Master's thesis

NTNU
Norwegian University of Science and Technology

Magnus Myrmo Osberg

LSTM Hybrid Model for Water Reservoir Inflow Forecasting, a Comparison Between Black Box- and Interpretable Hybrid Models

Master's thesis in Computer Science
Supervisor: Odd Erik Gundersen
June 2019

NTNU
Norwegian University of
Science and Technology

Magnus Myrmo Osberg

LSTM Hybrid Model for Water Reservoir Inflow Forecasting, a Comparison Between Black Box- and Interpretable Hybrid Models

master project, spring 2020

Artificial Intelligence Group
Department of Computer and Information Science
Faculty of Information Technology, Mathematics and Electrical Engineering

Abstract

Norway's inland waters powered over 31GW installed hydropower capacity, and produced over 144 TWh of clean power in 2016 according to the International Hydropower Association. This thesis investigates 9 hybrid models in order to improve the predictions of the much used HBV model for inflow forecasting. The hybrid models are combinations of two different HBV implementations, one closed source commercial product and one open source model, in conjunction with a deep LSTM network to be able to provide better forecasts for a 40 hour time window into the future, in order to improve power producing companies ability to predict how much power they should produce hour by hour in the Northern European power market's day-ahead market and also save environmental and societal cost.

This study is conducted with a basis in a systematic literature review analysing the state of the art within hydrological inflow modelling using hybrid models with neural networks, with emphasis on investigating which neural network architectures are in focus, which multiclassifier methods are in focus and which tools are necessary to conduct good experiments in the field. The review is followed by experiments with 9 hybrid models and 2 pure LSTM networks on the water power plant inflow field, Søa, with a discussion and evaluation of the effect of using closed source vs open source HBV models with access to a limited or full set of state variables, the effect of having access to previous inflow as input, the effect of using two HBV model's predictions, and the effect of using no physical model at all. The results present multiple good models and the difference between their predictions. The problem is defined as predicting water reservoir inflow utilizing historical data of precipitation and temperature together with previous reservoir inflow. The thesis ends by outlining suggested future work to be done in the field.

Sammendrag

Norges innsjøer oppnådde over 31GW innstallert vannkraftskapasitet, og produserte over 144 TWh ren kraft i 2016 i følge International Hydropower Association. Denne avhandlignen vurderer 9 hybridmodeller for prediksjon av tilsig til vannreservoirer for å forbedre prediksjonene til den mye brukte HBV-modellen for tilsigsprediksjon. Hybridmodellene er kombinasjoner av to ulike HBV implementasjoner, en med lukket kildekode som er et kommersielt produkt som selges og en med åpen kildekode, sammen med et dypt LSTM nettverk for å kunne tilby bedre prediksjoner for en 40 timers tidshorisont fremover, for å kunne bedre vannkraftprodusenters evne til å predikere hvor mye kraft de burde produsere time for time i det Nordiske kraftmarkedets day-ahead market og også spare miljømessige og samfunnsmessige kostnader.

Studiet er gjennomført med grunnlag i et systematisk litteratursøk som analyserer de fremste teknikkene innen hydrologisk tilsigsmodellering ved hjelp av hybridmodeller med nevrale nettverk, med fokus på å undersøke hvilke nevrale nettverksarkitekturer som er i fokus, hvilke multi-klassifikator-metoder som er i fokus og hvilke støtteverktøy som er nødvendige for å gjennomføre gode eksperimenter i feltet. Litteratursøket følges opp med eksperimenter med 9 hybridmodeller og 2 rene LSTM-nett vannkraftverket på Søa's tilsigsfelt, med diskusjon og evaluaering av effekten av å bruke en HBV modell med lukket- eller åpen kildekode med tilgang til en begrenset eller komplett del av modellens interne tilstand, av effekten av å ha tilgang til tidlige tilsigsdata som input, av effekten av å bruke to HBV modeller sine prediksjoner, og effekten av å ikke bruke noen fysik modell sammen med LSTM-nettet no HBV model at all. Resultatene presenterer flere gode modeller og forskjellene på deres prediksjoner. Problemet er definert som prediksjon av tilsig ved hjelp av historisk data om nedbør, temperatur sammen med tidligere reservoir-tilsig. Avhandlingen avlsuttes ved å beskrive anbefalt fremtidig arbeid innen feltet.

