HydroCen

NORWEGIAN RESEARCH CENTRE
FOR HYDROPOWER TECHNOLOGY
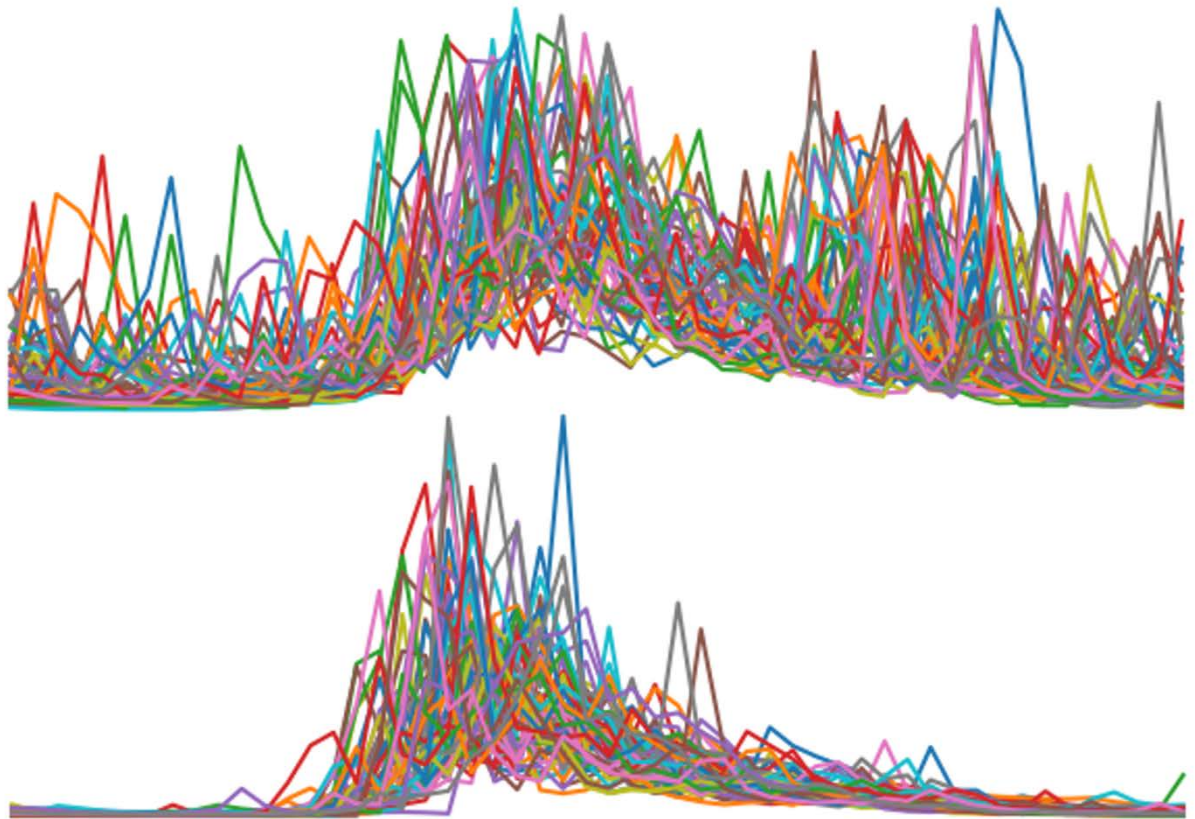
HydroCen Report   nr. 30

Improved Inflow Modelling
For the SDDP Algorithm

Siri Mathisen

NTNU   SINTEF   NINA

HydroCen

The main objective of HydroCen (Norwegian Research Centre for Hydropower Technology) is to enable the Norwegian hydropower sector to meet complex challenges and exploit new opportunities through innovative technological solutions.

The research areas include:

- Hydropower structures
- Turbine and generators
- Market and services
- Environmental design

The Norwegian University of Science and Technology (NTNU) is the host institution and is the main research partner together with SINTEF Energy Research and the Norwegian Institute for Nature Research (NINA).

HydroCen has about 50 national and international partners from industry, R&D institutes and universities.

HydroCen is a Centre for Environment-friendly Energy Research (FME). The FME scheme is established by the Norwegian Research Council.

The objective of the Research Council of Norway FME-scheme is to establish time-limited research centres, which conduct concentrated, focused and long-term research of high international calibre in order to solve specific challenges in the field.

The FME-centres can be established for a maximum period of eight years HydroCen was established in 2016.

Improved Inflow Modelling

For the SDDP Algorithm

Siri Mathisen

HydroCen

Abstract

This report is written as part of HydroCen WP3, associated with WP3.3 and WP3.4. Hydropower scheduling enables optimal use of the flexible energy stored in hydropower reservoirs. Taking inflow and power price prognoses as input, a detailed operation strategy for a river system can be made. When an operation strategy for a period of analysis spanning several years, there are no accurate inflow forecasts available, and a stochastic inflow model can be used instead. The accuracy of the strategy depends on the accuracy of the inflow model. This report focuses on stochastic inflow models suitable for the group of hydropower scheduling models based on the stochastic dual dynamic programming (SDDP) algorithm, which has a convexity requirement for the optimization problem and thus requires a convex inflow model. The state-of-the-art inflow model used in the SDDP algorithm is some variation of the autoregressive model, which is a linear model estimated from a set of observed samples. One drawback of the basic autoregressive model when used to synthesize inflow is that they can occasionally produce negative inflow, which can influence the hydropower production strategy badly. Another drawback is the insufficient ability to capture slow changing trends in the inflow, to model long-lasting extreme inflow values. The aim of this research is to study available literature on stochastic inflow modelling for SDDP with specific focus on avoiding negative inflow and capturing low-frequency trends in the inflow. The more promising inflow models are implemented and tested to review their statistical properties and to compare them with the statistical properties of the historical inflow observations.

The literature study describes the basic principles of the SDDP algorithm and how the inflow is modelled in SINTEF's implementation of the algorithm for medium-term hydropower scheduling found in the ProdRisk model. Then the report moves chronologically through the literature describing the development in SINTEF's SDDP-based algorithm. The next sections describe how the general literature treats the subject of negative inflow produced by the inflow models in SDDP algorithms, and whether the inflow models in the literature can capture low-frequent trend changes. A comparison of the implementations of the more promising inflow models shows that a 3-parameter log-normally distributed noise for the autoregressive inflow model can prevent the model from producing negative inflows while still conserving the statistical properties of the observations. Furthermore, if a term containing the average of the previous year's inflow is added to the model, then the standard deviation of the generated inflow gets much closer to the standard deviation of the observed inflow and the error of the generated average annual inflow is decreased. The resulting model can still be used in an algorithm demanding convexity and is suited for implementation in an SDDP algorithm. The alternative method to generate strictly non-negative inflow samples is to fit an autoregressive model to logarithm transformed inflow samples and transform the generated inflow back using an exponential function. However, the resulting generated inflow will be strongly non-linear, and a linearization process is needed to make the model suitable for the SDDP algorithm. Therefore, the conclusion is to use the autoregressive model with a 3-parameter log-normally distributed noise and an annual component to capture the low-frequent trend changes.

Siri Mathisen, SINTEF Energi, Trondheim, siri.mathisen@sintef.no

Sammendrag

Mathisen, S. 2022. Improved Inflow Modelling. HydroCen rapport 30. Norwegian Research Centre for Hydropower Technology

Denne rapporten er skrevet som en del av HydroCen WP3, i samarbeid med WP3.3 og WP3.4. Vannkraft-planlegging muliggjør optimal bruk av den fleksible energien som er lagret i vannmagasiner. En detaljert strategi for drift av et vassdrag kan lages med tilsig og kraftprisprognoser som input. Når en produksjons-strategi for en periode på flere år skal lages, er det ikke presise tilsigsvarsler tilgjengelig så en stokastisk tilsigsmodell kan brukes istedenfor. Nøyaktigheten av en strategi avhenger av nøyaktigheten av tilsigs-modellen. Denne rapporten setter søkelys på stokastiske tilsigsmodeller tilpasset en gruppe vannkraft-planleggingsmodeller basert på stokastisk dual dynamisk programmering, som har et konveksitetskrav til optimaliseringsproblemet og dermed krever en konveks tilsigsmodell. En velkjent tilsigsmodell i bruk i SDDP-algoritmen er en variant av den autoregressive modellen, som er en lineær modell estimert fra et sett av observasjoner. En ulempe ved den enkle autoregressive modellen når den brukes til å syntetisere tilsig er at den tidvis produserer negative tilsig, som kan påvirke produksjonsstrategien for vannkraftverk negativt. En annen ulempe er den utilstrekkelige evnen til å fange opp langtidstrender i tilsiget, for å modellere ekstreme tilsigsverdier som pågår over tid. Målet med denne forskningen er å studere tilgjeng-elig litteratur på stokastisk tilsigsmodellering for SDDP med søkelys på hvordan negativt tilsig kan unngås og hvordan lavfrekvente trender i tilsiget kan tas med i modellen. De mest lovende tilsigsmodellene im-plementeres og testes for å vurdere deres statistiske egenskaper og for å sammenlikne dem med de statistiske egenskapene til historiske observasjoner.

Litteraturstudiet beskriver basisprinsipper for SDDP-algoritmen og hvordan tilsiget modelleres i SIN-TEF's implementasjon av algoritmen for mellomtids vannkraftplanlegging, som kan finnes i ProdRisk-modellen. Rapporten beveger seg deretter kronologisk gjennom litteraturen som beskriver utvik-lingen til SINTEF's SDDP-baserte algoritme. De neste seksjonene beskriver hvordan den generelle litteraturen behandler produksjon av negativt tilsig av tilsigsmodeller i SDDP-algoritmer, og om til-sigsmodeller i litteraturen kan fange lavfrekvente endringer i tilsiget. En sammenlikning av imple-mentasjoner av de mest lovende tilsigsmodellene viser at en 3-parameter-log-normal fordeling av støyen i den autoregressive tilsigsmodellen kan forhindre modellen fra å produsere negativt tilsig, samtidig som de statistiske egenskapene i observasjonene bevares. Dessuten kan man legge til et ledd i modellen som inneholder gjennomsnittet av det siste årets tilsig, slik at standardavviket til det genererte tilsiget kommer nærmere standardavviket til det observerte tilsiget, og feilen i det gene-rerte gjennomsnittlige årlige tilsiget blir mindre. Denne modellen kan fremdeles brukes i en algo-ritme som krever konveksitet, og egner seg for implementering i en SDDP-algoritme. Den alternative metoden for å frembringe strengt ikke-negative tilsig er å tilpasse en autoregressiv modell til det logaritmetransformerte tilsiget, og transformere det genererte tilsiget tilbake med en eksponential-funksjon. Imidlertid vil det genererte tilsiget bli sterkt ikke-lineært med denne metoden, og må li-neariseres for kunne brukes i en SDDP-algoritme. Konklusjonen er derfor å bruke den autoregressive modellen med en 3-parameter-log-normal fordeling av støyen og en årlig komponent som fanger opp de lavfrekvente tilsigsendringene.

Siri Mathisen, SINTEF Energi, Trondheim, siri.mathisen@sintef.no

Content

Foreword

The main purpose of this report is to document the work done in HydroCen Open Calls project Inflow Modelling, which is part of WP3 and associated with WP3.3 and WP3.4. The main aim of the study is to study available literature on inflow models used in SDDP algorithms with special focus on non-negative inflow and the ability to capture a longer trend. The HydroCen research centre is financed by the Norwegian Research Council, the Norwegian hydropower industry, NVE and the Norwegian environment agency.

Trondheim December 2022, Siri Mathisen

Foreword

1 Introduction

1.1 Motivation

Optimal scheduling of hydropower resources is important in the transition towards a zero-emission society as it enables hydropower producers to produce more power when the demand is higher and allows them to save water when the demand is lower. It may also reduce spillage from reservoirs and prevent overflow that can be destructive on settlements or infrastructure downstream the reservoir. As hydropower scheduling is a large and complex problem, it is divided spatially and temporally into subtasks: The long-term power market problem that makes price prognoses for a large geographical area, the medium-term hydropower scheduling problem that takes the price prognoses as input and provides a production strategy, and the short-term operational scheduling problem that uses the production strategy to provide an operational plan for the impending future. The medium-term scheduling typically covers a period of analyses longer than we are able to accurately forecast inflow for. Thus, historical records or stochastic models derived from historical data are typically used to predict the inflow, with the addition of forecasts for the short-term horizon.

SINTEF has developed a model called ProdRisk intended for medium-term scheduling on local hydropower systems, assuming the perspective of a price taker. The model uses a combination of stochastic dynamic programming (SDP) and stochastic dual dynamic programming (SDDP) to generate a production strategy describing the future expected profit as a function of reservoir storages, price, and inflow. The discretized price process is treated as an exogenous state in the SDDP algorithm, whereas inflow is a continuous state variable in the SDDP algorithm and therefore subject to the algorithm's convexity requirements [1]. The stochastic inflow used as input to ProdRisk is modelled as a first-order vector autoregressive (VAR1) process with weekly time steps, as it is simple to implement, contains few states and captures the observed causality of the inflow. This inflow model produces occasional negative inflow values, which can violate requirements to the reservoir volume. To avoid a negative reservoir level, a slack variable is introduced and penalized severely in the objective function. This penalty does not reflect a reality without negative inflow samples. To overcome this problem, ProdRisk penalises constraint violations, but the negative inflows can influence the strategy more than realistically necessary and higher inflow variation due to climate changes magnify the problem. Another weakness with the autoregressive process is its underrepresentation of prolonged extreme inflow values lasting for several weeks. The deterministic part of the process tends towards zero, producing average inflow samples, and the stochastic part of the process has an average value of zero.

A model-based control strategy is dependent on a well-modelled process. The inflow variation is intended to create robustness in the production strategy, and an inflow model that is inconsistent with the reality will not produce the optimal strategy for a real hydropower production system. This motivates research on stochastic inflow models intended for use in SDDP-based algorithms. Specifically, the inflow model used in ProdRisk must be improved.

1.2 About the project

The aim of this project is to study available literature on stochastic inflow modelling for SDDP algorithms, assess promising inflow models with special focus on non-negative generated inflow and the ability to capture a longer trend, and give a recommendation of which inflow model should be implemented in ProdRisk. The suitability of the proposed inflow models will be assessed using the following criteria:

- Is the model linear?
- Is the model suited for inclusion in the SDDP algorithm?
- Does the model produce negative inflow?
- Does the model on average reproduce the annual inflow from the observed scenarios?
- Is the model able to reproduce the distribution of the observed inflow?

Measuring the impact that the change in the inflow model has on the ProdRisk computation time is beyond the scope of this project. The literature search is originated in the current VAR process and focuses

on strategies to avoid negative inflow and to maintain the representation of prolonged extreme values that we find in the observed inflow. It will build competence on inflow modelling for SDDP algorithms and promising inflow models will be implemented and tested in a simplified SDDP algorithm. The result of this will be a report containing the literature survey, the assessment of inflow models suitable SDDP algorithms and a recommendation for which inflow model should be implemented in ProdRisk.

This project will join forces with an industrial project called *Improved Stochastic Inflow Modelling in ProdRisk* (IMPRO), which started in Q1 2022. The idea of IMPRO is to implement improved inflow modelling in the SDDP algorithm in ProdRisk. This project will contribute to the IMPRO project by building competence on the newest research results and perform own testing of possible statistical methods, supporting the choice of method to be implemented in IMPRO.

2 Inflow series used for testing of inflow model implementations

For implementation testing purposes, the NVE inflow series 630-E og 956-E for the locations Øye and Risefoss, used for the Aura River system in hydropower scheduling, will be used. The series contain 58 historical weather years from 1958.01.01 with daily time resolution.

To test the performance of the inflow models, NVE inflow series from the following river systems will be used; all having daily time resolution and 58 weather years starting on 1958.01.01:

Table 1: Inflow series used in the testing of stochastic inflow models

River system	Inflow Series
Aura	630-E, 956-E
Hallingdal	2478-E, 1364-E, 1736-E, 2723-E
Aurland	606-E, 1420-E
Mandal	2024-E, 1150-E, 556-E, 530-E
Arendal	2024-E, 530-E, 2155-E, 1958-E, 2304-E
Sira-Kvina	1804-E, 1950-E, 2629-E, 556-E, 1802-E
Lysebotn	1950-E
Rjukan	1687-E, 2155-E, 1206-E, 2301-E, 1958-E, 592-E, 2478-E, 2152-E, 1128-E
Tyin	630-E, 956-E

3 The SDDP algorithm and requirements for the inflow model

The standard SDDP algorithm [2] consists of subsequent backward iterations generating the production strategy and forward iterations testing the same strategy in simulations, until the objective from the forward and backward iterations converge within a statistical confidence interval. Each iteration of the SDDP algorithm consists of stage-wise solutions of linear programming (LP) problems where the initial conditions of one stage are the end states of the previous stage. The production strategy describes the upper limit of the future expected costs for the hydropower plants in the system and can be illustrated as a set of hyperplanes or cuts along the dimensions of the states.
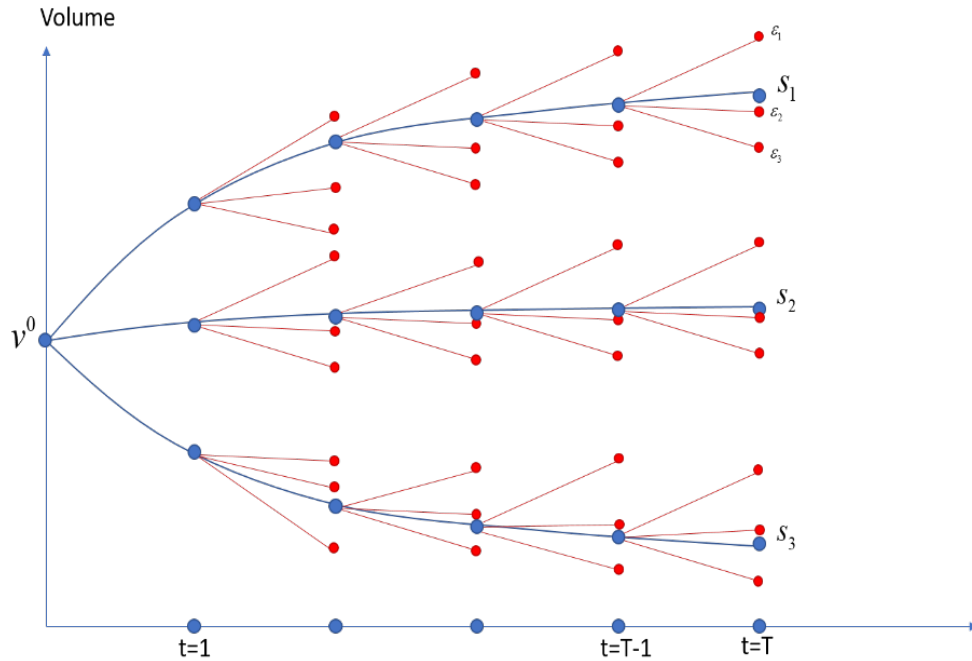

Figure 1: Illustration of the forward and backward iterations in the SDDP algorithm. The state of the system is illustrated with volume v for time t and the forward inflow uses observed inflow scenarios s_i. The backward inflow contains stochastic error terms.

