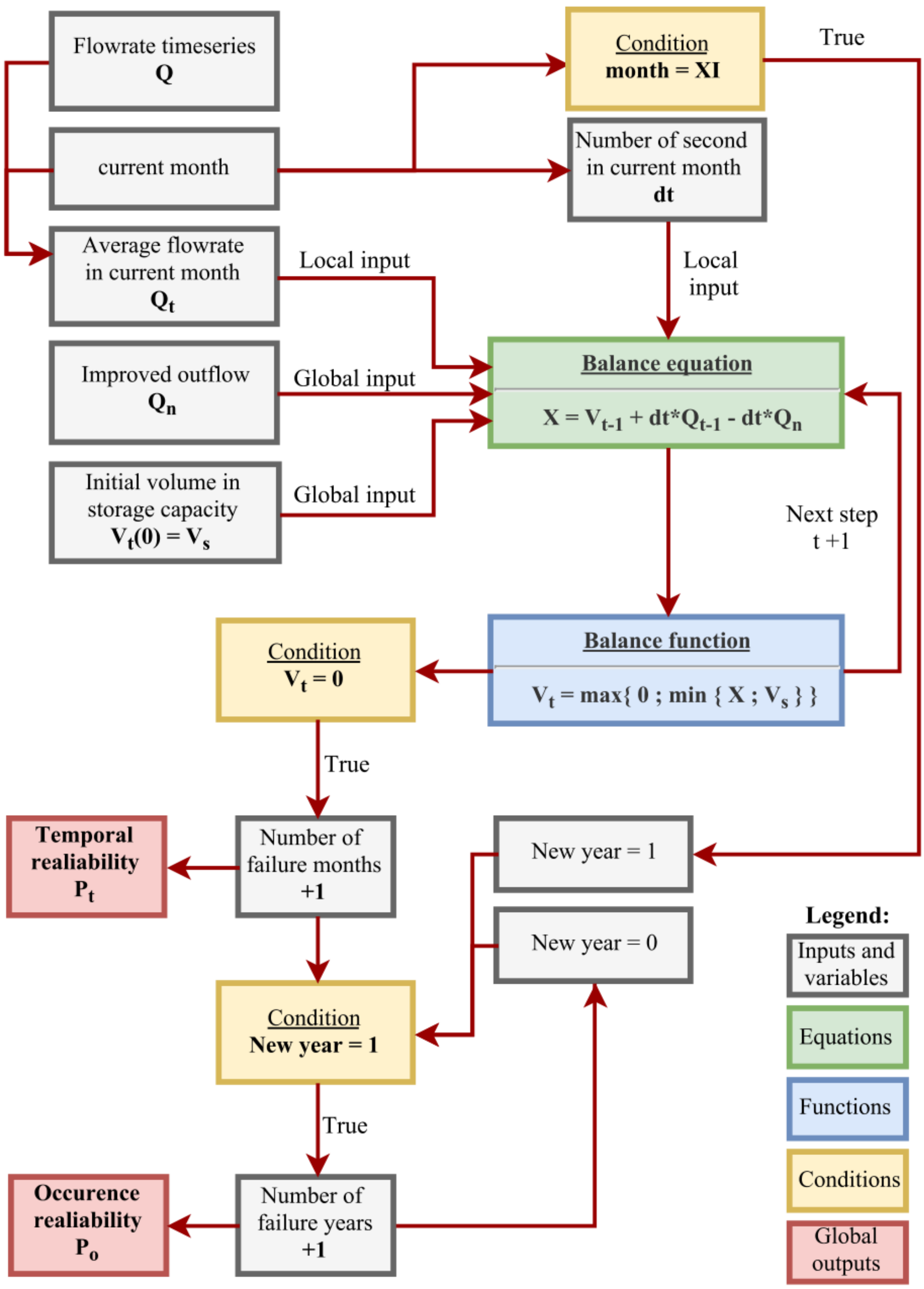
Following table show requirements for both historical and synthetic data, while the synthetic time series used as an input were 100 years long and the values of required storage capacities were averaged throughout 50 runs.

Table 7.1

Water Supply Storage Capacity requirements													
Stations	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
Historical data [mil.m^3]	85.10	31.50	81.50	14.70	8.80	17.60	3.00	4.00	77.90	34.60	9.20	72.70	2.00
Synthetic data [mil.m^3]	83.09	34.56	83.08	19.40	9.74	19.30	2.90	3.49	74.05	39.91	8.03	74.23	1.70

Picture 7.1

WATER SUPPLY RELIABILITY FUNCTION (WSR)

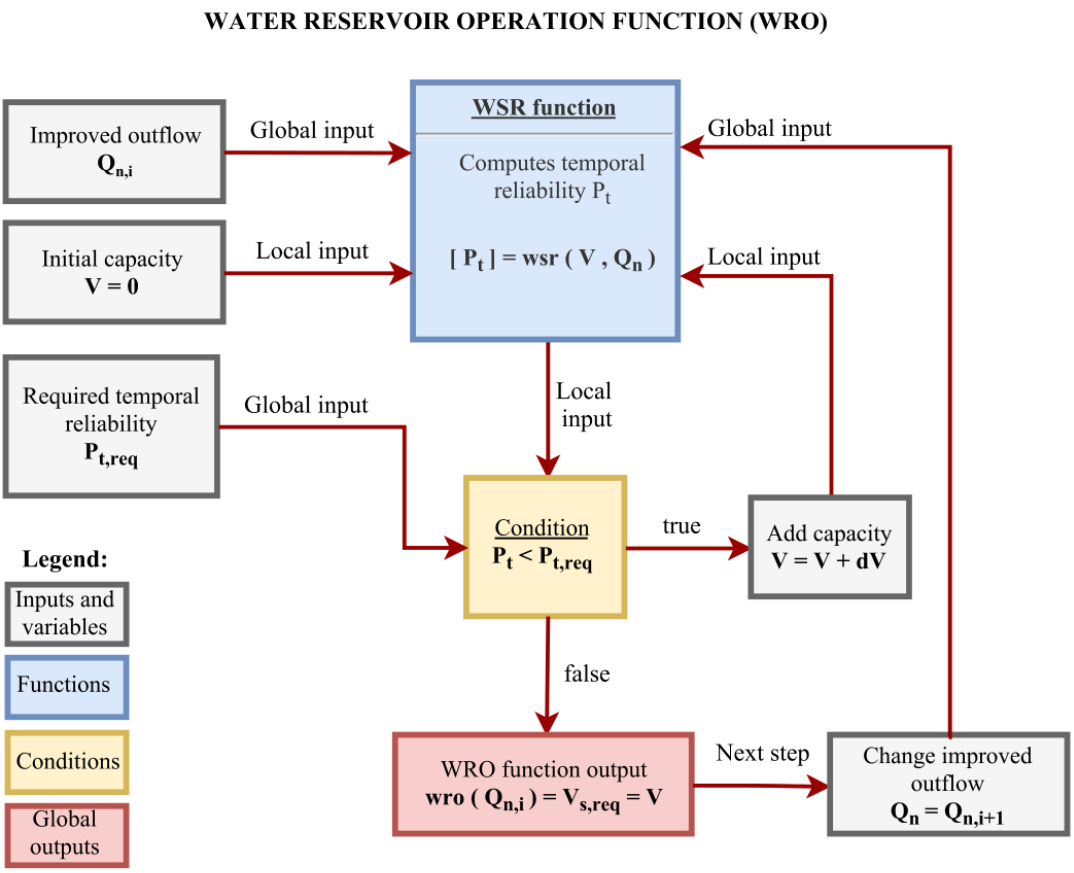


7.1.2 Water Reservoir Operation function – WRO

One of the exercises that has been used to study the behaviour of a time series, is Water Reservoir Operation function. It gives us relation between required storage capacity and different demanded improved outflows secured at given constant temporal reliability. It is a complex water reservoir management solution and elementary characteristic of any water reservoir.

Computing reservoir function involves iteration process, where for selected improved outflows, small volumes of storage capacity are added with each iteration to the reservoir characteristic, until the required reliability is secured. The smallest storage capacity which will satisfy that condition is the function output for that given improved outflow. The process is described in its computational diagram in Picture 7.2.

Picture 7.2



Operation function is very useful tool for evaluation process of model's performance and an important indicator of its correctness. Classic time series models are usually performing very good in terms of reproducing time series with the same sets of statistical moments. Especially low order moments are easily determinable and quite simply manageable, for the transformations built within the model are based on the identification of these characteristics and in case of their insignificant deviation from desired values, due to the unreliability of the stochastic component, uncomplicated correction techniques are available to deterministically achieve output with desired properties. However not only the preservation of overall statistics is important, but also accurate simulation of long-term and short-term behaviour of the time series represented by fair appearance of characteristic events like drought periods or flood flows and their realistic succession. This is difficult to express mathematically. The behaviour of synthetic time series can be examined by experienced eye of a researcher directly from a plot of the time series, but this can be very subjective and time consuming and the results are difficult to quantify. One way to examine realisticness of the time series behaviour is through operation function. The function plot can represent several scenarios in scaled and compact figures, where the behaviour of time series is represented by simple exponential-like curve based on the ability of the time series to satisfy certain conditions. This way we can easily compare the "performance" of synthetic time series with the one of the historical time series, judge the deviations and make conclusions about the performance of the model.

It is necessary to mention that the structure of the WSR function was substantially simplified compared to the real solutions in practice. For example, the evaporation element was left out of the function, simply because its computation would require Depth-Area-Volume function data, which is a fundamental characteristic of any water reservoir. This characteristic is obviously unavailable for the stations where there is no existent reservoir and therefore implementation of the evaporation element would require design of possible reservoir solutions and that was beyond the reach of this thesis. However, while this would certainly hold an impact on the applicability of the WSR function in practice, it does not affect herein application, because we omit the respective elements in both historical and synthetic data processing and this simplification can be safely used without the loss of representativeness of the method.

For improved flows a vector of 20 values was used, starting at 20% and ending at 80% of long-term annual average of historical data of corresponding station.

7.2 Testing tools

In this section, tools used to evaluate behaviour of both synthetic and historical time series are presented and their mathematical background is explained. It includes commonly used parameters of water reservoirs, like temporal and occurrence reliability or required supply storage capacity. Several significance tests which were used to determine whether statistical characteristics of the modelled water system were preserved in the synthetic system, are also presented.

7.3.1 Temporal and Occurrence Reliability

Temporal reliability is the primary parameter of water reservoir that is being used here to determine sufficient storage capacity. To determine temporal reliability, empirical probability estimate is used, based on the number of months the reservoir fails to provide sufficient water supply with given supply storage capacity. As from the definition of empirical probability, which is defined by the ratio of the number of outcomes in which the given event occurs to the total number of trials [37], ratio of the number of months the reservoir did not fail to meet the water demand to the total number of months is used to assess the temporal reliability. The ratio is being expressed through Čegodajev empirical probability formula [14], as in Equation 6.1, where l is the total number of months and f_m number of failure months.[29]

$$\mathbb{P}_t = \frac{(l - f_m) - 0{,}3}{l + 0{,}4} \tag{6.1}$$

Occurrence reliability is another measure used to evaluate the storage capacity requirement. It is very similar to temporal reliability, but instead of months it uses years in which the reservoir does not meet the requirements - ratio of non-failure years to total number of years is used. A year is considered as failed if it contains at least one failure month. The formula is expressed in Equation 6.1, where f_y is the number of failure years.[29]

$$\mathbb{P}_o = \frac{(\frac{l}{12} - f_y) - 0{,}3}{\frac{l}{12} + 0{,}4} \tag{6.1}$$

Level of reliability required differs and depends on size of the reservoir and its purpose. For large water reservoir, supplying water for civil usage, the demanded temporal reliability is usually 99.5%.

7.3.2 Two sample T-test for sample means and variances

Other test conducted on the simulated data was evaluation of deviation of sample statistics between synthetic and historical time series. Classic two sample t-test was used for this evaluation, testing the null hypothesis that both observation samples have equal means and variances and they are coming from independent random sample spaces. T-test is a type of parametric test that uses the significance level α upon which it rejects or accepts the null hypothesis. Second output of the t-test is the p-value, which is being compares to the limit level of significance, at which we still rejects the null hypothesis – we reject the hypothesis, if p is smaller than α.[47] MATLAB function *[h,p] = ttest2(x,y)*, was used for this test. The result h is 1 if the test rejects the null hypothesis at the significance level α (implicitly 5%), and 0 otherwise.[35]

The test was to determine whether the model preserves sample means and variances in whole time series but also across month samples. For example, sample mean of all monthly flowrate averages in Novembers in historical data were compared with those in synthetic data and so it was done for all months and for each station. Finally, whole sample means were compared.

Zero values signify that the null hypothesis was not rejected and means and the variances were not significantly different for the historical data and the modelled time series. Non-zero values mean rejection of the null hypothesis and at the same time they express the p-values upon which the hypothesis was rejected. They are therefore all smaller than the 5% significance level α.

7.3.3 Two sample correlation significance test using Fisher's procedure

Following method was used to compare pairs of correlation coefficients. The two sample correlation significance test is testing null hypothesis that a correlation of two samples from one time series is the same as a correlation of the same corresponding two samples from second time series. Procedure introduced by R. A. Fisher in 1921, explained in following equations, was implemented in the test.[12] *as cited in* [55]

First the two correlation coefficients being compared are transformed by the Fisher's transformation as in Equation 6.1, then the searched statistic is computed by Equation 6.1, where n is the size of the sample - number of observations, used to compute corresponding coefficient. [55]

$$\rho' = \frac{1}{2} \cdot \ln\left(\frac{1+\rho}{1-\rho}\right) \tag{6.1}$$

$$z = \frac{|\rho_1' - \rho_2'|}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}} \tag{6.1}$$

The p value is then computed standardly through normal distribution function Normal CDF with zero mean and variance of 1, as in Equation 6.1.[53]

$$p = 2 \cdot \left(1 - F(z|0,1)\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp\left(\frac{-z^2}{2}\right) dt \tag{6.1}$$

The value is than compared with the chosen significance level – here 0.05. If p is smaller than the the significance level, the null hypothesis is rejected.

7.3 Preservation of basic sample statistics

Because of the employment of the correction techniques, which directly reapply the sample mean and variance of the historical data to the synthetic time series, these two statistics are always preserved without error, if we compare only the end data. Following table proves that. It shows the percentages of cases in which the mean and variance were significantly different between the historical and synthetic time series, which is tested by the two-sample t-test from section 7.3.2. All of the results are zero, signifying that the test's null hypothesis was never rejected.

Table 7.1

Mean and Variance comparation - two sample Student Test's percentage results													
Stations	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
XI	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
XII	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
I	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
II	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
III	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
IV	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
V	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
VI	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
VII	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
VIII	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
IX	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
X	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Q	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Same percentages are shown in Table 7.2, signifying cases in which the synthetic time series had significantly different probability distribution, which is tested by the classic Kolmogorov-Smirnov two sample test.

The percentages higher than 15% are highlighted by red colour. The 15% limit was chosen arbitrarily. As you can see, the model preserves fairly the distributions of values in vectors comprised of the same month. The least pleasant behaviour can be seen in July’s data. This is probably due to the higher variance of July’s averages across years, even in the historical data. In July, both major flood events and severe droughts can occur. However, the model does not perform very well in preserving the distributions of the time series as a whole. This can be caused by the structure of the model itself, as it is applying individual transformations separately on every month.

Table 7.2

Distribution comparation - two sample Kolmogorov-Smirnov Test's percentage results													
Stations	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
XI	0	0.1%	0	0	0	0	0.1%	0	0	0	0	0	0
XII	0	0.1%	0.4%	0	0	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	0	0	0	0	0	0	0
II	0	0	0	0	0	0	0	0	0	0	0	0	0
III	0	0	0	0	0	0	0	0	0	0	0	0	0
IV	0.1%	0	0	0	0	0	0.1%	0	0	0	0.1%	0	0
V	10.3%	0	0.3%	0.5%	0.2%	0.3%	14.5%	0.9%	3.8%	4.9%	0.2%	15.7%	0.3%
VI	0.1%	2.1%	1.9%	0	0	0	0.1%	0	0	0	0	0	0
VII	80.2%	61.8%	76.5%	11.6%	13.9%	11.7%	55.1%	2.3%	18.4%	8.7%	1.0%	4.9%	3.9%
VIII	39.8%	0	0	0.5%	1.2%	0.2%	28.0%	1.2%	0.3%	0	15.0%	0.4%	0
IX	8.7%	6.6%	0.5%	3.6%	4.7%	5.4%	1.1%	1.5%	0.9%	10.8%	1.0%	0.7%	0.4%
X	3.7%	22.7%	3.5%	0	0	0	0.3%	0.3%	0.2%	0	0	0	0
Q	99.8%	91.3%	80.0%	66.9%	34.7%	49.5%	99.5%	76.1%	41.2%	45.2%	77.5%	34.7%	8.0%

Table 7.3

Distribution comparation - two sample Kolmogorov-Smirnov Test's *p* values mean													
Stations	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
XI	0	2.8%	0	0	0	0	2.8%	0	0	0	0	0	0
XII	0	2.8%	2.8%	0	0	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	0	0	0	0	0	0	0
II	0	0	0	0	0	0	0	0	0	0	0	0	0
III	0	0	0	0	0	0	0	0	0	0	0	0	0
IV	2.8%	0	0	0	0	0	2.8%	0	0	0	2.8%	0	0
V	2.3%	0	2.8%	2.8%	2.8%	2.8%	2.2%	2.6%	2.5%	2.5%	2.8%	2.3%	2.3%
VI	2.8%	2.7%	2.4%	0	0	0	0.6%	0	0	0	0	0	0
VII	1.3%	1.7%	1.3%	2.4%	2.2%	2.2%	1.8%	2.4%	2.2%	2.4%	2.3%	2.4%	2.3%
VIII	1.9%	0	0	2.8%	2.3%	2.0%	2.0%	2.8%	1.8%	0	2.2%	2.8%	0
IX	2.4%	2.4%	2.8%	2.4%	2.5%	2.4%	2.8%	2.6%	2.8%	2.3%	2.8%	2.8%	2.8%
X	2.5%	2.2%	2.6%	0	0	0	2.8%	2.8%	2.8%	0	0	0	0
Q	0.4%	1.5%	1.8%	2.3%	2.9%	2.7%	0.3%	1.9%	2.7%	2.7%	1.9%	3.0%	3.4%

Following table – Table 7.2 a), shows summary of the Table 7.2, making averages of the percentages firstly only for monthly values, secondly for the full series comparison and finally it shows total error of the table by making an average of the entire table.

Table 7.2 a)

Total error in equal distribution test within months	**3.66%**
Total error in equal distribution test within full series	**61.88%**
Total error for entire table	**8.13%**

7.4 Examination of correlation coefficients

Two basic sets of correlation trends were examined:

1) **Autocorrelations of months with previous months** - The first set show the inner autocorrelation structure of individual stations. By examining the set, it is aimed to determine whether the model preserved inner time dependent and periodical behaviour of the individual historical time series. This could be also done separately for each station.

2) **Cross-correlations between stations –** The second set represents matter of higher interest in this thesis. It shows us both temporal and spatial dependence of stations with each other, which the very thing the model aims to preserve and replicate. By temporal dependence it is meant time-coordinated behaviour - if an event occurs in one of the stations, the other stations, being in its proximity, behave accordingly always with similar time response, depending on the extent of the event of course. By spatial dependence, it is meant the similarity of behaviour based on their mutual distances and their topographic relationships – an event cause by natural phenomenon, like precipitation, in one of the stations, probably affects by certain measure other stations in its proximity, depending on the extent of the phenomenon.

7.4.1 Autocorrelations of months with previous months – ad 1)

At first were tested coefficients expressing correlations of each month average flowrates with previous month average flowrates throughout the whole observed period. For example, vector of flowrate averages in all Novembers in station 1 were taken and compared with vector of flowrate averages in all Octobers. Finally, autocorrelation with lag 1 for vector with all months was computed, i.e. relationship between a times series and latter itself, shifted by one month (one value). This was done for the historical time series, then for the synthetic one and then were these pairs of correlation coefficients compared by the two-sample correlation significance test using Fisher's procedure, described in section 7.3.3.

Table 7.4 shows autocorrelation structure for the historical data and Table 7.5 shows averages of correlation coefficients for 1000 generated time series. Row labels with months represent correlation between the month on the label and its previous month. Final row represents full time series autocorrelation. The tables are colour-scaled with red colour for positive correlation, blue colour for negative correlation and white cells are those with very small or no correlation at all. Higher saturation means stronger relationship.

It is clear from the tables, that the structure for historical data is much more chaotic, while the structure for synthetic data is much more homogeneous for individual months. This is obviously by main part caused by the fact that the table for historic data represents one single scenario, while the table for synthetic data is smoothed by 1000 runs making the values to converge to correlation figures of the modelled population. Also there are no negative correlations in the synthetic data, but that can be of no significance, as the very few negative correlations in historical data are very week and their reproduction in synthetic time series can be assessed as unimportant. However, some considerable difference can be spotted and their significance will be assessed in following paragraph.

Table 7.4

Autocorrelation coefficients for Historical Time Series													
Stations	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
XI	0.56	0.87	0.87	0.43	0.44	0.40	0.47	0.62	0.51	0.45	0.57	0.47	0.39
XII	0.33	0.15	0.45	-0.02	0.05	0.08	0.50	0.30	0.21	0.10	0.29	0.20	0.11
I	0.33	0.53	0.55	0.20	0.14	0.09	0.21	0.09	0.08	-0.01	0.10	0.13	0.01
II	-0.04	0.08	0.10	0.07	0.08	0.13	-0.04	0.09	0.04	0.11	0.06	0.07	0.05
III	0.18	0.23	0.18	0.21	0.13	0.16	0.13	0.09	0.17	0.02	0.12	0.07	0.15
IV	0.27	0.21	0.32	0.03	0.08	0.23	0.15	0.16	0.31	0.35	0.21	0.33	0.27
V	0.38	0.58	0.48	0.01	-0.01	-0.01	0.11	0.05	0.02	-0.02	0.07	-0.03	0.01
VI	0.53	0.50	0.49	0.51	0.17	0.46	0.11	0.27	0.29	0.38	0.13	0.32	0.26
VII	0.05	0.06	0.02	0.36	0.37	0.21	0.07	0.13	0.17	0.19	0.18	0.13	0.09
VIII	0.13	0.38	0.20	0.02	0.07	0.10	0.07	0.10	0.09	0.08	0.10	0.08	0.08
IX	0.08	0.37	0.20	-0.06	0.00	0.02	-0.03	0.09	-0.01	0.10	-0.01	0.07	0.08
X	0.46	0.47	0.49	0.31	0.43	0.39	0.20	0.14	0.30	0.28	0.19	0.32	0.32
Q	0.27	0.45	0.35	0.20	0.20	0.23	0.13	0.19	0.19	0.21	0.16	0.19	0.21

Table 7.5

Autocorrelation coefficients for Synthetic Time Series													
Stations	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
XI	0.49	0.52	0.53	0.48	0.51	0.53	0.41	0.51	0.54	0.53	0.51	0.54	0.52
XII	0.22	0.25	0.24	0.22	0.22	0.22	0.21	0.21	0.23	0.22	0.23	0.22	0.22
I	0.12	0.07	0.10	0.10	0.11	0.11	0.11	0.13	0.14	0.13	0.13	0.14	0.12
II	0.09	0.12	0.11	0.08	0.08	0.08	0.10	0.09	0.06	0.06	0.07	0.06	0.06
III	0.17	0.17	0.18	0.16	0.16	0.17	0.14	0.15	0.18	0.18	0.16	0.18	0.17
IV	0.23	0.25	0.27	0.20	0.21	0.24	0.19	0.19	0.26	0.25	0.19	0.25	0.25
V	0.10	0.11	0.11	0.10	0.11	0.11	0.09	0.10	0.11	0.10	0.10	0.11	0.11
VI	0.28	0.25	0.26	0.26	0.26	0.30	0.27	0.28	0.31	0.29	0.27	0.30	0.30
VII	0.22	0.20	0.19	0.18	0.21	0.23	0.21	0.21	0.23	0.22	0.21	0.23	0.21
VIII	0.33	0.30	0.32	0.33	0.36	0.37	0.32	0.35	0.36	0.36	0.34	0.37	0.36
IX	0.18	0.27	0.21	0.17	0.18	0.17	0.16	0.18	0.17	0.17	0.17	0.17	0.17
X	0.35	0.32	0.35	0.34	0.36	0.36	0.33	0.34	0.37	0.36	0.35	0.36	0.34
Q	0.26	0.38	0.30	0.25	0.26	0.28	0.22	0.26	0.28	0.26	0.25	0.26	0.30

Next two tables, Table 7.6 and Table 7.7, represent the two-sample comparation test's results. By the same logic as in the two-sample test of equal means and variances, or distributions, the first one (7.6) gives us percentages of runs when the null hypothesis was rejected, and the second one gives us average p values upon which the hypothesis was rejected, counting only cases when it was actually rejected. For example, the autocorrelation coefficient of Novembers with preceding Decembers in station 1, was significantly different in 1.7% of cases (first row, first column). Zero values signify that the null hypothesis was never rejected for that particular relation and the correlation coefficients were never significantly different for the historical data and the modelled time series. The table 7.6 is arguably the most important one, because it shows something about how similarly to the historical data the synthetic time series behave and therefore it tells us something about the model's performance.

The percentages higher than 15% are highlighted by red colour. As you can see, the autocorrelation structure of the series was not disrupted by the model too often in many places as substantial majority of values is smaller than 1% or even 0. It however highlighted problematic periods especially in stations 2 and 3, namely autocorrelation at lag 1 for Novembers, Januaries and Mays, where the structure was disrupted significantly. Cross-referencing the locations of these deviations with correlation coefficient in tables 7.4 and 7.5, it is clear, that this has been caused by sort of systematic error of the model, rather than being a result of weak autocorrelation or unpredictability of relevant periods.

Table 7.7, as expected, corresponds to results in Table 7.6, and top of it, it shows that the p values are not critically small, most of them being between 3-4% meaning that the hypothesises were not far from the rejection limit, set at 5% significance level. Values below 1% are highlighted by yellow colour, pointing to spots with largest probability of different values in terms of autocorrelation.

Table 7.6

Autocorrelation coefficients comparation - two sample Fisher's Test's percentage results													
Stations	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
XI	1.7%	88.9%	87.3%	0	0.2%	0	0.6%	4.0%	0.1%	0	1.3%	0	0.1%
XII	1.6%	0	7.7%	0	0	0	19.7%	0.8%	0.1%	0	0.3%	0.2%	0
I	6.9%	58.6%	60.0%	1.3%	0.2%	0.2%	1.6%	0.1%	0	0	0	0.1%	0
II	0	0.2%	0.4%	0.4%	0.5%	1.1%	0	0.1%	0.3%	0.5%	0.2%	0.2%	0.2%
III	0.2%	0.8%	0.1%	0.5%	0.3%	0.3%	0.2%	0	0.3%	0.1%	0.2%	0	0.2%
IV	0.2%	0.6%	0.8%	0	0.1%	0.5%	0	0	0.3%	1.9%	0.3%	0.4%	0.6%
V	15.1%	62.8%	36.0%	0.1%	0	0	0.2%	0.1%	0	0	0	0	0
VI	15.6%	15.6%	11.4%	14.5%	0.1%	4.2%	0	0	0.1%	1.4%	0	0	0
VII	0	0	0	4.2%	3.2%	0.2%	0	0.1%	0	0	0.1%	0	0.1%
VIII	0	1.2%	0	0	0	0	0	0	0	0	0	0	0
IX	0	1.6%	0.2%	0	0	0	0	0	0	0	0	0	0
X	1.3%	5.2%	3.9%	0	0.6%	0.3%	0	0	0	0	0	0	0.2%
Q	1.1%	11.6%	5.3%	0	0	0	0	0	0	0	0	0	0

Table 7.7

Autocorrelation coefficients comparation - two sample Fisher's Test's *p* values mean													
Stations	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
XI	3.2%	0.6%	0.8%	0	3.4%	0	2.9%	2.9%	3.8%	0	3.1%	0	2.6%
XII	3.5%	0	3.0%	0	0	0	2.4%	2.7%	2.6%	0	2.0%	3.1%	0
I	3.0%	1.7%	1.7%	3.0%	2.4%	3.8%	3.3%	4.3%	0	0	0	4.9%	0
II	0	2.8%	3.0%	3.5%	4.1%	3.6%	0	3.5%	4.4%	4.1%	3.4%	4.6%	4.2%
III	2.0%	3.0%	0.5%	2.3%	3.2%	1.1%	2.6%	0	2.5%	4.8%	4.5%	0	1.2%
IV	1.9%	3.2%	2.6%	0	4.8%	2.9%	0	0	3.3%	3.5%	3.8%	3.9%	3.5%
V	2.6%	1.6%	2.1%	4.4%	0	0	3.3%	3.8%	0	0	0	0	0
VI	2.4%	2.8%	2.8%	2.5%	3.5%	3.0%	0	0	4.4%	3.8%	0	0	0
VII	0	0	0	3.1%	3.2%	2.8%	0	4.0%	0	0	2.2%	0	4.3%
VIII	0	3.4%	0	0	0	0	0	0	0	0	0	0	0
IX	0	3.0%	2.3%	0	0	0	0	0	0	0	0	0	0
X	3.6%	3.0%	2.9%	0	3.3%	3.0%	0	0	0	0	0	0	4.5%
Q	3.1%	2.9%	3.0%	0	0	0	0	0	0	0	0	0	0

Following the same measure done for the equal distributions test, Table 7.6 a), shows summary of the Table 7.6, making averages of the percentages firstly only for monthly values, secondly for the full series comparison and finally it shows total error of the table by making an average of the entire table.