Preface

This work compose a master's thesis in Computer Science at the Norwegian University of Science and Technology (NTNU). The structured literature review was conducted during a specialization project during the fall of 2019, where parts of the content and knowledge behind chapter 2 and 3 were accumulated. During the spring of 2020, the duration of the Master's Project the LSTM and hybrid models with surrounding testing framework were designed, implemented experiments regarding the models were conducted.

This thesis was conducted with the supervision of professor Odd Erik Gundersen at the institute of Computer Science (IDI) and with the cooperation of TrønderEnergi Kraft AS. I would like to address my substantial gratitude for the guidance from Odd Erik throughout the project. By being a great and knowledgeable mentor, he has allowed me to work independently and steered me in the right direction towards the project goal throughout the semester. I would also like to thank Frode Vassenden, hydrologist at TrønderEnergi Kraft AS, who has been great help for understanding the hydrology of an inflow field and the challenges of predicting future inflow in fields connected to water power plants. Finally i would like to give thanks to my partner, friends and family for feedback during the work on this thesis.

Magnus Myrmo Osberg
Trondheim, July 29, 2020

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis implements and tests different hybrid machine learning models for predicting reservoir inflow for a water power plant in Norway and compares their loss/NSE-efficiency in order to create a model with better performance than commercially available tools today that are solely based on numerical models.

1.1 Introduction and Problem Statement

Norway's inland waters powered over 31GW installed hydropower capacity, and produced over 144 TWh of clean power in 2016 according to the International Hydropower Association (House et al., 2017). One of the mail challenges faced by hydro power producers is the scheduling task (Seim and Thorsnes, 2007), of which one of the mail factors that contribute uncertainty is water reservoir inflow.

Predicting water's movement through soil is a complex problem. Rainfall from a long backwards timespan affects the water content in an inflow field's surrounding soil, which affects both inflow amount and latency in accordance to rainfall. This problem is complex and difficult to model accurately, and for climates with yearly snowfall the complexity is even greater due to the need for modelling the building and melting of a snowpack.

The ability to improve accuracy of inflow prediction models is however of great interest and would lead to great societal, environmental and economic benefit for water power plant operation and surrounding society and environment. There are multiple reasons for this, and they are described in section 2.5 Motivation. The economic benefits have their foundation in the fact that deciding how much water to use in order to generate power at any given point in time is a complex problem where one need to consider the amount of water left in the reservoir, the free capacity before water overflows, the power consumption at the given time in the consuming region and the price per Megawatt of generated power for the future 36 hours(due to the Nordic power market system described in section 2.1), and the predicted inflow into the power plant's water reservoir.

Having more accurate inflow predictions simplifies this problem and minimizes chances of reservoir overflow and reservoir depletion due to inaccurate predictions, which will both help improve income optimization per liters of inflow for power plant operations.

This thesis implements and tests different hybrid machine learning models for predicting reservoir inflow for a water power plant in Norway and compares their loss/NSE-efficiency in order to create a model with better performance than commercially available tools today that are solely based on numerical models. Then we evaluate the effect of having access to previous inflow data as model input, and the importance of a non black-box model where the inner state of the numerical sub-model, HBV, is accessible to the machine learning architecture compared to a black-box hybrid where only the output is accessible to the machine learning architecture.

1.2 Research Motivation

The underlying motivation for this thesis is to improve water inflow forecasting. This can however be specified and concretized by categorizing it into societal, ecological, economical, technological and personal motivation in the following list, which is further elaborated in section 2.5

- Societal motivation - Being able to accurately predict inflow to water reservoirs accurately can help predict floods, especially during spring, and ensure time for a safe temporary evacuation for citizens of a flood prone residential area.

- Environmental motivation - Being able to accurately predict inflow can help power plant operators ensure that enough water is available in the future to help ecosystems around water power plants downstream to remain stable (Suen and Eheart, 2006)

- Economic motivation - Being able to accurately predict inflow enables water power plant operators to maximize revenue per water available
- Personal motivation - It has been a personal motivation to learn more about neural networks through this thesis, because the amount of data available (155304 points) is a relatively large dataset capable of tuning neural networks
- Technical motivation - Being able to more accurately predict inflow by using modern machine learning techniques in conjunction with domain specific models in order to verify if deep learning can provide value in the field of hydrology

1.3 Goals and Research Questions

This section presents the goal of this thesis and related research questions which need to be answered to meet the goal. They are the following:

Goal *To measure the effectiveness of HBV - Neural Network hybrid models for timeseries prediction of inflow in hydrology and pinpoint the importance of access to previous inflow data, access to advanced HBV models and access to inner states of HBV models in conjunction with a neural network*

Research question 1 *Will a machine learning model that has access to previous inflow data and the state of an HBV-model be able to predict inflow more accurately than the HBV-model?*

Research question 1.1 *Will a machine learning model that does not have access to the state of an HBV-model be able to predict inflow more accurately than the HBV-model?*

Research question 2 *Will a machine learning model that has access to previous inflow be able to predict inflow more accurately than the HBV-model?*

These research questions are set into the context of a climate with yearly snowfall, which makes the prediction problem more complex to solve.