The SDDP-algorithm has a convexity (or concavity for maximization problems) requirement, stating that the expected future profit should be convex (or concave for maximization) in all state variables. In ProdRisk, the state variables are reservoir volumes, inflows, and the market price. Although the inflow is not a decision variable in the optimization problem, it is a state variable in the SDDP problem (due to the time correlation) formulated in ProdRisk and must therefore fulfil the concavity requirement [3]. For each set of initial states of the system, the set of cuts will provide a suitable production strategy. This is described in e.g. [2], [4]. The inflow in the backward iterations of the SDDP algorithm is given by a stochastic model where the inflow of the current week has a linear relationship with the inflow of the previous week, with a number of variations given by the noise vector ε. The inflow of the SDDP algorithm is illustrated in Figure 1.

The novelty of the SDDP algorithm compared to the SDP algorithm was its ability to overcome the curse of dimensionality, which makes the SDP algorithm computationally intractable for multi-reservoir systems. As the cuts can be used for any set of initial states in a specific decision stage, they can be shared among the simulated scenarios in the forward simulations. Cut sharing is a requirement for sensible convergence properties in the SDDP algorithm, and the concavity is a condition for cut sharing. [4] states that although the SDDP algorithm is intended for linear problems, it could be extended to the nonlinear

case, but is dependent on a convex problem (or a concave relationship between the state variables and the objective function) to guarantee convergence to a global optimum. In [5], the concavity requirements for cut sharing in the SDDP algorithm are elaborated and formalized.

In [1], the combination of SDP and SDDP used in ProdRisk is introduced, where the inflow and reservoir volume are continuous state variables in the SDDP algorithm and the price process is discretized and treated as an outer layer based on SDP. ProdRisk uses historical observed inflow and price scenarios for the forward iterations of the SDDP algorithm to preserve correlations between the price and inflow. In the backward recursions, however, a first-order autoregressive inflow model is used, estimated from observed inflow.

As the functional relationship between the inflow and the future expected profit is concave, it is possible to model the inflow so that it does not violate the concavity requirement of the SDDP algorithm. The functional relationship between the reservoir volume and the expected future profit is concave, whereas the functional relationship between the price and the expected future profit is convex. The price is therefore treated in an outer SDP-algorithm outside the SDDP algorithm in ProdRisk. This topic is further discussed in [1] .

4 The inflow model used in ProdRisk

In ProdRisk, the inflow of the forward simulation is given by the observed historical inflow scenarios. The inflow of the backward recursion is modelled with a first-order vector autoregressive (VAR1) model. A VAR1 model is linear and time invariant, and therefore the simplest way to model a process. It captures the relationship between the process state in one time step z_t and the process state in the previous time step z_{t-1} through the transition constant ϕ and with a standard normally distributed noise input $\varepsilon \sim N(0,1)$ that is independent of the process state z_t:

$$z_t = \phi z_{t-1} + \varepsilon \tag{1}$$

The VAR1 is a special case of the more general ARMA – autoregressive moving average – family of models, with zero moving average. To fit a series of observations to a VAR1 model, the series must be a weakly stationary process, with time invariant mean value and an autocorrelation (or autocovariance) that is dependent only on the time difference between t_1 and t_2 [6]:

$$
\begin{aligned}
&E[X(t_1)] = E[X(t_2)], \\
&E[X(t_1)X(t2)] = E[X(t_1 + \Delta)X(t_2 + \Delta)]
\end{aligned}
\tag{2}
$$

Most inflow series are not stationary, but the standardization of the observations will most often make the series weakly stationary, given that the autocovariance is independent of the time t. This approach can be extended to several inflow series, making a VAR1-model. Starting with N inflow series containing S historical inflow records with observations from T weeks, $q_{t,n}^i$ is the observed inflow of week t for inflow scenario i and series n. The weekly observations are standardized to remove seasonal variation, enabling us to estimate one transition matrix (or correlation matrix) for the whole set of observations. An alternative is to divide the observations into seasons and estimating one transition matrix per season. The standardized observed inflow $z_{t,n}^{i,obs}$ is obtained by subtracting the sample mean value $\overline{q}_{t,n}$ and dividing by the sample standard deviation $s_{t,n}$, for week t and series n. Omitting the serial index n, this gives:

$$z_t^{i,obs} = \frac{q_t^i - \overline{q}_t}{s_t} \quad \text{for} \quad i = 1,..,S, \quad t = 1,..,T, \tag{3}$$

The sample standard deviation for sample values q_t^i with Bessel's correction is given by:

$$s_t = \sqrt{\sum_{i=1}^{S} \frac{(q_t^i - \overline{q}_t)^2}{S-1}} \tag{4}$$

The observed and standardized inflow scenarios are shown in Figure 2. Then a VAR1 model is fitted to the standardized inflow scenarios $\mathbf{z}_t^j$. It is possible to split the observations in seasons and estimate one transition matrix for each season. M forecasted inflow scenarios are generated for the N inflow series by:

$$\mathbf{z}_t^j = \phi \mathbf{z}_{t-1}^j + \boldsymbol{\varepsilon}_t^j \quad , t = 1,..,T, \quad j = 1,..,M, \tag{5}$$

Where ε_t^j is a vector of size N containing the noise of week *t* for forecasted scenario *j*, ϕ is the transition matrix of size $N \times N$ describing the correlation between the inflow of two time steps and the cross correlation between series. $\mathbf{z}_{t-1}^i$ is the modelled inflow of week *t-1* and forecasted scenario *j* for *N* inflow series. The initial values are given by the model user, for simplicity 0 can be used. ε_t^j is assumed independent of $\mathbf{z}_t^j$, with a distribution function independent of t. To estimate the transition matrix ϕ, a usually multivariate linear regression and least-squares estimation are used. The input to the least-squares estimation consists of all standardized observations for each season. As it is a multivariate linear regression to gain one common transition matrix for all inflow series, all inflow series are used as input to the estimation. As a result, there is one transition matrix and *N* sets of residuals for each season, but common for all weeks within this season. To find the transition matrix for a spatially uncorrelated autoregressive model (AR), the Yule-Walker equations can be used [7]. We omit the vector notation in the rest of the section.
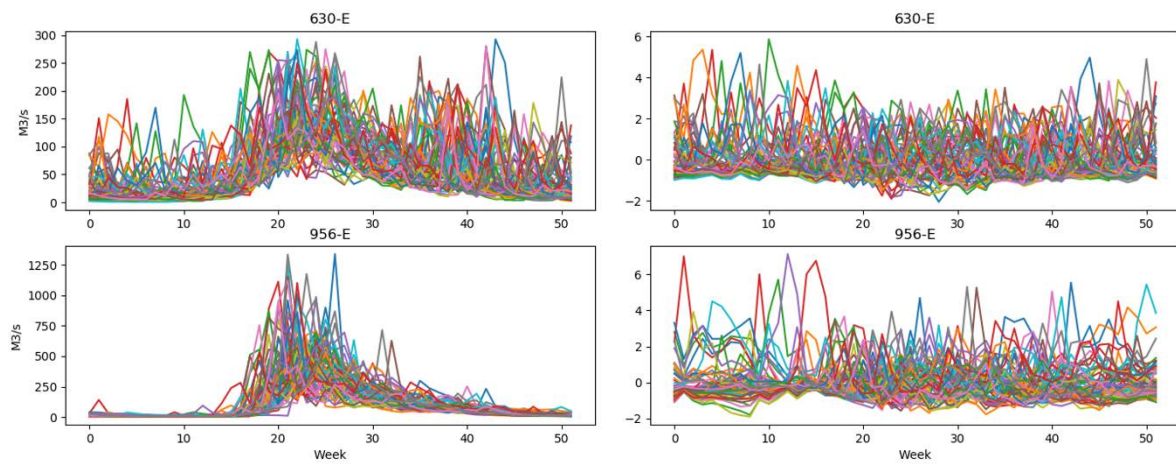

Figure 2: Observed (left) and standardized inflow (right) for the inflow series 630-E and 956-E.
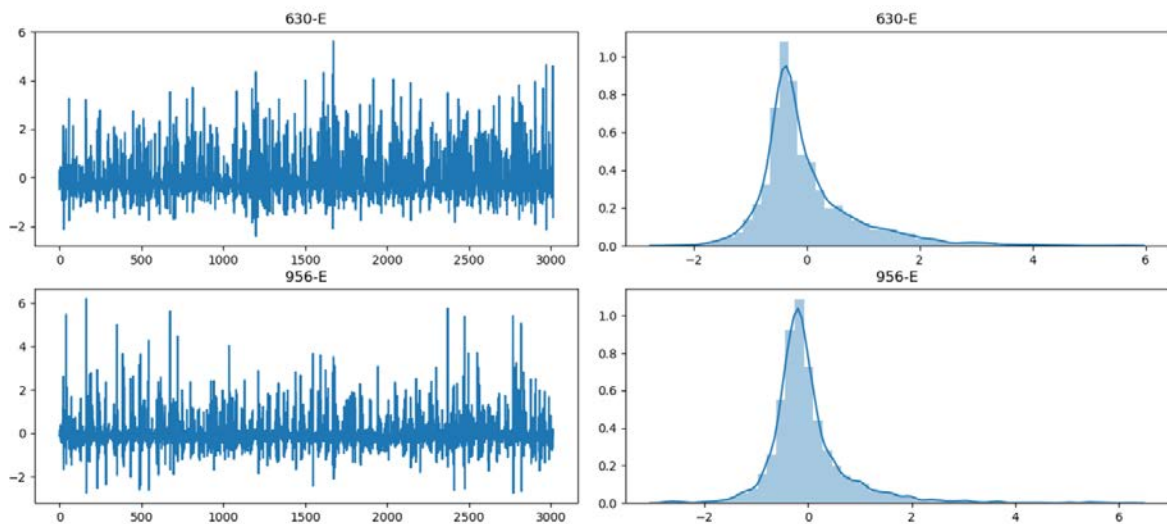

Figure 3: Residuals from the fitting of the AR1 model (left) and the distribution of the residuals (right) for the inflow series 630-E and 956-E.

After the autoregressive model is fitted to the standardized observations, the residuals are used to form a probability distribution for the noise. One can either sample directly from the residuals or reduce the cardinality by e.g. a clustering method. In ProdRisk, the alternative to direct sampling from the residuals is to use principal component analysis (PCA) on the residuals. The PCA is not treated further in this work, but literature on how the PCA is used in ProdRisk can be found in e.g. [8], [9], [10]. The residuals, and the distribution of the residuals, are shown in Figure 3. In this report, the noise is sampled randomly from the set of residuals belonging to the week that is currently being generated. For instance, when the inflow for the tth week of the year is synthesized, the noise is sampled from the residuals coming from the inflow observations of week t. With S available observed inflow scenarios, a random number s between 1 and S is chosen and the residuals from inflow series 1 through N for inflow scenario s are used to generate week t for N inflow series. This is shown in Figure 4, where the random number 3 is chosen and the residuals from inflow series 1 through N are marked. When week t+1 is synthesized, a new random number is chosen and the residuals from the corresponding historical inflow scenario are used.

Values for week t:	Inflow series n				
	1	2	3	...	N
1	0.4364	-0.876	0.325	...	-0.904
2	0.5364	0.001	-0.661	...	-0.0231
3	-0.0164	0.421	-0.025	...	0.002
...	...	...	...	...	...
S	0.0020	-0.516	0.4793	-0.0317	-0.0412
Inflow scenarios s					

Figure 4: To generate the inflow of week t for inflow series 1 through N, the residuals from the corresponding inflow series from the observed scenario s are chosen, where s is a random number between 1 and S.
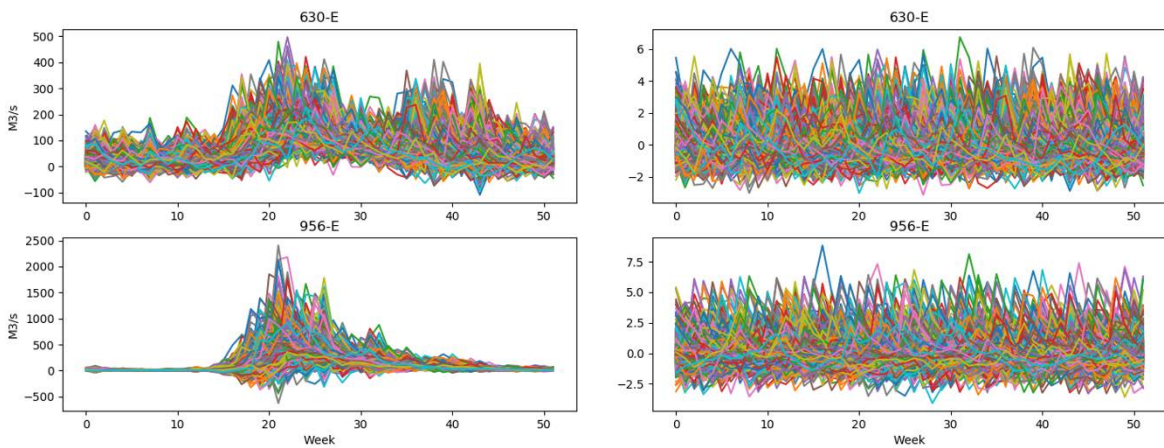
Figure 5: Generated, standardized inflow (right) and generated inflow (left) for the inflow series 630-E and 956-E, using theory from the ProdRisk implementation, M=1000.

The forecasted inflow scenario samples are then de-standardized using the mean value and standard deviations from the weekly observations (omitting vector notation):

$$\hat{q}_t^j = s_t z_t^j + \overline{q}_t = s_t(\phi \frac{q_{t-1}^j - \overline{q}_{t-1}}{s_{t-1}} + \varepsilon_t^j) + \overline{q}_t \qquad (6)$$

The standardized forecasted inflow scenarios and the generated forecasted inflow scenarios are shown in Figure 5. The variations of the observed and forecasted inflow series using the VAR1 inflow model that is currently used in ProdRisk, are shown in the percentiles and boxplot in Figure 6. The variations of the observed and forecasted inflow series are of the same size, but the forecasted inflows generally produce lower minimum values than the observed ones, due to the negative generated inflow values. The other percentiles are generally higher for the modelled inflow than for the observed inflow.

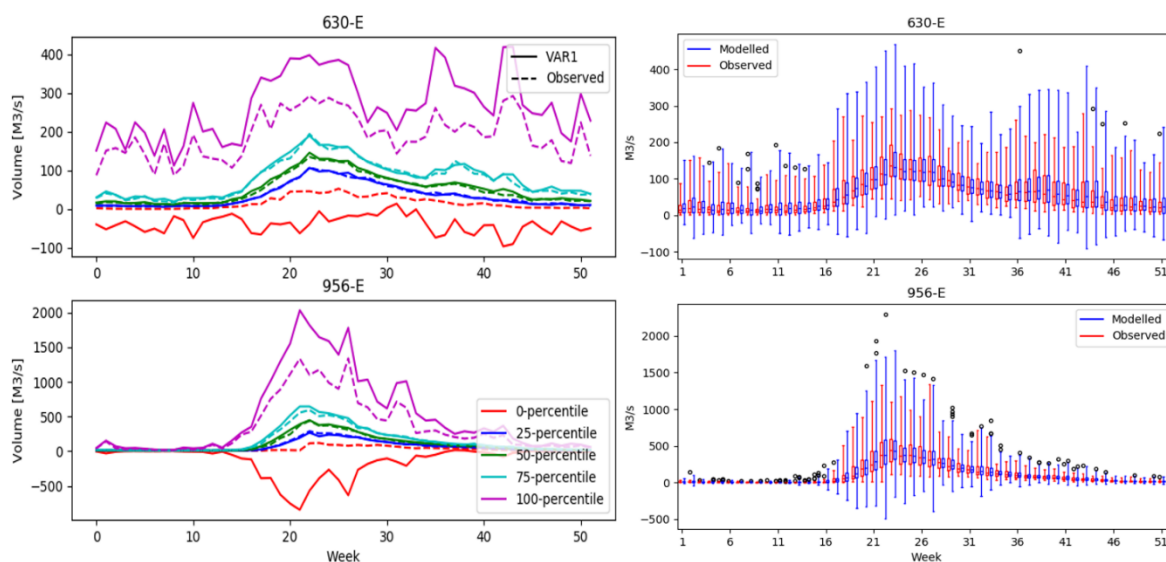Figure 6: Percentiles (left) and boxplot (right) of the observed and generated inflow series for each week for the inflow series 630-E and 956-E, using the current ProdRisk implementation and M=1000.