Table 7.6 a)

Total error in autocorrelation within months	**3.57%**
Total error in autocorrelation within full series	**1.38%**
Total error for entire table	**3.40%**

7.4.2 Cross-correlations between stations

As a second measure, cross correlations between stations were computed and compared. Again, this was first done for separate months, i.e. correlation of Novembers in Station 1 with Novembers in Station 2, and then for the time series as a whole, expressing correlation of development of flowrates between the two examined stations throughout he whole observed period. These coefficients were again computed for the historical time series, for 1000 of synthetic time series, construction average values of them and then were these pairs of correlation coefficients compared by the two-sample correlation significance test from section 7.3.3.

Four tables were constructed by the same manner as in previous section 7.4.1, but due to the large size of the tables (with 13 stations there is $\binom{13}{2}$ possible combinations, that is 78 relationships), they are included as appendices – namely in Appendix 5 (pages XIV – XVI).

Table 7.8 shows cross-correlation structure for the historical data and Table 7.9 shows averages of cross-correlation coefficients for 1000 generated time series. Row labels tell between which two stations is the corresponding relationship, columns represent months in which the relationship is examined. Final column represents correlation between full time series. The table is colour-scaled, but differently than in Tables 7.4 and 7.5, as here, in Table 7.8 and 7.9 the correlations are always positive, which is understandable considering they represent spatial dependence in one watershed and being strongly tied to weather behaviour in the area. Red colour is for strongest positive correlations, blue colour for least strong relationships and white colour is used as transition colour, signifying average strength of dependence. Higher saturation of red or blue colour means higher proximity to maximum or minimum values respectively.

There is very clearly observable similarity between the structures for historical data and synthetic data. Overall, most of the correlations are very strong. There are observable patterns, showing for example weaker correlations for winter and spring months especially among first few stations and particularly strong correlations in months May, July and September in all of the profile relationships. While all values remain positive, naturally all strengths can appear, as some stations used here for verification are quite distant from each other and their relationship can lack the spatial dependence at all, leaving only the temporal dependence, which is with greater distance arguably also weaker. This means that no strength of dependence is suspicious, unless it is preserved (not reduced nor boosted) in the synthetic data. Altogether, we can safely say, that the structures are not apparently different, signifying fair performance of the PCA method. Evaluation of significance of discovered difference follows in next two paragraphs.

The next two tables in Appendix 5, Tables 7.10 and 7.11 show percentages of times series where the particular correlations were significantly different between the historical and synthetic data and the average p values upon which the null hypothesis assuming insignificant difference was usually rejected, if so, counting again only the cases it was rejected. Table 7.10 is colour-scaled (unlike in Table 7.6) with more saturated red colour signifying higher number of different outcomes. It is very clear from the table that the most "bad-behaving" month is July, and May follows. The results for July are unpleasant, showing that the cross correlations were significantly different in majority of cases. This can again signify systematic error of the model and/or certain unpredictability of the flows in month of July. It was previously discussed in section 7.3, that July is the month most likely to be unpredictable, along with discussion of results in 7.2, where July was also showing highest rate of error. There is however another thing in play, which is the nature of Fisher's transformation, because of which the differences between high correlation coefficients close to 1 are rated as much more significant than the differences between weaker correlations. For example, the Fisher's test will reject the null hypothesis for comparation of correlations 0.90 and 0.97, but it will not reject the hypothesis for values 0.2 and 0.5, assuming all values were computed from vectors of the same length. This is arguably right approach considering the definition of correlation coefficient [53], but it might be in some applications, including this one, overestimating the significance of difference between strong correlation relationships.

By making average summary of the table 7.10 we get a total error of the structure. Three values were produced - first as a total error in cross-correlations within months, second as total error in cross-correlations between full time series and third as a total error of the complete structure, making an average from entire table. Results are in Table 7.11 a), and are showing that the total errors are within reasonable limits.

Table 7.11 a)

Total error in cross correlations within months	**7.19%**
Total error in cross correlations between full series	**15.90%**
Total error for entire table	**7.86%**

In Table 7.11 the p values lower than 1% are highlighted with yellow colour, making not very extensive list. This tells us that even if the null hypothesis of the test was rejected in majority of cases in those particular relationships, it was mostly rejected upon p value not very far from the significance level.

7.5 Influence of particular transformations

7.5.1 Importance of the correction techniques

If the definitions of sample mean and sample variance is taken into consideration, it is clear that the correction techniques suppress only the outcome by-product of the stochastic component, not its desired feature directly, and only from the long-term point of view, where for example the variance could differ slightly in realistic solution of 1000 years long and longer synthetic time series , but would also be autonomously corrected by the convergence process, refining the variance, bringing it closer to the population mean, the longer the series is.

8. CONCLUSION

Based on its applicability and features, two approaches for multivariate time series modelling were discussed. The first, method based Principal Component Analysis is much more simple and direct method, having the advantage of closed form computational processes and therefore holding much smaller computational burden. Its disadvantage is, that it theoretically destroys part of the mutual information that the multivariate data contain, because it preserves only raw mutual correlations between stations but not higher order dependencies. The basis is that it searches for transformation that has been designed based on the covariance matrix, which is a low order statistical characteristic of data. The second, method based on Independent component analysis, theoretically preserves even those higher order dependencies, because it extracts from the data more mutual information and is therefore able to reapply this information to independent univariate synthetic time series that were generated individually.

The practical part of this thesis involved construction of the PCA method based multivariate model and evaluation of its performance. Regarding the preservation of the correlation structure the model performed arguably quite well, having total error as a performance measure explained in section 7.4 around 3.40% for the autocorrelation structure of lag 1 of the data set and total error of 7.86% for the cross-correlation structure describing mutual relationships of the multivariate data.

In traditional applications of streamflow data the generated time series did not deviate extensively from expected outcomes, making the model's output usable in some classical water management solutions. However, there were some drawback of the model's performance especially in water reservoir operation solutions, where the model produced data that underestimated storage capacity requirements for longer time series.

LIST OF MATHEMATICAL NOTATIONS

x	random variable
$\mathbf{x}$	vector of random variables
$\mathbf{X}$	matrix
$\mathbf{x}^T/\mathbf{X}^T$	transpose of a vector/matrix
./	element wise division operator, for other operations analogically
E	expected value operator
μ_x	mean of a sample of a population of variable x
$\bar{x}$	mean of a sample of a variable x
σ_x	standard deviation of variable x
κ_x	skewness of variable x
ζ_x	kurtosis of variable x
$\mathbb{P}(\mathrm{A})$	probability of event A
$p(x)$	marginal probability density function of variable x
$p(x,y)$	joint probability density function of variables x and y
$F(x)$	cumulative distribution function of variable x
$\rho_{x,y}$	Pearson correlation coefficient between variables x and y
$\sigma_{x,y}$	covariance between variables x and y, also $cov(x,y)$
$\gamma_{x,x+k}$	covariance of variable x with latter itself, also γ_k
$\mathbf{\Sigma_x}$	covariance matrix of a random vector $\mathbf{x}$
ε_t	white noise element
φ_p	autocorrelation coefficient of order p

BIBLIOGRAPHY

[1] Y. Ainhoren, S. Engelberg, and S. Friedman, “The Cocktail Party Problem [instrumentation notes],” *IEEE Instrum. Meas. Mag.*, vol. 11, no. 3, pp. 44–48, 2008.

[2] F. Aires, A. Chédin, and J.-P. Nadal, “Independent component analysis of multivariate time series: Application to the tropical SST variability,” *J. Geophys. Res. Atmos.*, vol. 105, no. D13, pp. 17437–17455, 2000.

[3] H. Allende, C. Moraga, and R. Salas, “Artificial Neural Networks In Time Series Forecasting: A Comparative Analysis,” *Kybernetika*, vol. 88, no. 6, pp. 685–707, 2002.

[4] J. Anmala, B. Zhang, and R. S. Govindaraju, “Comparison of ANNs and Empirical Approaches for Predicting Watershed Runoff,” *J. Water Resour. Plan. Manag.*, vol. 126, no. 3, pp. 156–166, 2000.

[5] B. Ans, J. Hérault, and C. Juten, “Adaptive neural architectures: detection of primitives. *From original: ‘Architectures neuromimétiques adaptatives: Détection de primitives’* ,” in *COGNITIVA 85*, 1985, pp. 593–597.

[6] A. Baghaie, R. M. D’Souza, and Z. Yu, “Application of Independent Component Analysis techniques in speckle noise reduction of retinal OCT images,” *Opt. - Int. J. Light Electron Opt.*, vol. 127, no. 15, pp. 5783–5791, 2016.

[7] J. Bobin, J. Starck, Y. Moudden, and J. Fadili M, “Blind Source Separation : the Sparsity Revolution,” *Adv. Imaging Electron Phys.*, vol. 152, pp. 221--306, 2008.

[8] M. C. Bryson, “Heavy-Tailed Distributions: Properties and Tests,” *Technometrics*, vol. 16, no. 1, pp. 61–68, 1974.

[9] P. Comon, “Independent component analysis, A new concept?,” *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.

[10] T. M. Cover and J. a. Thomas, *Rate Distortion Theory*. 1991.

[11] R. B. D’Agostino and A. Belanger, “A Suggestion for Using Powerful and Informative Tests of Normality,” *Am. Stat.*, vol. 44, no. 4, pp. 316–321, 1990.

[12] R. A. Fisher, “On the probable error of a coefficient of correlation deduced from a small sample,” *Metron*, vol. 1, pp. 3–32, 1921.

[13] P. Fošumpaur, “Diploma project consultation,” 2017.

[14] P. Fošumpaur, “Water Reservoir Storage Capacity Solution,” *Lecture of masters program subject Water Schemes at Faculty of Civil Engineering, Czech Technical University*, Prague, 2016.

[15] S. A. Frank, “The common patterns of nature,” *Journal of Evolutionary Biology*, vol. 22, no. 8. pp. 1563–1585, 2009.

[16] F. E. Grubbs, “Procedures for Detecting Outlying Observations in Samples,” *Technometrics*, vol. 11, no. issue 1, pp. 1–21, 1969.

[17] J. Hérault and B. Ans, “Circuits neuronaux à synapses modifiables: décodage demessages composites par apprentissage non supervisé,” *C.-R. l’Académie des Sci.*, vol. III, no. 13, pp. 525–528, 1984.

[18] J. Hérault, C. Jutten, and B. Ans, “Détection de grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé,” in *Actes du Xème colloque GRETSI*, 1985, pp. 1017–1022.

[19] K. W. Hipel and A. I. McLeod, *Time series modelling of water resources and environmental systems*. New York: Elsevier, 1994.

[20] P. J. Huber, *Robust Statistics*. 1981.

[21] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: J. Wiley, 2001.

[22] A. Hyvärinen and E. Oja, “Independent Component Analysis: Algorithms and Applications,” *Neural Networks*, vol. 13, no. 45, pp. 411–430, 2000.

[23] E. C. Cherry, “Some experiments on the recognition of speech, with one and with 2

ears," *J. Acoust. Soc. Am.*, vol. 25, no. 5, pp. 975–979, 1953.

[24] A. T. Christian Jutten, "Source Separation: From Dusk Till Dawn," in *Proc. ICA*, 2000, pp. 15–26.

[25] J.R.M. Hosking and J.R.Wallis, *Regional frequency analysis*. 1997.

[26] I. T. Jolliffe, "Principal Component Analysis, Second Edition," *Springer Ser. Stat.*, vol. 98, p. 487, 2002.

[27] C. J. Keylock, "A resampling method for generating synthetic hydrological time series with preservation of cross-correlative structure and higher-order properties," *Water Resour. Res.*, vol. 48, no. 12, 2012.

[28] C. J. Keylock, "Characterizing the structure of nonlinear systems using gradual wavelet reconstruction," *Nonlinear Process. Geophys.*, vol. 17, no. 6, pp. 615–632, 2010.

[29] D. Koutsoyiannis, "Reliability Concepts in Reservoir Design," *Water Encycl. Surf. Agric. Water*, vol. 4, pp. 256–259, Apr. 2005.

[30] E. Kreyszig, H. Kreyszig, and E. J. Norminton, *Advanced Engineering Mathematics*, 10th ed. Hoboken, NJ: John Wiley.

[31] J. Lin, Y. Li, and C. Liu, "Building Time Series Forecasting Model By Independent Component Analysis Mechanism," *Proc. World Congr. Eng.*, vol. II, pp. 4–9, 2007.

[32] D. Machiwal and M. K. Jha, *Hydrologic Time Series Analysis: Theory and Practice*, vol. 53, no. 9. 2012.

[33] A. Marantz, "Basic Principles of Magnetoencephalography." [Online]. Available: http://web.mit.edu/kitmitmeg/whatis.html. [Accessed: 10-Dec-2017].

[34] D. Marton, M. Starý, and P. Menšík, "Water management solution of reservoir storage function under condition of measurement uncertainties in hydrological input data," in *Procedia Engineering*, 2014, vol. 70, pp. 1094–1101.

[35] Mathworks, "Statistics and Machine Learning Toolbox: User's Guide (R2015a)," 2011. [Online]. Available: https://uk.mathworks.com/help/stats/ttest2.html.

[Accessed: 01-Jan-2018].

[36] J. H. McDermott, “The cocktail party problem,” *Curr. Biol.*, vol. 19, no. 22, 2009.

[37] A. M. Mood, F. A. Graybill, and D. C. Boes, “Introduction to the Theory of Statistics,” *Book*, vol. 3, pp. 540–541, 1974.

[38] K. Nacházel, *Stochastické metody ve vodním hospodářství*, 2nd ed. Prague: Czech Technical University, 1993.

[39] K. Nacházel, *Estimation theory in hydrology and water systems*. New York: Elsevier, 1993.

[40] J. C. Ochoa-Rivera, R. Garcia-Bartual, and J. Andreu, “Multivariate synthetic streamflow generation using a hybrid model based on artificial neural networks,” *Hyrology Earth Syst. Sci.*, vol. 6, no. 4, pp. 641–654, 2002.

[41] K. Pearson, “LIII. *On lines and planes of closest fit to systems of points in space*,” *Philos. Mag. Ser. 6*, vol. 2, no. 11, pp. 559–572, 1901.

[42] H. Raman and N. Sunilkumar, “Multivariate modelling of water resources time series using artificial neural networks,” *Hydrol. Sci. J.*, vol. 40, no. 2, pp. 145–163, 1995.

[43] B. P. Sangal and A. K. Biswas, “The 3-Parameter Lognormal Distribution and Its Applications in Hydrology,” *Water Resour. Res.*, vol. 6, no. 2, pp. 505–515, Apr. 1970.

[44] J. V Stone, *Independent Component Analysis: A Tutorial Introduction*. Cambridge, Massachusetts: MIT Press, 2004.

[45] G. Strang, *Introduction to Linear Algebra*, 5th ed. Wellesley: Wellesley - Cambridge Press, 2016.

[46] The National Institute of Standards and Technology, “Kolmogorov-Smirnov Goodness-of-Fit Test,” *NIST/SEMATECH e-Handbook of Statistical Methods*. [Online]. Available: http://www.itl.nist.gov/div898/handbook/eda/section3/eda35g.htm. [Accessed: 06-Dec-2017].

[47] The National Institute of Standards and Technology, “Two-Sample t-Test for Equal Means,” *NIST/SEMATECH e-Handbook of Statistical Methods*. [Online]. Available: http://www.itl.nist.gov/div898/handbook/eda/section3/eda353.htm. [Accessed: 01-Jan-2018].

[48] R. Vidal, Y. Ma, and S. S. Sastry, *Generalized Principal Component Analysis*. New York, NY: Springer, 2016.

[49] L. Votruba and V. Broža, *Hospodaření s vodou v nádržích*. Prague: Praha: SNTL - Nakladatelství technické literatury, 1966.

[50] J. Wang and C. I. Chang, “Applications of Independent Component Analysis in Endmember Extraction and Abundance Quantification for Hyperspectral Imagery,” *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 9, pp. 2601–2616, 2006.

[51] R. Wang, *Introduction to Orthogonal Transforms*. New York: Cambridge University Press, 2012.

[52] L. Wasserman, *All of Statistics*, Corrected. New York: Springer, 2010.

[53] P. H. Westfall and K. S. S. Henning, *Understanding Advanced Statistical Methods*. Boca Raton: CRC Press, Taylor & Francis roup.

[54] S. Westra, C. Brown, U. Lall, and A. Sharma, “Modeling multivariable hydrological series: Principal component analysis or independent component analysis?,” *Water Resour. Res.*, vol. 43, no. 6, pp. 1–11, 2007.

[55] K. L. Wuensch, “Comparing Correlation Coefficients, Slopes, and Intercepts,” 2016. [Online]. Available: http://core.ecu.edu/psyc/wuenschk/docs30/CompareCorrCoeff.pdf. [Accessed: 01-Jun-2018].

[56] F. Yao, J. Coquery, and K.-A. Lê Cao, “Independent Principal Component Analysis for biologically meaningful dimension reduction of large biological data sets,” *BMC Bioinformatics*, vol. 13, p. 24, Feb. 2012.

[57] “Levels and flows on watercourses,” *Povodí Odry, státní podnik*, 2018. [Online]. Available: https://www.pod.cz/portal/SaP/cz/pc/CelkovaMapa.aspx. [Accessed: 06-

Jan-2018].

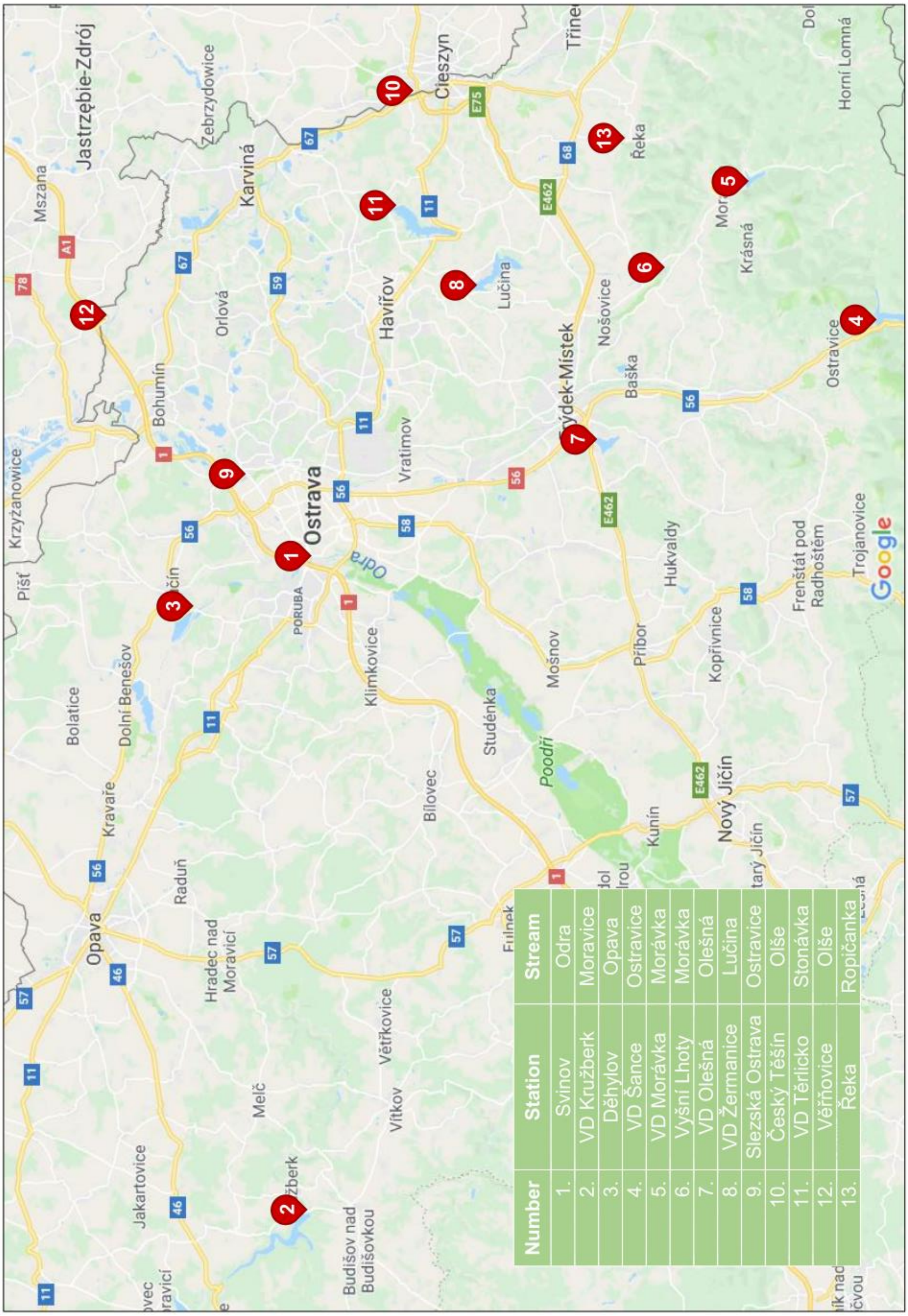
[58] "Water Reservoirs; Přehrady," *Povodí Odry, státní podnik*, 2018. [Online]. Available: https://www.pod.cz/stranka/kruzberk.html(/sance.html,/moravka.html,/olesna.html,/zermanice.html,/terlicko.html)%0A. [Accessed: 06-Jan-2018].

[59] "VYHODNOCENÍ POVODNÍ V KVĚTNU A ČERVNU 2010 NA ÚZEMÍ ČR Evaluation of floods on CZ territory in May and June of 2010," *T. G. Masaryk Water Research Institute, public research institution (VÚV TGM); Czech Hydrometeorological Institute (ČHMÚ)*, 2010. [Online]. Available: http://voda.chmi.cz/pov10/.

[60] "Map of Moravian-Silesian Region," *Google Maps*. [Online]. Available: https://www.google.cz/maps/@49.751634,18.0981573,10.08z. [Accessed: 01-May-2018].

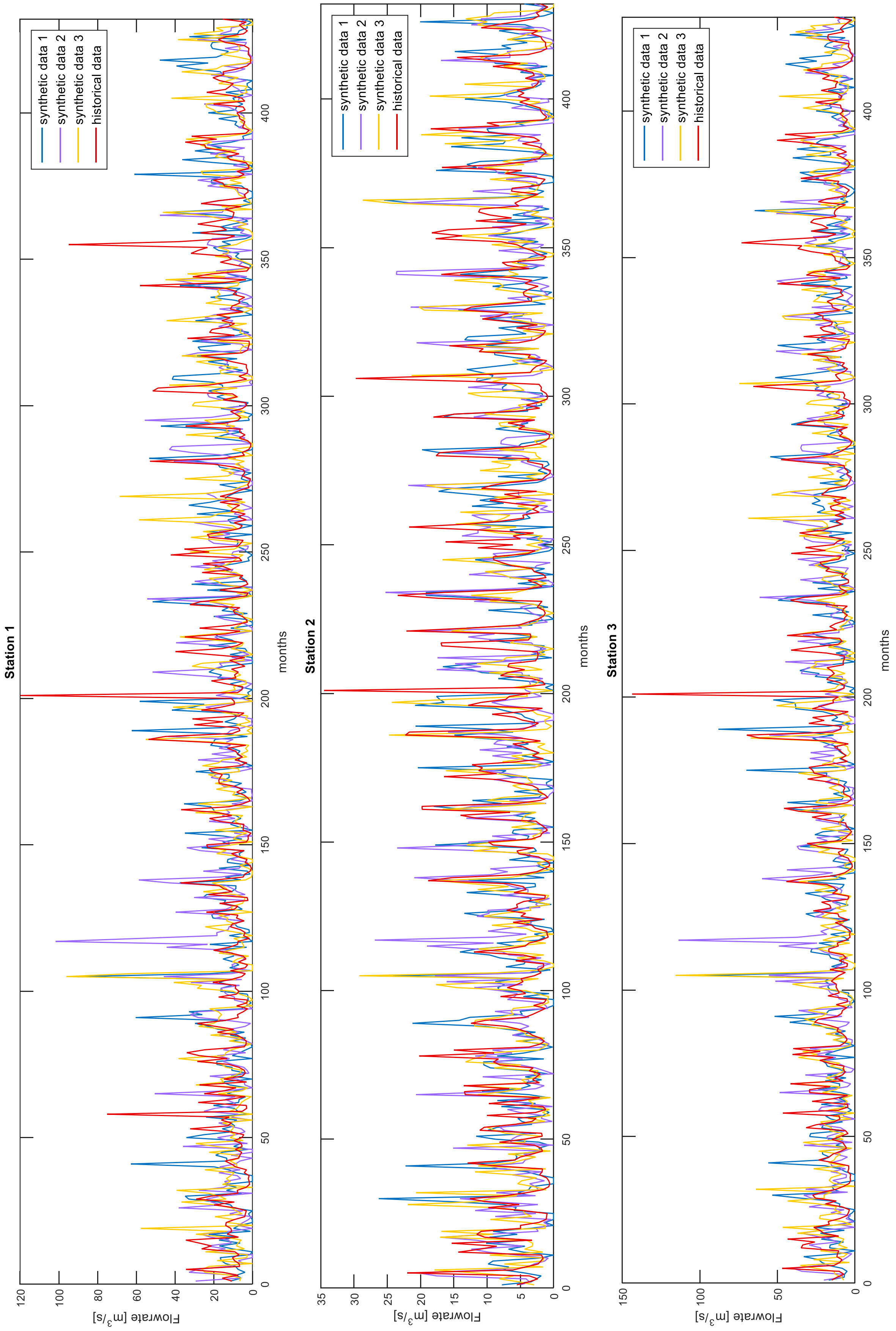
End

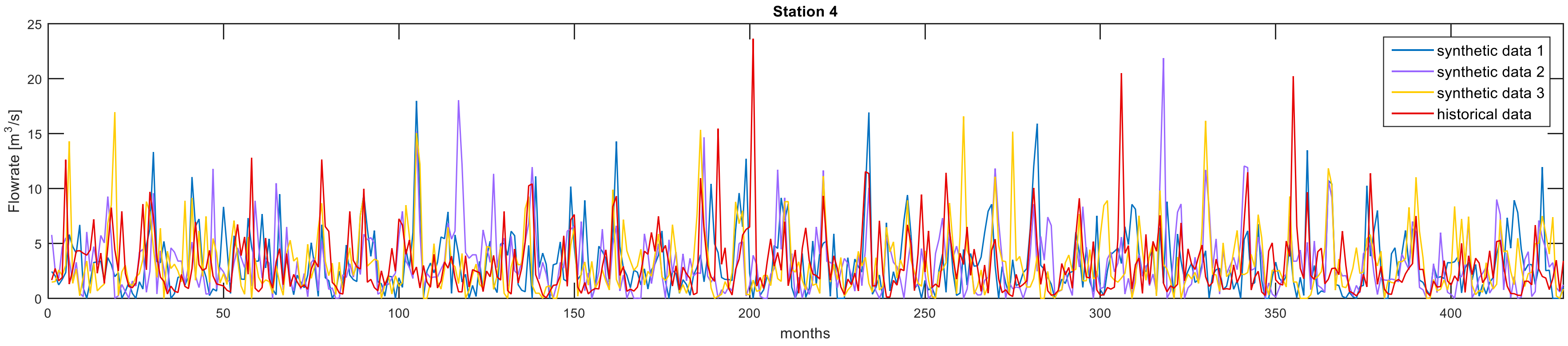
APPENDIX 1; Map with locations of examined stations; [60]