1.4 Research Method

The research method in this thesis consist of 3 phases: First an analytical process consisting of a literature review of research in the field of hydrological inflow forecasting. The knowledge from the analytical phase is used to design and implement algorithmic models and supporting systems based on untouched models in the field. The final phase consist of conducting experiments and evaluating the performance of the tested models, where statistical analysis is used to compare the models performance against both each other and benchmark models.

1.5 Thesis Structure

The following parts of the thesis is structured as follows:

- Chapter 1 - Introduction, gives the aforementioned intro to the thesis and answers why this study is of importance.
- Chapter 2 - Background Theory and Motivation, describes the necessary concepts related to this thesis, such as inflow forecasting, the Power Market, LSTM networks; and presents a structured literature review in the field of inflow forecasting with neural networks, including research objectives, research questions, inclusion critera, query selection and database selection; as well as describing the motivation behind this thesis with context to the background theory.
- Chapter 3 - Architecture / Model, This section describes the model and architecture used for the experiments in this thesis by presenting figures for each model's input.
- Chapter 4 - Experiments and Results, sketches initial plans for models to build, how to evaluate their performance and describes the data that is the foundation for the experiments together with each model's performance measured by MSE
- Chapter 5 - Evaluation and Conclusion, presents an evaluation and discussion around the current results and describes the tools needed for good future work.
- Appendix - Contains visualization of the results using the other error metrics: MAE, Smooth L1 loss and NSE, and tuned hyperparameters for the models

Chapter 2

Background Theory and Motivation

2.1 Background Theory

The Identification of the relavant background was carried out in the project preceding this thesis (Osberg, 2019). This is amended with additional information adapted to the full thesis such as: Background related to the Power market, LSTM networks, additional evaluation metrics, review of all M-Competitions up until M-4 and their findings. As it is not expected that the readers of this thesis will have read the pre-project report, its contents have been re-stated here.

The section describes the theoretical background related to water inflow forecasting. The topics that build the foundations of this domain are hydrological inflow modelling, time series forecasting methods and recurrent neural networks, which are all described in this chapter.

2.1.1 Inflow forecasting

Hydrological inflow modelling is the process of modelling the inflow to a water reservoir. The process can be done by using weather data such as precipitation and temperature as input to the model and reservoir inflow as output. The precipitation and temperature data will either be measured or forecasted data from the field, or the general region the field is in. Water reservoir inflow forecasting is an important field in hydrology with the focus on understanding how water from rain, hail, sleet and snow moves in soil and ground water, and a field with great societal, environmental, economical and technical interest as described in 2.5.

2.1.2 Hydrological time series forecasting

As an initial background description of the field, a short review is conducted for a arbitrarily selected hybrid model used for hydrological inflow modelling in order to map out the architecture in focus and the experiment structure.

In the field of hydrological time series forecasting, (Wang and Lou, 2019) show how a hybrid machine learning model consisting of the statistical time series forecasting method ARIMA-modelling and an LSTM recurrent neural network can be utilized for improved forecasting of daily water level in a River Basin. Here the LSTM is used to forecast the error produced by the ARIMA model and then correct the given error. The authors of this recent paper (March 2019) describe that their experiments shows that this model can be well adapted to the hydrological time series forecast and that the version using a BP-ANN-ARIMA model has the best forecast effect, compared to experiments with EMD-ANN-ARIMA. ”BP” being short for ”back propagation” and ”EMD” for ”empirical mode decomposition”, both being adaptive techniques for time series decomposition (Chengzhao et al., 2015)

The proposed forecasting technique uses ARIMA as it's main forecasting tool, and states that ARIMA is not suited for long term forecasting, so the test set is therefore predicted using the one-step forecasting method. This method uses all data up until the Nth day as training data for predicting the value of time N, and then uses data up until inclusive N as training data to predict time N+1. When predicting the value for hte N+1'th datapoint it is critical to mention that the authors did not use the predicted values for N as training data, but rather the exact measured data. This is a key limitation to their model, as the tests the authors have performed are all predictions of 1 day into the future, however with this limitation the authors achieve a very low error for their predictions. The MSE of their experiments with 2129 days of water level data was at minuscule 0.0078. Here the 2000 first datapoints were always used as training data and the last 129 used as test data.