4.1 Literature review of ProdRisk

The AR1 inflow model is first introduced for the SINTEF models in [11] with a single inflow series. Here, the model is not aimed specifically at the SDDP model with its requirement for convexity. This report introduces the logarithmic transformation as a solution to the cases where the stochastic inflow model produces negative samples. Then, the inflow model is described in [10], using several inflow series and with an SDDP model in mind. This report is more specific and practical than the previous one, explaining in detail how the computations for the inflow model is done. The report also describes how the noise term of (1) is found: In a Monte Carlo simulation, the noise can be sampled from the set of equiprobable residuals from the fitting of the transition matrix. However, if only a small number of scenarios are generated, the sampling does not cover the probability range sufficiently, and principal components with corresponding probability are used. The report also doubts the weakly stationary property of the standardized inflow series, and states that logarithmic transformation cannot be used in operation scheduling with SDDP, as that algorithm demands certain properties of the inflow model. In fact, the report remarks that as the inflow model is intended for operation scheduling and not general inflow synthesizing, a simpler model is chosen. In [5], the concavity requirements for cut sharing in the SDDP algorithm are elaborated and formalized, laying the foundation for the later papers.

In [1], the combination of SDP and SDDP used in ProdRisk is introduced, where the inflow and reservoir volume are continuous state variables in the SDDP algorithm and the price process is discretized and treated in the SDP algorithm. As in modern ProdRisk, the paper describes the observed inflow and price scenarios used in the forward iterations of the SDDP algorithm, intended to preserve long-term couplings between price and inflow. The paper also touches the convergence gap that this inconsistency between stochastic models and observed values for inflow and price introduces.

Then in [12], SDDP inflow models and their statistic properties are studied. The noise generation in the VAR1 model is considered, and where originally only a PCA was used in the earlier versions of the ProdRisk inflow model, the residual method is now introduced to increase the variance of the inflow. Also here, the lack of the weak stationarity is commented on, but the report offers no solution to this problem. In [8], the inflow model used in the current version of ProdRisk is explained, providing an introduction to the stochastic inflow modelling used today, its weaknesses and examples to understand them.

In [13] and [14], an inflow model without convexity requirements was studied, focusing on improving and validating univariate models and attempt to extend this model to multivariate inflow series, then validate this. Although the reports claim good results, they remark that the time dependencies only hold for one time period, and an inclusion of several past weeks will complicate the computations significantly without improving the modelled inflow significantly. The last report suggests using covariates such as snow in the model to help build the seasonal trend, but this would require a model for them as well.

In [9], an SDDP-based model is used on a case study where the Norwegian hydropower system as a whole is scheduled, containing 500 hydropower modules, 105 unique inflow series and using weekly stochastic time resolution, to test the model on a larger system than it is usually used on. The paper concludes that the SDDP method is relevant on larger power system with a high share of hydropower, although better results can be obtained with smaller computation times using other models. The paper remarks that the VAR1 inflow model represents fewer inflow models over a smaller geographical area better than more inflow series over a larger geographical area.

5 Focus on negative inflow generated in SDDP algorithms in the literature

The Inflow model used in ProdRisk was last revised in 2005, with the introduction of noise generation with residuals. However, in some SDDP implementations used elsewhere, procedures for dealing with negative inflows have been successfully applied. Moreover, recent literature (in this context later than 2005) has been presented to improve inflow modelling in SDDP. Note that most of the technical literature on this topic apply monthly decision stages, while ProdRisk applies weekly decision stages.

The three-parameter log-normal distribution fitted to the residuals will be studied closely, as will the box-cox transformation of the observed inflow. The other relevant literature will be accounted for more briefly.

[15] discussed how to deal with negative inflows when the inflow model does in fact generate negative inflow samples. The discussed strategies are to penalize slack variables for the negative inflow samples in the objective function, to truncate the negative inflow and to find an optimal inflow value to maintain feasibility of the long-term horizon scheduling problem. The last approach is only meant for the forward iteration of the SDDP algorithm.

In [16], an inflow model developed for the Brazilian hydroelectric system is described. The inflow model is intended for use in SDDP and is therefore bound by convexity requirements, which is solved using a periodic AR (PAR) model of different orders for each month, with monthly transition matrices and standardization within each month. To avoid negative inflow, the paper describes the fitting of a three-parameter log-normal distribution to the residuals from the AR model fitting. The paper describes approaches for both univariate and multivariate time series. The three-parameter log-normal distribution fitting for the residuals from the AR model fitting is used and explained in [17] and [18] (together with other methods) based on parameter estimates from [19] and [20], and from [21] respectively. The latter does not explain all its parameter estimations, and thus the PAR Model Partially Truncated as described in [18], [19] and [20] are more reliable.

In [22], a method called *selective sampling* is used within the operative Brazilian SDDP-based long-term scheduling model NEWAVE. This method is used to reduce the cardinality of the samples, going from a large equiprobable set to a non-equiprobable set by use of the K-means clustering method [23]. Selective sampling is used together with the three-parameter log-normal distribution described above, and the whole inflow generation process is described thoroughly in e.g. [24], [25], the latter including wind sample generation.

5.1 Residuals modelled as a three-parameter log-normal distribution

The inflow model described in [16] and in [17] is based on principles described in [26]–[28]: A PAR model intended for use in an SDDP algorithm, where the residuals from the PAR model fitting are fitted to a three-parameter log-normal distribution. The papers assume that originally, the standard noise-term of the AR-model in (1) is sampled from the residuals, and that the residuals should be normally distributed [7], [29] with 0 mean and sample standard deviation equal to σ_{κ}^{2}. However, as the residuals produced from the historical inflow are in fact not normally distributed, the inflow model does not get a log-normal distribution. The use of a Box-Cox transformation would overcome this obstacle and remove negative inflow in the forecasts. However, a Box-Cox transformation of the inflow and an inverse transformation of the forecasted inflow would produce a nonlinear inflow model, which would have to be linearized to be suited for an SDDP algorithm. Instead, the papers suggest exchanging sampling directly from the residuals with sampling from a three-parameter log-normal distribution, with parameters that are estimated to keep the inflow above zero and conserve the first and second statistical moments from the historical inflow. The two-parameter log-normal distribution does not produce negative samples, but the three-parameter log-normal distribution may produce negative samples depending on the choice of the

δ_t -parameter [26]. The principles of this implementation were also compared to other implementations in [18].

If a series of observations x are three-parameter log-normally distributed, then the data can be translated to a normally distributed variable y using the relationship:

$$y = \ln(x - \delta) \sim N(\mu_y, \sigma_y) \tag{7}$$

The distribution of x is given by

$$x = e^{\mu_y + \sigma_y \xi} + \delta \tag{8}$$

Where ξ is a standard normal variable $\xi \sim N(0,1)$. The three-parameter log-normal distribution contains three parameters that must be estimated. With these three parameters, three properties of the distribution can be decided. To maintain the first two moments of the observations, these can be used as criteria for the parameter estimations. However, other criteria could also be used. Knowing the statistical properties of the observations x, we can find the moments μ_y, σ_y using the equations for the mean value and standard deviation for a 3-parameter log-normal distribution, which are derived from the definitions of mean value and standard deviation (e.g. [30]):

$$\mu_x = e^{\mu_y + \sigma_y^2 / 2} + \delta \tag{9}$$

$$\sigma_x^2 = e^{2\mu_y + \sigma_y^2} (e^{\sigma_y^2} - 1) \tag{10}$$

Along with a third equation. The method suggested in [26] is to use the skewness as the third equation:

$$\gamma_x = \frac{\exp\left[3\sigma_y^2\right] - 3\exp\left[\sigma_y^2\right] + 2}{\left(\exp\left[\sigma_y^2\right] - 1\right)^{3/2}} \tag{11}$$

Substituting

$$\varphi = \exp\left[\sigma_y^2\right] \tag{12}$$

Into (10) and solving for μ_y gives us:

$$\mu_y = \frac{1}{2}\ln\left(\frac{\sigma_x^2}{\varphi^2 - \varphi}\right) \tag{13}$$

while

$$\sigma_y^2 = \ln\varphi \tag{14}$$

And substituting (12) and (13) into (9) and solving for δ gives us:

$$\delta = \mu_x - \sqrt{\frac{\sigma_x^2}{\varphi - 1}} \tag{15}$$

Finally, substituting (12) into (11) and solving for φ gives us:

$$\varphi = \left(\left[1 + \frac{\gamma_x^2}{2} \right] + \left[\gamma_x^2 + \frac{\gamma_x^4}{4} \right]^{1/2} \right)^{1/3} + \left(\left[1 + \frac{\gamma_x^2}{2} \right] - \left[\gamma_x^2 + \frac{\gamma_x^4}{4} \right]^{1/2} \right)^{1/3} - 1 \qquad (16)$$

[18] describes an implementation with a periodic autoregressive model (PAR) with residuals modelled by a three-parameter log-normal distribution given by (8), with parameters given by (13)-(16). However, this produces occasional negative inflows; $\delta < 0$ when

$$\left(\varphi - 1 \right)^{1/2} < \frac{\sigma_x}{\mu_x} \qquad (17)$$

5.1.1   Variation of three-parameter log-normal distribution for noise sampling: Requirement for positive inflow

The other publications using a three-parameter log-normal distribution for the residuals do not use the skewness as the third equation. In [16], [20] and in [17], the observations x are assumed having zero mean, which changes (15). The first two omit the mean value of the translated variable y as well, however, this is kept by [17]. These publications only use the first and second moment of the historical data are preserved with the parameters (13), (14) and:

$$\varphi = \frac{\sigma_x^2}{\delta^2} + 1 \qquad (18)$$

The skewness is not used as a parameter estimation criterion, but a requirement is also set here to keep the generated inflow positive. This requirement (from [17], not as plainly written in [16] and [20]) is demonstrated in our own notation, using the VAR1-model described in (5). Omitting vector notation for the inflow series, we get, for week *t* and forecasted scenario *j*:

$$z_t^j = \phi z_{t-1}^j + x_t^j \qquad (19)$$

The generated inflow described in (6) will then be:

$$\hat{q}_t^j = s_t (\phi z_{t-1}^j + x_t^j) + \overline{q}_t \qquad (20)$$

Where x_t^j is drawn from a 3-parameter log normal distribution, fitted to the residuals from the AR1-fitting, and given by:

$$x_t^j = e^{\xi_t^j \sigma_{y,t}^j + \mu_{y,t}^j} + \delta_t^j \qquad (21)$$

To keep the generated inflow non-negative, we require that (20) is greater or equal to zero and get:

$$\phi z_{t-1}^j + e^{\xi_t^j \sigma y_{,t}^j + \mu_{y,t}^j} + \delta_t^j + \frac{\overline{q}_t}{s_t} \geq 0 \qquad (22)$$

The sample standard deviation s_t and the exponential term are always non-negative. Then, the condition in (22) gives:

$$\delta_t^{\,j} \geq -\frac{\overline{q}_t}{s_t} - \phi z_{t-1}^{\,j} \tag{23}$$

to ensure a positive generated inflow. Using the boundary condition with an equality-sign in (23), (19) becomes:

$$z_t^{\,j} = \phi z_{t-1}^{\,j} + \mathrm{e}^{\xi_t^{\,j} \sigma_{y,t}^{\,j} + \mu_{y,t}^{\,j}} - \frac{\overline{q}_t}{s_t} - \phi z_{t-1}^{\,j} \tag{24}$$

$$\Rightarrow z_t^{\,j} = \mathrm{e}^{\xi_t^{\,j} \sigma_{y,t}^{\,j} + \mu_{y,t}^{\,j}} - \frac{\overline{q}_t}{s_t} \tag{25}$$

And the generated inflow is:

$$\frac{\hat{q}_t^{\,j}}{s_t} = \mathrm{e}^{\xi_t^{\,j} \sigma_{y,t}^{\,j} + \mu_{y,t}^{\,j}} - \frac{\overline{q}_t}{s_t} + \frac{\overline{q}_t}{s_t} \tag{26}$$

$$\Rightarrow \hat{q}_t^{\,j} = s_t\, \mathrm{e}^{\xi_t^{\,j} \sigma_{y,t}^{\,j} + \mu_{y,t}^{\,j}} \tag{27}$$

The paper sets the variable $\xi_t^{\,j}$ to be normally distributed with zero mean and standard deviation equal to 1. All parameters of (21) are thus given by:

$$\xi_t^{\,j} \sim N(0,1) \tag{28}$$

$$\delta_t^{\,j} = -\frac{\overline{q}_t}{s_t} - \phi z_{t-1}^{\,j} \tag{29}$$

$$\sigma_{y,t}^{\,j} = \sqrt{\ln(\varphi_t^{\,j})} \tag{30}$$

$$\mu_{y,t}^{\,j} = \ln\left(\frac{\sigma_{x,t}}{\sqrt{\varphi_t^{\,j}(\varphi_t^{\,j} - 1)}} \right) \tag{31}$$

Where:

$$\varphi_t^{\,j} = 1 + \frac{(\sigma_{x,t})^2}{(\delta_t^{\,j})^2} \tag{32}$$

There will be one ϕ for each season, and the parameters $\delta_t^{\,j}$, $\sigma_{y,t}^{\,j}$, $\mu_{y,t}^{\,j}$ and $\varphi_t^{\,j}$ are calculated for each week and for each generated inflow scenario, as $\delta_t^{\,j}$ is state dependent. In Figure 4, the sampling from the residuals is illustrated and described. For the 3-parameter log-normal distribution of noise, the noise sampling procedure is implemented in Algorithm 1. The inputs to the algorithm are the correlation matrix ϕ, the standard deviation of the residuals for each week t and each inflow series n $\sigma_{x,t,n}$, the mean inflow for each week t and inflow series n $\overline{q}_{t,n}$, the sample standard deviation of the inflow for each week t and inflow series n $s_{t,n}$ and the inflow of the previous time step $z_{t-1,n}^{\,j}$ of the generated inflow scenario j. The index y in Algorithm 1 refers to the 3-parameter lognormal distribution, see (28)-(32). The last line of the algorithm is done with matrix multiplication, thus leaving out the index n. The parameters

$\delta_{t,n}^{j}$, $\sigma_{y,t,n}^{j}$, $\mu_{y,t,n}^{j}$, $\varphi_{t,n}^{j}$ and $x_{t,n}^{j}$ are endogenous variables in the algorithm, computed for each generated inflow scenario *j* for each week *t* and for each inflow series *n*, as $\delta_{t,n}^{j}$ is dependent on $z_{t-1,n}^{j}$. As [17] points out, this state dependency implies that the model has a nonlinear relationship with the inflow of the previous time step. The paper suggests two solutions to overcome this problem; either to compute a week dependent, state independent δ_t or to ignore the interstage dependency and use a correlation matrix equal to 0. The first solution is tested in Section 5.1.2, and the second is considered insufficient by [17] as it does not reproduce the periodic characteristics of the inflow.

For the Aura river, the generated inflow using (19), (21) and (28)-(32) is shown together with the observed inflow in Figure 7. The boxplot of the modelled inflow is shown in Figure 8.

Algorithm 1: Noise sampling for 3-parameter log-normal distribution of noise

```
for j in range(nGenerated):
    for t in range(nWeeks):
```
$$\xi_t^j \sim N(0,1)$$
```
        for n in range(nSeries):
```
$$\delta_{t,n}^{j} = -\frac{\overline{q}_{t,n}}{s_{t,n}} - \phi z_{t-1,n}^{j}$$

$$\varphi_{t,n}^{j} = 1 + \frac{(\sigma_{x,t,n})^2}{(\delta_{t,n}^{j})^2}$$

$$\sigma_{y,t,n}^{j} = \sqrt{\ln(\varphi_{t,n}^{j})}$$

$$\mu_{y,t,n}^{j} = \ln\left(\frac{\sigma_{x,t,n}}{\sqrt{\varphi_{t,n}^{j}(\varphi_{t,n}^{j}-1)}}\right)$$

$$x_{t,n}^{j} = e^{\xi_{t,n}^{j}\sigma_{y,t}^{j}+\mu_{y,t,n}^{j}} + \delta_{t,n}^{j}$$

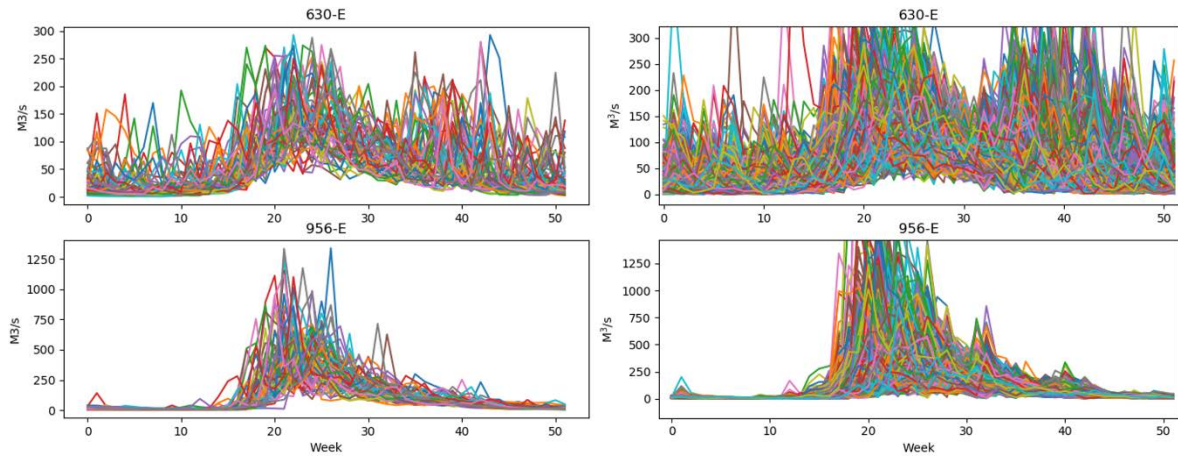$$z_t^j = \phi z_{t-1}^j + x_t^j$$

Figure 7: Observed inflow (left) and 1000 samples of generated inflow (right) for the inflow series 630-E and 956-E, using 3-parameter log-normal noise and parameters described in [17], M=1000.
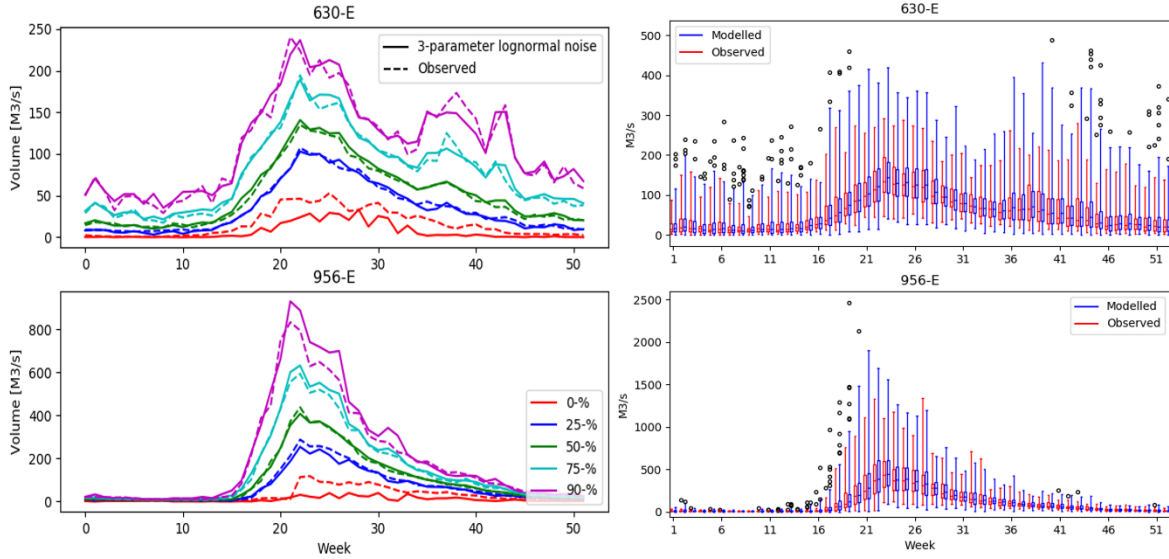
Figure 8: Percentiles (left) and boxplot (right) of the observed and generated inflow series for each week for the inflow series 630-E and 956-E, using 3-parameter log-normal noise and parameters described in [17], M=1000.