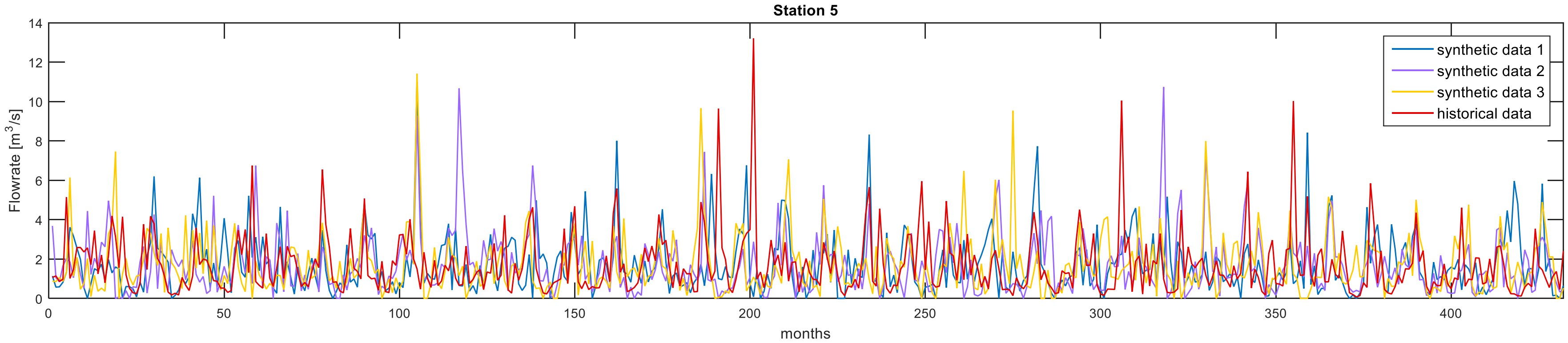


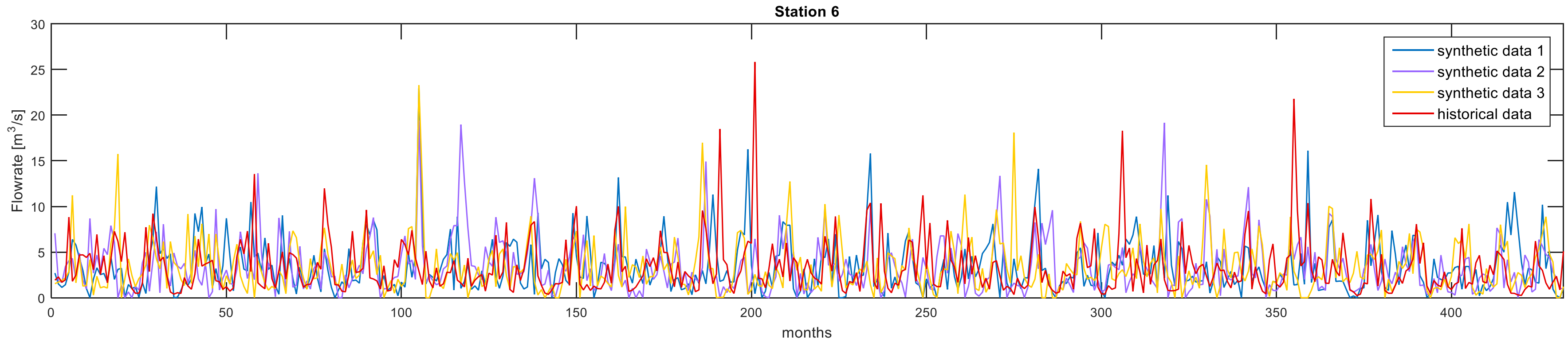
Number	Station	Stream
1.	Svinov	Odra
2.	VD Kružberk	Moravice
3.	Děhylov	Opava
4.	VD Šance	Ostravice
5.	VD Morávka	Morávka
6.	Vyšní Lhoty	Morávka
7.	VD Olešná	Olešná
8.	VD Žermanice	Lučina
9.	Slezská Ostrava	Ostravice
10.	Český Těšín	Olše
11.	VD Těrlicko	Stonávka
12.	Věřňovice	Olše
13.	Řeka	Ropičanka

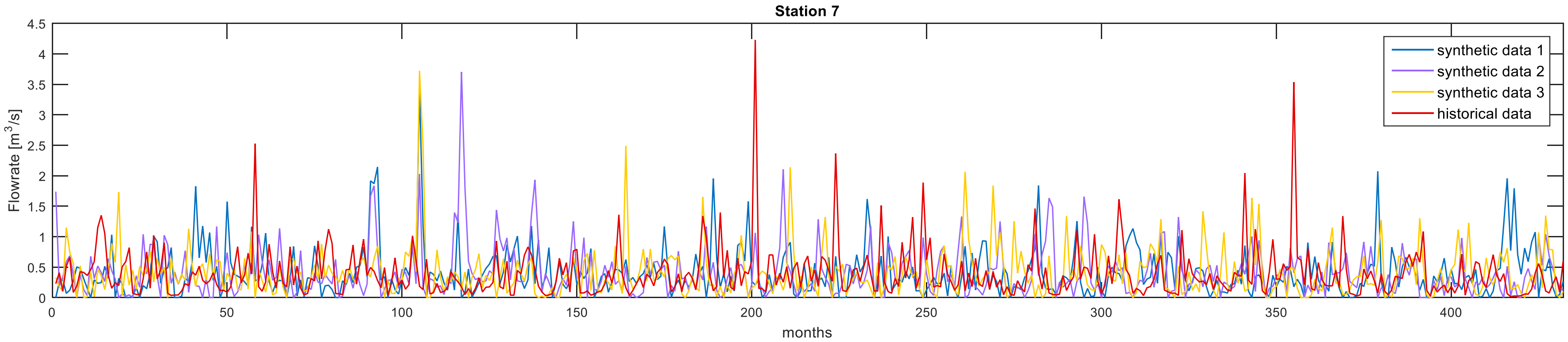
APPENDIX 2; 36 years long synthetic time series

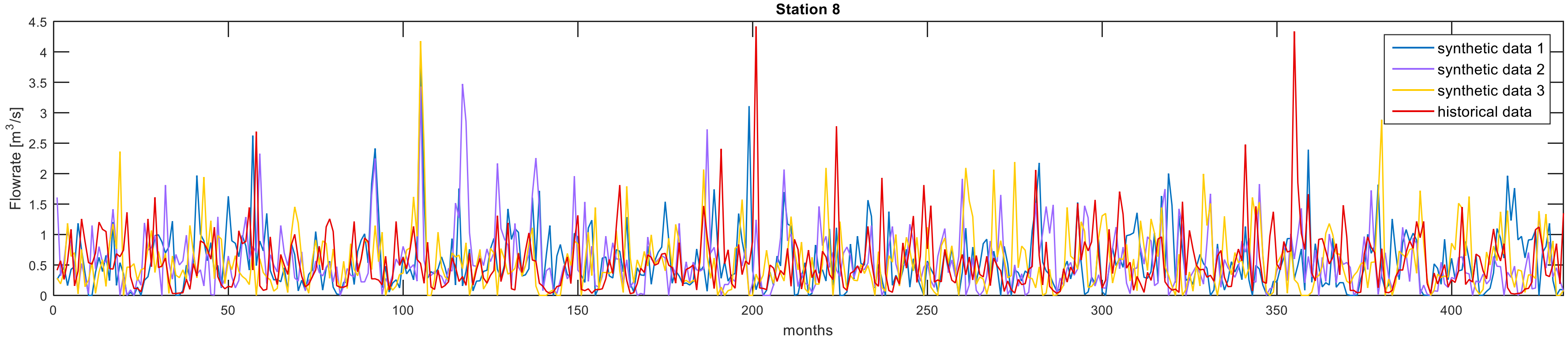


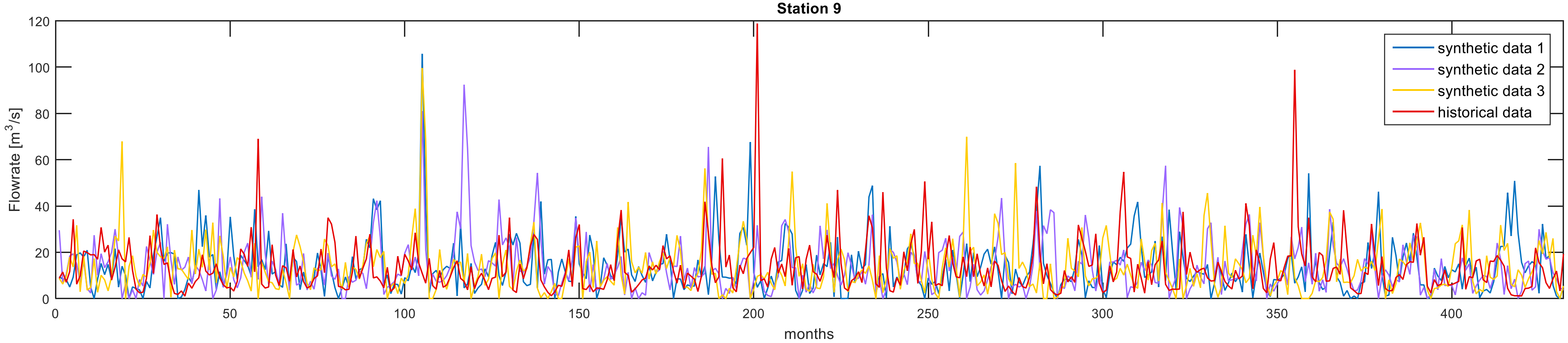

Station 4
Flowrate [m3/s]
months
synthetic data 1
synthetic data 2
synthetic data 3
historical data

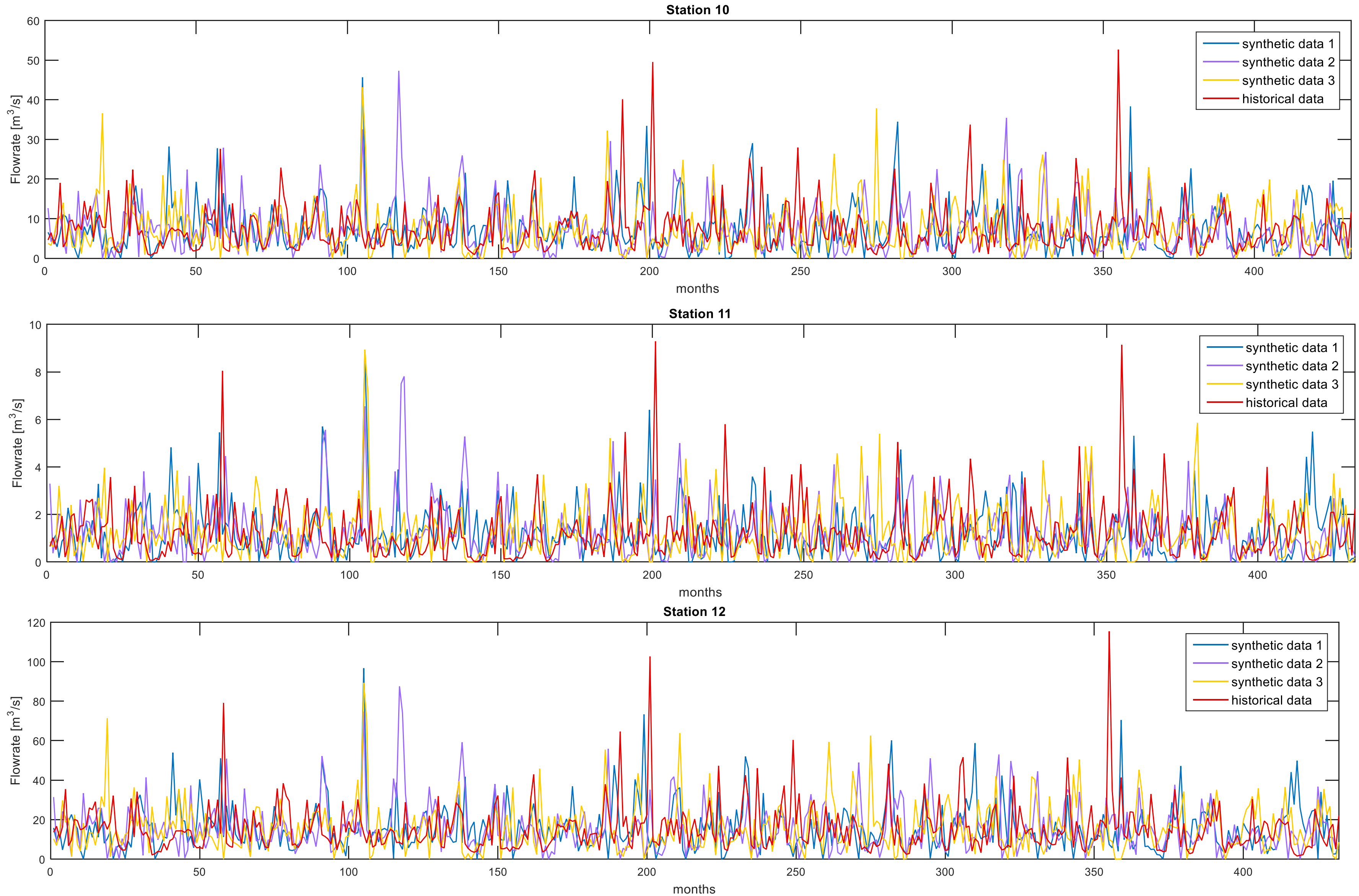

Station 5
Flowrate [m3/s]
months
synthetic data 1
synthetic data 2
synthetic data 3
historical data

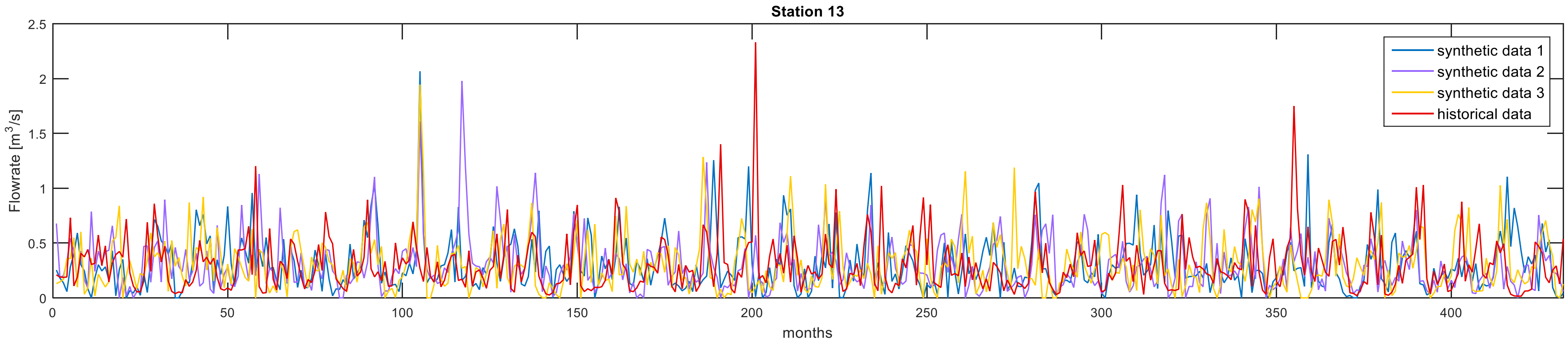

Station 6
Flowrate [m3/s]
months
synthetic data 1
synthetic data 2
synthetic data 3
historical data


Station 7
Flowrate [m^3/s]
months
synthetic data 1
synthetic data 2
synthetic data 3
historical data


Station 8
Flowrate [m^3/s]
months
synthetic data 1
synthetic data 2
synthetic data 3
historical data


Station 9
Flowrate [m^3/s]
months
synthetic data 1
synthetic data 2
synthetic data 3
historical data


Station 10
Flowrate [m^3/s]
months
synthetic data 1
synthetic data 2
synthetic data 3
historical data
Station 11
Flowrate [m^3/s]
months
synthetic data 1
synthetic data 2
synthetic data 3
historical data
Station 12
Flowrate [m^3/s]
months
synthetic data 1
synthetic data 2
synthetic data 3
historical data


Station 13
Flowrate [m3/s]
months
synthetic data 1
synthetic data 2
synthetic data 3
historical data
0
0.5
1
1.5
2
2.5
0
50
100
150
200
250
300
350
400

APPENDIX 4; Maximum monthly averages comparison

Maximum monthly averages for 36 years long simulations [m³/s]													
Stations	**St. 1**	**St. 2**	**St. 3**	**St. 4**	**St. 5**	**St. 6**	**St. 7**	**St. 8**	**St. 9**	**St. 10**	**St. 11**	**St. 12**	**St. 13**
Historic TS	**119.82**	**34.50**	**143.65**	**23.64**	**13.20**	**25.83**	**4.22**	**4.42**	**118.82**	**52.65**	**9.28**	**115.37**	**2.33**
Synthetic TS (Run 1)	84.10	31.16	70.28	17.91	11.02	21.27	3.01	4.04	89.45	45.26	7.10	86.79	1.81
Synthetic TS (Run 2)	100.06	29.45	112.58	15.15	9.34	16.28	3.70	2.89	87.96	35.24	6.95	74.19	2.06
Synthetic TS (Run 3)	96.24	29.24	104.98	22.39	12.64	24.31	3.64	4.40	113.14	49.61	9.50	102.16	2.10
Synthetic TS (Run 4)	69.77	24.12	74.29	18.60	10.27	19.22	2.35	3.03	71.54	39.60	6.19	77.99	1.64
Synthetic TS (Run 5)	75.61	23.10	75.74	17.94	7.97	15.41	2.73	3.20	78.73	32.22	7.12	70.01	1.40
Synthetic TS (Run 6)	103.65	35.04	123.42	22.40	11.36	22.94	3.69	4.51	113.54	42.41	10.34	95.42	1.96
Synthetic TS (Run 7)	67.42	28.75	94.88	19.99	10.32	19.05	2.73	3.47	75.10	41.04	6.84	79.07	1.71
Synthetic TS (Run 8)	95.73	28.23	126.17	20.65	11.88	21.78	3.34	3.83	94.67	46.12	9.24	95.92	1.87
Synthetic TS (Run 9)	65.49	24.64	84.30	17.94	9.86	16.59	2.09	2.82	63.29	28.68	6.59	62.80	1.31
Synthetic TS (Run 10)	82.54	27.74	84.82	21.53	11.29	23.55	3.36	3.84	100.88	52.44	8.60	106.98	2.05
E [Synthetic/Historical]	**70%**	**82%**	**66%**	**82%**	**80%**	**78%**	**73%**	**82%**	**75%**	**78%**	**85%**	**74%**	**77%**

Maximum monthly averages for **100** years long simulations [m³/s]													
Stations	**St. 1**	**St. 2**	**St. 3**	**St. 4**	**St. 5**	**St. 6**	**St. 7**	**St. 8**	**St. 9**	**St. 10**	**St. 11**	**St. 12**	**St. 13**
Historic TS	**119.82**	**34.50**	**143.65**	**23.64**	**13.20**	**25.83**	**4.22**	**4.42**	**118.82**	**52.65**	**9.28**	**115.37**	**2.33**
Synthetic TS (Run 1)	119.82	38.11	148.30	21.79	11.17	21.23	4.08	4.97	120.98	53.29	11.04	105.46	2.05
Synthetic TS (Run 2)	119.77	35.50	123.07	30.30	15.97	35.58	4.41	6.25	142.96	81.85	13.29	159.15	2.76
Synthetic TS (Run 3)	139.37	32.00	110.25	18.44	9.56	21.88	4.47	4.93	109.55	44.91	11.32	123.03	1.64
Synthetic TS (Run 4)	108.90	35.84	110.49	19.83	12.69	26.00	3.62	3.67	102.89	55.00	9.07	110.46	1.87
Synthetic TS (Run 5)	124.72	32.62	102.46	24.93	14.51	29.01	4.27	4.24	106.58	66.90	9.20	121.27	2.69
Synthetic TS (Run 6)	121.91	42.10	148.38	21.74	12.04	21.67	4.49	5.11	120.29	48.28	12.30	107.46	2.28
Synthetic TS (Run 7)	107.49	30.36	117.24	27.74	13.24	25.97	3.08	3.59	84.11	50.51	8.03	83.02	1.63
Synthetic TS (Run 8)	166.33	43.24	175.67	30.61	13.26	25.66	5.67	5.02	153.42	53.77	11.96	120.99	2.00
Synthetic TS (Run 9)	119.95	30.80	109.56	33.61	17.97	38.30	4.45	6.68	161.18	81.94	14.07	166.52	2.98
Synthetic TS (Run 10)	89.58	37.64	117.93	25.16	16.67	35.26	4.00	5.35	138.94	59.79	11.45	130.18	2.75
E [Synthetic/Historical]	**102%**	**104%**	**88%**	**108%**	**104%**	**109%**	**101%**	**113%**	**104%**	**113%**	**120%**	**106%**	**97%**

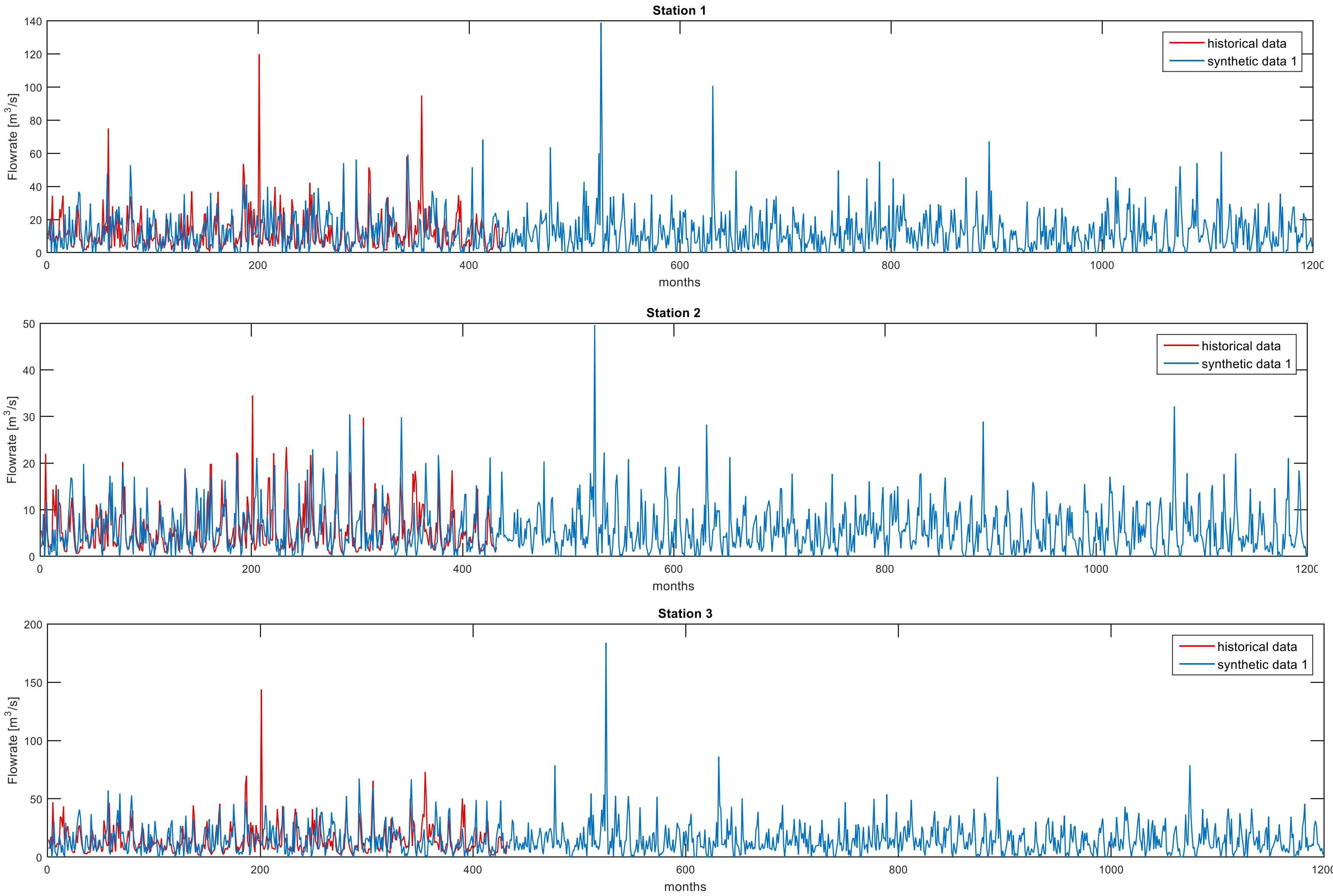
Maximum monthly averages for **1000** years long simulations [m³/s]													
Stations	**St. 1**	**St. 2**	**St. 3**	**St. 4**	**St. 5**	**St. 6**	**St. 7**	**St. 8**	**St. 9**	**St. 10**	**St. 11**	**St. 12**	**St. 13**
Historic TS	**119.82**	**34.50**	**143.65**	**23.64**	**13.20**	**25.83**	**4.22**	**4.42**	**118.82**	**52.65**	**9.28**	**115.37**	**2.33**
Synthetic TS (Run 1)	167.79	48.98	138.72	40.78	24.33	49.89	6.02	7.93	172.69	88.31	15.23	170.25	4.05
Synthetic TS (Run 2)	259.81	46.03	237.81	40.16	24.97	45.59	8.10	7.86	176.13	108.99	14.69	178.57	4.20
Synthetic TS (Run 3)	175.00	42.28	194.08	47.24	21.85	38.08	6.72	8.55	179.07	89.76	25.25	238.84	3.76
Synthetic TS (Run 4)	195.49	59.13	202.51	43.96	18.11	39.66	6.59	7.81	179.03	87.06	16.98	168.09	3.88
Synthetic TS (Run 5)	163.79	40.95	147.90	29.68	17.90	39.49	5.71	7.53	152.55	64.58	15.13	162.53	3.61
Synthetic TS (Run 6)	156.97	51.92	164.61	41.03	18.28	38.06	6.18	6.80	167.36	79.33	16.37	163.56	2.94
Synthetic TS (Run 7)	172.47	43.23	172.48	40.41	19.18	31.95	6.15	6.34	168.51	79.26	13.59	155.83	2.67
Synthetic TS (Run 8)	197.35	59.24	227.65	33.86	17.50	37.18	6.92	7.68	203.82	70.68	19.19	157.26	3.41
Synthetic TS (Run 9)	137.50	43.14	176.67	34.89	21.32	43.26	5.58	7.48	189.60	82.01	15.08	181.98	3.28
Synthetic TS (Run 10)	174.80	46.06	183.23	39.99	24.11	50.51	6.74	9.01	203.65	103.26	17.93	195.71	3.96
E [Synthetic/Historical]	**150%**	**139%**	**129%**	**166%**	**157%**	**160%**	**153%**	**174%**	**151%**	**162%**	**183%**	**154%**	**154%**