2.1.3 HBV - Hydrological inflow model

The HBV model is a rainfall-runoff model and includes both conceptual numerical descriptions of hydrological processes of water catchment, which is the process of collecting water.

There has been made different versions of the model and these versions has been applied to over 40 countries throughout the world and the models for the countries have been ranging from lysimeter plots (Lindström and Rodhe, 1992), a plot for a device at the size of a flower pot for measuring water percolation through soil, to the entire baltic sea drainage basin (Bergström, 1976) and (Mason et al., 1999). The HBV model being using in this thesis uses the same features as the model used by the Norwegian Water Resources and Energy Directorate (Langsholt and Beldring, 2020) which is based on the work of Bergström and the Swedish Meteorological and Hydrological Institute (Bergström, 1976).

This model is built around the concepts of water basins/buckets for a snowpack, soil water and the remaning parts of the ground which is where most water is moving in the field and contributing to water runoff and inflow. These remaining buckets can be separated into one for the upper parts of groundwater and one for lower parts of groundwater or many more buckets for separating the surface are of an inflow field by field height with lower temperatures for the higher placed basins, as is the case for the commercial HBV model used in this thesis.

HBV can be turned into a semi-distributed model by dividing the entirety of the catchment area into sub-basins, where each sub-basing again can be divided into zones according to altitude within the basin, lake area and vegetation.

The state of the art of the HBV model is based on the research of the HBV-96 model that was made after a comprehensive re-evaluation of the model during the 1990's. The objective then was to improve the potential for making use of spatially distributed data in the model to make it match the physical attributes of water catchment and thus improve the models performance.

2.1.4 Power Market

To understand the motivation behind this thesis, it is essential to understand the Norwegian power market. The system is a market based power production and turnover system and a strictly regulated power grid monopoly. The grid monopoly together with proposed regulation changes requires the separation of power grid companies from power producing companies (The Norwegian Water Resources and Energy Directorate, 2019), with different regulation and motivations as described below.

2.1.4.1 Power grid monopoly and strict regulation

A power grid is an expensive system to build and it is not societally rational to build multiple competing power grids (Energifakta Norge - Ministry of Petroleum and Energy, 2019). The power grid business is therefore a monopoly in Norway, and the monopoly is subject to strict regulation in order to avoid the exploitation of a power grid company's customers. NVE, The Norwegian Water Resources and Energy Directorate, even dictates an annual allowed revenue for each power grid company which should cover costs of operation, depreciation of material and give a reasonable return on invested capital (Energifakta Norge - Ministry of Petroleum and Energy, 2019). It is therefore an upper limit to how much revenue can be gained from a grid company.

2.1.4.2 Tariffs

Grid customers pay a so-called point tariff for transfer and distribution of power and get access to the entire power market. The tariff amount is subject to variation due to differences in topology making the building of the power grid more difficult, in differences in density of homes in the region and differences in how efficiently the grid company operates.

2.1.4.3 Integrated market & market organisation

The Norwegian power production and distribution market is integrated with the market for Sweden, Denmark and Finland, which then again is integrated with the european power production and distribution market via transfer-connections to the Netherlands, Germany, The Baltic's, Poland and Russia (Norwegian Ministry of Petroleum and Energy, 2019b).

This market is organised as an engross market and end user market, where the engross market is where large power volumes are traded and the actors in the market are power production companies, brokers, power delivery companies and large industry customers. The market consists of the following three organised sub-markets where market players place bids and prices are determined:

- Day-ahead market
- Balance market

The end user market on the other hand is where end users sign a deal with a specific power supplier. In this market the end user is roughly $\frac{1}{3}$ private housing, $\frac{1}{3}$ industry and $\frac{1}{3}$ medium-size end users, such as hotels or brand stores (Norwegian Ministry of Petroleum and Energy, 2019b).

2.1.4.4 Day-Ahead Market

The day-ahead market is the main power market in the Nordics, where most of the power volume is traded on the exchange "Nord Pool" (Norwegian Ministry of Petroleum and Energy, 2019b). This market deals in contracts of delivery of power per hour for the next full day and sales and buy offers are places between 8 am and 12pm the day before (Norwegian Ministry of Petroleum and Energy, 2019b), meaning that the contracts are placed for a timeframe of minimum 12 hours into the future and maximum 40 hours into the future. This higher and lower boundary for the time-frame is therefore used as the output window for the model predictions in the experiments that follow.