5.1.2 Variation of three-parameter log-normal distribution for noise sampling: Using a deterministic value for Delta

The described method using a 3-parameter log-normal distribution to sample from has a weakness that is pointed out by [17]: As the state z_{t-1}^j is included in δ_t^j, the model is no longer strictly linear. The suggested solution is to exchange expression (29), which includes the state, with one that does not include the state. Choosing a high enough number will guarantee positivity, but might not give a satisfactory performance, as the statistical properties of the model are affected by the choice of $\delta_{\max}$ as well. Using the choice (for each series):

$$\delta_{t,\max} = \max_i \left(-\frac{\overline{q}_t}{s_t} - \phi q_{t-1}^i \right) \tag{33}$$

will give the generated inflow and variation as shown in Figure 9 and Figure 10.
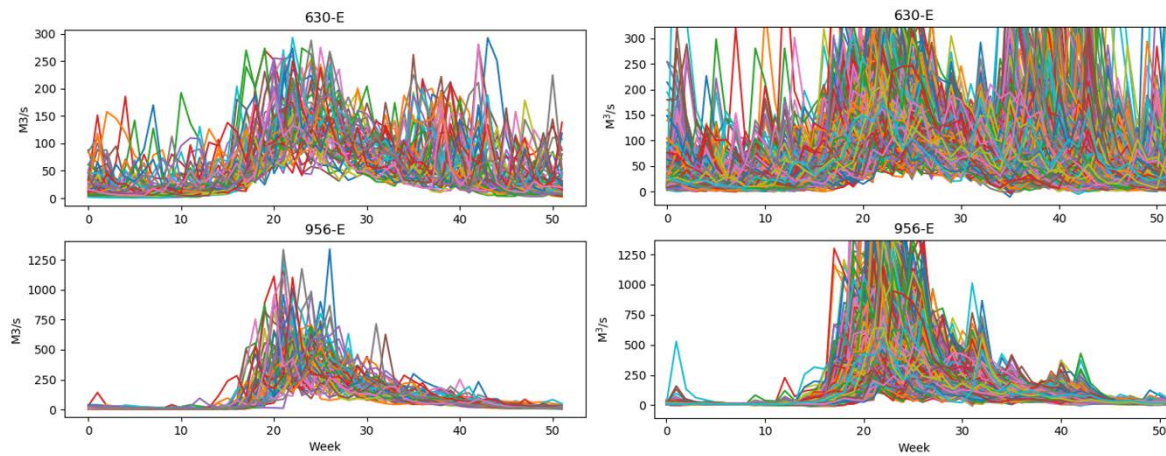

Figure 9: Observed inflow (left) and generated inflow (right) for the inflow series 630-E and 956-E, using the theory from [17] and $\delta_{t,\max}$ defined by (33), M=1000.
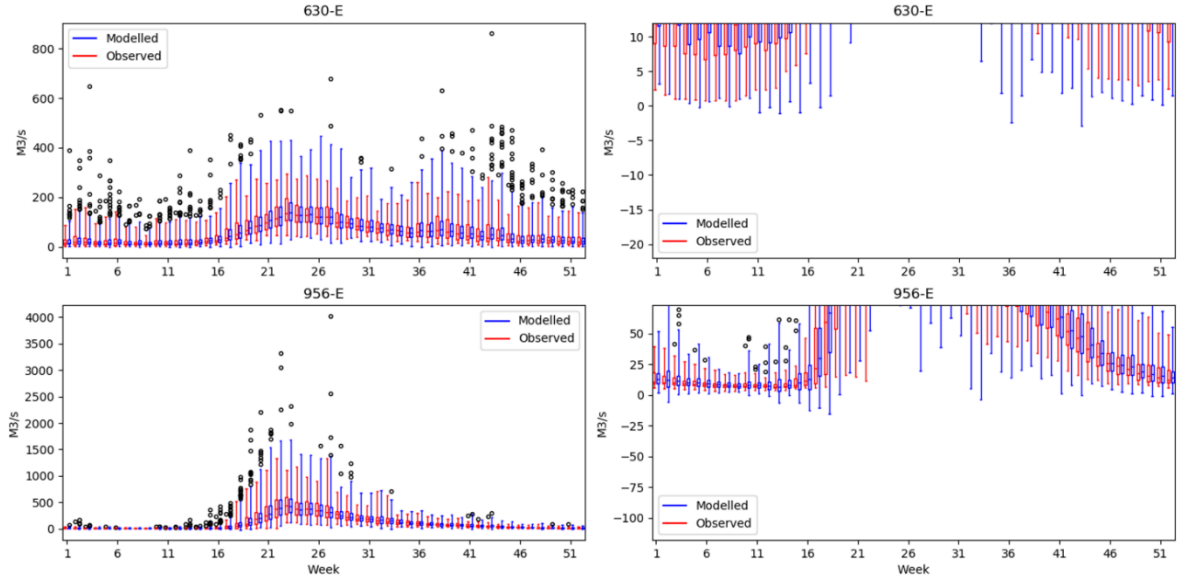
Figure 10: Boxplot (left) and enlargement of boxplot (right) of the observed and generated inflow series for each week for the inflow series 630-E and 956-E, using the theory from [17] and $\delta_{t,\max}$ defined by (33), M=1000.

This implementation gives a better variation of the generated inflow, but as it produces occasional negative inflow sample, it will not be considered for a final solution. The $\delta_{\max}$-values in (33) and δ-values in (29) are shown in Figure 11.
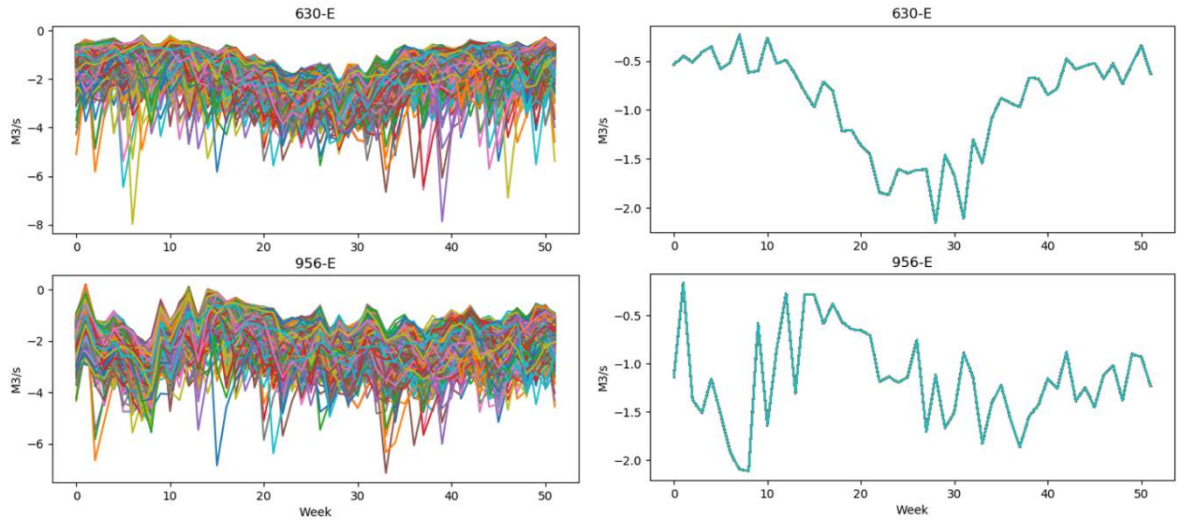

Figure 11: δ_t-values (left) and $\delta_{t,\max}$-values (right) for the inflow series 630-E and 956-E.

5.2 Spatial correlation

The parameters of (21) can be calculated using estimates of the residuals from the inflow observations, like (28)-(32). However, random numbers ξ sampled from (28) will be spatially uncorrelated. In [16] and in [20], an approach to generate multivariate, spatially correlated noise W from spatially uncorrelated random signals ξ is described. To perform this transformation, the following relationship can be used:

$$W_t = B_t y_t \tag{34}$$

Where the square matrix B has the dimensions N, the number of inflow series, and can be estimated as

$$BB^T = \hat{U} \tag{35}$$

Where $\hat{U}$ is estimated as the spatial correlation matrix of the observed residuals from the inflow series, $E\left[\hat{x}_t \hat{x}_t^T\right]$. This is backed by standard theory about simulation of sampled continuous-time random processes [6, p. 211], which suggests finding B using Cholesky factorization. However, [31] argues that this requires the matrix $\hat{U}$ to be positive definite, and that this cannot be guaranteed. Therefore, they recommend using spectral decomposition with the Jacobi method.

5.3   Selective Sampling

To generate a scenario three as the one shown in Figure 1 for the SDDP algorithm, using a VAR-model with 3-parameter log-normal noise distribution, a small set of inflow noise samples must represent a larger probability range. The method used for this purpose in literature is the Selective Sampling. The publications [16], [20], [24], [25] describe that normal spatial uncorrelated noise is randomly sampled before the matrix B is estimated according to (35) and the three-parameter log-normal distribution is fitted to the spatial correlated residuals given by (34). In [22], [32], it is explained how the very large number of normal spatial uncorrelated noise samples can be reduced using *selective sampling*, if the K-means clustering method [23] is applied to reduce the cardinality of the sample. The clustering method means that the large, equiprobable set of residuals is reduced to smaller, non-equiprobable set. This process is named the Selective Sampling and is summed up as follows[24]:

- Generate a large original sample of spatially uncorrelated residuals
- Use the K-means clustering method to reduce the cardinality, resulting in a smaller, non-equiprobable set of residuals.
- Fit a three-parameter log-normal distribution to the spatially correlated residuals
- Generate inflow using (27)

5.4   Box-Cox-transformed inflow

The second inflow model to be tested is based on the assumption in [8] that negative inflows could be avoided if the sampled inflow is logarithm transformed. This approach is mentioned in e.g. [16], [17], [19] as well, but not implemented. In [33], the Box-Cox transformation is implemented on a hydrological model and in [34] and [35], a logarithm transformation is used, although the intention of [34] is to make model identification easier in the process of shaping a periodic autoregressive model with a moving average (PARMA), using a multiplicative error model. Here, a Taylor expansion is also used to linearize the nonlinear inflows. Also [36] uses a logarithmic transformation with a linear approximation. There are other publications showing logarithm transformations (e.g. [37] also use a logarithm transformation of the observed inflow), but in other optimisation models than the SDDP algorithm. The approach transforms the observed inflow using theory from [38], before an inflow model is fitted to the transformed observations. This approach is highly nonlinear and requires a linear approximation of the inflow model to be used in the SDDP algorithm, which requires convexity. [33] describes how the convexity requirement of SDDP is met for a statistic inflow model that is fit to a time series transformed by Box-Cox transformation. The proposed approach uses a time series of observed inflow that is transformed by Box-Cox transformation and thus non-convex. Applying a proposed linear approximation, this transformed time series may still be used for scenario generation in SDDP.

The Box-Cox transformation [38] is defined as follows:

$$\tilde{q}_t = \begin{cases} \dfrac{q_t^{\lambda} - 1}{\lambda} & \lambda \neq 0 \\ \ln(q_t) & \lambda = 0 \end{cases} \tag{36}$$

The optimal choice of λ is the value that optimizes the log-likelihood (or likelihood) function for the generated series (using maximum likelihood estimation). Then, the VAR1 model described in (5) can be fitted to the transformed inflow. The Box-Cox-transformation is intended to correct a lack of normality in the distribution of samples. Therefore, the transformation does not guarantee positivity, although the special case for $\lambda = 0$ guarantees positivity in the generated inflow. Therefore, the logarithm transformation has potential to be used as a benchmark for inflow models that seek to remove negative generated inflows. The Box-Cox transformation applies to observed samples that are greater than zero. To overcome this challenge, the implementations correct the observed inflow that is equal to zero by setting it to a small constant Δ.

5.4.1 Implementation with optimal λ

For the Aura river, the generated inflow using (5) and (36) with the optimal choice for λ is shown together with the observed inflow in Figure 12. The boxplot and percentiles of the modelled inflow are shown in Figure 13. The optimal choice of λ for the inflow series 630-E and 956-E are 0.18707257 and -0.09833567, respectively. The 100-percentile of the modelled inflow is considerably higher than the 100-percentile of the observed value (a trend that is also noticeable in the 90-percentile. For better visibility of the information in the plots, the 100-percentile is not shown in Figure 13.
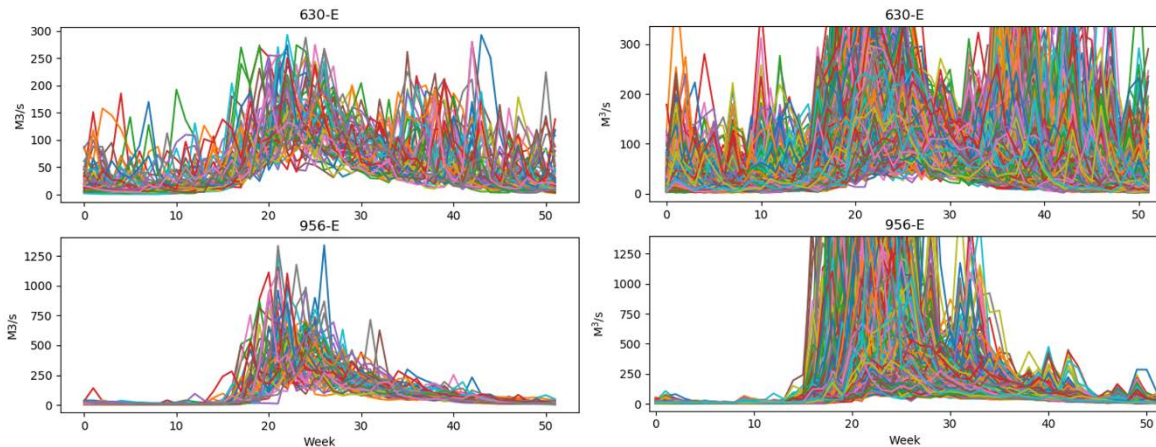

Figure 12: Observed inflow (left) and generated inflow (right) for the inflow series 630-E and 956-E, using the optimal Box-Cox transformation as defined in (36) and the VAR1 model (5), M=1000.
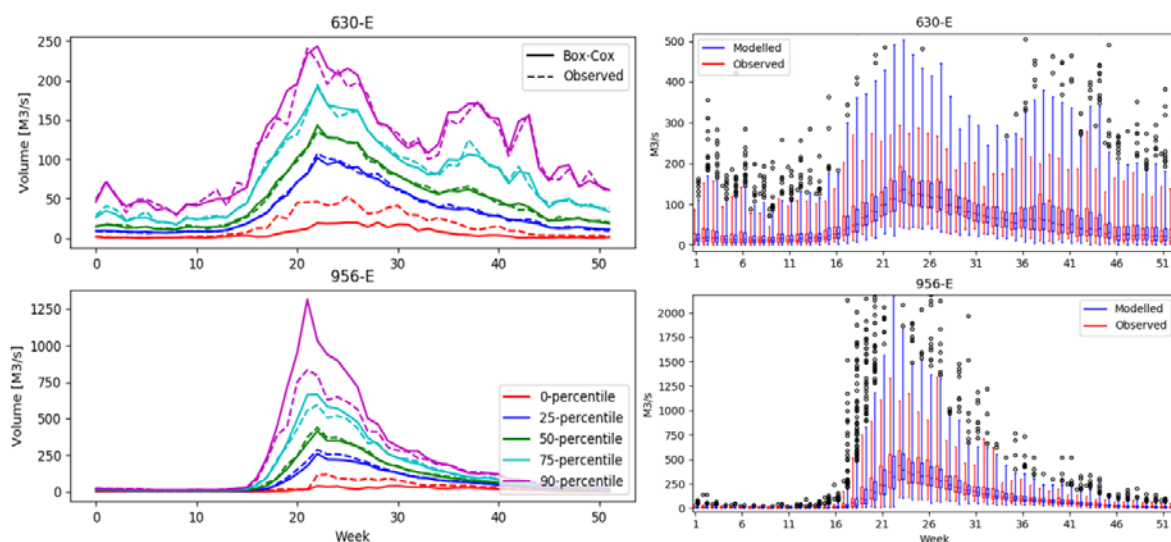
Figure 13: Percentiles (left) and boxplot (right) of the observed and generated inflow series for each week for the inflow series 630-E and 956-E, using the optimal Box-Cox transformation as defined in (36) and the VAR1 model (5), M=1000. To see the boxes clearly, the axes of the figure are manipulated, and the upper outliers are not visible.