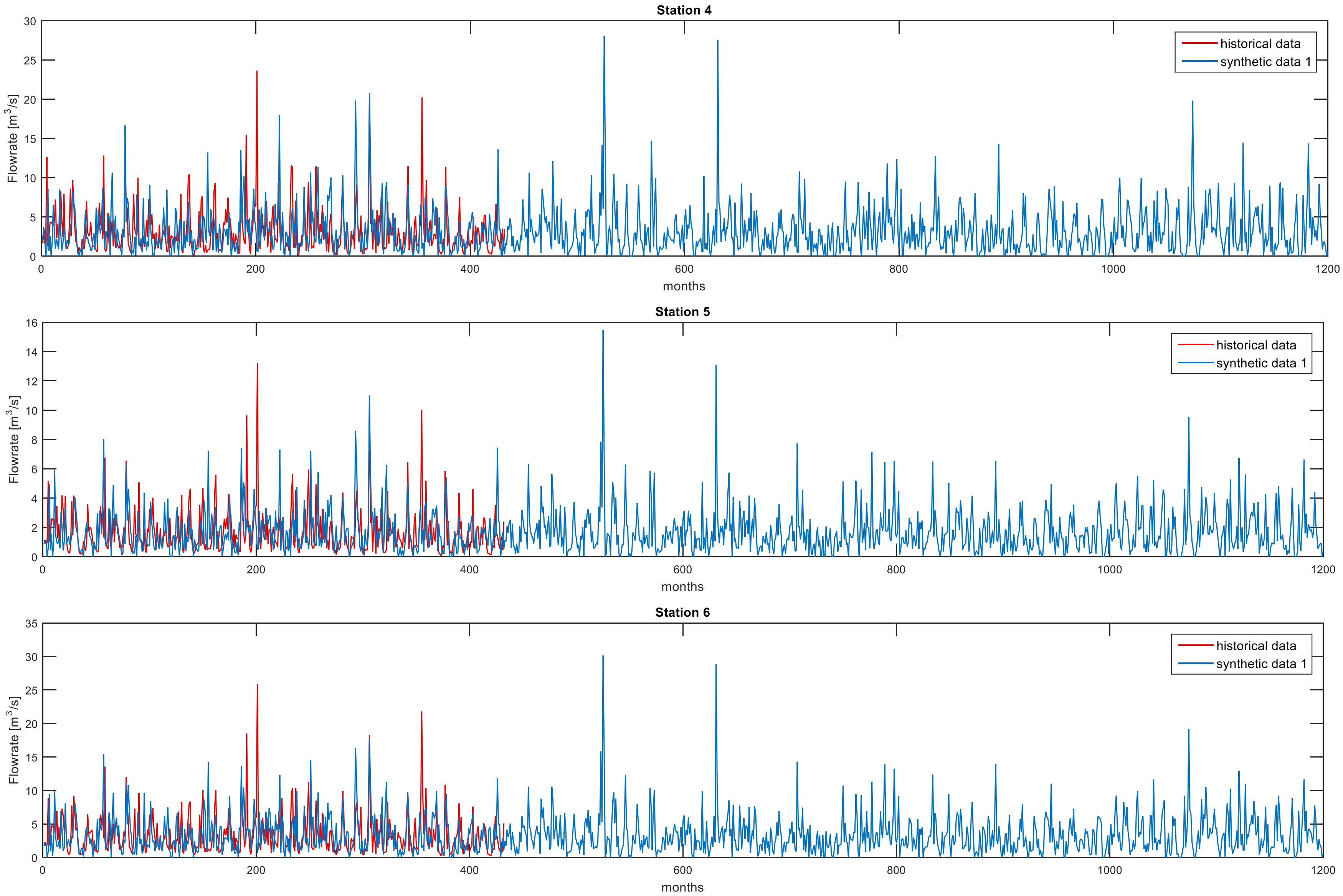
Maximum monthly averages for **10 000** years long simulations [m³/s]													
Stations	**St. 1**	**St. 2**	**St. 3**	**St. 4**	**St. 5**	**St. 6**	**St. 7**	**St. 8**	**St. 9**	**St. 10**	**St. 11**	**St. 12**	**St. 13**
Historic TS	**119.82**	**34.50**	**143.65**	**23.64**	**13.20**	**25.83**	**4.22**	**4.42**	**118.82**	**52.65**	**9.28**	**115.37**	**2.33**
Synthetic TS (Run 1)	193.98	64.83	232.00	46.78	24.56	49.11	7.16	11.09	211.81	108.92	20.57	205.51	5.45
Synthetic TS (Run 2)	220.91	74.64	262.48	59.66	24.30	51.98	8.60	9.29	220.76	140.92	18.38	228.48	4.42
Synthetic TS (Run 3)	271.49	84.01	334.20	60.40	30.82	56.57	11.00	8.95	289.40	99.32	19.63	217.89	4.25
Synthetic TS (Run 4)	230.53	68.59	248.45	52.28	32.22	59.71	8.31	9.56	257.71	125.66	24.36	252.93	4.29
Synthetic TS (Run 5)	205.62	59.55	172.31	44.69	31.04	58.44	6.93	11.53	203.72	139.62	23.49	226.62	4.85
Synthetic TS (Run 6)	380.90	94.04	422.44	74.46	37.30	67.32	14.87	10.66	374.94	125.45	25.51	296.66	5.61
Synthetic TS (Run 7)	180.58	52.67	204.28	49.98	24.63	51.89	8.88	9.93	241.52	137.25	20.93	292.37	5.71
Synthetic TS (Run 8)	198.52	78.07	301.19	52.62	29.28	65.00	10.16	10.99	260.21	129.69	22.29	218.71	6.07
Synthetic TS (Run 9)	174.43	56.65	216.34	49.50	23.08	49.07	8.76	9.30	219.42	122.42	24.07	221.16	4.64
Synthetic TS (Run 10)	419.34	77.57	292.72	76.39	36.20	67.74	15.10	17.13	418.08	182.48	35.21	365.50	5.89
E [Synthetic/Historical]	**207%**	**206%**	**187%**	**240%**	**222%**	**223%**	**236%**	**245%**	**227%**	**249%**	**253%**	**219%**	**220%**

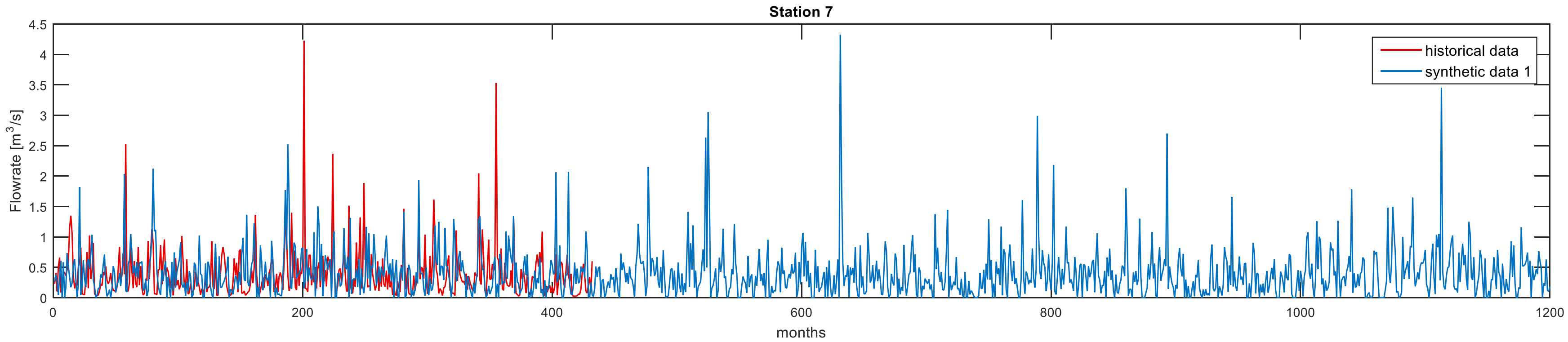
APPENDIX 5; Cross-correlation test results; *Table 7.8*

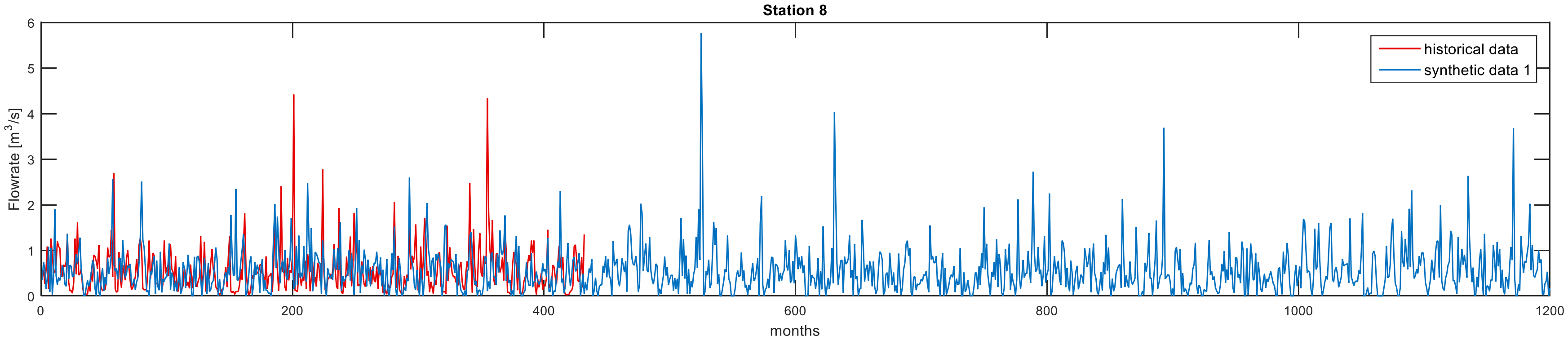
CROSS-CORRELATIONS for HISTORICAL TIME SERIES															
Compared st.			**XI**	**XII**	**I**	**II**	**III**	**IV**	**V**	**VI**	**VII**	**VIII**	**IX**	**X**	**Q**
1	x	2	0.67	0.76	0.81	0.68	0.60	0.83	0.79	0.84	0.98	0.86	0.84	0.93	0.77
1	x	3	0.86	0.92	0.91	0.83	0.88	0.92	0.90	0.91	0.98	0.95	0.96	0.96	0.92
1	x	4	0.68	0.60	0.58	0.42	0.49	0.66	0.92	0.60	0.95	0.83	0.82	0.86	0.76
1	x	5	0.64	0.63	0.57	0.36	0.41	0.66	0.89	0.50	0.95	0.84	0.80	0.70	0.73
1	x	6	0.66	0.67	0.60	0.44	0.59	0.70	0.92	0.73	0.94	0.83	0.82	0.71	0.78
1	x	7	0.69	0.83	0.84	0.71	0.93	0.87	0.96	0.73	0.97	0.94	0.94	0.78	0.89
1	x	8	0.74	0.79	0.82	0.63	0.85	0.80	0.91	0.75	0.92	0.86	0.87	0.68	0.84
1	x	9	0.76	0.84	0.82	0.71	0.84	0.85	0.95	0.78	0.96	0.92	0.89	0.84	0.88
1	x	10	0.70	0.76	0.73	0.65	0.79	0.77	0.92	0.72	0.91	0.85	0.79	0.73	0.83
1	x	11	0.76	0.83	0.85	0.78	0.86	0.81	0.90	0.73	0.90	0.93	0.89	0.76	0.85
1	x	12	0.78	0.83	0.85	0.80	0.87	0.83	0.94	0.77	0.93	0.94	0.85	0.76	0.88
1	x	13	0.72	0.77	0.65	0.63	0.69	0.72	0.89	0.80	0.93	0.85	0.85	0.60	0.81
2	x	3	0.90	0.90	0.93	0.93	0.88	0.95	0.95	0.95	0.99	0.94	0.82	0.97	0.90
2	x	4	0.51	0.47	0.73	0.81	0.78	0.83	0.68	0.56	0.91	0.71	0.60	0.77	0.77
2	x	5	0.45	0.50	0.66	0.76	0.70	0.81	0.67	0.38	0.90	0.73	0.54	0.55	0.70
2	x	6	0.46	0.48	0.65	0.76	0.71	0.82	0.67	0.52	0.90	0.75	0.59	0.54	0.69
2	x	7	0.31	0.40	0.52	0.21	0.41	0.63	0.65	0.39	0.91	0.75	0.70	0.63	0.54
2	x	8	0.34	0.40	0.65	0.27	0.44	0.58	0.57	0.40	0.88	0.75	0.65	0.47	0.54
2	x	9	0.49	0.50	0.72	0.74	0.69	0.83	0.65	0.47	0.91	0.81	0.65	0.68	0.68
2	x	10	0.49	0.51	0.69	0.81	0.71	0.81	0.60	0.53	0.86	0.74	0.53	0.56	0.70
2	x	11	0.33	0.41	0.63	0.38	0.41	0.56	0.56	0.36	0.84	0.79	0.66	0.58	0.52
2	x	12	0.42	0.48	0.70	0.68	0.66	0.76	0.60	0.52	0.87	0.79	0.60	0.62	0.65
2	x	13	0.37	0.57	0.62	0.69	0.69	0.64	0.60	0.48	0.89	0.74	0.60	0.39	0.64
3	x	4	0.67	0.50	0.61	0.71	0.73	0.79	0.77	0.49	0.92	0.85	0.81	0.81	0.80
3	x	5	0.62	0.55	0.60	0.66	0.67	0.79	0.77	0.35	0.92	0.87	0.80	0.66	0.78
3	x	6	0.64	0.57	0.61	0.69	0.76	0.81	0.78	0.54	0.91	0.87	0.83	0.66	0.80
3	x	7	0.58	0.66	0.65	0.38	0.72	0.72	0.79	0.49	0.92	0.87	0.93	0.70	0.75
3	x	8	0.60	0.60	0.69	0.41	0.72	0.70	0.73	0.50	0.88	0.83	0.86	0.59	0.73
3	x	9	0.72	0.68	0.75	0.77	0.86	0.88	0.78	0.54	0.92	0.93	0.88	0.77	0.83
3	x	10	0.66	0.63	0.67	0.79	0.85	0.82	0.73	0.54	0.87	0.86	0.78	0.66	0.79
3	x	11	0.61	0.64	0.70	0.54	0.70	0.68	0.72	0.49	0.85	0.88	0.89	0.69	0.71
3	x	12	0.65	0.67	0.76	0.76	0.86	0.81	0.75	0.57	0.88	0.90	0.83	0.71	0.79
3	x	13	0.61	0.69	0.62	0.71	0.81	0.74	0.74	0.61	0.90	0.88	0.85	0.53	0.79
4	x	5	0.95	0.94	0.95	0.98	0.96	0.98	0.97	0.87	0.98	0.97	0.98	0.88	0.96
4	x	6	0.95	0.94	0.95	0.98	0.95	0.97	0.98	0.85	0.97	0.97	0.98	0.88	0.95
4	x	7	0.73	0.55	0.51	0.18	0.36	0.53	0.94	0.45	0.96	0.87	0.90	0.85	0.69
4	x	8	0.82	0.63	0.72	0.27	0.45	0.57	0.94	0.61	0.95	0.90	0.93	0.83	0.74
4	x	9	0.94	0.85	0.87	0.86	0.81	0.91	0.98	0.75	0.98	0.97	0.97	0.96	0.90
4	x	10	0.95	0.88	0.90	0.90	0.80	0.92	0.98	0.82	0.96	0.94	0.98	0.89	0.92
4	x	11	0.79	0.67	0.67	0.30	0.41	0.53	0.91	0.55	0.94	0.91	0.95	0.89	0.72
4	x	12	0.90	0.77	0.77	0.70	0.69	0.82	0.97	0.75	0.97	0.93	0.96	0.90	0.86
4	x	13	0.87	0.84	0.88	0.80	0.81	0.76	0.94	0.62	0.96	0.95	0.94	0.75	0.86
5	x	6	0.99	0.98	0.98	0.98	0.95	0.98	0.99	0.85	1.00	0.99	0.99	0.98	0.98
5	x	7	0.75	0.58	0.51	0.16	0.32	0.56	0.92	0.49	0.96	0.87	0.90	0.80	0.71
5	x	8	0.87	0.64	0.71	0.27	0.45	0.64	0.93	0.56	0.97	0.90	0.94	0.88	0.78
5	x	9	0.94	0.87	0.87	0.85	0.78	0.93	0.97	0.75	0.98	0.97	0.98	0.93	0.92
5	x	10	0.95	0.91	0.88	0.90	0.77	0.94	0.96	0.82	0.98	0.97	0.99	0.95	0.93
5	x	11	0.82	0.70	0.66	0.31	0.41	0.60	0.92	0.56	0.95	0.91	0.95	0.90	0.76
5	x	12	0.91	0.79	0.77	0.69	0.66	0.86	0.94	0.73	0.99	0.95	0.97	0.93	0.87
5	x	13	0.91	0.89	0.93	0.82	0.82	0.83	0.96	0.69	0.98	0.96	0.97	0.87	0.91
6	x	7	0.76	0.63	0.58	0.27	0.53	0.61	0.95	0.72	0.97	0.87	0.92	0.83	0.78
6	x	8	0.88	0.69	0.75	0.37	0.62	0.68	0.96	0.84	0.98	0.93	0.96	0.90	0.84
6	x	9	0.96	0.91	0.90	0.91	0.91	0.95	0.99	0.94	0.99	0.98	0.98	0.95	0.96
6	x	10	0.97	0.93	0.93	0.93	0.90	0.95	0.97	0.95	0.98	0.96	0.98	0.97	0.95
6	x	11	0.86	0.76	0.70	0.40	0.60	0.64	0.94	0.79	0.96	0.92	0.97	0.94	0.82
6	x	12	0.93	0.84	0.81	0.76	0.82	0.89	0.96	0.92	0.99	0.94	0.97	0.95	0.91
6	x	13	0.94	0.92	0.95	0.87	0.90	0.84	0.96	0.88	0.99	0.96	0.97	0.92	0.94
7	x	8	0.85	0.85	0.82	0.74	0.94	0.89	0.96	0.88	0.96	0.89	0.95	0.84	0.91
7	x	9	0.87	0.87	0.82	0.61	0.81	0.80	0.98	0.88	0.99	0.93	0.97	0.92	0.91
7	x	10	0.77	0.77	0.75	0.50	0.79	0.72	0.96	0.76	0.96	0.90	0.89	0.89	0.83
7	x	11	0.87	0.92	0.91	0.91	0.94	0.94	0.95	0.89	0.95	0.94	0.97	0.91	0.93
7	x	12	0.81	0.87	0.85	0.76	0.88	0.85	0.98	0.85	0.97	0.95	0.92	0.90	0.91
7	x	13	0.79	0.69	0.68	0.56	0.68	0.68	0.93	0.82	0.96	0.85	0.94	0.76	0.83
8	x	9	0.93	0.87	0.90	0.64	0.86	0.81	0.97	0.94	0.98	0.94	0.97	0.91	0.93
8	x	10	0.89	0.83	0.86	0.52	0.84	0.75	0.97	0.84	0.97	0.93	0.94	0.95	0.89
8	x	11	0.96	0.94	0.91	0.85	0.97	0.95	0.99	0.97	0.98	0.97	0.99	0.97	0.97
8	x	12	0.94	0.93	0.91	0.68	0.90	0.86	0.98	0.92	0.97	0.95	0.96	0.90	0.93
8	x	13	0.92	0.76	0.81	0.59	0.80	0.80	0.96	0.88	0.98	0.91	0.96	0.85	0.90
9	x	10	0.97	0.96	0.95	0.96	0.97	0.96	0.98	0.93	0.97	0.96	0.97	0.97	0.96
9	x	11	0.93	0.94	0.90	0.71	0.85	0.81	0.96	0.92	0.97	0.96	0.98	0.96	0.92
9	x	12	0.95	0.96	0.95	0.94	0.96	0.95	0.99	0.97	0.98	0.98	0.98	0.96	0.97
9	x	13	0.94	0.91	0.92	0.94	0.92	0.88	0.96	0.92	0.98	0.96	0.97	0.85	0.95
10	x	11	0.88	0.88	0.87	0.63	0.82	0.74	0.96	0.82	0.95	0.92	0.93	0.97	0.87
10	x	12	0.96	0.95	0.93	0.91	0.96	0.95	0.99	0.96	0.99	0.97	0.99	0.98	0.97
10	x	13	0.94	0.95	0.94	0.94	0.92	0.85	0.96	0.89	0.98	0.95	0.96	0.88	0.93
11	x	12	0.93	0.95	0.95	0.84	0.91	0.88	0.96	0.92	0.96	0.97	0.96	0.95	0.94
11	x	13	0.91	0.83	0.79	0.70	0.75	0.78	0.97	0.91	0.95	0.91	0.96	0.89	0.88
12	x	13	0.95	0.88	0.85	0.91	0.87	0.85	0.95	0.92	0.97	0.94	0.97	0.86	0.92

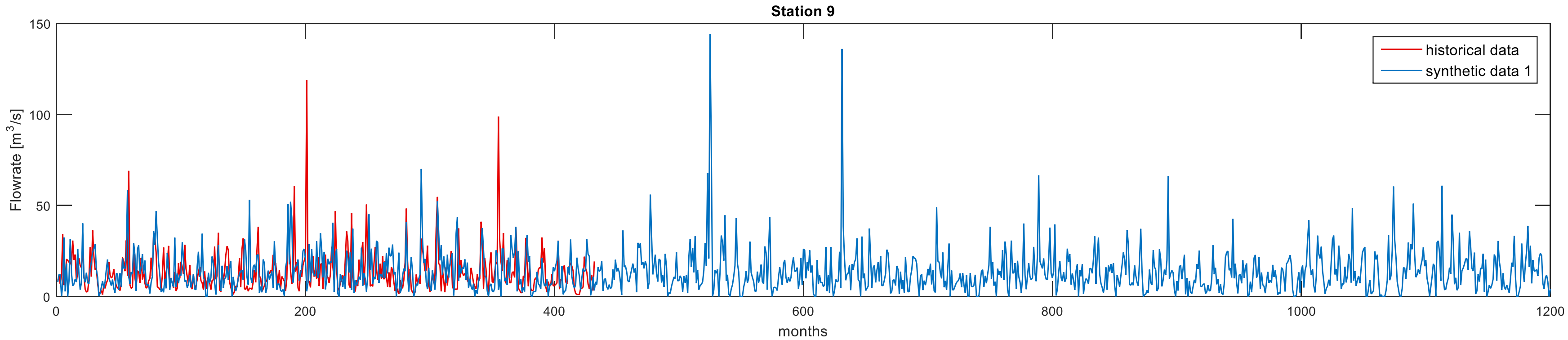
APPENDIX 3; 100 years long synthetic time series

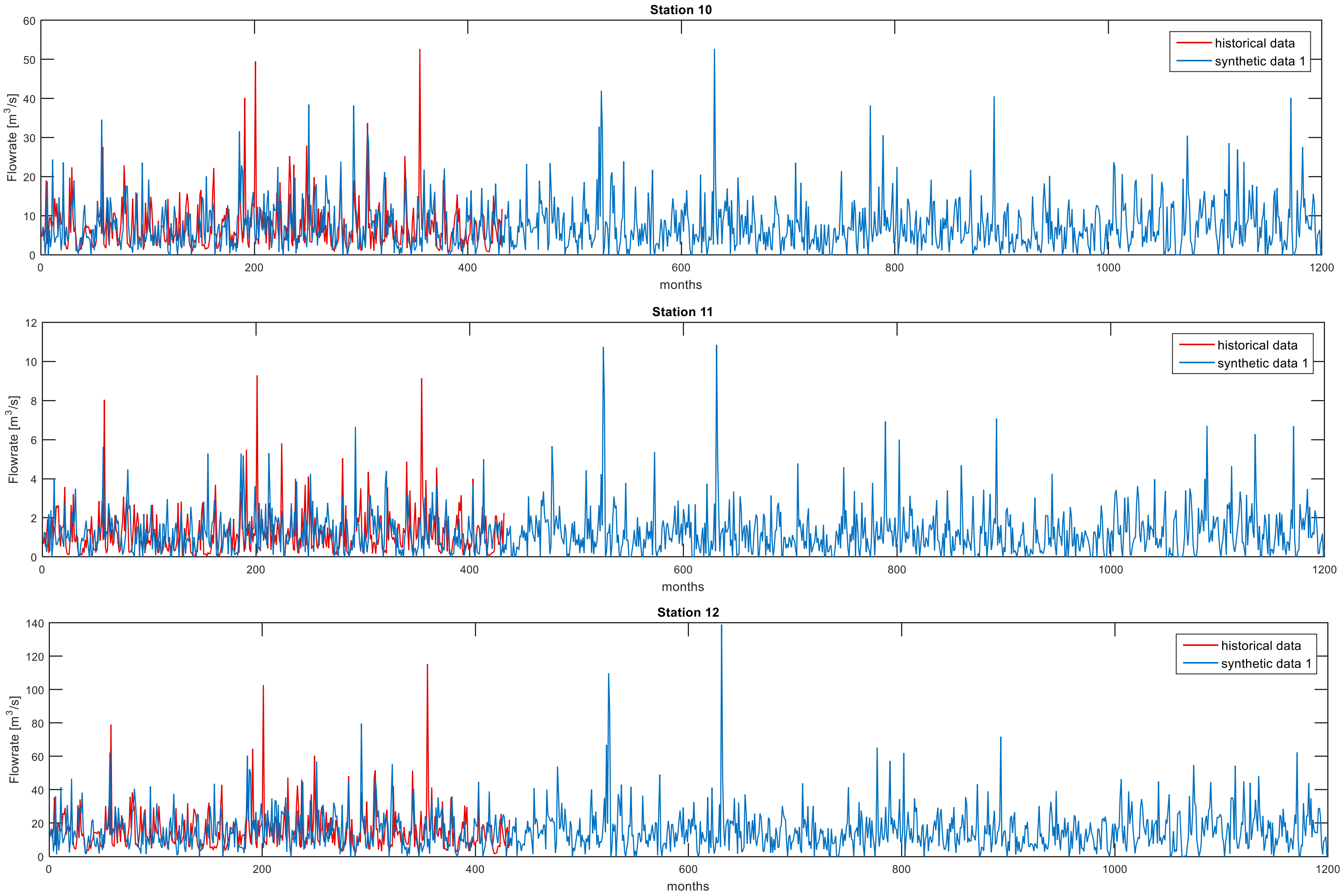


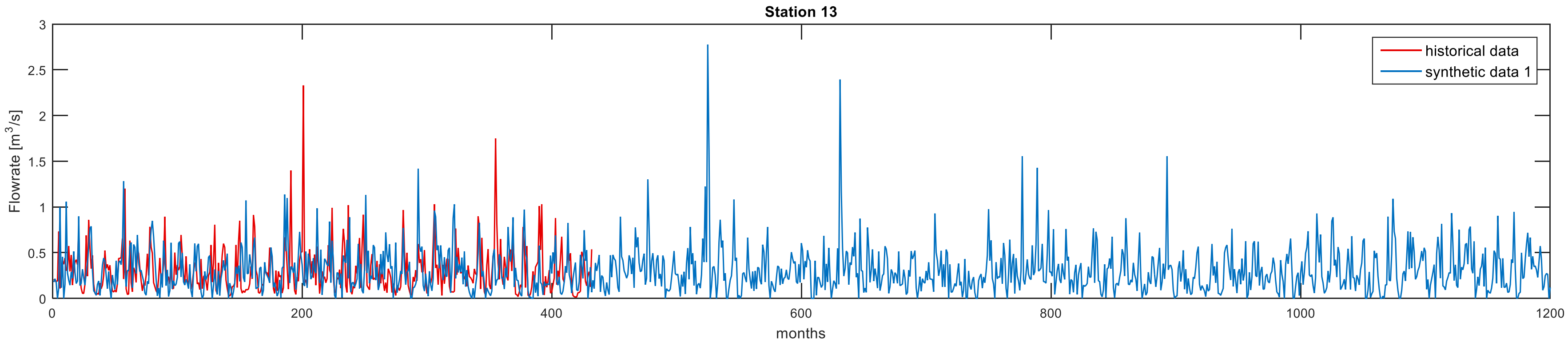

Station 4
historical data
synthetic data 1
Flowrate [m^3/s]
months
Station 5
historical data
synthetic data 1
Flowrate [m^3/s]
months
Station 6
historical data
synthetic data 1
Flowrate [m^3/s]
months