2.1.4.5 Intra-Day Market

There can occur changes that can affect the prognosis of the day ahead-market after the offers are frozen at 12pm, such as changed weather forecasts, which can affect the market players actual production and consumption. To cope with this, there exists an intra-day market where trades are done between the freeze time of the day-ahead market and until 1 hour before the operation hour (Norwegian Ministry of Petroleum and Energy, 2019b) This gives the market players the opportunity to trade into balance to adjust for the changes that have occurred.

2.1.4.6 Balance Market

To ensure the momentary balance in the power grid, there is also a balance market where Statnett, the designated transmission system operator(TSO) of the Norwegian power system (Norwegian Ministry of Petroleum and Energy, 2019b), regulates production and consumption up or down (Norwegian Ministry of Petroleum and Energy, 2019b). It is often more expensive for a power producer to purchase production quotas from Statnett if a producer cannot meet their day-ahead predictions and haven't traded accordingly in the intra-day market. It is therefore of great value to the company to be able to make accurate predictions to how much power they will produce.

2.1.4.7 Relevance for this thesis

The northern european power market and how it operates is highly relevant for the motivation of this thesis. This is due to the necessity of reporting good prognosis for how much power will be produced 12-40 hours into the future. The process of deciding how much power to produce is a complex problem of maximising the profit per inflow amount, which again means minimizing the chance for reservoir overflow due to larger inflow amounts than anticipated or reservoir dry out due to smaller inflow amount than anticipated.

In order to minimize these two unwanted scenarios, it is of great value to have accurate inflow predictions throughout the day-ahead market time window of 12 to 40 hours into the future, and accurate predictions for 1 to 12 hours into the future for trades in the intra-day market.

2.1.5 Time Series Analysis

A time series is a set of observations x_t, each one being recorded at a specified time t. A discrete-time series is one in which the set T_0 of times at which observations are made is a discrete set, as is the case for example when observation are made at fixed time intervals. On the other hand continuous-time series are obtained when observations are recorded continuously over some time interval, e.g. when $T_0 = [0, 1]$ (Brockwell et al., 1998).

The analysis of a given time series is primarily aimed at studying it's internal structure (autocorrelation, trend, seasonality), to gain a better understanding of the dynamic process by which the time series data are generated (Palit and Popovic, 2016).

2.1.6 Makridakis Competitions

The Makridakis Competitions, also called M-Competitions are forecasting competitions led by Spyros Makridakis, a researcher in the field of forecasting. The competitions aim at evaluating different forecasting methods accuracy. In 2020 there has been arranged 4 M-Competitions and M5 is being organized from march through june 2020 and a conference in desember 2020 where findings from M-5 will be presented.

The reviews of the M-Competitions provide empirical evidence and comparisons of performance of different methods for time series prediction. Therefore, the paper discussing results, conclusions and implications of the M-Competitions until and including M-3 has been reviewed, as well as the winning solution of the recent M-4 competition.

2.1.7 M-4 competition winner: ESRNN - Exponential Smoothing Recurrent Neural Network

An exciting recent advancement in time series forecasting by machine learning is the 2018 winner of the M4 competition, Makridakis (M) Competition a challenge for time series forecasting. Here where Slawek Smyl, one of Uber's self proclaimed leading data scientists won by a solid margin with his hybrid model. (Smyl, 2018) The algorithm is split into two connected layers: a pre-processing layer which aims to normalize and deseasonalize the data and does so by exponential smoothing in the algorithm itself rather than a preprocessing step; and an

LSTM layer that aims to update parameters for each series of the Holts-Winter model

2.1.7.1 Preprocessing

Slavek Smyl states that when preprocessing the data the output window size was chosen to be the same as the prediction horizon, and the input size was set partly by experimentation, but also noting that it should be at least the size of the seasonality, e.g. 12 for a monthly series. Normalization is also a key part of the preprocessing, of which Slavek normalized per input/output windows by dividing the values in the timeseries by some constant, of which he chose to divide by level. Little argumentation is given to the choice, although the level is a smoothed estimate of the value of the data at the end of each period (Kalekar, 2004)