5.4.2 Implementation with logarithmic transformation

Exploring the special case $\lambda = 0$, this has also been implemented and the results for the inflow series 630-E and 956-E are shown in Figure 14. The boxplot and percentiles of the modelled inflow are shown in Figure 15. As in Figure 13, Figure 15 contains the 90-percentile instead of the 100-percentile to preserve visibility.
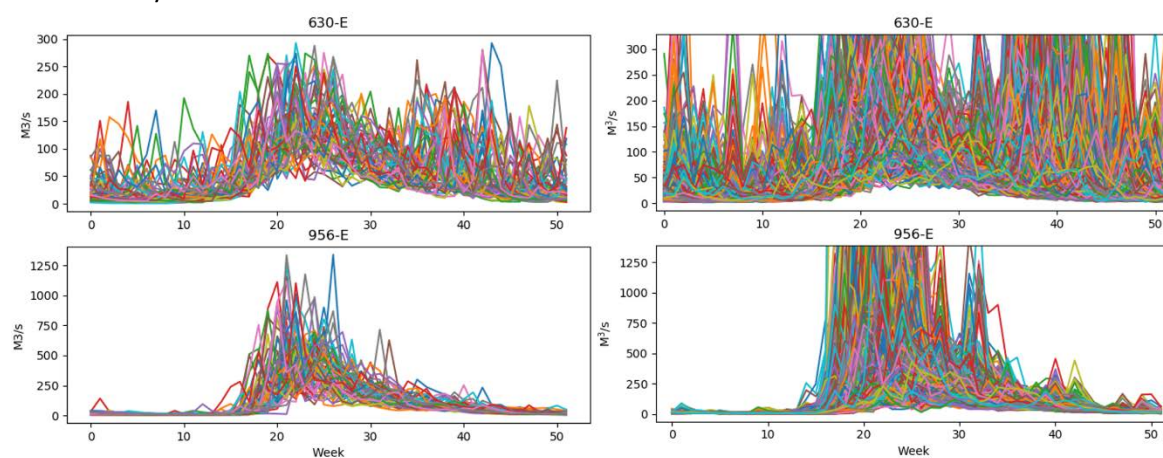

Figure 14: Observed inflow (left) and generated inflow (right) for the inflow series 630-E and 956-E, using logarithmic transformation and the VAR1 model (5), M=1000.
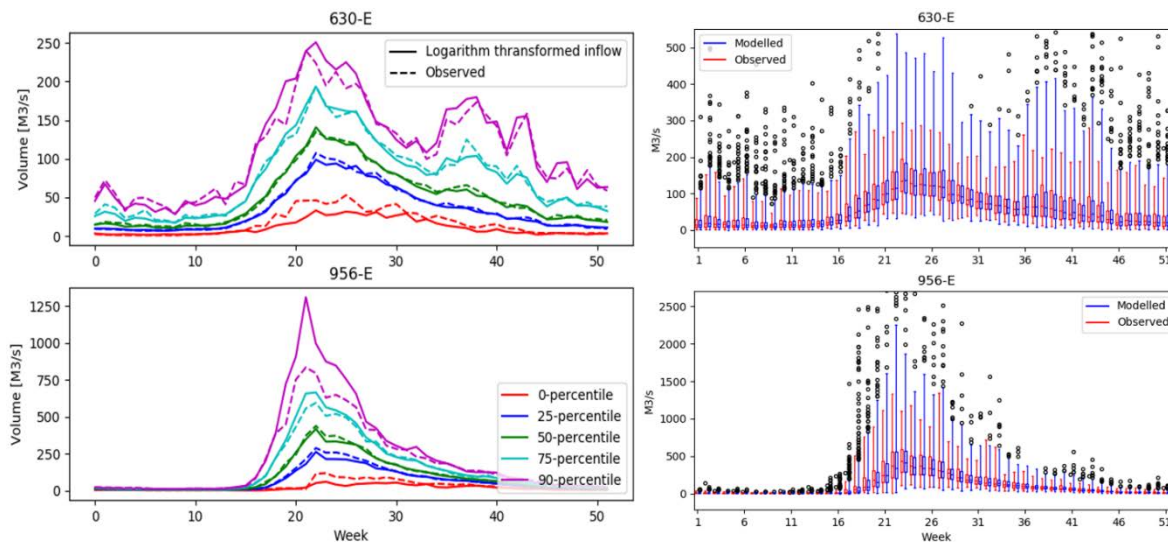
Figure 15: Percentiles (left) and boxplot (right) of the observed and generated inflow series for each week for the inflow series 630-E and 956-E, using the logarithmic transformation as defined in (36) with $\lambda = 0$ and the VAR1 model (5), M=1000. To see the boxes clearly, the axes of the figure are manipulated, and the upper outliers are not visible.

6 Literature on how the low-frequent change in the inflow can be modelled

6.1 Annual component

In [24], the PAR(p) inflow model for the Brazilian hydroelectric system is extended with an annual component (PAR-A(p)) to extend the memory of the model to better represent longer periods of very dry or wet weather. The annual component is the average inflow over the last year, standardized in the same way as the inflow of the current stage. In our notation (omitting vector indices), with weekly inflow and 52 weeks a year, for a first-order autoregressive model, this is interpreted to be:

$$z_t = \phi z_{t-1} + \psi z_{t-1}^{AV} + \varepsilon_t \tag{37}$$

Where:

$$z_{t-1}^{AV} = \frac{\hat{q}_{t-1}^{AV} - \mu_{t-1}^{AV}}{\sigma_{t-1}^{AV}} \tag{38}$$

$$\hat{q}_{t-1}^{AV} = \sum_{\tau=1}^{52} \frac{\hat{q}_{t-\tau}}{52} \tag{39}$$

$$\mu_t^{AV} = \frac{1}{S} \sum_{i=1}^{S} q_t^{AV,i} \tag{40}$$

$$\sigma_t^{AV} = \frac{1}{S-1} \sum_{i=1}^{S} \left(q_t^{AV,i} - \mu_t^{AV} \right)^2 \tag{41}$$

The correlation matrix ψ for the PAR-model described in the paper is found by extending the Yule-Walker equations that are originally used to calculate ϕ. According to [24], this model presents results with annual autocorrelation closer to the historical annual autocorrelation than the PAR(p) model but use in the SDDP algorithm is not yet found in the literature. Changes to the SDDP algorithm are necessary for the inclusion of an annual component. The annual component can be included as an endogenous state or as an exogenous state. Note that this additional component could also cover a shorter ('seasonal') period. The annual component z_t^{AV} could be realized as an annual component per inflow series, or an annual component per hydropower system. In this report, only the first approach has been tested, which means that the number of inflow variables per week is doubled; the size of z_t^{AV} equals the size of z_t.

6.1.1 Variations in the implementation of the annual component

The annual component can be included as an endogenous state in the parameter estimation by writing (37) as:

$$\begin{bmatrix} z_t^{obs} \\ z_t^{obs,AV} \end{bmatrix} = \begin{bmatrix} \phi & \psi \\ c_1 & c_2 \end{bmatrix} \begin{bmatrix} z_{t-1}^{obs} \\ z_{t-1}^{obs,AV} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ 0 \end{bmatrix} \tag{42}$$

In the inflow generation, the simulated variable z_t will be calculated as in (37), with a three-parameter log-normal distribution fitted to the residuals from the estimation of ϕ. The parameters μ_t^{AV} and σ_t^{AV} are calculated from q_t^{AV} as defined in (39), but based on observed inflow. The correlations c_1 and c_2

will not be used in the inflow generation, as the generation of z_t^{AV} is deterministic once the latest realizations of $\hat{q}_t$ are known. The problem is however that this method will include an estimation of the correlations c_1 and c_2 as well, which means that the least-squares estimation must balance the error between twice as many variables. The generated inflow using the formulation in (42) is shown in Figure 16, and the percentiles and boxplot of the modelled inflow are shown in Figure 17.
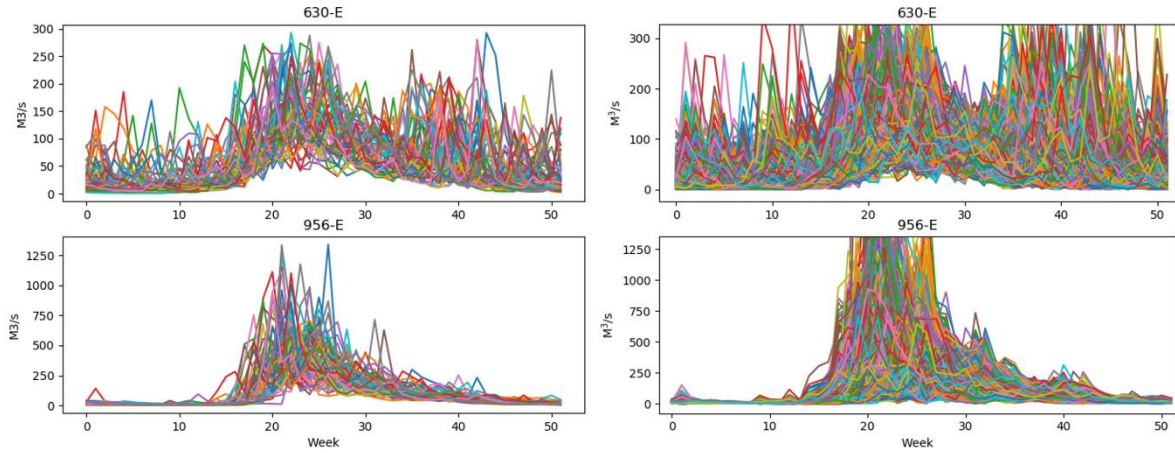

Figure 16: Observed inflow (left) and generated inflow (right) for the inflow series 630-E and 956-E, using annual component modelled as endogenous variables in the parameter estimation.
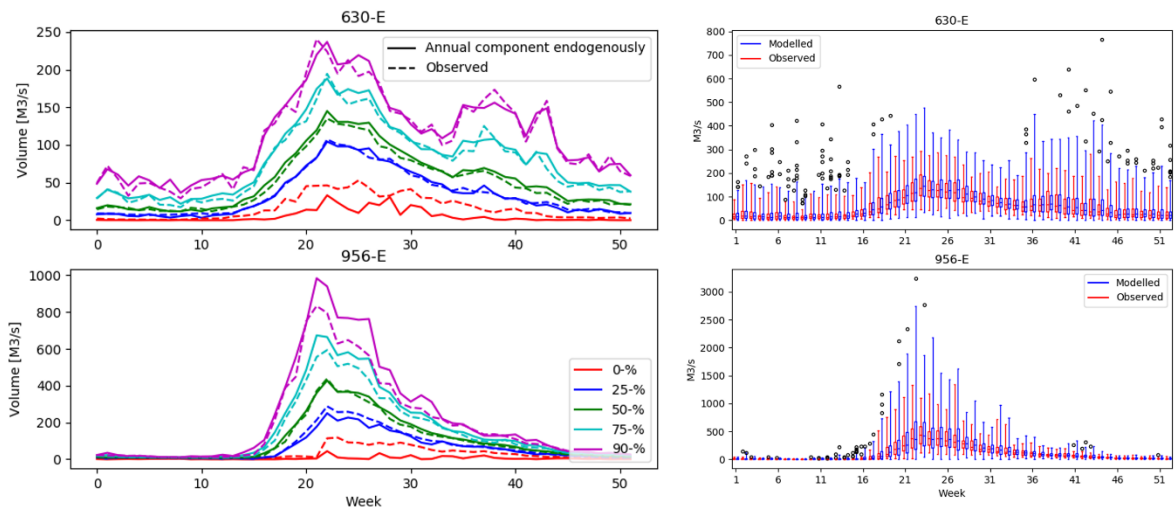

Figure 17: Percentiles (left) and boxplot (right) of the observed and generated inflow series for each week for the inflow series 630-E and 956-E, using annual component modelled as endogenous variables in the parameter estimation.

Alternatively, the annual component in (37) can be included as an exogenous state, similar to [39]. For this approach, the generated inflow is shown in Figure 18 and Figure 19. The model's ability to predict and its correctness when compared to the observed inflow is assessed in Section 8.
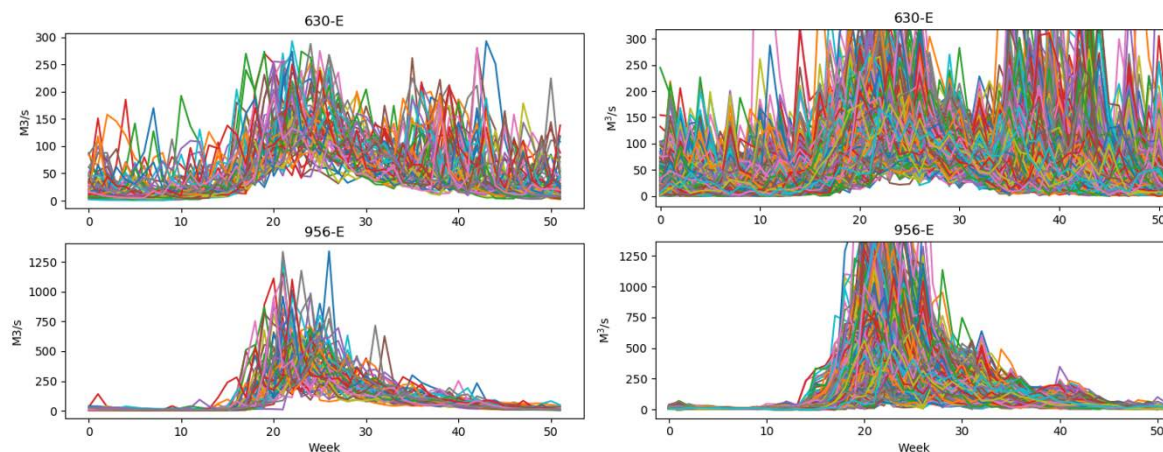
Figure 18: Observed inflow (left) and generated inflow (right) for the inflow series 630-E and 956-E, using annual component modelled as exogenous variables in the parameter estimation.
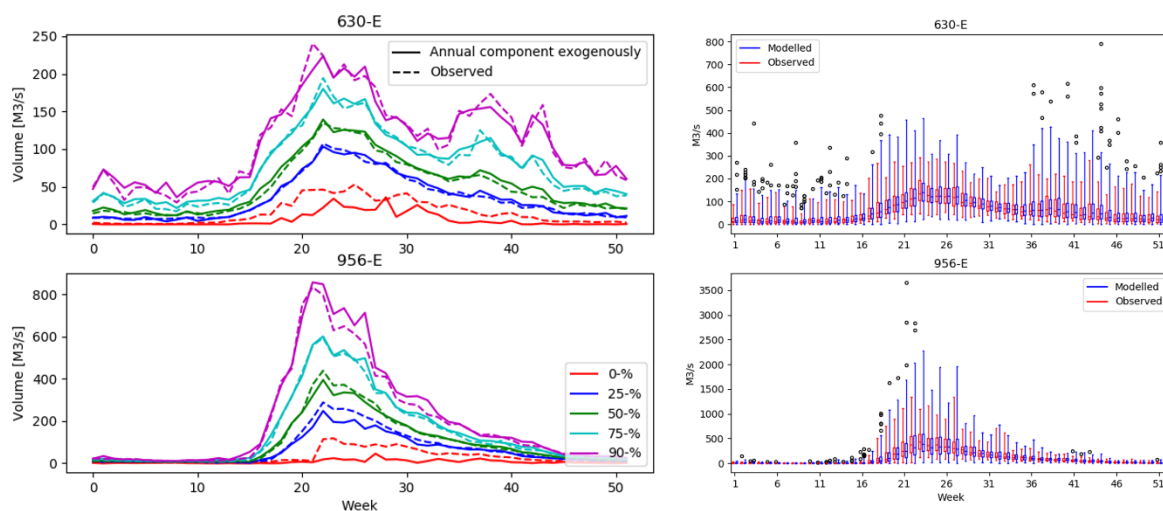

Figure 19: Percentiles (left) and boxplot (right) of the observed and generated inflow series for each week for the inflow series 630-E and 956-E, using annual component modelled as endogenous variables in the parameter estimation.

6.1.2   Correlation matrices with annual component

The correlation matrix for a VAR1-model with no annual component, as in (5), can be estimated with a maximum likelihood estimator. When an annual component is included, the inflow of each week is dependent on not only a vector of inflow from the previous week, but also on a vector of the same size containing the annual component for each inflow series. Therefore, the correlation matrix varies between the VAR1-model, the VAR1 with an annual component modelled as an endogenous variable and the VAR1 with and annual component modelled as an exogenous variable. For inflow series 630-E and 956-E, the correlation matrices are shown in Table 2.

To test the argument that the annual component leads to a more correct annual autocorrelation for the generated inflow (or, in our case, a correlation matrix closer to the historical correlation matrix), the generated inflow from three different VAR1 variations have been used to fit a VAR1 model for inflow series 630-E and 956-E. Each VAR1 variation has generated 10 000 inflow years, and this inflow is used

as input to the estimation of a VAR1 correlation matrix ϕ, see Table 3. The error of the correlation matrices for estimated by the generated inflows is calculated by subtracting the correlation matrix from the historical inflow.