Station 7
Flowrate [m3/s]
months
historical data
synthetic data 1
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
0
200
400
600
800
1000
1200


Station 8
Flowrate [m3/s]
months
historical data
synthetic data 1
0
1
2
3
4
5
6
0
200
400
600
800
1000
1200


Station 9
Flowrate [m3/s]
months
historical data
synthetic data 1
0
50
100
150
0
200
400
600
800
1000
1200


Station 10
Flowrate [m³/s]
months
historical data
synthetic data 1
Station 11
Flowrate [m³/s]
months
historical data
synthetic data 1
Station 12
Flowrate [m³/s]
months
historical data
synthetic data 1


Station 13
historical data
synthetic data 1
Flowrate [m3/s]
months

APPENDIX 4; Maximum monthly averages comparison

Maximum monthly averages for 36 years long simulations [m^3/s]													
Stations	**St. 1**	**St. 2**	**St. 3**	**St. 4**	**St. 5**	**St. 6**	**St. 7**	**St. 8**	**St. 9**	**St. 10**	**St. 11**	**St. 12**	**St. 13**
Historic TS	**119.82**	**34.50**	**143.65**	**23.64**	**13.20**	**25.83**	**4.22**	**4.42**	**118.82**	**52.65**	**9.28**	**115.37**	**2.33**
Synthetic TS (Run 1)	84.10	31.16	70.28	17.91	11.02	21.27	3.01	4.04	89.45	45.26	7.10	86.79	1.81
Synthetic TS (Run 2)	100.06	29.45	112.58	15.15	9.34	16.28	3.70	2.89	87.96	35.24	6.95	74.19	2.06
Synthetic TS (Run 3)	96.24	29.24	104.98	22.39	12.64	24.31	3.64	4.40	113.14	49.61	9.50	102.16	2.10
Synthetic TS (Run 4)	69.77	24.12	74.29	18.60	10.27	19.22	2.35	3.03	71.54	39.60	6.19	77.99	1.64
Synthetic TS (Run 5)	75.61	23.10	75.74	17.94	7.97	15.41	2.73	3.20	78.73	32.22	7.12	70.01	1.40
Synthetic TS (Run 6)	103.65	35.04	123.42	22.40	11.36	22.94	3.69	4.51	113.54	42.41	10.34	95.42	1.96
Synthetic TS (Run 7)	67.42	28.75	94.88	19.99	10.32	19.05	2.73	3.47	75.10	41.04	6.84	79.07	1.71
Synthetic TS (Run 8)	95.73	28.23	126.17	20.65	11.88	21.78	3.34	3.83	94.67	46.12	9.24	95.92	1.87
Synthetic TS (Run 9)	65.49	24.64	84.30	17.94	9.86	16.59	2.09	2.82	63.29	28.68	6.59	62.80	1.31
Synthetic TS (Run 10)	82.54	27.74	84.82	21.53	11.29	23.55	3.36	3.84	100.88	52.44	8.60	106.98	2.05
E [Synthetic/Historical]	**70%**	**82%**	**66%**	**82%**	**80%**	**78%**	**73%**	**82%**	**75%**	**78%**	**85%**	**74%**	**77%**

Maximum monthly averages for **100** years long simulations [m^3/s]													
Stations	**St. 1**	**St. 2**	**St. 3**	**St. 4**	**St. 5**	**St. 6**	**St. 7**	**St. 8**	**St. 9**	**St. 10**	**St. 11**	**St. 12**	**St. 13**
Historic TS	**119.82**	**34.50**	**143.65**	**23.64**	**13.20**	**25.83**	**4.22**	**4.42**	**118.82**	**52.65**	**9.28**	**115.37**	**2.33**
Synthetic TS (Run 1)	119.82	38.11	148.30	21.79	11.17	21.23	4.08	4.97	120.98	53.29	11.04	105.46	2.05
Synthetic TS (Run 2)	119.77	35.50	123.07	30.30	15.97	35.58	4.41	6.25	142.96	81.85	13.29	159.15	2.76
Synthetic TS (Run 3)	139.37	32.00	110.25	18.44	9.56	21.88	4.47	4.93	109.55	44.91	11.32	123.03	1.64
Synthetic TS (Run 4)	108.90	35.84	110.49	19.83	12.69	26.00	3.62	3.67	102.89	55.00	9.07	110.46	1.87
Synthetic TS (Run 5)	124.72	32.62	102.46	24.93	14.51	29.01	4.27	4.24	106.58	66.90	9.20	121.27	2.69
Synthetic TS (Run 6)	121.91	42.10	148.38	21.74	12.04	21.67	4.49	5.11	120.29	48.28	12.30	107.46	2.28
Synthetic TS (Run 7)	107.49	30.36	117.24	27.74	13.24	25.97	3.08	3.59	84.11	50.51	8.03	83.02	1.63
Synthetic TS (Run 8)	166.33	43.24	175.67	30.61	13.26	25.66	5.67	5.02	153.42	53.77	11.96	120.99	2.00
Synthetic TS (Run 9)	119.95	30.80	109.56	33.61	17.97	38.30	4.45	6.68	161.18	81.94	14.07	166.52	2.98
Synthetic TS (Run 10)	89.58	37.64	117.93	25.16	16.67	35.26	4.00	5.35	138.94	59.79	11.45	130.18	2.75
E [Synthetic/Historical]	**102%**	**104%**	**88%**	**108%**	**104%**	**109%**	**101%**	**113%**	**104%**	**113%**	**120%**	**106%**	**97%**

Maximum monthly averages for **1000** years long simulations [m^3/s]

Stations	St. 1	St. 2	St. 3	St. 4	St. 5	St. 6	St. 7	St. 8	St. 9	St. 10	St. 11	St. 12	St. 13
Historic TS	**119.82**	**34.50**	**143.65**	**23.64**	**13.20**	**25.83**	**4.22**	**4.42**	**118.82**	**52.65**	**9.28**	**115.37**	**2.33**
Synthetic TS (Run 1)	167.79	48.98	138.72	40.78	24.33	49.89	6.02	7.93	172.69	88.31	15.23	170.25	4.05
Synthetic TS (Run 2)	259.81	46.03	237.81	40.16	24.97	45.59	8.10	7.86	176.13	108.99	14.69	178.57	4.20
Synthetic TS (Run 3)	175.00	42.28	194.08	47.24	21.85	38.08	6.72	8.55	179.07	89.76	25.25	238.84	3.76
Synthetic TS (Run 4)	195.49	59.13	202.51	43.96	18.11	39.66	6.59	7.81	179.03	87.06	16.98	168.09	3.88
Synthetic TS (Run 5)	163.79	40.95	147.90	29.68	17.90	39.49	5.71	7.53	152.55	64.58	15.13	162.53	3.61
Synthetic TS (Run 6)	156.97	51.92	164.61	41.03	18.28	38.06	6.18	6.80	167.36	79.33	16.37	163.56	2.94
Synthetic TS (Run 7)	172.47	43.23	172.48	40.41	19.18	31.95	6.15	6.34	168.51	79.26	13.59	155.83	2.67
Synthetic TS (Run 8)	197.35	59.24	227.65	33.86	17.50	37.18	6.92	7.68	203.82	70.68	19.19	157.26	3.41
Synthetic TS (Run 9)	137.50	43.14	176.67	34.89	21.32	43.26	5.58	7.48	189.60	82.01	15.08	181.98	3.28
Synthetic TS (Run 10)	174.80	46.06	183.23	39.99	24.11	50.51	6.74	9.01	203.65	103.26	17.93	195.71	3.96
E [Synthetic/Historical]	**150%**	**139%**	**129%**	**166%**	**157%**	**160%**	**153%**	**174%**	**151%**	**162%**	**183%**	**154%**	**154%**

Maximum monthly averages for **10 000** years long simulations [m^3/s]

Stations	St. 1	St. 2	St. 3	St. 4	St. 5	St. 6	St. 7	St. 8	St. 9	St. 10	St. 11	St. 12	St. 13
Historic TS	**119.82**	**34.50**	**143.65**	**23.64**	**13.20**	**25.83**	**4.22**	**4.42**	**118.82**	**52.65**	**9.28**	**115.37**	**2.33**
Synthetic TS (Run 1)	193.98	64.83	232.00	46.78	24.56	49.11	7.16	11.09	211.81	108.92	20.57	205.51	5.45
Synthetic TS (Run 2)	220.91	74.64	262.48	59.66	24.30	51.98	8.60	9.29	220.76	140.92	18.38	228.48	4.42
Synthetic TS (Run 3)	271.49	84.01	334.20	60.40	30.82	56.57	11.00	8.95	289.40	99.32	19.63	217.89	4.25
Synthetic TS (Run 4)	230.53	68.59	248.45	52.28	32.22	59.71	8.31	9.56	257.71	125.66	24.36	252.93	4.29
Synthetic TS (Run 5)	205.62	59.55	172.31	44.69	31.04	58.44	6.93	11.53	203.72	139.62	23.49	226.62	4.85
Synthetic TS (Run 6)	380.90	94.04	422.44	74.46	37.30	67.32	14.87	10.66	374.94	125.45	25.51	296.66	5.61
Synthetic TS (Run 7)	180.58	52.67	204.28	49.98	24.63	51.89	8.88	9.93	241.52	137.25	20.93	292.37	5.71
Synthetic TS (Run 8)	198.52	78.07	301.19	52.62	29.28	65.00	10.16	10.99	260.21	129.69	22.29	218.71	6.07
Synthetic TS (Run 9)	174.43	56.65	216.34	49.50	23.08	49.07	8.76	9.30	219.42	122.42	24.07	221.16	4.64
Synthetic TS (Run 10)	419.34	77.57	292.72	76.39	36.20	67.74	15.10	17.13	418.08	182.48	35.21	365.50	5.89
E [Synthetic/Historical]	**207%**	**206%**	**187%**	**240%**	**222%**	**223%**	**236%**	**245%**	**227%**	**249%**	**253%**	**219%**	**220%**

APPENDIX 5; Cross-correlation test results; *Table 7.8*

CROSS-CORRELATIONS for HISTORICAL TIME SERIES															
Compared st.			XI	XII	I	II	III	IV	V	VI	VII	VIII	IX	X	Q
1	x	2	0.67	0.76	0.81	0.68	0.60	0.83	0.79	0.84	0.98	0.86	0.84	0.93	0.77
1	x	3	0.86	0.92	0.91	0.83	0.88	0.92	0.90	0.91	0.98	0.95	0.96	0.96	0.92
1	x	4	0.68	0.60	0.58	0.42	0.49	0.66	0.92	0.60	0.95	0.83	0.82	0.86	0.76
1	x	5	0.64	0.63	0.57	0.36	0.41	0.66	0.89	0.50	0.95	0.84	0.80	0.70	0.73
1	x	6	0.66	0.67	0.60	0.44	0.59	0.70	0.92	0.73	0.94	0.83	0.82	0.71	0.78
1	x	7	0.69	0.83	0.84	0.71	0.93	0.87	0.96	0.73	0.97	0.94	0.94	0.78	0.89
1	x	8	0.74	0.79	0.82	0.63	0.85	0.80	0.91	0.75	0.92	0.86	0.87	0.68	0.84
1	x	9	0.76	0.84	0.82	0.71	0.84	0.85	0.95	0.78	0.96	0.92	0.89	0.84	0.88
1	x	10	0.70	0.76	0.73	0.65	0.79	0.77	0.92	0.72	0.91	0.85	0.79	0.73	0.83
1	x	11	0.76	0.83	0.85	0.78	0.86	0.81	0.90	0.73	0.90	0.93	0.89	0.76	0.85
1	x	12	0.78	0.83	0.85	0.80	0.87	0.83	0.94	0.77	0.93	0.94	0.85	0.76	0.88
1	x	13	0.72	0.77	0.65	0.63	0.69	0.72	0.89	0.80	0.93	0.85	0.85	0.60	0.81
2	x	3	0.90	0.90	0.93	0.93	0.88	0.95	0.95	0.95	0.99	0.94	0.82	0.97	0.90
2	x	4	0.51	0.47	0.73	0.81	0.78	0.83	0.68	0.56	0.91	0.71	0.60	0.77	0.77
2	x	5	0.45	0.50	0.66	0.76	0.70	0.81	0.67	0.38	0.90	0.73	0.54	0.55	0.70
2	x	6	0.46	0.48	0.65	0.76	0.71	0.82	0.67	0.52	0.90	0.75	0.59	0.54	0.69
2	x	7	0.31	0.40	0.52	0.21	0.41	0.63	0.65	0.39	0.91	0.75	0.70	0.63	0.54
2	x	8	0.34	0.40	0.65	0.27	0.44	0.58	0.57	0.40	0.88	0.75	0.65	0.47	0.54
2	x	9	0.49	0.50	0.72	0.74	0.69	0.83	0.65	0.47	0.91	0.81	0.65	0.68	0.68
2	x	10	0.49	0.51	0.69	0.81	0.71	0.81	0.60	0.53	0.86	0.74	0.53	0.56	0.70
2	x	11	0.33	0.41	0.63	0.38	0.41	0.56	0.56	0.36	0.84	0.79	0.66	0.58	0.52
2	x	12	0.42	0.48	0.70	0.68	0.66	0.76	0.60	0.52	0.87	0.79	0.60	0.62	0.65
2	x	13	0.37	0.57	0.62	0.69	0.69	0.64	0.60	0.48	0.89	0.74	0.60	0.39	0.64
3	x	4	0.67	0.50	0.61	0.71	0.73	0.79	0.77	0.49	0.92	0.85	0.81	0.81	0.80
3	x	5	0.62	0.55	0.60	0.66	0.67	0.79	0.77	0.35	0.92	0.87	0.80	0.66	0.78
3	x	6	0.64	0.57	0.61	0.69	0.76	0.81	0.78	0.54	0.91	0.87	0.83	0.66	0.80
3	x	7	0.58	0.66	0.65	0.38	0.72	0.72	0.79	0.49	0.92	0.87	0.93	0.70	0.75
3	x	8	0.60	0.60	0.69	0.41	0.72	0.70	0.73	0.50	0.88	0.83	0.86	0.59	0.73
3	x	9	0.72	0.68	0.75	0.77	0.86	0.88	0.78	0.54	0.92	0.93	0.88	0.77	0.83
3	x	10	0.66	0.63	0.67	0.79	0.85	0.82	0.73	0.54	0.87	0.86	0.78	0.66	0.79
3	x	11	0.61	0.64	0.70	0.54	0.70	0.68	0.72	0.49	0.85	0.88	0.89	0.69	0.71
3	x	12	0.65	0.67	0.76	0.76	0.86	0.81	0.75	0.57	0.88	0.90	0.83	0.71	0.79
3	x	13	0.61	0.69	0.62	0.71	0.81	0.74	0.74	0.61	0.90	0.88	0.85	0.53	0.79
4	x	5	0.95	0.94	0.95	0.98	0.96	0.98	0.97	0.87	0.98	0.97	0.98	0.88	0.96
4	x	6	0.95	0.94	0.95	0.98	0.95	0.97	0.98	0.85	0.97	0.97	0.98	0.88	0.95
4	x	7	0.73	0.55	0.51	0.18	0.36	0.53	0.94	0.45	0.96	0.87	0.90	0.85	0.69
4	x	8	0.82	0.63	0.72	0.27	0.45	0.57	0.94	0.61	0.95	0.90	0.93	0.83	0.74
4	x	9	0.94	0.85	0.87	0.86	0.81	0.91	0.98	0.75	0.98	0.97	0.97	0.96	0.90
4	x	10	0.95	0.88	0.90	0.90	0.80	0.92	0.98	0.82	0.96	0.94	0.98	0.89	0.92
4	x	11	0.79	0.67	0.67	0.30	0.41	0.53	0.91	0.55	0.94	0.91	0.95	0.89	0.72
4	x	12	0.90	0.77	0.77	0.70	0.69	0.82	0.97	0.75	0.97	0.93	0.96	0.90	0.86
4	x	13	0.87	0.84	0.88	0.80	0.81	0.76	0.94	0.62	0.96	0.95	0.94	0.75	0.86
5	x	6	0.99	0.98	0.98	0.98	0.95	0.98	0.99	0.85	1.00	0.99	0.99	0.98	0.98
5	x	7	0.75	0.58	0.51	0.16	0.32	0.56	0.92	0.49	0.96	0.87	0.90	0.80	0.71
5	x	8	0.87	0.64	0.71	0.27	0.45	0.64	0.93	0.56	0.97	0.90	0.94	0.88	0.78
5	x	9	0.94	0.87	0.87	0.85	0.78	0.93	0.97	0.75	0.98	0.97	0.98	0.93	0.92
5	x	10	0.95	0.91	0.88	0.90	0.77	0.94	0.96	0.82	0.98	0.97	0.99	0.95	0.93
5	x	11	0.82	0.70	0.66	0.31	0.41	0.60	0.92	0.56	0.95	0.91	0.95	0.90	0.76
5	x	12	0.91	0.79	0.77	0.69	0.66	0.86	0.94	0.73	0.99	0.95	0.97	0.93	0.87
5	x	13	0.91	0.89	0.93	0.82	0.82	0.83	0.96	0.69	0.98	0.96	0.97	0.87	0.91
6	x	7	0.76	0.63	0.58	0.27	0.53	0.61	0.95	0.72	0.97	0.87	0.92	0.83	0.78
6	x	8	0.88	0.69	0.75	0.37	0.62	0.68	0.96	0.84	0.98	0.93	0.96	0.90	0.84
6	x	9	0.96	0.91	0.90	0.91	0.91	0.95	0.99	0.94	0.99	0.98	0.98	0.95	0.96
6	x	10	0.97	0.93	0.93	0.93	0.90	0.95	0.97	0.95	0.98	0.96	0.98	0.97	0.95
6	x	11	0.86	0.76	0.70	0.40	0.60	0.64	0.94	0.79	0.96	0.92	0.97	0.94	0.82
6	x	12	0.93	0.84	0.81	0.76	0.82	0.89	0.96	0.92	0.99	0.94	0.97	0.95	0.91
6	x	13	0.94	0.92	0.95	0.87	0.90	0.84	0.96	0.88	0.99	0.96	0.97	0.92	0.94
7	x	8	0.85	0.85	0.82	0.74	0.94	0.89	0.96	0.88	0.96	0.89	0.95	0.84	0.91
7	x	9	0.87	0.87	0.82	0.61	0.81	0.80	0.98	0.88	0.99	0.93	0.97	0.92	0.91
7	x	10	0.77	0.77	0.75	0.50	0.79	0.72	0.96	0.76	0.96	0.90	0.89	0.89	0.83
7	x	11	0.87	0.92	0.91	0.91	0.94	0.94	0.95	0.89	0.95	0.94	0.97	0.91	0.93
7	x	12	0.81	0.87	0.85	0.76	0.88	0.85	0.98	0.85	0.97	0.95	0.92	0.90	0.91
7	x	13	0.79	0.69	0.68	0.56	0.68	0.68	0.93	0.82	0.96	0.85	0.94	0.76	0.83
8	x	9	0.93	0.87	0.90	0.64	0.86	0.81	0.97	0.94	0.98	0.94	0.97	0.91	0.93
8	x	10	0.89	0.83	0.86	0.52	0.84	0.75	0.97	0.84	0.97	0.93	0.94	0.95	0.89
8	x	11	0.96	0.94	0.91	0.85	0.97	0.95	0.99	0.97	0.98	0.97	0.99	0.97	0.97
8	x	12	0.94	0.93	0.91	0.68	0.90	0.86	0.98	0.92	0.97	0.95	0.96	0.90	0.93
8	x	13	0.92	0.76	0.81	0.59	0.80	0.80	0.96	0.88	0.98	0.91	0.96	0.85	0.90
9	x	10	0.97	0.96	0.95	0.96	0.97	0.96	0.98	0.93	0.97	0.96	0.97	0.97	0.96
9	x	11	0.93	0.94	0.90	0.71	0.85	0.81	0.96	0.92	0.97	0.96	0.98	0.96	0.92
9	x	12	0.95	0.96	0.95	0.94	0.96	0.95	0.99	0.97	0.98	0.98	0.98	0.96	0.97
9	x	13	0.94	0.91	0.92	0.94	0.92	0.88	0.96	0.92	0.98	0.96	0.97	0.85	0.95
10	x	11	0.88	0.88	0.87	0.63	0.82	0.74	0.96	0.82	0.95	0.92	0.93	0.97	0.87
10	x	12	0.96	0.95	0.93	0.91	0.96	0.95	0.99	0.96	0.99	0.97	0.99	0.98	0.97
10	x	13	0.94	0.95	0.94	0.94	0.92	0.85	0.96	0.89	0.98	0.95	0.96	0.88	0.93
11	x	12	0.93	0.95	0.95	0.84	0.91	0.88	0.96	0.92	0.96	0.97	0.96	0.95	0.94
11	x	13	0.91	0.83	0.79	0.70	0.75	0.78	0.97	0.91	0.95	0.91	0.96	0.89	0.88
12	x	13	0.95	0.88	0.85	0.91	0.87	0.85	0.95	0.92	0.97	0.94	0.97	0.86	0.92

Table 7.9

CROSS-CORRELATIONS for SYNTHETIC TIME SERIES															
Compared st.			**XI**	**XII**	**I**	**II**	**III**	**IV**	**V**	**VI**	**VII**	**VIII**	**IX**	**X**	**Q**
1	x	2	0.75	0.66	0.81	0.72	0.66	0.80	0.78	0.84	0.86	0.86	0.87	0.86	0.76
1	x	3	0.89	0.88	0.90	0.84	0.89	0.89	0.86	0.89	0.91	0.92	0.94	0.93	0.89
1	x	4	0.68	0.62	0.68	0.49	0.53	0.66	0.78	0.67	0.85	0.82	0.85	0.84	0.73
1	x	5	0.66	0.63	0.66	0.42	0.46	0.67	0.77	0.60	0.84	0.83	0.82	0.80	0.71
1	x	6	0.70	0.68	0.67	0.51	0.62	0.71	0.79	0.77	0.85	0.84	0.86	0.83	0.76
1	x	7	0.67	0.84	0.80	0.74	0.91	0.86	0.90	0.82	0.94	0.90	0.94	0.83	0.87
1	x	8	0.76	0.82	0.86	0.69	0.83	0.77	0.81	0.78	0.85	0.85	0.87	0.84	0.82
1	x	9	0.78	0.84	0.86	0.75	0.83	0.83	0.87	0.83	0.93	0.90	0.92	0.89	0.87
1	x	10	0.73	0.76	0.81	0.70	0.80	0.75	0.82	0.75	0.83	0.84	0.83	0.85	0.81
1	x	11	0.79	0.83	0.86	0.82	0.83	0.77	0.82	0.76	0.85	0.89	0.90	0.87	0.83
1	x	12	0.78	0.84	0.87	0.83	0.86	0.80	0.86	0.80	0.87	0.89	0.87	0.85	0.86
1	x	13	0.73	0.76	0.70	0.67	0.71	0.75	0.81	0.80	0.83	0.85	0.85	0.78	0.79
2	x	3	0.93	0.88	0.93	0.94	0.89	0.95	0.94	0.96	0.96	0.92	0.89	0.94	0.90
2	x	4	0.58	0.50	0.81	0.81	0.77	0.79	0.64	0.53	0.74	0.67	0.69	0.70	0.75
2	x	5	0.57	0.51	0.74	0.77	0.71	0.77	0.65	0.42	0.72	0.70	0.66	0.64	0.68
2	x	6	0.58	0.47	0.73	0.78	0.72	0.78	0.64	0.53	0.71	0.73	0.71	0.64	0.68
2	x	7	0.36	0.33	0.52	0.28	0.49	0.61	0.64	0.54	0.72	0.70	0.78	0.62	0.52
2	x	8	0.46	0.35	0.66	0.38	0.48	0.52	0.56	0.46	0.72	0.70	0.70	0.59	0.53
2	x	9	0.60	0.45	0.76	0.76	0.72	0.79	0.66	0.53	0.77	0.78	0.76	0.69	0.66
2	x	10	0.58	0.46	0.75	0.81	0.73	0.72	0.58	0.54	0.68	0.70	0.65	0.63	0.68
2	x	11	0.45	0.35	0.65	0.48	0.46	0.51	0.53	0.42	0.71	0.73	0.74	0.65	0.51
2	x	12	0.53	0.42	0.72	0.72	0.69	0.70	0.61	0.55	0.72	0.74	0.70	0.66	0.63
2	x	13	0.49	0.50	0.66	0.73	0.71	0.62	0.59	0.51	0.70	0.69	0.67	0.55	0.63
3	x	4	0.66	0.52	0.76	0.74	0.73	0.79	0.67	0.49	0.77	0.78	0.80	0.76	0.76
3	x	5	0.65	0.55	0.74	0.69	0.68	0.80	0.70	0.39	0.77	0.81	0.81	0.75	0.75
3	x	6	0.68	0.57	0.74	0.73	0.77	0.81	0.70	0.54	0.77	0.83	0.85	0.76	0.77
3	x	7	0.54	0.62	0.60	0.44	0.75	0.71	0.74	0.60	0.79	0.79	0.91	0.70	0.71
3	x	8	0.63	0.59	0.72	0.49	0.73	0.66	0.64	0.55	0.77	0.76	0.82	0.71	0.69
3	x	9	0.73	0.65	0.80	0.79	0.87	0.86	0.73	0.58	0.82	0.88	0.89	0.80	0.80
3	x	10	0.68	0.60	0.78	0.82	0.86	0.79	0.65	0.56	0.73	0.80	0.79	0.75	0.76
3	x	11	0.65	0.60	0.72	0.61	0.71	0.64	0.63	0.53	0.77	0.78	0.87	0.77	0.68
3	x	12	0.69	0.64	0.79	0.79	0.87	0.78	0.68	0.59	0.78	0.83	0.82	0.76	0.76
3	x	13	0.64	0.65	0.70	0.75	0.82	0.74	0.68	0.60	0.76	0.82	0.84	0.70	0.76
4	x	5	0.93	0.93	0.95	0.97	0.97	0.97	0.96	0.89	0.91	0.95	0.95	0.91	0.94
4	x	6	0.94	0.91	0.96	0.97	0.95	0.97	0.95	0.88	0.89	0.94	0.95	0.91	0.93
4	x	7	0.68	0.58	0.54	0.27	0.42	0.58	0.80	0.63	0.85	0.85	0.87	0.80	0.66
4	x	8	0.82	0.63	0.71	0.36	0.48	0.62	0.86	0.70	0.86	0.88	0.87	0.88	0.72
4	x	9	0.92	0.82	0.87	0.86	0.84	0.92	0.95	0.83	0.93	0.96	0.95	0.95	0.89
4	x	10	0.94	0.86	0.89	0.91	0.82	0.90	0.94	0.84	0.89	0.91	0.95	0.91	0.90
4	x	11	0.78	0.63	0.68	0.41	0.44	0.58	0.79	0.66	0.89	0.89	0.91	0.88	0.70
4	x	12	0.91	0.73	0.79	0.75	0.72	0.83	0.91	0.81	0.92	0.90	0.93	0.88	0.84
4	x	13	0.86	0.85	0.86	0.84	0.81	0.78	0.91	0.71	0.83	0.92	0.92	0.86	0.84
5	x	6	0.98	0.98	0.98	0.98	0.95	0.98	0.98	0.88	0.99	0.98	0.99	0.98	0.98
5	x	7	0.70	0.62	0.54	0.22	0.37	0.61	0.80	0.66	0.86	0.84	0.87	0.78	0.68
5	x	8	0.84	0.67	0.72	0.32	0.49	0.68	0.85	0.68	0.94	0.87	0.90	0.89	0.76
5	x	9	0.94	0.86	0.88	0.83	0.81	0.93	0.95	0.82	0.95	0.96	0.96	0.94	0.91
5	x	10	0.95	0.90	0.89	0.88	0.79	0.93	0.93	0.85	0.96	0.95	0.97	0.94	0.92
5	x	11	0.80	0.69	0.67	0.38	0.44	0.65	0.80	0.66	0.91	0.88	0.93	0.88	0.74
5	x	12	0.90	0.79	0.79	0.71	0.69	0.87	0.91	0.79	0.97	0.93	0.95	0.92	0.86
5	x	13	0.90	0.89	0.92	0.84	0.83	0.84	0.91	0.75	0.92	0.95	0.97	0.91	0.89
6	x	7	0.73	0.69	0.58	0.33	0.57	0.66	0.83	0.80	0.87	0.85	0.91	0.81	0.74
6	x	8	0.87	0.73	0.73	0.43	0.65	0.72	0.89	0.86	0.96	0.91	0.92	0.91	0.83
6	x	9	0.96	0.90	0.89	0.90	0.93	0.95	0.97	0.95	0.96	0.97	0.98	0.96	0.95
6	x	10	0.97	0.93	0.91	0.93	0.92	0.93	0.94	0.95	0.96	0.94	0.96	0.96	0.94
6	x	11	0.85	0.76	0.69	0.48	0.61	0.69	0.84	0.82	0.93	0.90	0.95	0.92	0.80
6	x	12	0.93	0.84	0.81	0.79	0.83	0.89	0.93	0.93	0.96	0.93	0.96	0.94	0.90
6	x	13	0.94	0.93	0.93	0.89	0.91	0.86	0.93	0.90	0.94	0.95	0.97	0.94	0.93
7	x	8	0.82	0.88	0.84	0.75	0.91	0.86	0.85	0.87	0.88	0.86	0.93	0.87	0.87
7	x	9	0.81	0.90	0.82	0.65	0.81	0.81	0.90	0.91	0.95	0.90	0.95	0.90	0.88
7	x	10	0.73	0.80	0.77	0.55	0.79	0.74	0.86	0.81	0.88	0.86	0.87	0.89	0.80
7	x	11	0.83	0.93	0.91	0.91	0.92	0.90	0.87	0.88	0.90	0.89	0.96	0.91	0.90
7	x	12	0.77	0.90	0.83	0.77	0.87	0.83	0.92	0.87	0.91	0.89	0.90	0.88	0.88
7	x	13	0.75	0.76	0.68	0.58	0.70	0.74	0.87	0.86	0.87	0.83	0.90	0.77	0.81
8	x	9	0.93	0.90	0.90	0.68	0.85	0.82	0.93	0.94	0.95	0.92	0.94	0.94	0.91
8	x	10	0.89	0.85	0.86	0.59	0.83	0.79	0.94	0.86	0.95	0.89	0.90	0.95	0.87
8	x	11	0.95	0.94	0.91	0.86	0.97	0.94	0.96	0.97	0.97	0.97	0.98	0.96	0.96
8	x	12	0.94	0.94	0.91	0.72	0.89	0.86	0.95	0.92	0.95	0.92	0.94	0.92	0.92
8	x	13	0.91	0.80	0.81	0.62	0.81	0.84	0.92	0.92	0.94	0.89	0.94	0.89	0.89
9	x	10	0.96	0.96	0.96	0.96	0.98	0.96	0.96	0.93	0.94	0.93	0.96	0.97	0.95
9	x	11	0.92	0.94	0.90	0.76	0.83	0.82	0.90	0.92	0.96	0.93	0.97	0.96	0.90
9	x	12	0.96	0.96	0.95	0.95	0.96	0.95	0.97	0.97	0.97	0.95	0.97	0.95	0.96
9	x	13	0.93	0.93	0.91	0.94	0.93	0.90	0.95	0.94	0.91	0.94	0.96	0.90	0.93
10	x	11	0.86	0.88	0.88	0.68	0.82	0.79	0.91	0.84	0.93	0.88	0.91	0.95	0.86
10	x	12	0.96	0.95	0.94	0.92	0.96	0.96	0.98	0.97	0.98	0.97	0.98	0.97	0.96
10	x	13	0.93	0.94	0.93	0.95	0.93	0.88	0.95	0.91	0.93	0.94	0.95	0.91	0.92
11	x	12	0.93	0.95	0.95	0.85	0.90	0.88	0.95	0.92	0.95	0.92	0.94	0.92	0.93
11	x	13	0.89	0.85	0.79	0.72	0.75	0.83	0.91	0.93	0.90	0.89	0.94	0.89	0.87
12	x	13	0.93	0.89	0.85	0.91	0.88	0.88	0.94	0.94	0.92	0.93	0.96	0.89	0.91