2.1.7.2 Exponential Smoothing and the Holt Winter's model

Exponential smoothing in itself is a function from the time series forecasting field, being published in the 1950's by Brown and Holt, separately. Holts article has been republished in recent years (Holt, 2004). The formula for exponential smoothing is $\widehat{y_{t+1}} = \widehat{y_t} + a(y_t - \widehat{t_t})$, of which a is a smoothing factor in the range $[0, 1]$. As Slawek Smyl describes it, the algorithm says that the forecast of a next step is equal to the forecast of the previous step adjusted by part of the previous error. The characteristics and goal of this method has later been expanded to span other components like level, trend and seasonality. One of he most well known models describing these concepts together is the Holt-Winters model, published in 1960 by Peter R. Winters, a student of Holt, who improved upon Holt's work by adding seasonality and published here (Winters, 1960). It is also known as the triple exponential smoothing method.

It's formulas are:

$$l_t = a(y_t/s_t) + (1-a)(l_{t-1} + b_{t-1}) \tag{2.1}$$

$$b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1} \tag{2.2}$$

$$s_{t+m} = \gamma \frac{y_t}{l_t + b_t} + (1-\gamma)s_t \tag{2.3}$$

Here s are multiplicative seasonality coefficients, a, β and γ are smoothing factors in the weighted sums of respectively (3.1), (3.2) and (3.3), l is the level of the series and b is the trend of the series. (Taylor, 2003)

It is worth noting that the Holt Winters method was even extended in 2003, more than 40 years after being published, by James W Taylor adding methods for handling multiple seasonalities simultaneously, also called n-th exponential smoothing, here (Taylor, 2003).

2.1.8 The M-3 and older competitions

The paper reviewing results, conclusions and implications from the M-3 competition was written by Makridakis & Hibon in 2000 after the M-3 competition. Here they present the results of M3 with the accuracy of various methods compared to a benchmark, the four conclusions of the M-Competition, implications for the theory and practice of forecasting, and suggestions for further research.

The four conclusions were:

- Statistically sophisticated or complex methods do not necessarily produce more accurate forecasts than simpler ones
- The rankings of the performance of the various methods vary according to the accuracy measure being used
- The accuracy of the combination of various methods outperforms, on average, the specific methods being combined and does well in comparison with other methods
- The performance of the various methods depends upon the length of the forecast horizon

These conclusions are highly applicable to this thesis, where we evaluate models with different levels of complexity, from an LSTM network alone to a stacked hybrid model using data from two separate HBV models as input to the LSTM network. The depth and width of the LSTM is also varied throughout the parameter search.

The fact that rankings of performance varies with the performance measure being used is also a factor that will be used to these experiments' advantage. The choice of different performance measures enables us to choose one that is biased towards our ideal features, which in this case is to minimize large errors that occur less frequently. Since that are more important errors to minimize than smaller consistent errors, we choose a performance measure/error metric that enhances that behaviour. For our case that means choosing MSE or RMSE where larger loss values are penalized more than when using MAE.

The fact that combination methods on average outperform specific methods motivates us to create and evaluate good hybrid models in comparison to specific methods, which in these experiments are a standalone LSTM network and the predictions from both the HBV models alone.

The final conclusion about the length of the forecast horizon's effect on method performance highlights the importance of choosing an appropriate forecast horizon, which in this study was easy to pinpoint to 40 hours due to the NordPool Nordic power market's day-ahead market where energy production amounts per hour from midnight to midnight need to be submitted before 12pm the day before, such that forecasts need to be computed and predicted between 8am and 12pm each day.

2.1.9 ANN - Artificial Neural Networks

Artificial Neural Networks (ANN's) are computing systems that are inspired by the synapses and neurons in a biological brain. The computing system consists of a directed graph of nodes(neurons) and connections between them, edges, with an associated weight, whereas the nodes has a bias state. When data is propagated through the directed graph, it is passed through an activation function in each neuron and passed to neurons in the next network layer, where the value of a neuron is the sum of the inputs leading to the specific neuron, where the data is iteratively passed through another activation function, until it reaches the output layer, where we measure the output-layer's neurons values before tuning the network by backpropagation in order to improve performance.
Neural networks are a computer architecture capable of distorting the input space to make classes of data linearly separable (Lecun et al., 2015), this is done by the activation functions so that that the The chain rule of derivatives tells us how two small effects of x on y, and z on y; where x is the neural network input, y is a layers output from the activation function with z as its input, and z being the sum of weights multiplied by layer input. Small changes in Δx is transformed first into a small change Δy in y by being multiplied by $\frac{\delta y}{\delta x}$.