Table 2: Correlation matrices for inflow series 630-E and 956-E estimated from the historical inflow, using estimations for three different VAR1 variations.

	VAR1, no annual component	VAR1 annual component endogenously modelled	VAR1 annual component exogenously modelled
ϕ	$\begin{bmatrix} 0.49591642 & 0.02272521 \\ -0.02175612 & 0.60782719 \end{bmatrix}$	$\begin{bmatrix} 0.48007626 & 0.02145006 \\ -0.01626949 & 0.608946 \end{bmatrix}$	$\begin{bmatrix} 0.47523782 & -0.01275941 \\ 0.02010256 & 0.61159397 \end{bmatrix}$
ψ	NA	$\begin{bmatrix} 0.05921725 & 0.01634054 \\ 0.0053134 & 0.06156787 \end{bmatrix}$	$\begin{bmatrix} 0.0943007 & -0.06548665 \\ 0.01460528 & -0.03741968 \end{bmatrix}$

Table 3: VAR1 Correlation matrices for the historical and generated inflows for inflow series 630-E and 956-E, using inflow generated by various variations of VAR1 inflow models and 3-parameter log-normally distributed noise with parameters described in [17], with and without annual component modelled as an endogenous and exogenous variable respectively.

Historical inflow	VAR1 with 3-parameter log-normally distributed noise		
	No annual component	Annual component	
		endogenous	exogenous
$\begin{bmatrix} 0.4959164 & 0.0227252 \\ -0.0217561 & 0.6078272 \end{bmatrix}$	$\begin{bmatrix} 0.4988439 & -0.0232128 \\ 0.0177285 & 0.6104431 \end{bmatrix}$	$\begin{bmatrix} 0.3969094 & -0.1854064 \\ 0.1517852 & 0.840558 \end{bmatrix}$	$\begin{bmatrix} 0.5226990 & 0.0279004 \\ -0.0424452 & 0.61307124 \end{bmatrix}$
Determinant of error	0.000213	0.013077	0.000248

6.2 Periodic autoregressive and moving average model1

A periodic autoregressive and moving average (PARMA) model was described already in [7], but was rejected due to nonlinearities causing difficulties in the maximum likelihood parameter estimations. In [40], this was taken up again. Here, the moving average is implemented through a contribution from the noise samples of the *r* last weeks (here, *r* is the order of the moving average). In our notation, for a first-order autoregressive model, this would be:

$$z_t = \phi z_{t-1} + \varepsilon_t + \sum_{j=1}^{r} \theta_{j,t} \varepsilon_{t-j} \tag{43}$$

Where θ_j is the constant weighing the contribution from the noise ε_{t-j}. The residuals are modelled with a three-parameter log-normal distribution. The paper warns that solving the Yule-Walker equations can be difficult, especially for a periodic moving average problem, as the equations are no longer separable by periods, and suggests using a constrained least-squares estimation to find the parameters. This method involves an adjustment of the SDDP algorithm as the noise enters as a state in the inflow model. The paper concludes that the moving average model can better capture higher autocorrelation dynamics and annual mean of inflows, but at the same time increases the complexity in the SDDP and particularly in the calibration of the hydrological process.

Based on the conclusions from [13], [14], the inflow improvement introduced by the inclusion of more lags will not justify the increased complexity in the computation time of the SDDP algorithm. Based on this conclusion, the inclusion of several noise lags will be unsuited as well. This leaves the inclusion of a moving average based in the previous noise term, which will not make any significant change for the trend of the generated inflow model.

6.3   Other approaches for inclusion of slow-shifting trends

In [41], the SDDP-algorithm is extended to incorporate slowly shifting signal trends, like prolonged droughts or high-inflow periods. This is done by the inclusion of an additional climate state variable, which can assume a limited number of discrete states and transitions from one state to another by a hidden Markov model. By including an extra loop in the backward iterations, the transition from one climate state to another is included, providing a limited amount of possible climatic next-step inflows provided the inflow in the previous week. The paper shows an increase in run time of 300%, which is reasonable when including an extra loop with 3 nodes, and a more apparent improvement of the results for extreme weather.

7 Further literature on stochastic inflow models for the SDDP algorithm

In [42], the parameter estimation for a PAR(p) model is discussed, focusing on the uncertainty of these parameters as they are based on varying observations, and trying to assess the impact this uncertainty has on the resulting operation policy: As the parameter estimates are based on only one information source, the sample errors of the historical observations can influence the generated inflow series. The work concludes that this impact is significant and advises care in parameter estimation. The periodic autoregressive model uses sample average and sample standard deviation estimations for each month, just like the ProdRisk model uses sample average and sample standard deviation estimations for each week. However, [42] also uses one transition matrix for each week, whereas the ProdRisk model uses one transition matrix for each season of the year, usually only one or two seasons. The reduced number of seasons gives an increased number of samples compared to [42] and therefore more statistically significant estimates.

In [39], an SDDP-algorithm is extended with an additional exogenous hydrological state that is added in the inflow multisite PAR-model. This extension requires a modification of the cut generation as well, including the extra state of the hyperplanes. The paper also fits a three-parameter log-normal distribution to the residuals and shows two inflow model formulation with the snow water equivalent and a combination of the snow water equivalent and the accumulated winter precipitation, respectively, as the exogenous variable. The case studies show that the use of exogenous variables for snow water equivalent or the combination of snow water equivalent and winter precipitation lead to better anticipation of snowmelt and less spill.

In [31], an adjustment to the PAR(p) method using Markov-Switching was presented, intended to improve the generation of synthetic inflow scenarios in periods when the inflow is below-average. This model improvement appears to be rather academic.

8   Comparison of considered inflow models

The stochastic inflow models that have been tested in this project are the:

- VAR1 model used as benchmark and described in Section 4
- VAR1 model with noise modelled as a three-parameter log-normal distribution described in Section 5.1
- Box-Cox transformed inflow using a VAR1 model described in Section 5.4
- introduction of an annual component to better catch the long-term trends of the inflow series, described in Section 6.1.

Using the selection criteria for the inflow model to be recommended for use in ProdRisk, the parameter variation for the three-parameter log-normal noise distribution described in Section 5.1.2 must be excluded, as it produces negative inflow. The same can be said for the Box-Cox transformed inflow with optimal choice of λ. The logarithm transformed inflow is highly nonlinear but will be included in the comparison as the ideal choice when linearity is not required.

Table 4: Inflow models to be compared. The references show where the method is described in the literature

Model	Box-Cox	Model type	Noise component	Annual component	Ref.
VAR1	No	VAR1	White noise	No	[1]
3-parameter log-normal	No	VAR1	3-parameter log-normal	No	[17]
Annual component end	No	VAR1	3-parameter log-normal	Endogenous	[31]
Annual component ex	No	VAR1	3-parameter log-normal	Exogenous	[31], [39]
Logarithmic transformed	Logarithmic	VAR1	3-parameter log-normal	no	[35], [36]

The inflow models that will be compared in this section are shown in Table 4. The properties of the models that will be compared are the accumulated probability, the annual average error and the weekly mean and standard deviation.

8.1   Accumulated probability

Assuming stationarity, this report presents inflow models using one correlation matrix for all weeks of the inflow series. One alternative is to divide the year into seasons, or, as described in [12], to reset the autocorrelation in the weeks 18-25, but only one season is chosen as basis for this report. To show how the generated inflow fits the observed values for different periods, the accumulated probability for weeks 1-17 and for weeks 15-35 have been plotted in Figure 20 and Figure 21, using 10 000 generated inflow scenarios lasting 52 weeks. For both inflow series for weeks 1-17, it is hard to distinguish between the modelled inflow and rank them according to which is closer to the observed inflow, as most models give results close to the observed values, although the logarithmic transformed model is able to sample higher inflow values better than the other methods for series 956-E. For weeks 15-35, the inflow models struggle more to replicate the observed samples, especially for series 956-E. Recalling the characteristics of this inflow series in e.g. Figure 2, this inflow series has minimal inflow during the winter season and heavy melting from weeks 14-30. However, with this non-stationary behaviour, the logarithmic transformed inflow model appears to follow the observed values less accurately than the other models. For the series 630-E, the VAR model with 3-parameter log-normal noise distribution and an annual component modelled as an endogenous variable looks like it is able to follow the observed samples, as is the logarithm transformed inflow. For both inflow series and sum of inflow from week 15-35, the inflow models capture the inflow samples with around 50% probability better than the inflow samples with very high or very low probability. Based on the probability plots, a division of the year into seasons could be an alternative.
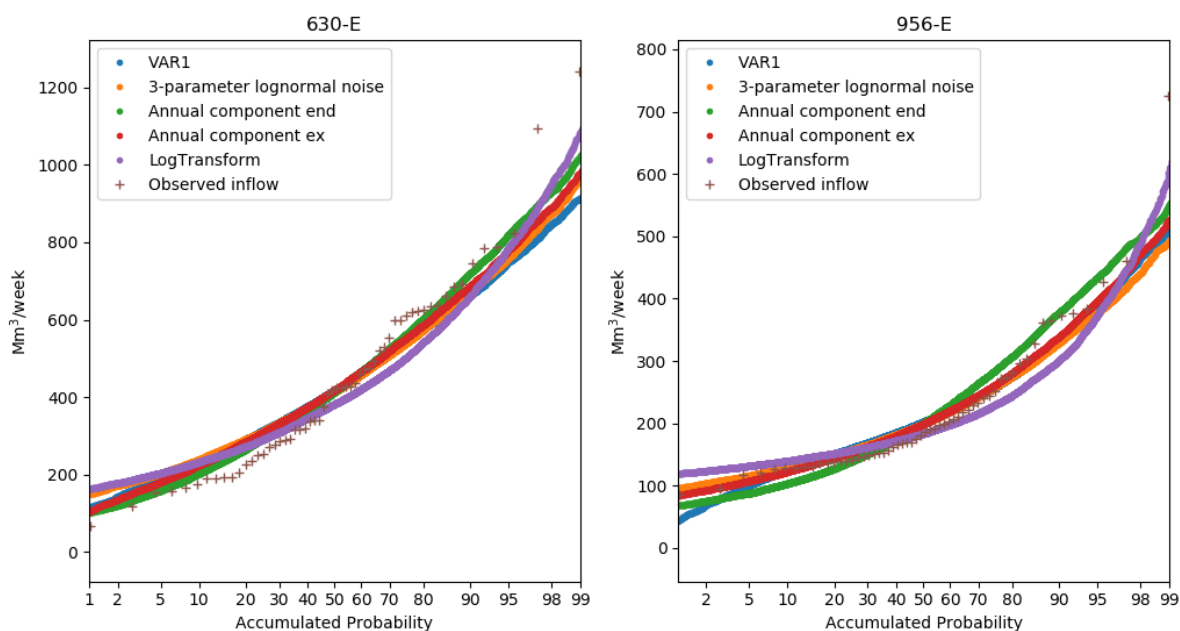
Figure 20: Sum of inflow from week 1 to week 17 for inflow series 630-E and 956-E using one season, shown as function of accumulated probability.
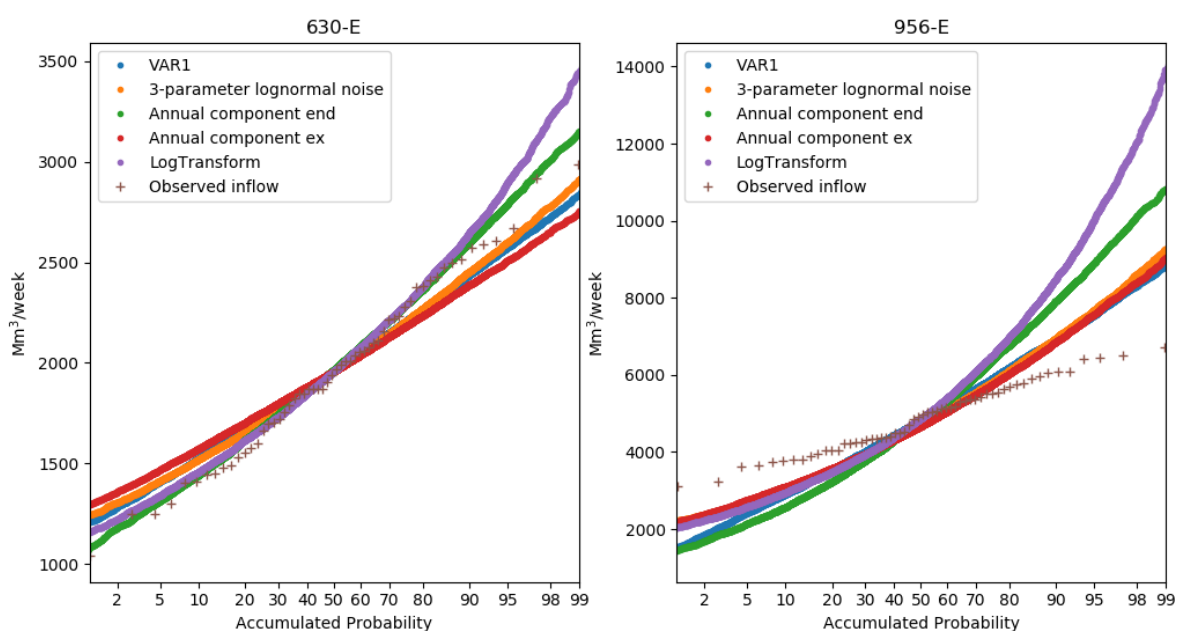

Figure 21: Sum of inflow from week 15 to week 35 for inflow series 630-E and 956-E using one season, shown as function of accumulated probability.

8.2 Annual average error

For all hydropower systems listed in Table 1, the average of the annual inflow is computed and compared with the average annual inflow of the historical observations. For the VAR1-model, the VAR1-model with 3-parameter log-normal noise and for the logarithm transformed inflow model, 100 000 inflow years are generated. The annual average is the average of all 100 000 annual sums of inflow. For the VAR1-model with 3-parameter log-normal noise and an annual component, both endogenous and exogenous modelled, only 10 000 inflow years are generated, as the computation time is higher for this alternative. The absolute value of the percentwise error between the generated annual average and the historical annual

average is shown in Figure 22. The average absolute error for all inflow series for all tested hydropower systems is shown in Figure 23.
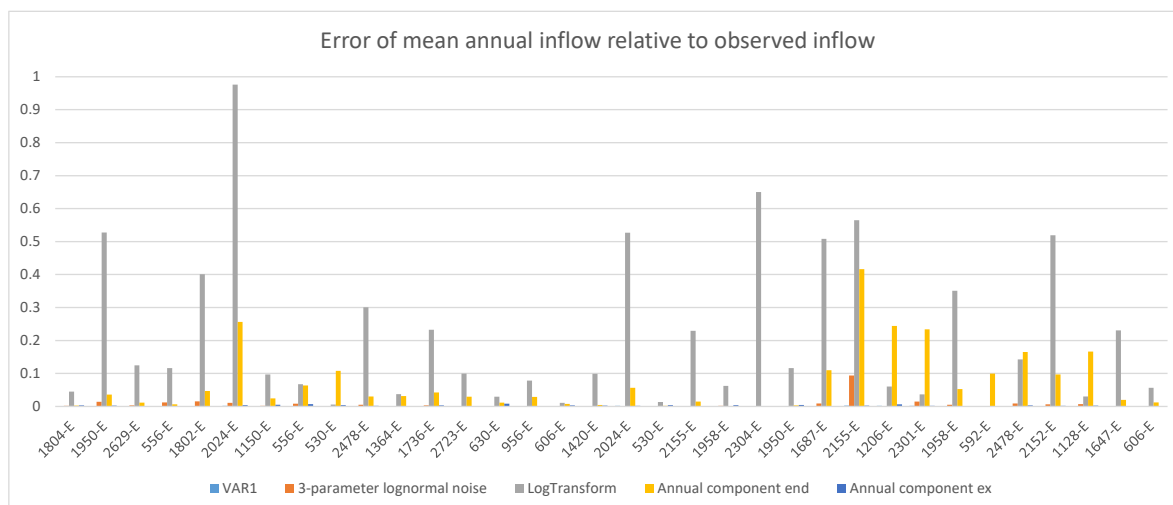

Figure 22: Absolute value of percentwise error between generated annual average and the historical annual average. From left to right, the inflow series belong to the hydropower systems Sira-Kvina, Mandal, Halling-dal, Aura, Aurland, Arendal, Lysebotn, Rjukan and Tyin.
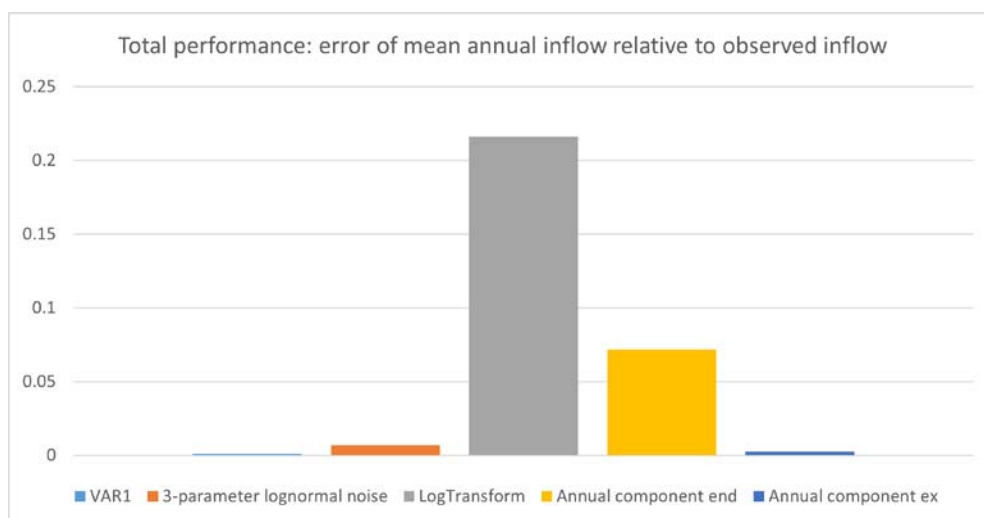

Figure 23: Average of absolute error for all inflow series in the tested hydropower systems.