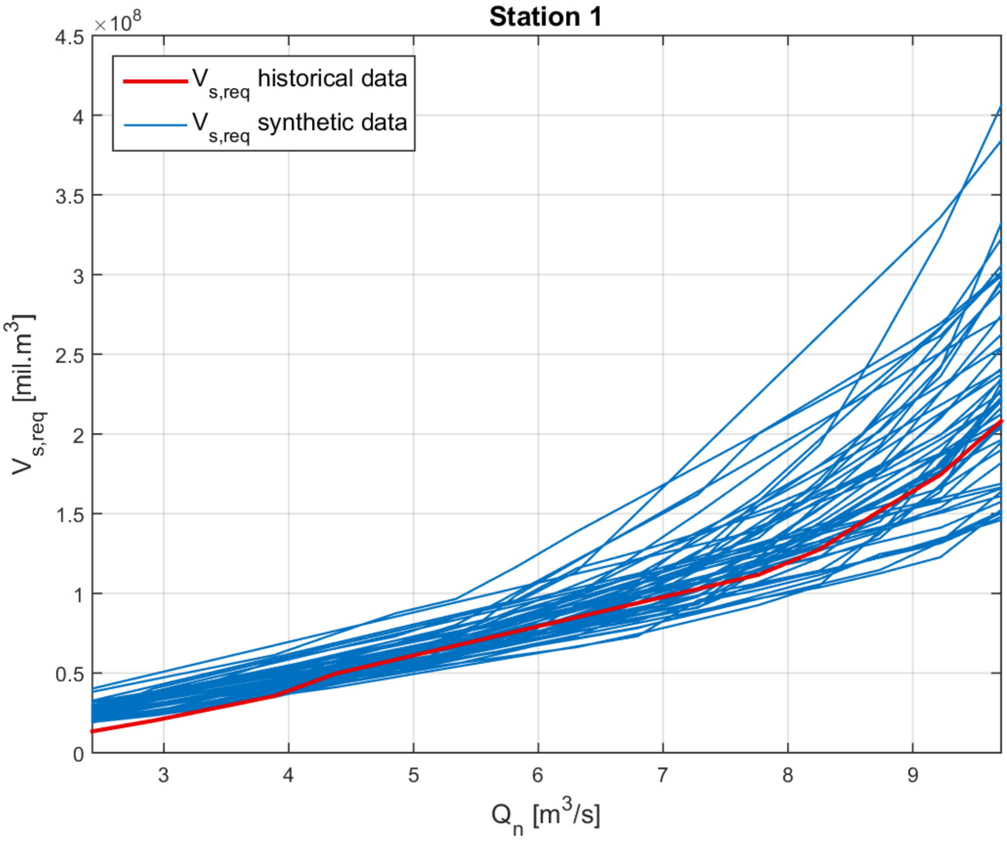
Table 7.10

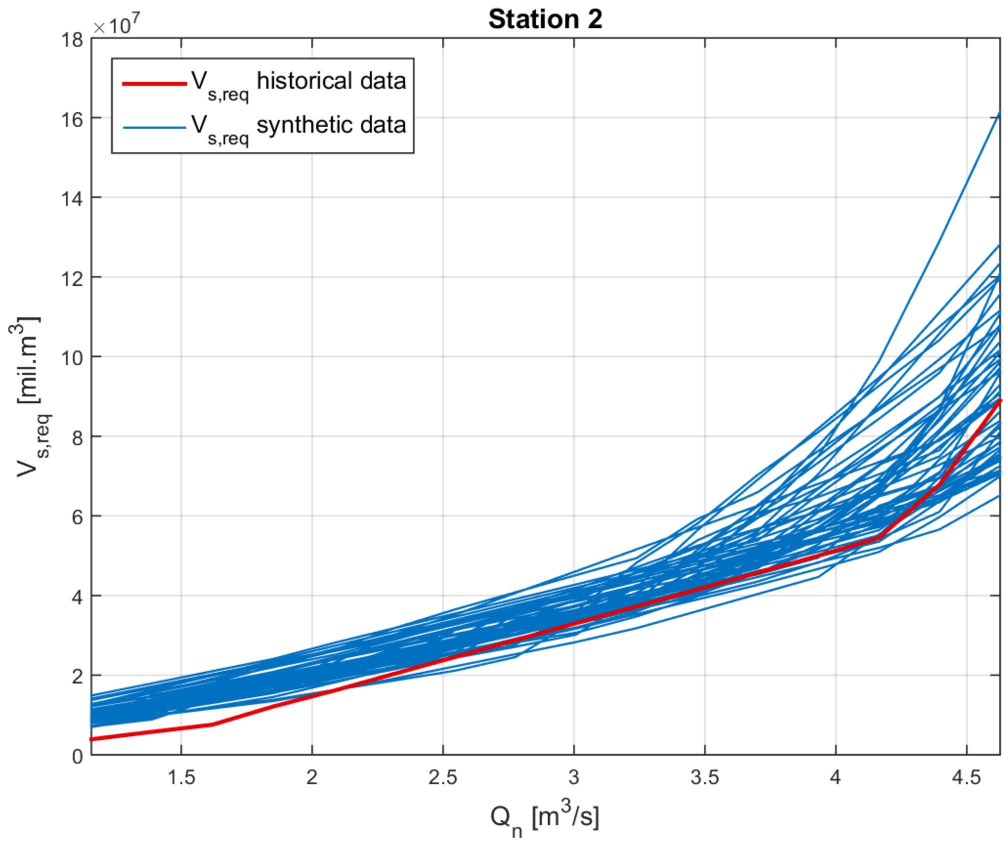
COMPARISON of CROSS-CORRELATION COEFFCIENTS															
Compared st.			XI	XII	I	II	III	IV	V	VI	VII	VIII	IX	X	Q
1	x	2	0	3.7%	0.1%	0	0	0.2%	0.4%	0.2%	93.5%	0.2%	0.1%	19.8%	0.8%
1	x	3	0	4.5%	0.1%	0	0	2.1%	5.3%	2.3%	92.4%	8.7%	5.1%	12.8%	58.7%
1	x	4	0	0	0	0	0	0	57.1%	0	60.8%	0.2%	0	0.5%	5.3%
1	x	5	0	0	0	0	0	0	30.9%	0	59.7%	0.2%	0	0	2.7%
1	x	6	0	0	0	0	0	0	46.7%	0	52.6%	0	0	0	4.5%
1	x	7	0	0	0.2%	0	0.4%	0.2%	44.7%	0	23.1%	12.3%	0.2%	0	14.0%
1	x	8	0	0.1%	0	0	0.1%	0	31.7%	0	25.9%	0.5%	0.1%	0	17.5%
1	x	9	0	0	0	0	0	0	38.9%	0	22.6%	1.8%	0	0	8.5%
1	x	10	0	0	0	0	0	0	37.0%	0	25.1%	0.3%	0	0	9.6%
1	x	11	0	0	0	0	0.6%	0.2%	19.9%	0	6.3%	7.4%	0	0	13.3%
1	x	12	0	0	0	0	0	0	29.1%	0	17.5%	15.1%	0	0	25.2%
1	x	13	0	0	0	0	0	0	13.1%	0	45.8%	0.1%	0	0	6.1%
2	x	3	0.1%	1.6%	0	0	0	0.2%	1.2%	0	95.0%	2.0%	0	34.1%	0
2	x	4	0	0	0	0	0	0	0.2%	0	67.9%	0	0	1.7%	2.0%
2	x	5	0	0	0	0	0	0	0	0	66.1%	0.1%	0	0	0.1%
2	x	6	0	0	0	0	0	0	0.2%	0	62.4%	0	0	0	0.1%
2	x	7	0	0.2%	0	0	0	0	0.1%	0	69.4%	0.7%	0	0	0.3%
2	x	8	0	0	0	0	0	0	0	0	40.5%	0.1%	0	0	0
2	x	9	0	0	0	0	0	0.3%	0	0	55.7%	0.1%	0	0	0.2%
2	x	10	0	0	0	0	0	1.0%	0	0	42.1%	0	0	0	0.1%
2	x	11	0	0.1%	0	0	0	0	0.1%	0.1%	18.9%	1.4%	0	0	0
2	x	12	0	0	0	0	0	0.3%	0	0.1%	38.8%	0.1%	0	0	0.1%
2	x	13	0	0.1%	0	0	0	0	0	0.1%	59.1%	0.3%	0	0	0.2%
3	x	4	0	0	0	0	0	0	6.1%	0	64.7%	4.0%	0.3%	0.7%	21.8%
3	x	5	0	0	0	0	0	0	1.8%	0	57.6%	4.1%	0	0	7.6%
3	x	6	0	0	0	0	0	0	3.5%	0	51.9%	1.8%	0	0	11.2%
3	x	7	0.1%	0	0	0	0	0	1.4%	0	55.1%	11.2%	1.2%	0	18.9%
3	x	8	0	0	0	0	0	0	1.4%	0	26.9%	5.2%	1.4%	0	8.5%
3	x	9	0	0	0	0	0	0.1%	0.9%	0	40.7%	11.8%	0.1%	0	19.3%
3	x	10	0	0	0	0	0	0.1%	1.4%	0	27.8%	4.9%	0	0	14.4%
3	x	11	0	0	0	0	0	0	2.3%	0	8.0%	19.4%	0.6%	0	5.5%
3	x	12	0	0	0	0	0	0	0.8%	0	21.9%	12.9%	0	0	12.3%
3	x	13	0	0	0	0	0	0	0.7%	0.1%	42.6%	4.3%	0.1%	0	15.0%
4	x	5	1.4%	0.7%	0	1.5%	0	0.1%	5.9%	0	84.9%	18.8%	40.5%	0	73.8%
4	x	6	0.6%	2.8%	0	0.1%	0.1%	0.2%	43.3%	0	79.0%	8.5%	39.4%	0	78.1%
4	x	7	0.4%	0	0	0	0	0	78.4%	0	70.8%	0.6%	1.9%	0.9%	0.9%
4	x	8	0	0	0	0	0	0	37.4%	0	61.5%	1.2%	13.3%	0	0.6%
4	x	9	0.1%	0.2%	0	0	0	0	43.1%	0	55.4%	2.3%	8.8%	0.6%	3.0%
4	x	10	0.1%	0.2%	0.1%	0	0	0.1%	53.4%	0	51.5%	4.3%	21.0%	0	39.6%
4	x	11	0	0	0	0	0	0	43.5%	0	22.7%	1.1%	9.8%	0	0.4%
4	x	12	0	0	0	0	0	0	55.7%	0	43.9%	2.3%	11.2%	0	3.1%
4	x	13	0	0.1%	0.2%	0	0	0	4.3%	0	84.4%	7.1%	4.3%	0	12.0%
5	x	6	1.5%	0.3%	0.1%	1.0%	0	0.1%	8.7%	0	55.9%	8.1%	18.6%	0	7.6%
5	x	7	0	0	0	0	0	0	43.7%	0	77.7%	1.8%	1.8%	0	2.1%
5	x	8	0.1%	0	0	0	0	0	30.9%	0	41.5%	1.6%	12.2%	0	0
5	x	9	0.2%	0.1%	0	0	0	0	10.1%	0	59.5%	9.8%	8.5%	0	3.0%
5	x	10	0	0.1%	0	0.3%	0	0.7%	11.9%	0	36.1%	5.3%	29.4%	0	8.4%
5	x	11	0.1%	0	0	0	0	0	53.0%	0	11.5%	1.4%	5.5%	0.1%	0.5%
5	x	12	0.1%	0	0	0	0	0	6.1%	0	32.7%	2.0%	6.1%	0	0.1%
5	x	13	0.2%	0	0.3%	0	0	0	25.9%	0	81.9%	3.5%	1.6%	0	9.8%
6	x	7	0	0	0	0	0	0	68.6%	0	83.0%	1.1%	0.4%	0	5.7%
6	x	8	0	0	0	0	0	0	46.6%	0	49.3%	1.2%	12.2%	0	0.6%
6	x	9	0	0.1%	0	0	0	0	34.5%	0	72.5%	0.7%	2.7%	0	17.3%
6	x	10	0	0.2%	0.3%	0	0	1.2%	27.8%	0	43.0%	9.6%	16.6%	0.1%	46.6%
6	x	11	0.1%	0	0	0	0	0	56.3%	0	9.3%	0.6%	3.6%	0.2%	1.1%
6	x	12	0	0	0	0	0	0	18.6%	0	38.4%	1.2%	2.4%	0.5%	2.0%
6	x	13	0.3%	0	1.9%	0	0	0	15.1%	0	90.1%	1.8%	1.2%	0	20.6%
7	x	8	0.9%	0	0	0	2.7%	0.7%	75.4%	0	59.0%	1.4%	3.8%	0	83.3%
7	x	9	2.7%	0	0	0	0	0	85.3%	0	90.3%	4.7%	2.4%	0.3%	46.0%
7	x	10	0.2%	0	0	0	0	0	80.4%	0	53.8%	3.5%	0.6%	0	24.2%
7	x	11	0.8%	0	0	0.1%	1.1%	5.4%	41.7%	0.2%	21.6%	18.2%	1.1%	0.1%	79.0%
7	x	12	0.1%	0	0.1%	0.2%	0.1%	0	78.7%	0	69.9%	29.2%	1.2%	0.2%	71.8%
7	x	13	0.3%	0	0	0	0	0	17.4%	0	68.9%	1.5%	6.7%	0	12.3%
8	x	9	0	0	0	0.1%	0	0	56.3%	0.1%	36.3%	3.0%	16.2%	0	28.7%
8	x	10	0	0	0	0	0	0	33.0%	0	19.1%	7.1%	4.8%	0	2.5%
8	x	11	0.2%	0	0	0	0.3%	0.4%	49.9%	0	3.8%	0.4%	26.8%	0.2%	21.5%
8	x	12	0	0	0	0.1%	0	0	31.9%	0	13.9%	4.7%	7.7%	0	16.8%
8	x	13	0.3%	0	0	0.1%	0	0	23.7%	0	73.4%	1.4%	9.1%	0	4.4%
9	x	10	0	0	0	0	0	0.1%	54.0%	0	28.4%	9.9%	3.7%	0	46.5%
9	x	11	0.6%	0	0	0	0.1%	0	48.2%	0	1.3%	15.8%	5.4%	0.2%	23.9%
9	x	12	0	0	0	0	0	0	44.7%	0	16.1%	19.5%	2.5%	0.1%	41.5%
9	x	13	0	0	0	0	0	0	3.4%	0	77.9%	5.8%	8.0%	0	43.0%
10	x	11	0	0	0	0	0	0	35.6%	0	4.4%	8.0%	2.4%	5.3%	3.6%
10	x	12	0	0	0	0	0	0	60.1%	0	42.3%	0.4%	3.7%	0.8%	5.8%
10	x	13	0	0	0.3%	0	0	0	2.9%	0	60.2%	1.8%	2.9%	0	4.0%
11	x	12	0.2%	0	0.2%	0	0.1%	0	7.1%	0	1.8%	47.5%	2.9%	1.7%	24.0%
11	x	13	0.3%	0	0	0	0	0	56.5%	0	23.8%	1.6%	2.3%	0.1%	12.5%
12	x	13	0.2%	0	0	0	0	0	0.5%	0	57.0%	0.4%	2.4%	0	4.2%