2.1.10 RNN - Recurrent neural networks

Recurrent neural networks is a class of neural networks where the nodes connection is structured as a directed graph. This forms a temporal graph along the network sequence and enables the network to be trained to recognise temporal dynamic behaviour. A common example of this is the research field of NLP, natural language processing. A sequence of words's semantic meaning is influenced not only by the set of words, but also by their order. Recognition of this meaning can be trained for a recurrent neural network. Another typical challenge where RNN's are applied is the field of speech recognition, also due to the semantic meaning of sequences of words being related to the order of the words.

RNN's process an input sequence one element at a time, and maintains a state vector containing all the neurons hidden units. This state vector contains information about the history of all the past sequence elements (Lecun et al., 2015).

2.1.11 LSTM - Long Short Term Memory networks

The LSTM(Long Short Term Memory) network, a specified RNN architecture, was first presented in 1997 by Hochreiter and Schmidhuber (Hochreiter and Schmidhuber, 1997) under the title "LSTM can solve hard long time lag problems" which was the prime motivation for creating a such a recurrent neural network. As they state in the article, traditional recurrent nets fail when data has a long minimum time lag between input and corresponding targets.

In order to improve upon these challenges for regular recurrent networks, functionality has been added that also carry signals from an LSTM cell to later cells while being adjusted less by cell input than a regular RNN cell. Each cell is called a memory block (Graves and Schmidhuber, 2005) and consists of three gates, or multiplicative units/activation functions, which control memory by an input, output and forget gate. The input gate affects how much of the current cell's input, x_t and h_{t-1} should be accepted onto the long term memory lane, where it after an activation function is combined with the current state, on top of the cell in figure 2.1, shown by the +sign; while the forget gate is the top leftmost x that decides the scalar between 0 and 1 that the state on the long term memory lane should be multiplied with, in order to make the state forget its stored values. The right bottom-most x is the output gate that decides the amount of the state in the long term memory should be used as input to the succeeding LSTM cell together with x_{t+1}.

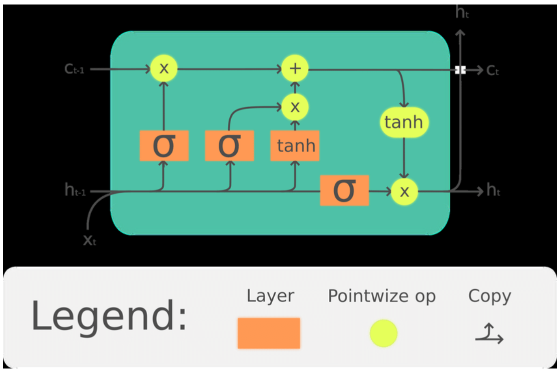

Figure 2.1: LSTM cell structure, created by (Chevalier, 2018)

An LSTM network is build up of many of these cells both in width and depth, as can be seen in section 3.2.

2.1.11.1 Hybrid machine learning models

One of the most used formal definitions of a hybrid system is the hybrid automaton (Niggemann et al., 2012) and a simplified and adapted version for learning tasks has been defined by (Vodenčarević et al., 2011) based on Niggemann's work. The defintion defines the system as a state machine with S finite states, σ as its alphabet and $T \subseteq S x \sigma x S$ giving the set of transitions in the system. The definition also contains timing constraints, function for counting number of observations per transition, and a set of functions with elements to compute value changes within S.

The types of hybrid systems in focus in this thesis consists of at least one system of non-linear machine learning methods such as neural networks and decision trees, in conjunction with another prediction system, such as a numerical forecasting method or another different type of non-linear machine learning method.

2.1.11.2 LSTM Hybrid Models

Since regular neural networks tend to struggle with learning seasonalities, as Slavek describes, and when using neural networks it most often is used tools to deseasonalize the data, he describes it as sensible to combine and merge the Holt Winters and NN models, where the NN was a recurrent neural network to be precise. When it becomes unnecessary to account for the linear trend, he describes that the forecasting formula becomes:

$$\widehat{y}_{t+1..t+h} = RNN(X_t) * l_t * s_{t+1..t+h} \tag{2.4}$$

here, X_t is a vector of preprocessed data and the multiplication is elementwise.