The dominating error in Figure 23 belongs to the logarithm transformed inflow model. If we remove this model due to the nonlinearity and the VAR1 (which indeed has the smallest error) due to the negative inflow samples, we get the results shown in Figure 24 and Figure 25. All the average errors are well below 0.1 %, but both the VAR1-model with 3-parameter log-normal noise without an annual component and the VAR1-model with 3-parameter log-normal noise with an annual component modelled as an exoge-nous variable are well below 0.01%. Considering every inflow series in Figure 25, only the VAR1-model with 3-parameter log-normal noise with an annual component modelled as an exogenous variable has annual errors below 0.01 for every inflow series, as the VAR1-model with 3-parameter log-normal noise without any annual component has a high error for 1 inflow series (2155-E for Rjukan).
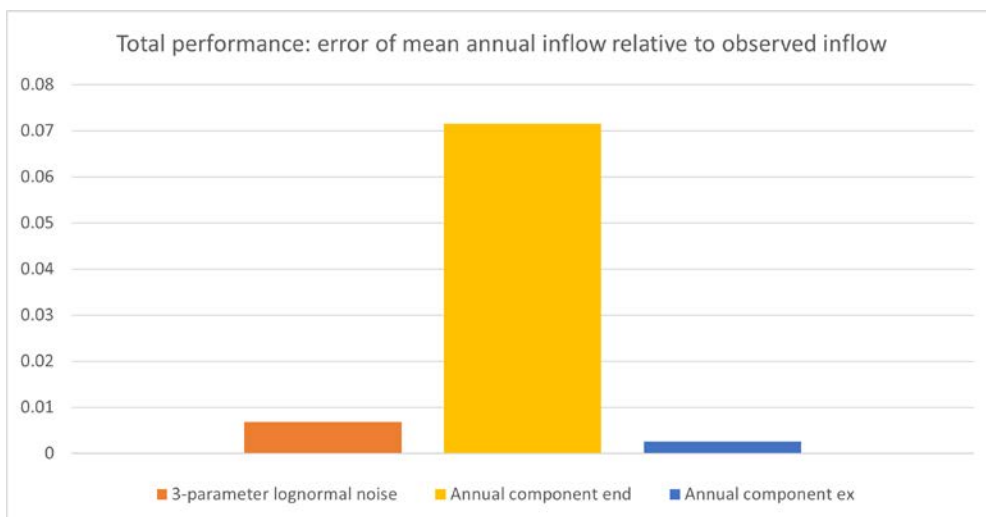
Figure 24: Average of absolute error for all inflow series in the tested hydropower systems without the logarithm transformed inflow model and the VAR1 model.
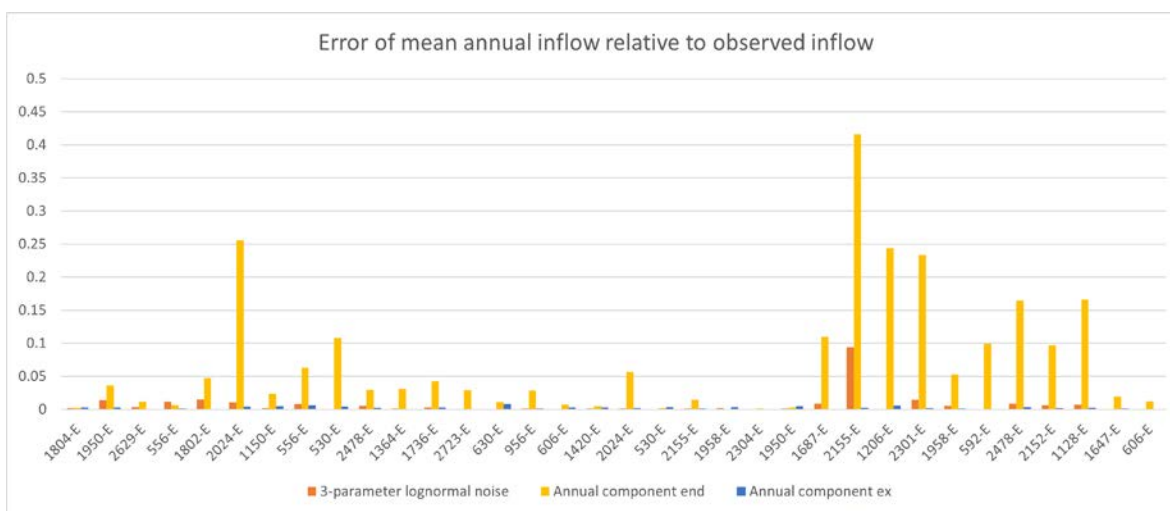

Figure 25: Absolute value of percentwise error between generated annual average and the historical annual average, without the logarithm transformed inflow model and the VAR1 model. From left to right, the inflow series belong to the hydropower systems Sira-Kvina, Mandal, Hallingdal, Aura, Aurland, Arendal, Lysebotn, Rjukan and Tyin.

9 Mean values and standard deviations

The weekly mean values and standard deviations for the inflow series 630-E and 956-E are shown in Figure 26 for all considered inflow models and in Figure 27 for only the VAR1-model with 3-parameter log-normal noise with and without annual component, with the observed value included as a reference. The logarithm transformation in Figure 26 has a visible error from the observations in both the standard deviation and the mean value. In Figure 27, only the VAR1-model with 3-parameter log-normal noise with annual component endogenously modelled shows a slight visible error for the mean value, and only for series 956-E. The standard deviations for the different models have more error when compared to the observations, and most for weeks 20-26 for series 956-E.
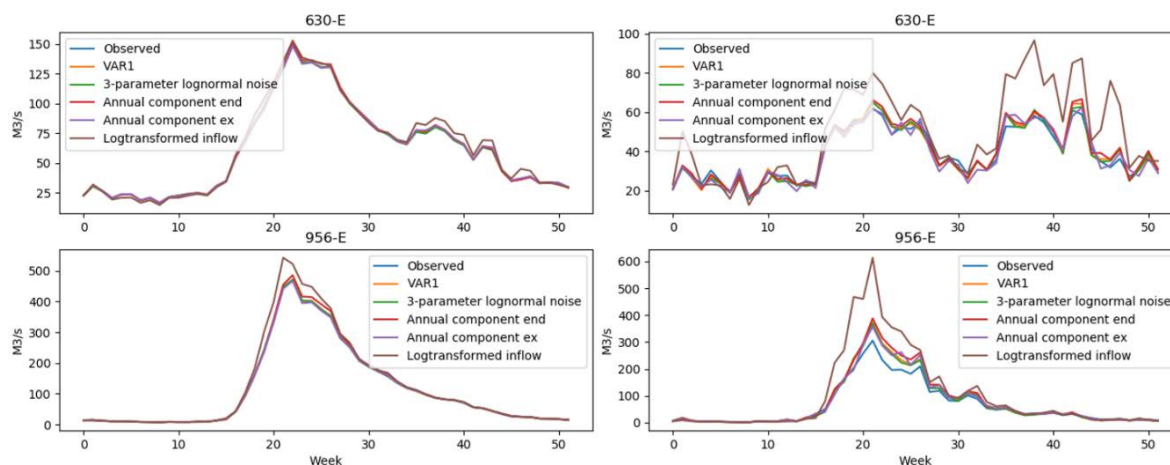

Figure 26: Weekly mean value (left) and standard deviation (right) of inflow for inflow series 630-E and 956-E, for all considered inflow models and observed inflow.
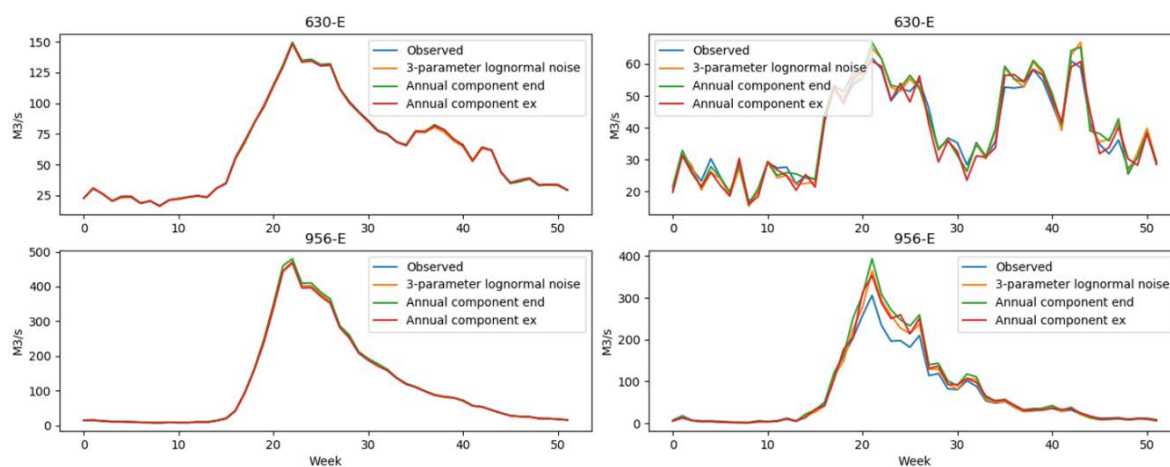

Figure 27: Weekly mean value (left) and standard deviation (right) of inflow for inflow series 630-E and 956-E, for observed inflow, VAR1-model with 3-parameter log-normal noise without annual component and with annual component endogenously and exogenously modelled

According to Figure 25, the Sira-Kvina River system and the Rjukan River system presents the higher error of mean annual inflow relative to observed inflow, and it is interesting to see how the weekly error will be. The weekly mean values and standard deviations for the inflow series from the Sira-Kvina River system are shown in Figure 28 for the VAR1-model with 3-parameter log-normal noise with and without annual component, with the observed value included as a reference. The weekly mean values and standard deviations for the inflow series from the Rjukan River system are shown in Figure 29 and Figure 30 for the VAR1-model with 3-parameter log-normal noise with and without annual component, with the observed value included as a reference. For the Sira-Kvina River system, all tested models except for the

VAR1-model with 3-parameter log-normal noise and endogenous annual component produce a weekly mean value with small error, but the VAR1-model with 3-parameter log-normal noise and exogenous annual component has a standard deviation that is clearly closed to the standard deviation of the observed inflow for most inflow series.

The same observation can be made for the Rjukan River system, except that the VAR1-model with 3-parameter log-normal noise without an annual produces a significant error on the mean value for inflow series 2155-E.
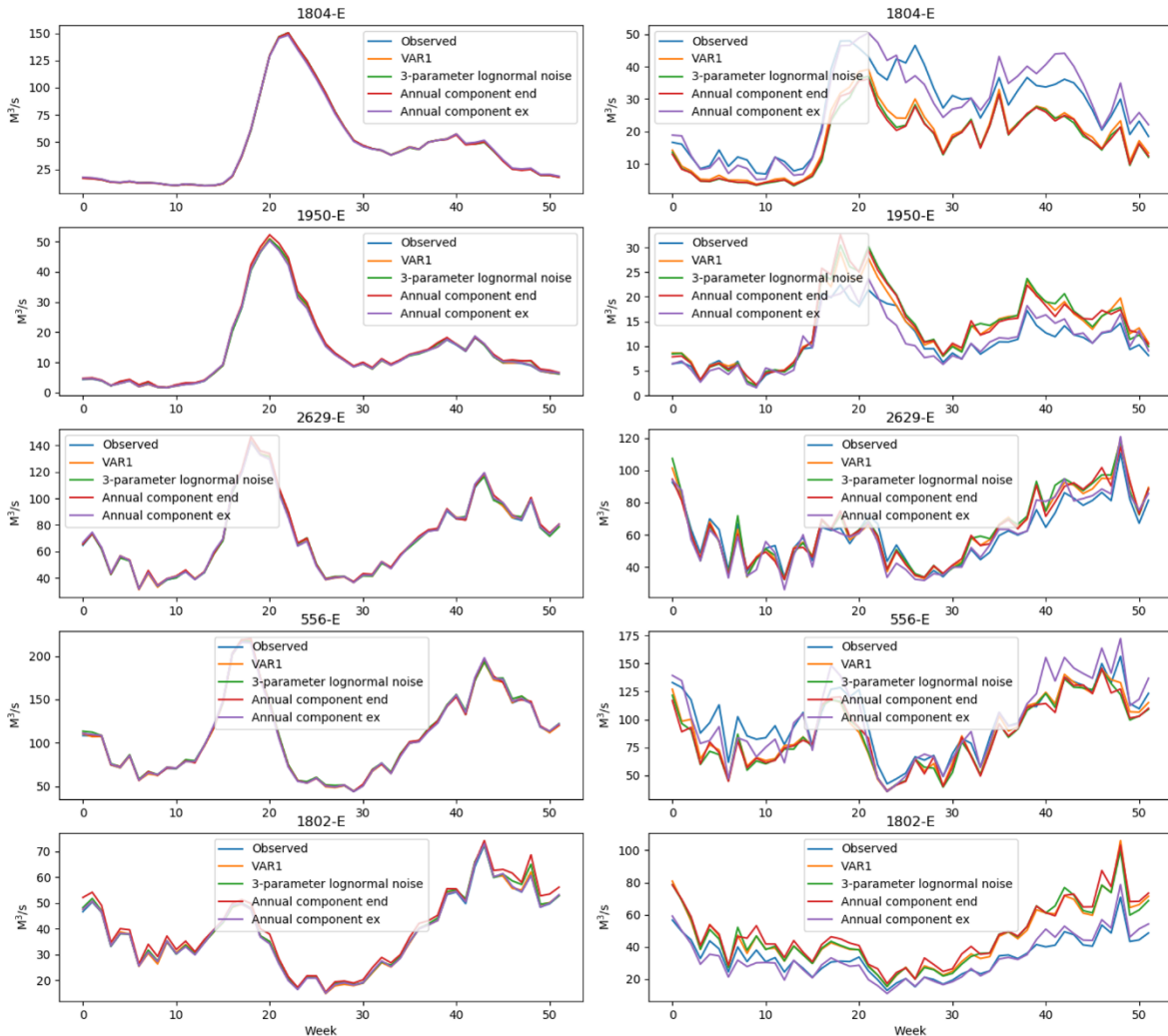

Figure 28: Weekly mean value (left) and standard deviation (right) of inflow for inflow series 630-E and 956-E, for observed inflow, VAR1-model, VAR1-model with 3-parameter log-normal noise without annual component and with annual component endogenously and exogenously modelled

Figure 29: Weekly mean value of inflow from inflow series in the Rjukan River system, for observed inflow, VAR1-model, VAR1-model with 3-parameter log-normal noise without annual component and with annual component endogenously and exogenously modelled