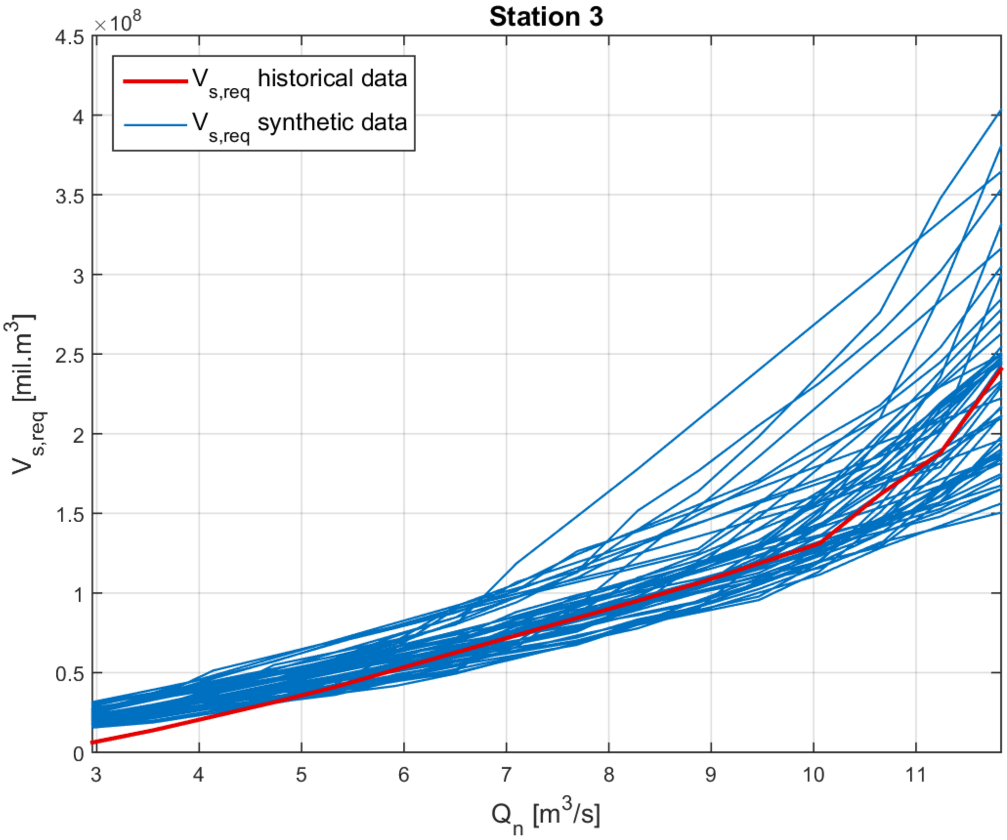
Table 7.11

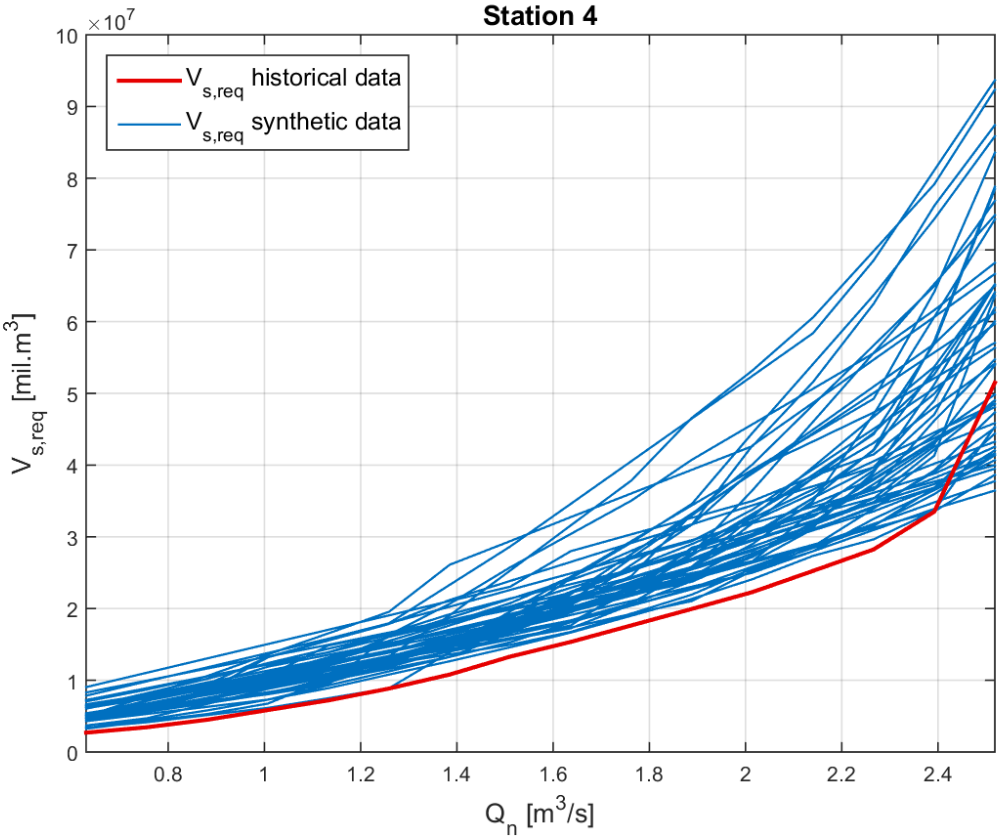
COMPARISON of CROSS-CORRELATION COEFFCIENTS															
Compared st.			**XI**	**XII**	**I**	**II**	**III**	**IV**	**V**	**VI**	**VII**	**VIII**	**IX**	**X**	**Q**
1	x	2	0	2.6%	4.0%	0	0	3.9%	2.6%	1.5%	0.4%	4.3%	2.3%	2.1%	3.5%
1	x	3	0	3.1%	4.7%	0	0	3.4%	3.1%	2.6%	0.4%	2.4%	3.0%	2.4%	1.4%
1	x	4	0	0	0	0	0	0	1.4%	0	1.2%	4.2%	0	4.3%	2.6%
1	x	5	0	0	0	0	0	0	2.0%	0	1.3%	4.1%	0	0	2.7%
1	x	6	0	0	0	0	0	0	1.6%	0	1.6%	0	0	0	2.5%
1	x	7	0	0	2.5%	0	2.1%	3.8%	1.6%	0	2.0%	2.3%	2.9%	0	2.5%
1	x	8	0	4.6%	0	0	4.9%	0	1.9%	0	2.3%	2.7%	3.3%	0	2.2%
1	x	9	0	0	0	0	0	0	1.9%	0	2.1%	3.5%	0	0	2.3%
1	x	10	0	0	0	0	0	0	1.8%	0	2.2%	3.1%	0	0	2.3%
1	x	11	0	0	0	0	3.4%	4.7%	2.3%	0	2.8%	2.8%	0	0	2.1%
1	x	12	0	0	0	0	0	0	2.1%	0	2.4%	2.5%	0	0	1.9%
1	x	13	0	0	0	0	0	0	2.4%	0	1.6%	3.2%	0	0	2.4%
2	x	3	4.6%	2.6%	0	0	0	4.3%	3.0%	0	0.3%	3.0%	0	1.9%	0
2	x	4	0	0	0	0	0	0	4.2%	0	1.2%	0	0	3.5%	3.5%
2	x	5	0	0	0	0	0	0	0	0	1.3%	4.2%	0	0	4.3%
2	x	6	0	0	0	0	0	0	4.1%	0	1.4%	0	0	0	4.1%
2	x	7	0	4.4%	0	0	0	0	2.5%	0	1.1%	3.9%	0	0	3.4%
2	x	8	0	0	0	0	0	0	0	0	1.9%	4.1%	0	0	0
2	x	9	0	0	0	0	0	3.8%	0	0	1.5%	4.8%	0	0	4.9%
2	x	10	0	0	0	0	0	3.4%	0	0	1.9%	0	0	0	3.8%
2	x	11	0	3.6%	0	0	0	0	3.8%	3.8%	2.4%	4.1%	0	0	0
2	x	12	0	0	0	0	0	3.7%	0	4.9%	2.0%	3.9%	0	0	4.8%
2	x	13	0	4.7%	0	0	0	0	0	4.0%	1.6%	3.9%	0	0	4.9%
3	x	4	0	0	0	0	0	0	3.0%	0	1.2%	3.0%	3.0%	3.9%	2.2%
3	x	5	0	0	0	0	0	0	3.6%	0	1.5%	3.3%	0	0	2.7%
3	x	6	0	0	0	0	0	0	3.4%	0	1.8%	3.6%	0	0	2.4%
3	x	7	4.2%	0	0	0	0	0	3.0%	0	1.4%	2.7%	3.6%	0	2.2%
3	x	8	0	0	0	0	0	0	3.6%	0	2.2%	3.1%	3.4%	0	2.7%
3	x	9	0	0	0	0	0	4.8%	3.3%	0	1.8%	2.7%	2.8%	0	2.2%
3	x	10	0	0	0	0	0	4.6%	3.4%	0	2.4%	2.8%	0	0	2.4%
3	x	11	0	0	0	0	0	0	3.3%	0	2.9%	2.4%	4.0%	0	2.8%
3	x	12	0	0	0	0	0	0	3.9%	0	2.5%	2.5%	0	0	2.7%
3	x	13	0	0	0	0	0	0	3.6%	3.7%	1.9%	2.9%	2.7%	0	2.5%
4	x	5	3.4%	3.6%	0	3.4%	0	4.8%	2.9%	0	0.7%	2.3%	1.8%	0	1.0%
4	x	6	3.3%	3.4%	0	3.2%	4.9%	3.9%	1.6%	0	0.9%	2.6%	1.9%	0	1.0%
4	x	7	3.7%	0	0	0	0	0	0.9%	0	1.1%	3.3%	3.0%	3.3%	3.1%
4	x	8	0	0	0	0	0	0	1.8%	0	1.3%	3.4%	2.5%	0	4.1%
4	x	9	2.9%	3.8%	0	0	0	0	1.6%	0	1.4%	3.0%	2.6%	2.5%	3.0%
4	x	10	4.2%	4.2%	3.7%	0	0	4.7%	1.4%	0	1.6%	3.3%	2.3%	0	2.0%
4	x	11	0	0	0	0	0	0	1.9%	0	2.3%	3.3%	2.6%	0	3.9%
4	x	12	0	0	0	0	0	0	1.4%	0	1.8%	3.2%	2.7%	0	3.1%
4	x	13	0	5.0%	4.2%	0	0	0	2.9%	0	0.7%	2.7%	3.0%	0	2.7%
5	x	6	3.7%	2.3%	1.3%	3.6%	0	4.8%	2.5%	0	1.3%	2.7%	2.2%	0	2.8%
5	x	7	0	0	0	0	0	0	1.7%	0	0.9%	2.8%	3.4%	0	3.4%
5	x	8	4.1%	0	0	0	0	0	1.9%	0	1.7%	3.1%	2.6%	0	0
5	x	9	4.6%	1.8%	0	0	0	0	2.5%	0	1.4%	2.4%	2.9%	0	3.0%
5	x	10	0	3.1%	0	3.9%	0	4.3%	2.6%	0	1.9%	2.7%	1.9%	0	2.8%
5	x	11	3.1%	0	0	0	0	0	1.7%	0	2.5%	3.0%	2.9%	3.8%	3.4%
5	x	12	2.9%	0	0	0	0	0	2.8%	0	2.1%	2.9%	2.5%	0	2.4%
5	x	13	3.3%	0	2.5%	0	0	0	2.1%	0	0.9%	3.1%	3.3%	0	2.7%
6	x	7	0	0	0	0	0	0	1.2%	0	0.8%	2.8%	3.4%	0	2.9%
6	x	8	0	0	0	0	0	0	1.6%	0	1.5%	2.7%	2.5%	0	3.4%
6	x	9	0	0.9%	0	0	0	0	1.7%	0	1.0%	2.8%	2.8%	0	2.2%
6	x	10	0	3.0%	3.3%	0	0	3.2%	2.2%	0	1.8%	2.6%	2.3%	4.7%	1.9%
6	x	11	3.0%	0	0	0	0	0	1.5%	0	2.9%	3.1%	2.7%	2.7%	3.6%
6	x	12	0	0	0	0	0	0	2.3%	0	1.9%	2.7%	2.6%	3.9%	3.2%
6	x	13	3.4%	0	3.0%	0	0	0	2.7%	0	0.5%	2.5%	3.2%	0	2.2%
7	x	8	3.6%	0	0	0	3.5%	3.4%	1.0%	0	1.4%	2.9%	2.9%	0	0.7%
7	x	9	2.9%	0	0	0	0	0	0.6%	0	0.5%	3.0%	3.0%	3.8%	1.8%
7	x	10	4.2%	0	0	0	0	0	0.8%	0	1.5%	2.8%	3.5%	0	2.3%
7	x	11	4.1%	0	0	3.0%	3.4%	2.8%	1.8%	3.6%	2.1%	2.6%	3.2%	3.6%	0.8%
7	x	12	3.5%	0	3.1%	4.1%	4.0%	0	0.8%	0	1.2%	1.8%	3.1%	4.1%	1.2%
7	x	13	4.1%	0	0	0	0	0	2.3%	0	1.1%	3.0%	2.7%	0	2.6%
8	x	9	0	0	0	4.1%	0	0	1.4%	4.6%	1.9%	3.1%	2.5%	0	2.1%
8	x	10	0	0	0	0	0	0	1.9%	0	1.9%	3.0%	2.8%	0	3.4%
8	x	11	3.7%	0	0	0	3.4%	2.9%	1.4%	0	2.9%	3.5%	2.1%	3.1%	2.1%
8	x	12	0	0	0	4.2%	0	0	1.9%	0	2.2%	2.8%	3.1%	0	2.2%
8	x	13	2.4%	0	0	2.0%	0	0	2.1%	0	1.0%	2.9%	2.6%	0	2.2%
9	x	10	0	0	0	0	0	1.3%	1.5%	0	1.9%	2.5%	3.5%	0	1.6%
9	x	11	4.2%	0	0	0	4.6%	0	1.5%	0	2.5%	2.5%	2.6%	2.8%	2.1%
9	x	12	0	0	0	0	0	0	1.7%	0	2.4%	2.3%	3.0%	3.6%	1.7%
9	x	13	0	0	0	0	0	0	3.0%	0	0.8%	3.0%	2.8%	0	1.5%
10	x	11	0	0	0	0	0	0	1.9%	0	2.8%	2.7%	3.1%	3.0%	2.9%
10	x	12	0	0	0	0	0	0	1.3%	0	1.5%	3.4%	2.9%	3.4%	3.0%
10	x	13	0	0	3.5%	0	0	0	3.1%	0	1.4%	3.1%	3.2%	0	2.2%
11	x	12	3.8%	0	2.9%	0	2.6%	0	2.6%	0	3.0%	1.6%	3.1%	2.9%	2.0%
11	x	13	3.3%	0	0	0	0	0	1.4%	0	2.0%	3.3%	3.3%	4.7%	2.5%
12	x	13	3.5%	0	0	0	0	0	2.9%	0	1.6%	3.4%	2.7%	0	2.4%

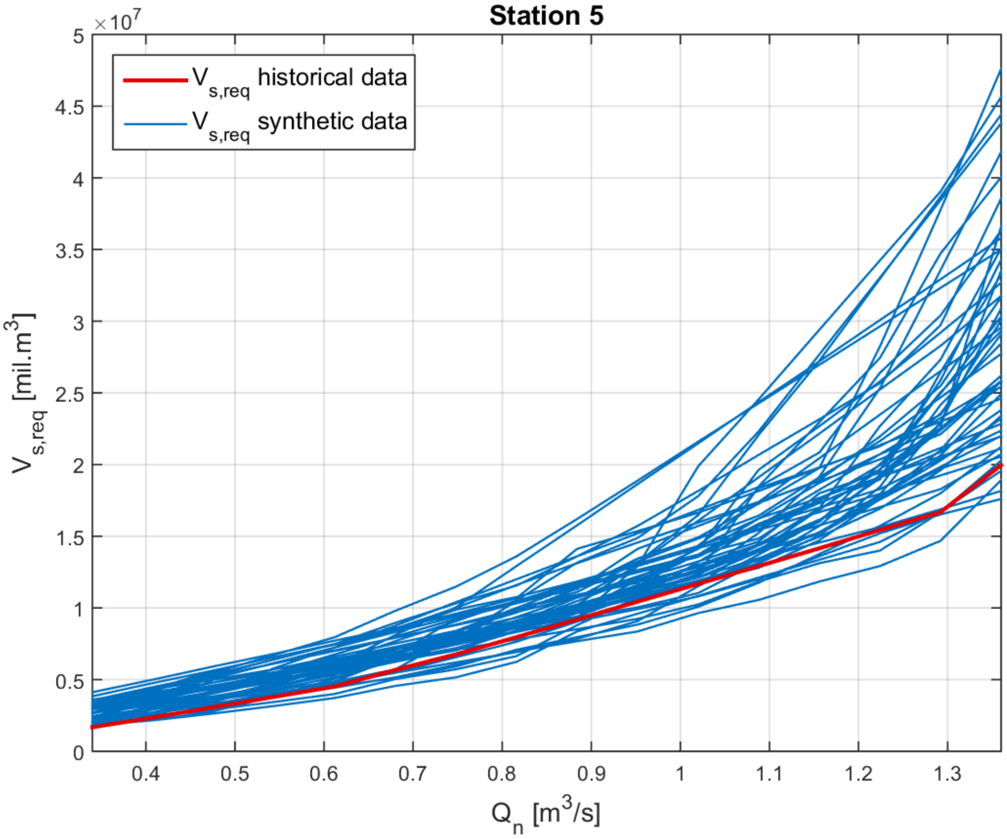
APPENDIX 6; Water Reservoir Operation Function plots for 36 years long synthetic TS



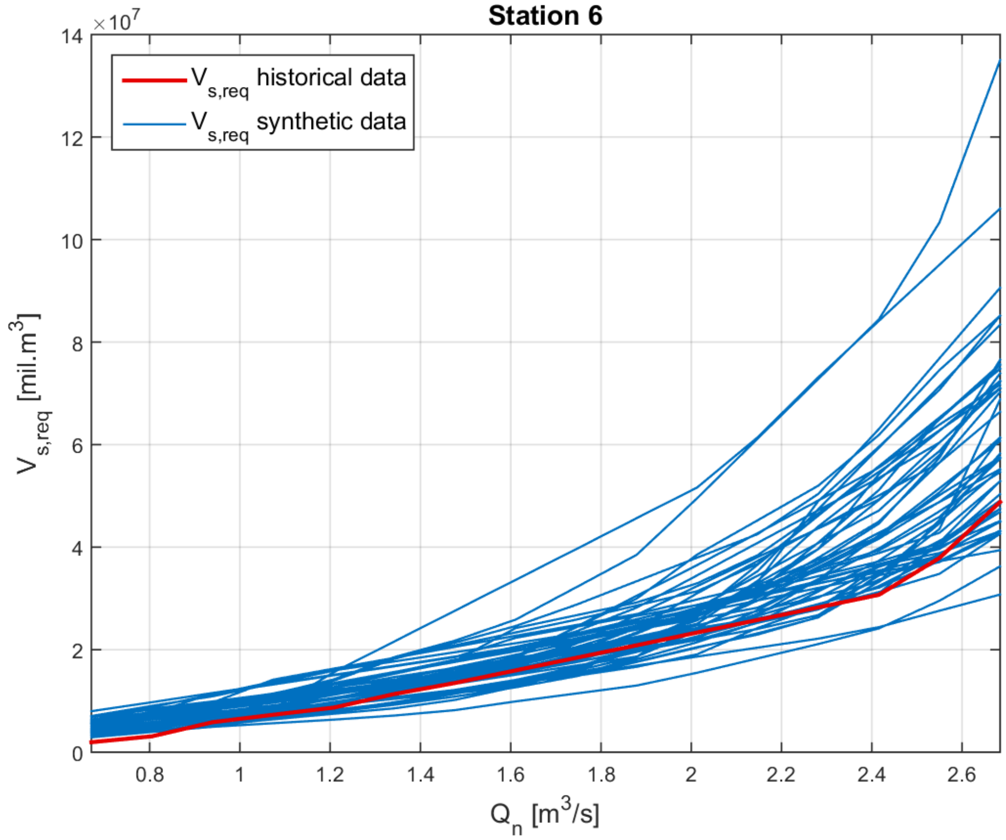



Station 3
×10^8
historical data
synthetic data
V s,req [mil.m3]
Q n [m3/s]

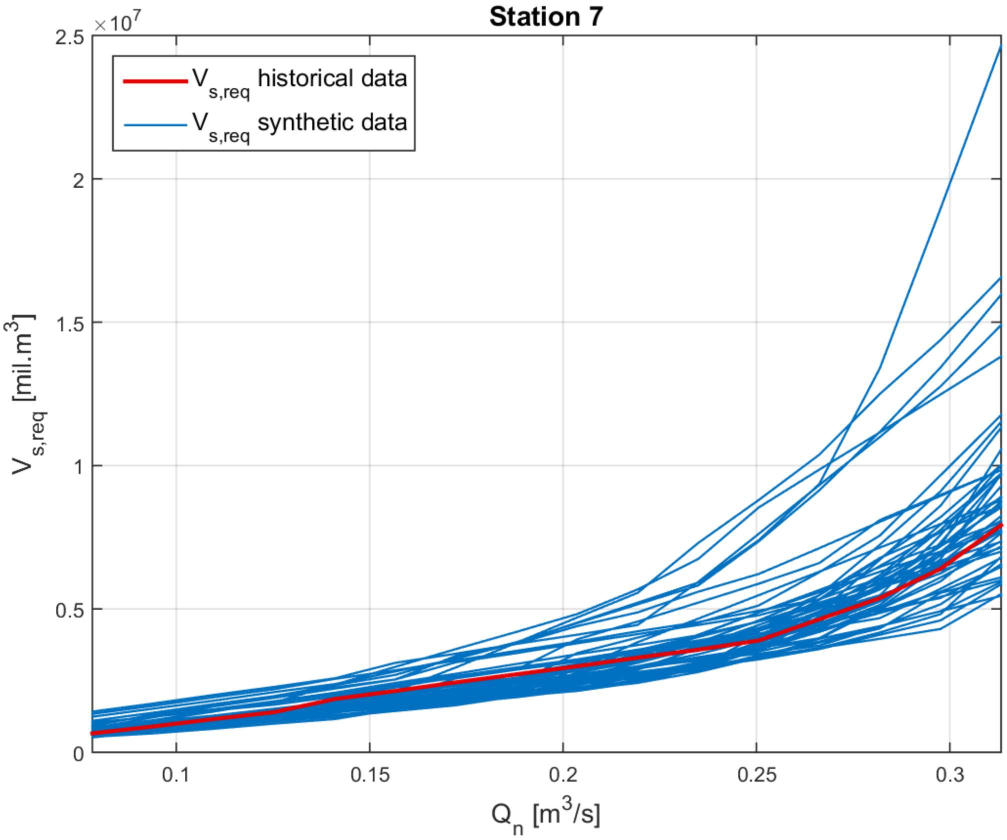

Station 4
×10^7
historical data
synthetic data
V s,req [mil.m3]
Q n [m3/s]

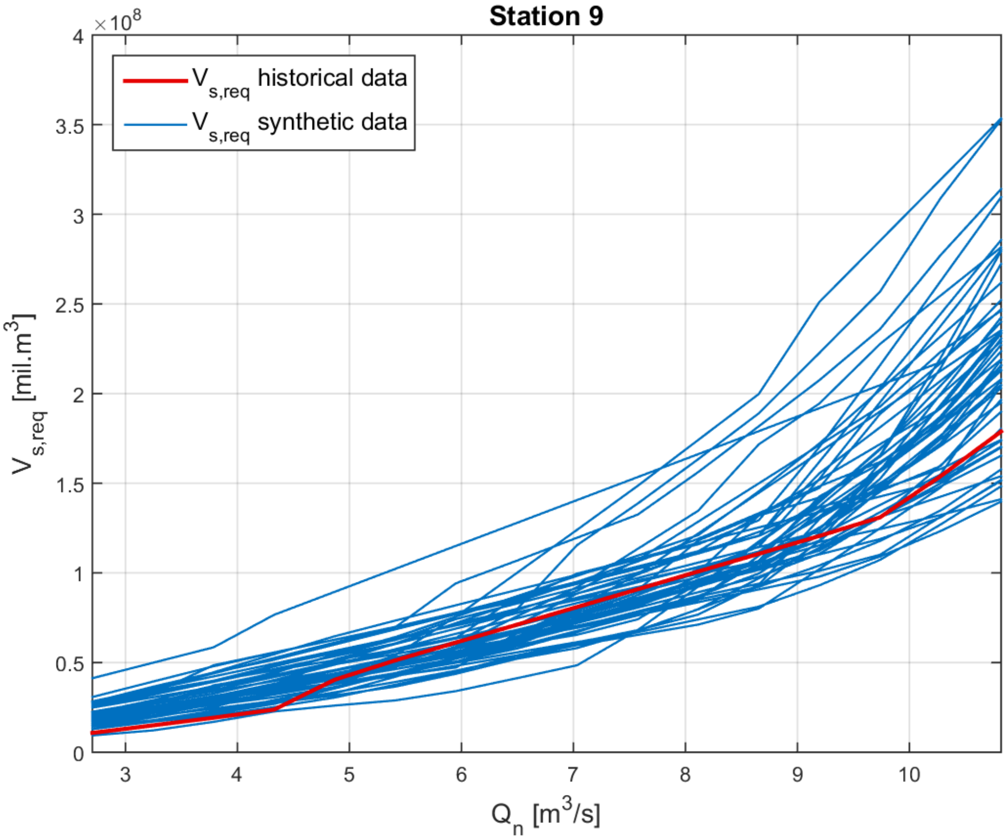


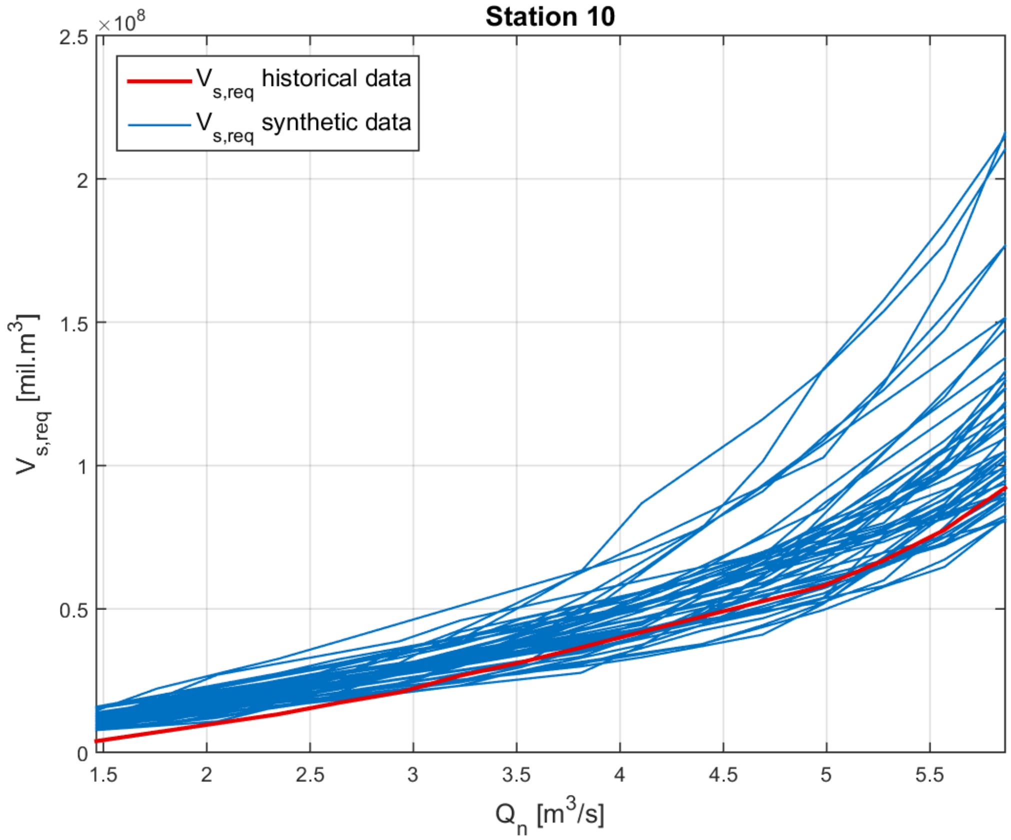
Station 5
×10^7
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
0.4
0.5
0.6
0.7
0.8
0.9
1
1.1
1.2
1.3
V s,req historical data
V s,req synthetic data
V s,req [mil.m3]
Q n [m3/s]



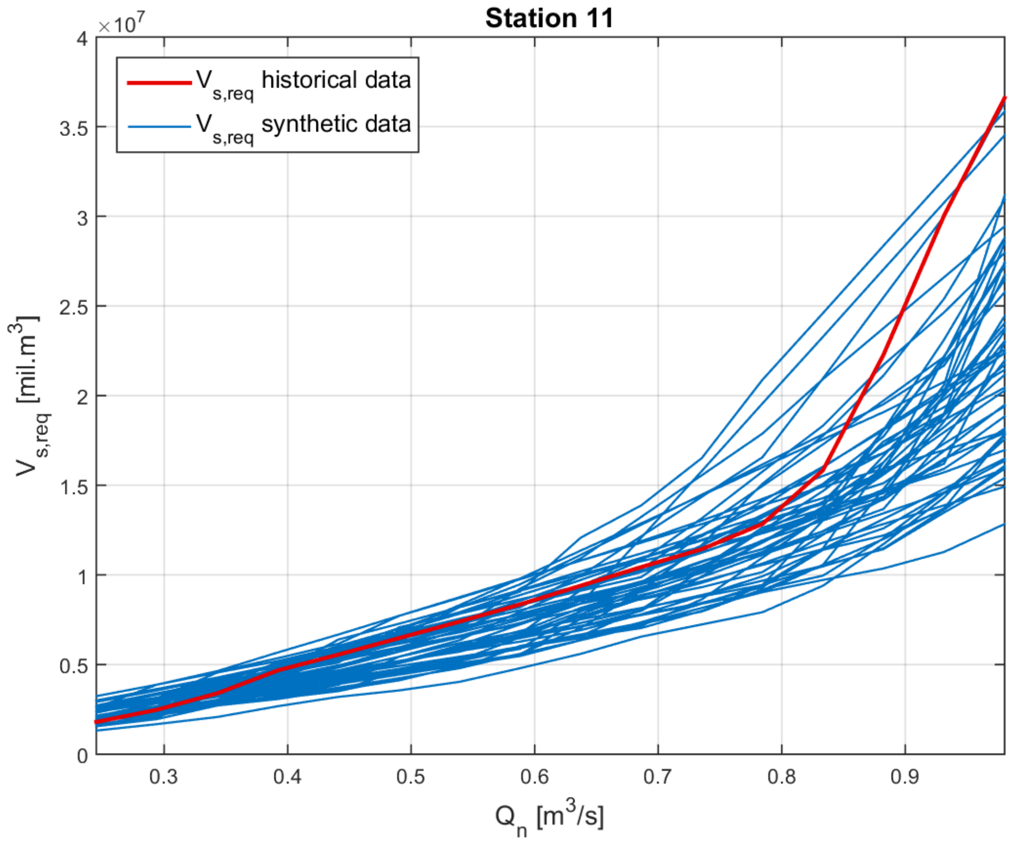
Station 6
×10^7
14
12
10
8
6
4
2
0
0.8
1
1.2
1.4
1.6
1.8
2
2.2
2.4
2.6
V s,req historical data
V s,req synthetic data
V s,req [mil.m3]
Q n [m3/s]


Station 7
$\times 10^7$
2.5
2
1.5
1
0.5
0
$V_{s,req}$ historical data
$V_{s,req}$ synthetic data
$V_{s,req}$ [mil.m^3]
0.1
0.15
0.2
0.25
0.3
Q_n [m^3/s]

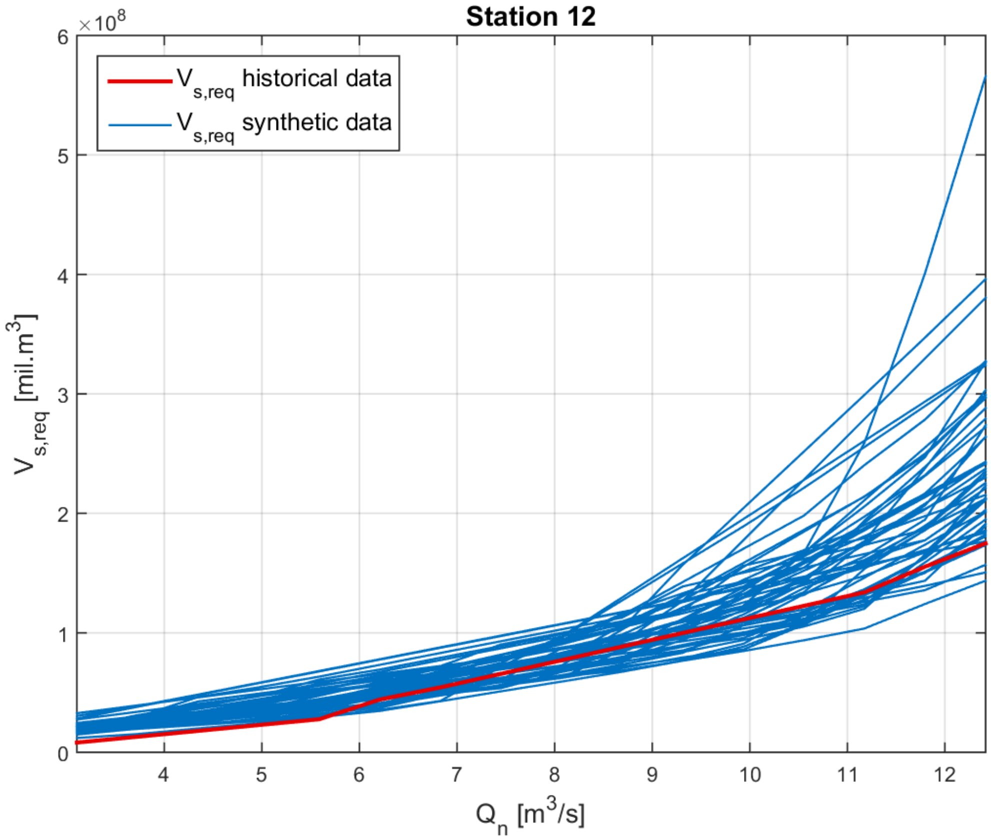

Station 9
$\times 10^8$
4
3.5
3
2.5
2
1.5
1
0.5
0
$V_{s,req}$ historical data
$V_{s,req}$ synthetic data
$V_{s,req}$ [mil.m^3]
3
4
5
6
7
8
9
10
Q_n [m^3/s]