Figure 30: Weekly mean value of inflow from inflow series in the Rjukan River system, for observed inflow, VAR1-model, VAR1-model with 3-parameter log-normal noise without annual component and with annual component endogenously and exogenously modelled
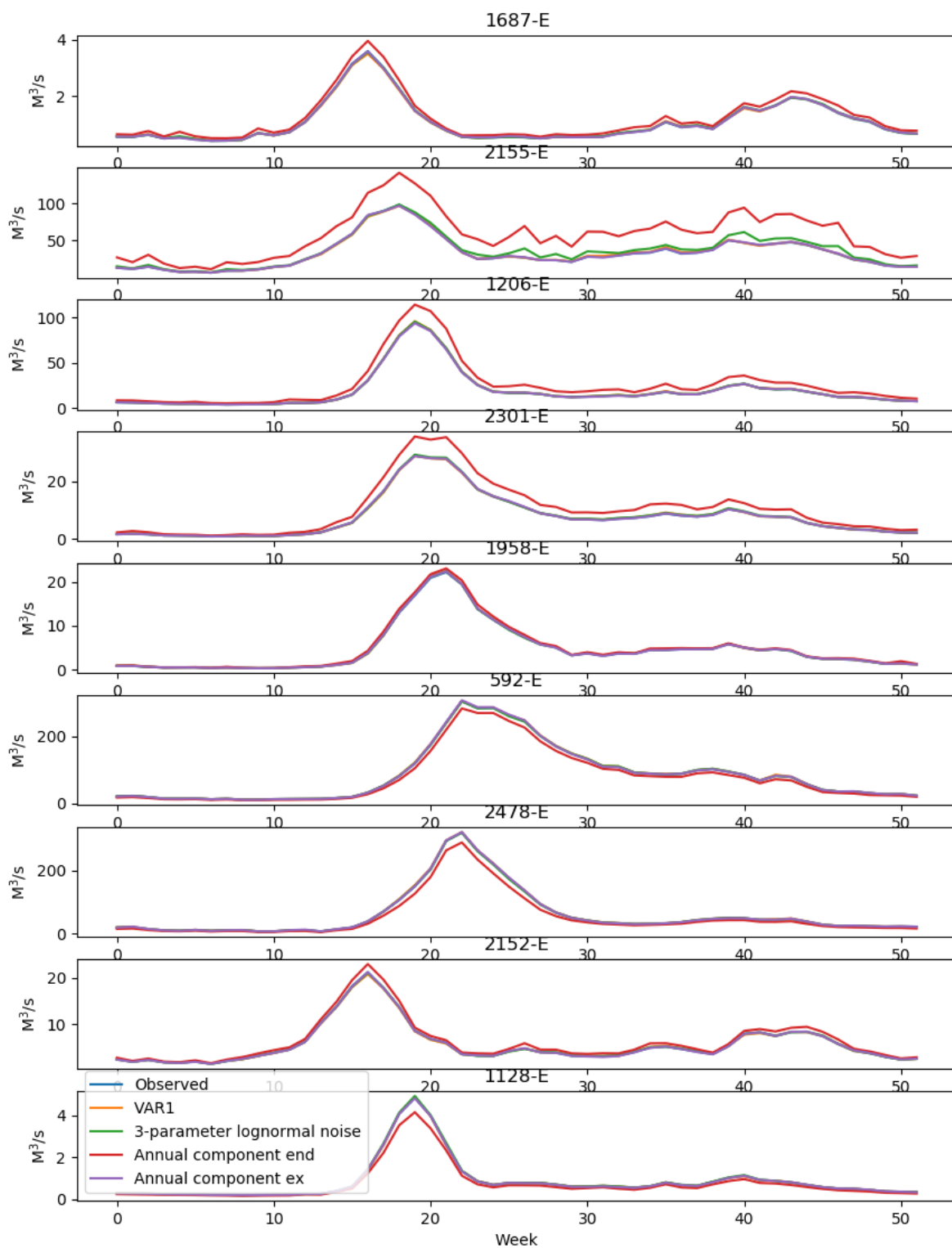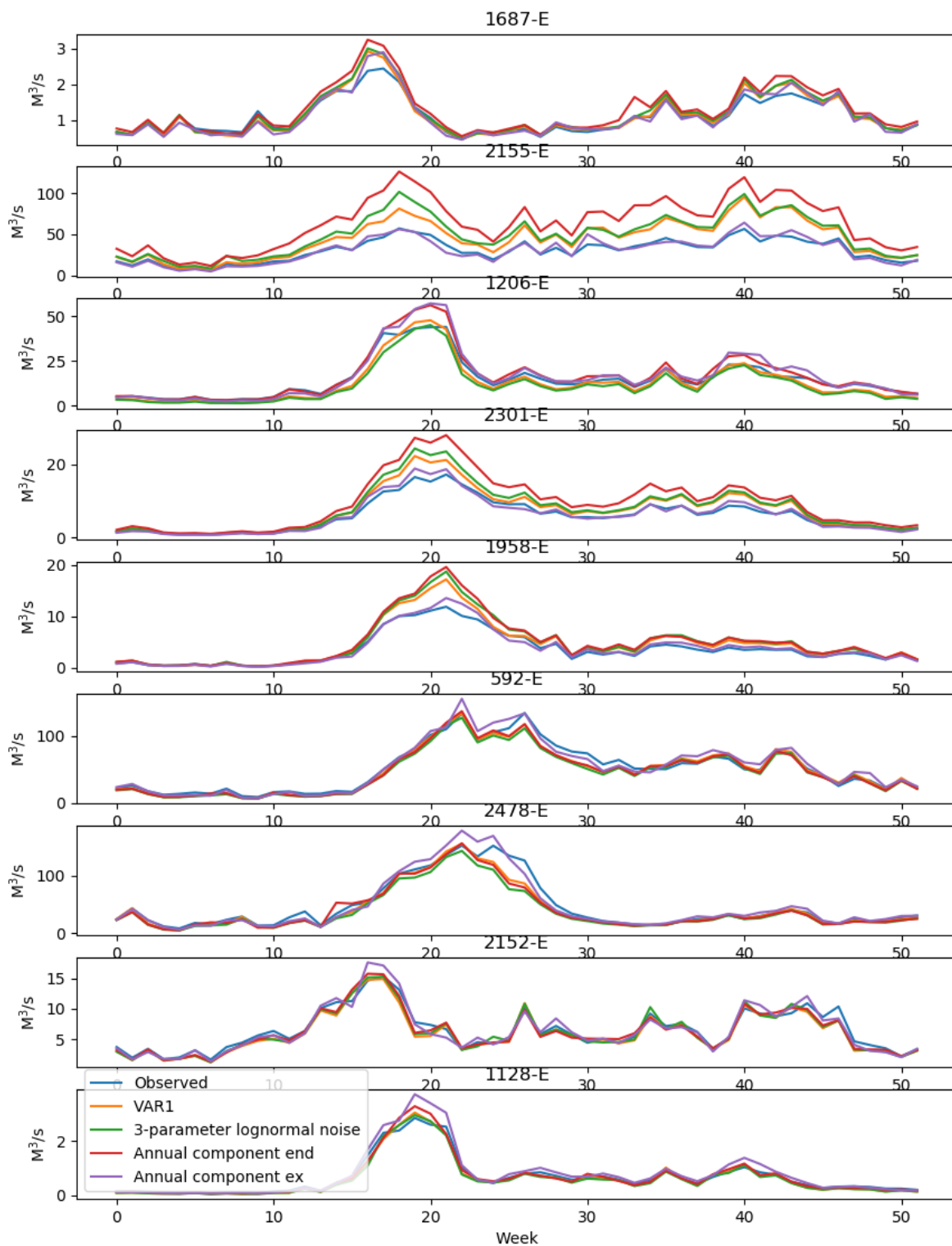
10 Recommendation of inflow model for use in ProdRisk

The inflow models from Table 4 were assessed according to the criteria mentioned in the introduction. The VAR model produces the lowest mean annual inflow error and is the method in use in ProdRisk today. However, the model generates negative inflow samples and does not show the weekly standard deviation that is closest to the observed inflow values.

The logarithm transformation is the benchmark model for how negative inflow samples in an autoregressive model can be avoided, although the 100-percentile for the logarithm transformed inflow series is too high for practical purposes, so a tuning of the model will be needed before it is taken into use. However, this method is highly nonlinear and therefore unsuited for the SDDP algorithm, and measures taken to include it includes a linearization process which will be unsuited for implementation in a complex, operative SDDP-based tool.

The results point towards using a VAR1-model with 3-parameter log-normal noise and an annual component modelled as an exogenous variable: It contains an adjustment of the noise distribution that guarantees positivity and presents an insignificant mean annual inflow error averaged on all inflow series, and for each single inflow series. It shows a weekly standard deviation that is closer to the observed inflow values than the other models for the studied inflow series, and an accumulated probability that is level with the VAR model. The model presents better results with an annual component than without, although the average annual inflow error is low without the annual component as well for most inflow series. When the annual component is modelled as an endogenous variable, it does not show as good results for the tested inflow series as when the annual component is modelled as an exogenous variable. The accumulated probability plots indicate that a division into more seasons could be beneficial.

11 Concluding remarks and further work

This report describes a literature survey and an assessment of promising stochastic inflow models intended for use in an SDDP algorithm, where the focus has been on removing negative inflow samples and preserving low-frequency signal trends. The literature survey begins with an introduction to the SDDP algorithm, reviews the current inflow model used in ProdRisk and continues with the recent work on stochastic inflow models intended for use in an SDDP algorithm in general, and autoregressive models in specific. The model that is best suited for implementation in the SDDP algorithm is a VAR1-model with a 3-parameter log-normal noise. The use of an annual component should be tested further, as it appears to capture the slow-shifting trends.

Remaining work is to implement the chosen inflow model in an SDDP algorithm and study the effects on the resulting change in profit for the hydropower scheduling problem, the effects on the convergence and the effects on the computation times. The 3-parameter log-normal noise sampling will be performed before the SDDP algorithm starts and will therefore increase the computation time of the full SDDP algorithm with a constant term. The model also contains an annual component, which includes a set of extra states to the SDDP-algorithm. The effect this has on the computation times must be studied before it can be included.

Further topics to investigate is the selective sampling, as this is not yet implemented in the test inflow models, and the use of snow data as external variables.

12 References

[1]   A. Gjelsvik, M. M. Belsnes, and A. Haugstad, 'An algorithm for stochastic medium-term hydrothermal scheduling under spot price uncertainty', in *13th Power System Computation Conference*, 1999.

[2]   M. V. F. Pereira and L. M. V. G. Pinto, 'Multi-stage stochastic optimization applied to energy planning', *Math. Program.*, vol. 52, 1991, doi: 10.1007/BF01582895.

[3]   A. Helseth, 'Environmental Constraints in Seasonal Hydropower Scheduling - Survey and Feasibility', SINTEF, HydroCen, Volume 12, 2019.

[4]   M. V. F. Pereira, 'Optimal stochastic operations scheduling of large hydroelectric systems', *Int. J. Electr. Power Energy Syst.*, vol. 11, 1989, doi: 10.1016/0142-0615(89)90025-2.

[5]   D. P. M. Gerd Infanger, 'Cut Sharing for multistage stochastic linear programs with interstage dependency', *Math. Program.*, vol. 75, pp. 241–256, 1996.

[6]   R. G. Brown and P. Y. C. Hwang, *Introduction to Random Signals and Applied Kalman Filtering, 3rd Edition*. John Wiley & Sons, 1996.

[7]   D. J. Noakes, A. I. McLeod, and K. W. Hipel, 'Forecasting Monthly Riverflow Time Series', *Int. J. Forecast.*, vol. 1, 1985.

[8]   A. Gjelsvik, B. Mo, and A. Haugstad, 'Long- and Medium-term Operations Planning and Stochastic Modelling in Hydro-dominated Power Systems Based on Stochastic Dual Dynamic Programming', in *Handbook of Power Systems*, Springer, Berlin, Heidelberg, 2010. [Online]. Available: DOI: 10.1007/978-3-642-02493-1_2

[9]   K. S. Gjerden, A. Helseth, B. Mo, and G. Warland, 'Hydrothermal scheduling in Norway using stochastic dual dynamic programming; a large-scale case study', in *IEEE PowerTech*, Eindhoven, 2015. doi: 10.1109/PTC.2015.7232278.

[10]   A. Gjelsvik, 'Stokastisk tilsigsmodell for driftsplanlegging', EFI SINTEF Gruppen, 1992.

[11]   Å. Killingtveit and S. Aam, 'Undersøkelser av varians i resultater fra kraftverdiberegninger og driftssimuleringer ved kombinerte vann-varmekraftsystemer ved bruk av en stokastisk dynamisk tilsigsmodell', Institutt for Vassbygging, NTH, Elektrisitetsforsyningens Forskningsinstitutt (EFI), 1974.

[12]   A. Gjelsvik, 'Studiar i tilsigsmodellering for SDDP', SINTEF Energiforskning AS, AN 05.12.75, 2005.

[13]   I. H. Haff, S. Martino, A. Løland, and T. Follestad, 'Statistical properties of historical inflow series for long-term models', Norsk Regnesentral, 2012.

[14]   S. Martino, I. H. Haff, A. Løland, and B. Mo, 'Statistical properties of historical inflow series for long-term models', Norsk Regnesentral, 2014.

[15]   R. P. Paulo Vitor Larroyd Felipe Beltrán, Gabriel Teixeira, Erlon Cristian Finardi, Lucas Borges Picarelli, 'Dealing with Negative Inflows in the Long-Term Hydrothermal Scheduling Problem', *Energies*, vol. 15, no. 3, p. 1115, 2022, doi: 10.3390/en15031115.

[16]   M. E. P. Maceira and C. V. Bezerra, 'STOCHASTIC STREAMFLOW MODEL FOR HYDROELECTRIC SYSTEMS', in *PMAPS 97*, Vancouver, 1997.

[17]   V. L. de Matos and E. C. Finardi, 'A computational study of a stochastic optimization model for long term hydrothermal scheduling', *Electr. Power Energy Syst.*, vol. 43, pp. 1443–1452, 2012, doi: 10.1016/j.ijepes.2012.06.021.

[18]   V. L. de Matos, P. V. Larroyd, and E. C. Finardi, 'Assessment of the Long-Term Hydrothermal Scheduling Operation Polices with Alternative Inflow Modeling', in *2014 Power Systems Computation Conference*, Aug. 2014. doi: 10.1109/PSCC.2014.7038350.

[19]   R. Charbeneau, 'Comparison of the Two- and Three-Parameter Log Normal Distributions Used in Streamflow Synthesis', *Water Resources Research*, Feb. 1978, doi: 10.1029/WR014i001p00149.

[20]   D. L. D. D. Jardim, M. E. P. Maceira, and D. M. Falcao, 'Stochastic streamflow model for hydroelectric systems using clustering techniques', in *2001 IEEE Porto Power Tech*, Sep. 2001. doi: 10.1109/PTC.2001.964916.

[21]   CEPEL, 'Manual de Referencia do Modelo NEWAVE', CEPEL, 2001.

[22] D. H. J. Penna, M. E. P. Maceira, and J. M. Damázio, 'Selective Sampling Applied to Long-Term Hydrothermal Generation Planning', in *17th Power Systems Computation Conference*, Stockholm, 2011.

[23] J. A. Hartigan and M. A. Wong, 'A K-Means Clustering Algorithm', *Journal of the Royal Statistical Society*, vol. 28, no. 1, pp. 100–108, 1979, doi: https://doi.org/10.2307/2346830.

[24] F. Treistman, M. E. P. Maceira, J. Damazio, and C. B. Cruz, 'Periodic Time Series Model with Annual Component Applied to Operation Planning of Hydrothermal Systems', in *2020 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, 2020. doi: 10.1109/PMAPS47429.2020.9183472.

[25] M. E. P. Maceira, A. C. G. Melo, J. F. M. Pessanha, C. B. Cruz, V. A. Almeida, and T. C. Justino, 'Wind Uncertainty Modeling in Long-Term Operation Planning of Hydro-Dominated Systems', in *2022 17th International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, Manchester, United Kingdom, Jun. 2022, pp. 1–6. doi: 10.1109/PMAPS53380.2022.9810576.

[26] R. J. Chaberneau, 'Comparison of the Two- and Three-Parameter Log Nrmal Distributions Used in Streamflow Synthesis', *Water Resources Research*, vol. 14, no. 1, 1978.

[27] B. P. Sangal and A. K. Biswas, 'The 3-Parameter Lognormal Distribution and Its Applications in Hydrology', *Water Resour. Res.*, vol. 6, no. 2, 1970, doi: 10.1029/WR006i002p00505.

[28] S. Burges, D. Lettenmaier, and C. Bates, 'Properties of the Three-Parameter Log Normal Probability Distribution', *Water Resour. Res.*, vol. 11, no. 2, 1975, doi: 10.1029/WR011i002p00229.

[29] V. L. de Matos, E. C. Finardi, and E. L. da Silva, 'Comparison between the Energy Equivalent Reservoir per Subsystem and per Cascade in the Long-Term Operational Planning in Brazil', in *EngOpt 2008 International Conference on Engineering Optimization*, Rio de Janeiro, Brazil, 2008.

[30] Walpole, Myers, Myers, and Ye, *Probability & Statistics for Engineers & Scientists, Seventh Edition*. Prentice Hall, 2002.

[31] F. Treistman, M. E. P. Maceira, D. D. J. Penna, J. M. Damázio, and O. C. R. Filho, 'Synthetic scenario generation of monthly streamflows conditioned to the El Niño–Southern Oscillation: application to operation planning of hydrothermal systems', *Stoch. Environ. Res. Risk Assess.*, p. 23, 2020.

[32] M. E. P. Maceiral *et al.*, 'Twenty Years of Application of Stochastic Dual Dynamic Programming in Official and Agent Studies in Brazil – Main Features and Improvements on the NEWAVE Model', in *2018 Power Systems Computation Conference (PSCC)*, Dublin, 2018. doi: 10.23919/PSCC.2018.8442754.

[33] H. Poorsepahy-Samian, V. Espanmanesh, and B. Zahraie, 'Improved Inflow Modeling in Stochastic Dual Dynamic Programming', *J. Water Resour. Plan. Manag.*, vol. 142, 2016, doi: 10.1061/(ASCE)WR.1943-5452.0000713.

[34] L. Raso, P.-O. Malaterre, and J.-C. Bader, 'Effective Streamflow Process Modeling for Optimal Reservoir Operation Using Stochastic Dual Dynamic Programming', *J. Water Resour. Plan. Manag.*, vol. 143, no. 4, 2017, doi: 10.1061/(ASCE)WR.1943-5452.0000746.

[35] A. Shapiro, W. Tekaya, J. P. da Costa, and M. P. Soares, 'Risk neutral and risk averse Stochastic Dual Dynamic Programming method', *Eur. J. Oper. Res.*, vol. 224, no. 2, pp. 375–391, 2013.

[36] G. Pritchard, 'Stochastic inflow modeling for hydropower scheduling problems', *Eur. J. Oper. Res.*, vol. 246, no. 2, pp. 496–504, Oct. 2015, doi: 10.1016/j.ejor.2015.05.022.

[37] C. Gauvin, E. Delage, and M. Genrdeay, 'A successive linear programming algorithm with non-linear time series for the reservoir management problem', *Comput. Manag. Sci.*, vol. 15, pp. 55–86, 2018, doi: 10.1007/s10287-017-0295-4.

[38] G. E. P. Box and D. Cox, 'An Analysis of Transformations (with Discussion)', *J. R. Stat. Soc.*, vol. 26, no. 2, pp. 211–252, 1964.

[39] J. Pina, A. Tilmant, and P. Côté, 'Optimizing Multireservoir System Operating Policies Using Exogenous Hydrologic Variables: SDDPX EXOGENOUS VARIABLES', *Water Resour. Res.*, vol. 53, no. 11, pp. 9845–9859, Nov. 2017, doi: 10.1002/2017WR021701.

[40] Y. Mbeutcha, M. Gendreau, and G. Emiel, 'Benefit of PARMA Modeling for Long-Term Hydroelectric Scheduling Using Stochastic Dual Dynamic Programming', *J. Water Resour. Plan. Manag.*, vol. 147, no. 3, 2021, doi: 10.1061/(ASCE)WR.1943-5452.0001333.

[41] V. Espanmanesh and A. Tilmant, 'Optimizing the Management of Multireservoir Systems Under Shifting Flow Regimes', *Water Resour. Res.*, vol. 58, no. 6, Jun. 2022, doi: 10.1029/2021WR030582.

[42] B. Bezerra, A. Veiga, L. A. Barroso, and M. V. F. Pereira, 'Assessment of parameter uncertainty in autoregressive streamflow models for stochastic long-term hydrothermal scheduling', in *Power and Energy Society General Meeting, 2012 IEEE*, Jul. 2012. doi: 10.1109/PESGM.2012.6345322.

www.hydrocen.no

@FMEHydroCen

HydroCen

HydroCen
v/ Vannkraftlaboratoriet, NTNU
Alfred Getz vei 4,
Gløshaugen, Trondheim

www.hydrocen.no
HydroCen
@FMEHydroCen