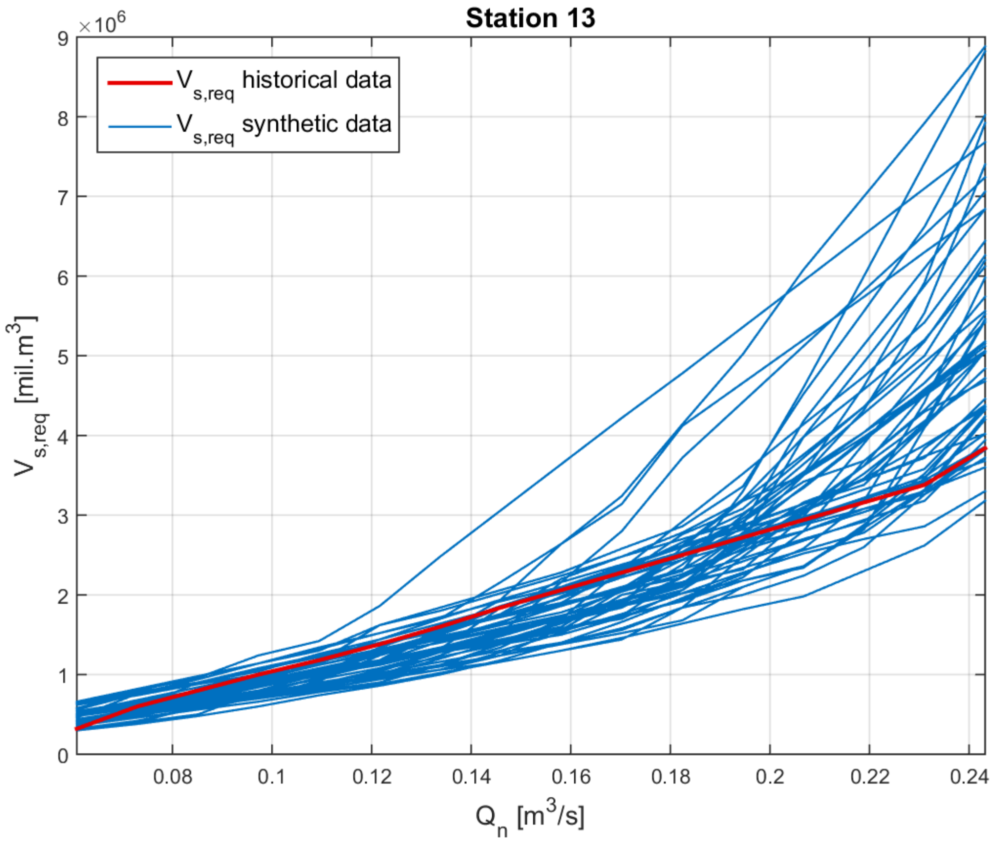


Station 10
×10^8
2.5
2
1.5
1
0.5
0
V_s,req historical data
V_s,req synthetic data
V_s,req [mil.m^3]
1.5
2
2.5
3
3.5
4
4.5
5
5.5
Q_n [m^3/s]

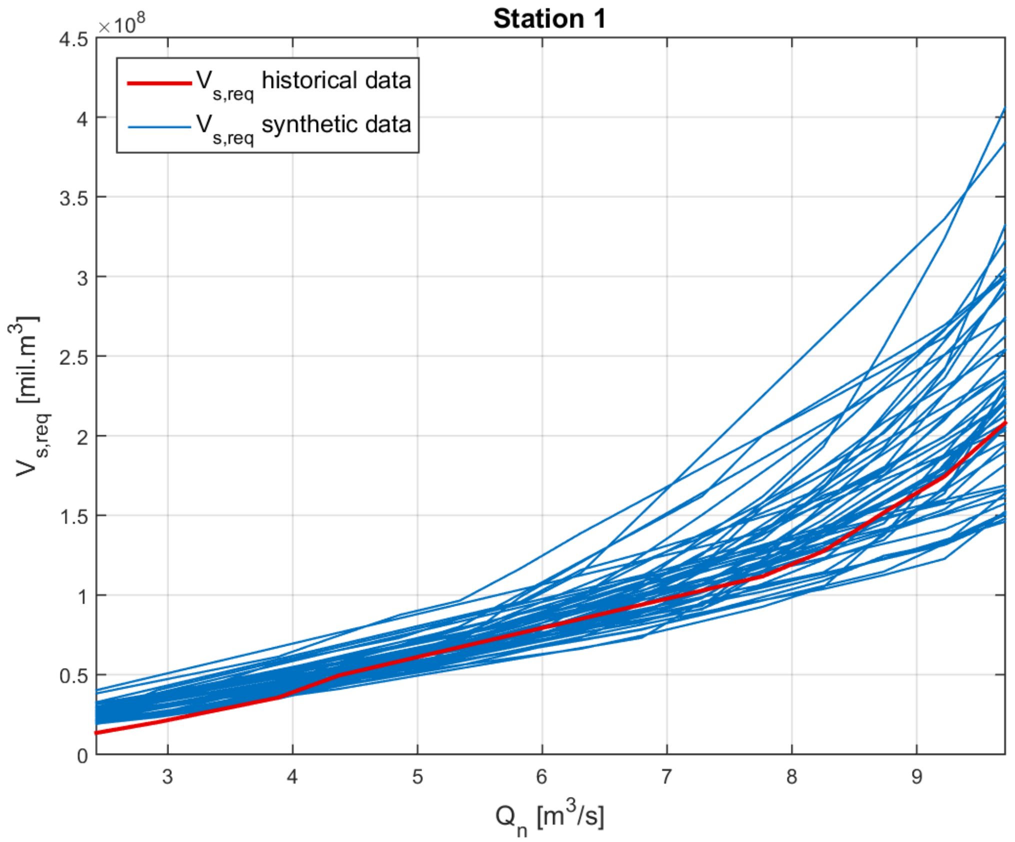


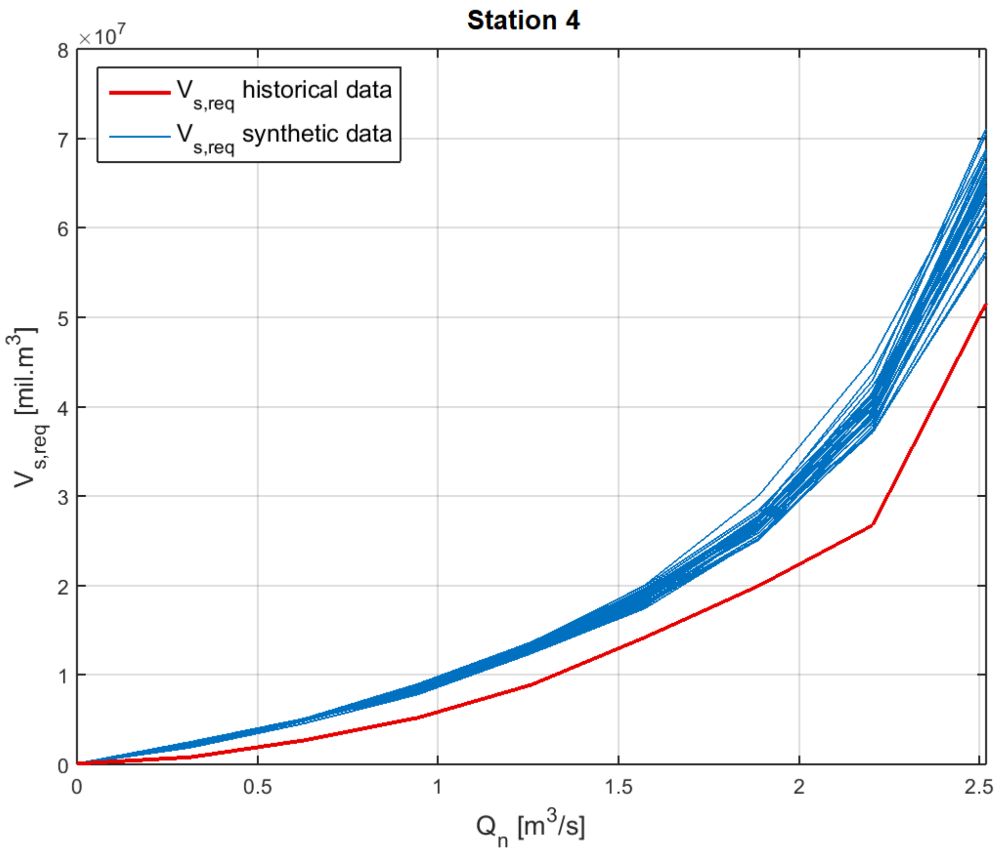
Station 11
×10^7
4
3.5
3
2.5
2
1.5
1
0.5
0
V_s,req historical data
V_s,req synthetic data
V_s,req [mil.m^3]
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Q_n [m^3/s]

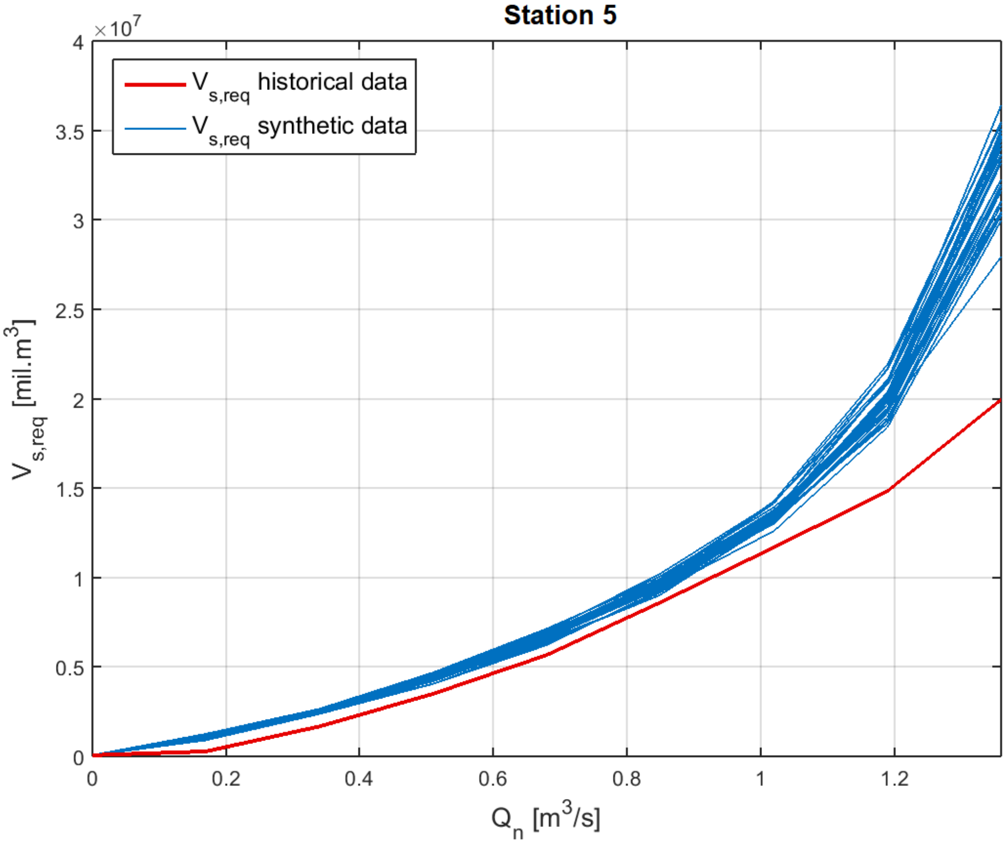

Station 12
×10^8
V s,req historical data
V s,req synthetic data
V s,req [mil.m3]
Q n [m3/s]
6
5
4
3
2
1
0
4
5
6
7
8
9
10
11
12

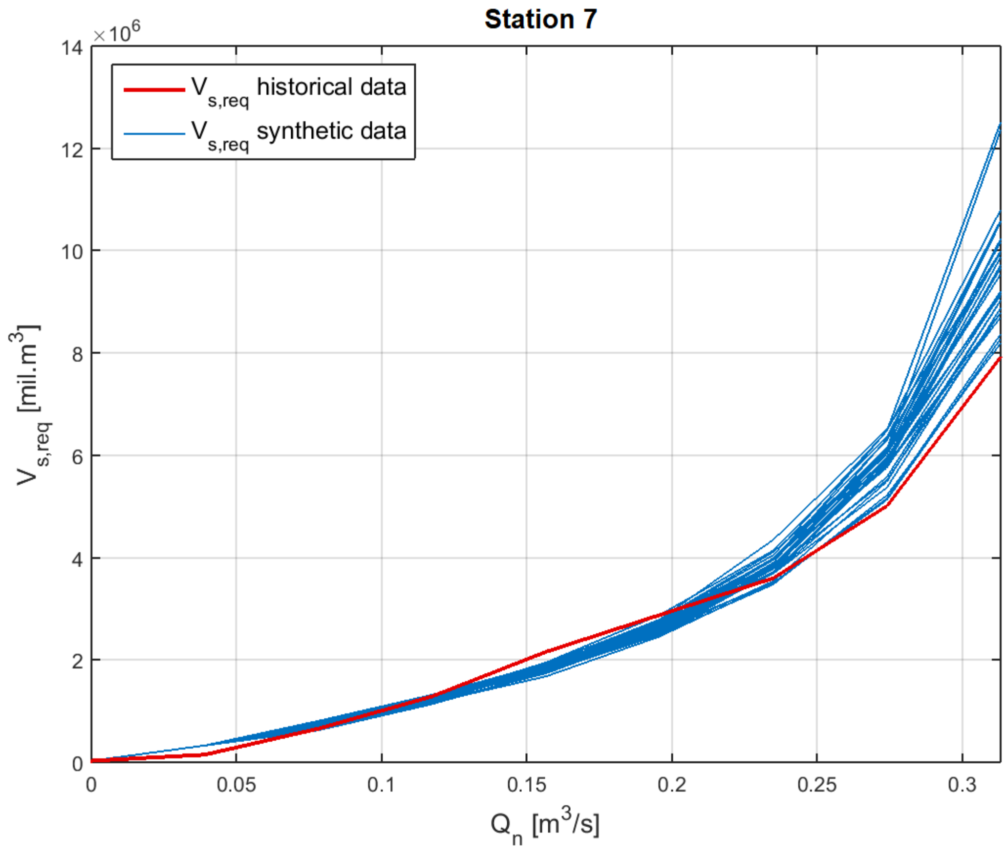

Station 13
×10^6
V s,req historical data
V s,req synthetic data
V s,req [mil.m3]
Q n [m3/s]
9
8
7
6
5
4
3
2
1
0
0.08
0.1
0.12
0.14
0.16
0.18
0.2
0.22
0.24

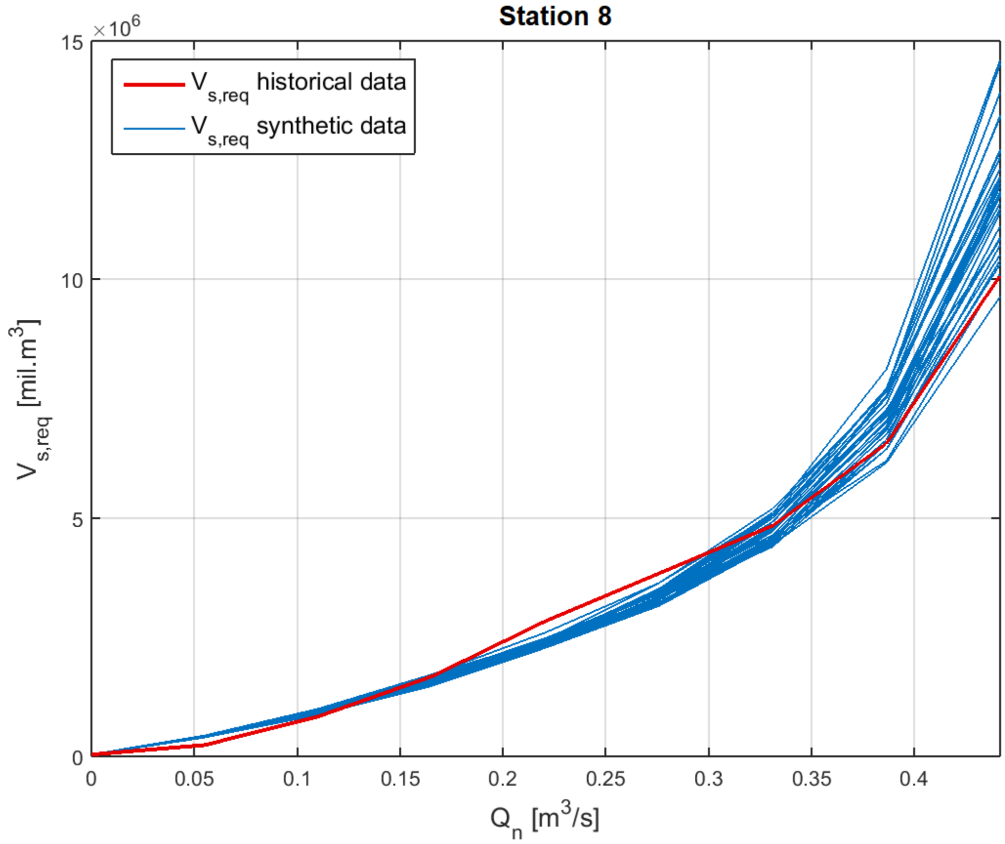
APPENDIX 7; Water Reservoir Operation Function plots for 100 years long synthetic TS

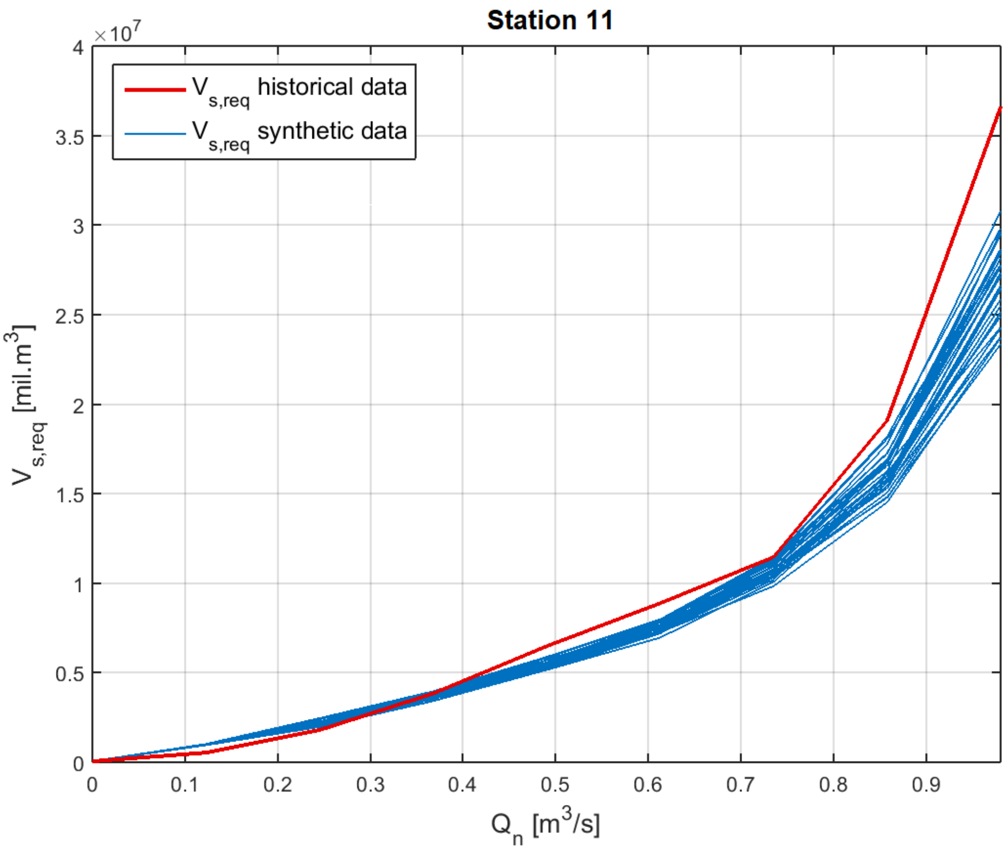





Station 5
×10^7
V_s,req historical data
V_s,req synthetic data
V_s,req [mil.m^3]
Q_n [m^3/s]
0
0.5
1
1.5
2
2.5
3
3.5
4
0
0.2
0.4
0.6
0.8
1
1.2


Station 7
×10^6
V_s,req historical data
V_s,req synthetic data
V_s,req [mil.m^3]
Q_n [m^3/s]
0
2
4
6
8
10
12
14
0
0.05
0.1
0.15
0.2
0.25
0.3


Station 8
×10^6
Vs,req historical data
Vs,req synthetic data
Vs,req [mil.m3]
Qn [m3/s]
15
10
5
0
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4


Station 11
×10^7
Vs,req historical data
Vs,req synthetic data
Vs,req [mil.m3]
Qn [m3/s]
4
3.5
3
2.5
2
1.5
1
0.5
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

APPENDIX 9; Matlab model

```
function [SYNT] = MTSM( X,T )
%%  MULTIVARIATE TIME SERIES MODEL
%%  performs PCA on any number of vectors and creates synthetic time series with
AR(1) model

%   Programmed by Waldemar Gresik
%       X - Input data - rows correspond observations and columns to
%       variables (stations)
%       T - number of synthetic years to generate
%   The function preserves the order of the months in the input
```

```
%% [1] Data preparation - measures dimensions, separate months and creates 12
matrices, one for each month
n=size(X,2);                    % n = number of stationsQmT
m=size(X,1);                    % m = lenght of the vectors - number of months
for i=1:12
    Qm{i}= X(i:12:end,:);       % creates cell array with 12 month matrices
End
```

```
%% [2] Normalize by 3LGN
CsN=zeros(12,n);
S0=zeros(12,n);
function[e]=cmin(c,kappax)          % Objective function
e=abs(c^3+3*c-kappax);              % Optimization criteria
end

for i=1:12
    N{i}=zeros(m/12,n);
    for j=1:n
        x = Qm{i}(:,j);         % Time series to normalize
        mux=mean(x);
        sigmax=std(x);
        kappax=skewness(x);
        initial=-0.5*std(x);    % initial guess for the optimization function

        % Optimization function - finds minimum of the Objective function
        c=fminsearch(@cmin,initial,[],kappax);

        x0=mux-1/c*sigmax;              % computes the shift parameter
        sigmay=(log(1+c^2))^0.5;        % second parameter of the distribution
                                        % first parameter of the distribution
        muy=log(sigmax)-log(abs(c))-1/2*log(1+c^2);

        y=log(x-x0);                    % Normalizing time series
        N{i}(:,j) = y;
        csy=skewness(y);                % Controls skewness of normalized data
        S0(i,j)=x0;                     % Assigns shift parameters
        CsN(i,j)=csy;                   % Assigns skewness coefficient
    end
end
assignin('base','S0',S0);
```

```
%% [3] 1. Standardization - Standardize data by substracting mean and dividing by
standard deviation
for i=1:12;
    S{i}=bsxfun(@minus,N{i},mean(N{i}));          % Makes the data zero mean
    S{i}=bsxfun(@rdivide,S{i},std(N{i}));         % Makes the data of standard
deviation of 1
end
```

```
%% [4] Kolmogorov-Smirnov test to Check if the probability distribution of the
vectors is standard normal
for i=1:12
    for j=1:n
```

```
        KS(i,j)=kstest(S{i}(:,j),'Alpha',0.05);     % K-S test with significance
level of Alpha (implicit value is 5%)
    end
end
disp('Kolmogorov-Smirnov I - before PCA - should be 0 and is:');
disp(sum(sum(KS)));
```

```
%% [5] PRINCIPAL COMPONENT ANALYSIS -
% finds the "mixing" matrix TRANSM such that the principal components are
uncorrelated
% and ordered by amount of variance they represent

for i=1:12;
   [TRANSM{i},Z{i},VAR{i}] = princomp(S{i});  % Singular Value Decomposition is
used to determine TRANSM
end
```

```
%% [6] 2. Standardization - smooths the redistribution of variance after PCA

for i=1:12;
    Zc{i}=bsxfun(@rdivide,Z{i},std(Z{i}));  % Makes data of variance of 1
end
```

```
%% [7] Kolmogorov-Smirnov test to Check if the probability distribution of the
vectors is standard normal
for i=1:12
    for j=1:n
        KSZ(i,j)=kstest(Zc{i}(:,j),'Alpha',0.05);    % K-S test with significance
level of Alpha (implicit value is 5%)
    end
end
display('Kolmogorov-Smirnov II - after PCA -  should be 0 and is:');
disp(sum(sum(KSZ)));
```

```
%% [8] AR(1) MODEL %% ---- %% AR(1) MODEL %%

Syn=zeros(T*12,n);             % empty matrix for synthetic data
Phi=zeros(12,n);                  % empty matrix for correlation coefficients
for j=1:n
    for i=2:12
        % Corr. coeff. for Novembers (one element shorter vectors)
        Phi(1,j)=corr(Zc{1}(2:m/12,j),Zc{12}(1:m/12-1,j));
        % Corr. coeff. for other months
        Phi(i,j)=corr(Zc{i}(:,j),Zc{i-1}(:,j));
    end
end
xm=zeros(T*12,1);          % empty vector for synthetic time series
for j=1:n
    epsilon=randn(T*12,1);% random numbers with N(0,1)
    xm(1,1)=epsilon(1);    % first observation (without deterministic component)
    for k=2:T*12;
        month=mod(k-1,12)+1;           % determination of month
        ssgm=sqrt(1-Phi(month,j).^2);% sigma coefficient for stochastic component

        % ´modelling of observations with AR(1) process
        xm(k,1)=(epsilon(k)*ssgm)+(xm(k-1,1)*Phi(month,j));
    end
    Syn(:,j)=xm;
end
assignin('base','Phi',Phi);
```

```
%% [9] Decomposition of Synthetic data for inverse transformations and correction
of statistics
for i=1:12
    Qmsyn{i}= Syn(i:12:end,:);
    % Correction of mean to 0
```

```
    Qmsyn01{i}=bsxfun(@minus,Qmsyn{i},mean(Qmsyn{i}));
    % Correction of variance to 1
    Qmsyn01{i}=bsxfun(@rdivide,Qmsyn01{i},std(Qmsyn01{i}));
end
```

```
%% [10] Reverse 2. Standardization from section [6]
for i=1:12;
    Zr{i}=bsxfun(@times,Qmsyn01{i},std(Z{i}));      % Incorporating PCA
redistributed variance
end
```

```
%% [11] Inverse PCA transformation - applying inverse of transformation matrix
TRANSM
for i=1:12;
    Sr{i}=Zr{i}*inv(TRANSM{i});       % Product with Inverse pca coefficients
matrix
end
```

```
%% [12] Reverse 1. Standardization from section [3]
for i=1:12;
    Nr{i}=bsxfun(@times,Sr{i},std(N{i}));     % Incorporating stadard deviation of
historical data
    Nr{i}=bsxfun(@plus,Nr{i},mean(N{i}));     % Incorporatin mean of historical
data
end
```

```
%% [13] Inverse normalization - reverses process from section [2]
for i=1:12
    for j=1:n
        Qmr{i}(:,j)=exp(Nr{i}(:,j))+S0(i,j);      % Inverse log transform into
Log-Normal distributed data
    end
end
```

```
%% [14] Correction of statistics
for i=1:12
    Qmr{i}=bsxfun(@times,Qmr{i},std(Qm{i})./std(Qmr{i}));
    Qmr{i}=bsxfun(@plus,Qmr{i},mean(Qm{i})- mean(Qmr{i}));
end
```

```
%% [15] Final Recomposition - creates synthetic data matrix
SYNT = zeros(T*12,n);
for i=1:12
    SYNT(i:12:end,:)= Qmr{i};        % puts the data in single matrix, following
the formatting of the input
end
```

```
%% [16] Check negatives
Negatives=zeros(T*12,13);
for i=1:(T*12)
    for j=1:n
        if SYNT(i,j)<0;
        Negatives(i,j)=1;
        end
    end
end
disp('Number of negatives for each station:');
disp(sum(Negatives));
```

```
%% [17] Replace negatives with minimum streamflow
correction=0.001*mean(X);
for i=1:n
    for j=1:T*12
        if SYNT(j,i)<0;
           SYNT(j,i)=correction(i);
```

```
        else SYNT(j,i)=SYNT(j,i);
        end
    end
end
assignin('base','SYNT',SYNT);
```

```
%% [18] Plot results
for i=1:n
    figure(i)
    set(gcf, 'Units', 'Normalized', 'OuterPosition', [0 0.3 1 (1/T^(1/6))]);
    plot (X(:,i),'Linewidth',1.0,'Color',[0.9020 0.0000 0.0000])
    hold on
    plot (SYNT(:,i),'Linewidth',1.2,'Color',[0.0000 0.4392 0.7529])
    xlim([0,T*12]);
    legend({'historical data','synthetic
data'},'location','northeast','FontSize',11);
    hold off
end

end
```

Appendix X; Diagram of the Time Series Model (TSM)