2.1.12 Bayesian Optimization

Bayesian optimization is founded in the method of bayesian inference, a method of continually improving a probability distributions by drawing observations from a system, in this case hourly measured weather and inflow data. The weight matrix in a regular neural network is regarded as a random latent variable drawn from a probability distribution. We want to learn a distribution for these weights that is adjusted according to the observations seen in the data, consisting of relations of output according to input, in this case output of water inflow amounts in relation to measured precipitation and temperature in the past. In the case of using bayesian optimization in combination with a numerical HBV-model the input could also consist of the amount of water stored in the layers in the soil, represented by state variables which again are updated according to previously measured precipitation, temperature, and a function for decay over time. (Springenberg et al., 2016)

The end goal is to be able to learn this posterior

$$p(W|X, y) \propto p(y|X, W)p(W)P(X)$$

However, it is computationally infeasible to perform exact inference on $p(W|X, y)$ (Louizos and Welling, 2017) resulting in the need for approximation of the posterior distribution. A common way of performing this approximation is to use variational inference, where an approximation of $p(W)$ is made as $q\lambda(W)$ with hyperparameters λ that are tuned to minimize the difference between the approximated distribution and the true posterior distribution. (Springenberg et al., 2016).

The difference between the distributions can be measured by calculating the Kullback-Leilbler divergence $KL(q||p)$ between the two distributions, and minimizing the divergence because $p = q \implies KL(q||p) = 0$.

2.1.13 Evaluation metrics

There are different evaluation metrics available to measure accuracy and loss, and these have different characteristics and behaviour

When performing inflow modelling, a focus on good evaluation is important. Some key principles that are taken into account when evaluating time series modelling are the following:

- Splitting the time series in a train and test partition for parameter tuning on the train partition and performing error calculations on the test set. The train-test split ratio is selected based on the situation, though common splits can be 70% to 90% of the full data set for training and the remaining 30% to 10% for testing and validation.
- Calculating the MAE(mean absolute error), MSE, the combination metric Smooth L1 Loss, NSE of the entire test partition of the inflow prediction are examples of ways to describe the error of a predicted time series.
- The forecast horizon for the predictions can be chosen according to the problem in focus. Examples include predictions by hour, day, month or 40 hours as in our case. Examples of prediction lead time rationale is that as correct hourly inflow predictions are important for calculating the optimal amount of water to spend in a water reservoir power plant and for reporting to the intra-day market described above. Daily correctness on the other hand is important for evacuating citizens in the case of potentially dangerous flood levels.
- Prediction results need context, and as such a baseline is important to provide context to the reader. The baseline in this thesis is the commercial HBV model by powel (Powel, 2019).

2.1.14 Loss Functions and Efficiency Coefficient

Loss Functions

- MSE - Mean squared error, also called L2 loss, is a metric that squares the difference between the prediction and target value. This has the effect of penalizing larger errors exponentially more than many smaller errors, so that for instance 5 errors of [1, 1, 1, 5] is significantly worse than [2, 2, 2, 2], which would receive the same loss value when using MAE. This property of assigning a higher loss value to larger errors is beneficial when more stability is wanted where larger errors also are of exponential importance, as described in chapter 3, Model. MSE's formula is $\frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y_i})^2$
- RMSE - Root mean squared error calculates the root of MSE which has the effect of changing the unit of the error to the same unit as the metric that is being measured, and otherwise show the same behaviour as described for MSE.

- MAE - Mean absolute error, also called L1 loss, is a metric where the absolute value of the difference between prediction and target value ensures that no loss is negative. As the loss scales linearly, this function evaluates many smaller losses with the same importance as fewer but larger losses, such as when 5 timesteps have 1 unit of error it has an equally negative outcome as 5 timesteps where 4 has an error of 0 and 1 has 5 units of error. It's formula is $\frac{1}{n}\sum_{i=1}^{n}|y_i - \tilde{y_i}|$

- Smooth L1 Loss - A combination metric of MSE and MAE, where it for values where $|x - y| > 1$ behaves exactly as MAE and for values where $|x - y| <= 1$ it behaves like MSE. This is beneficial compared to MAE when the error is small in a regression problem, where MAE can create instability; and beneficial compared to MSE as it is less sensitive to outliers.

- MAPE - Mean absolute percentage error, and sMAPE - symmetric MAPE. The mean absolute percentage error er is relative error measurement. It is defined as, when X is the observed, real value and F is the prediction/forecast: $\frac{1}{n}\sum_{i=1}^{n}\frac{|X-F|}{X} * 100$ and would have the effect in inflow prediction of penalizing heavily large errors when little water flows into the reservoir compared to when there's much inflow and the error is similarly high. This is not necessarily a correct prioritization for this domain, see chapter 5.5.

  An effect to consider when using MAPE as an error metric is that "it has a bias favoring estimates that are below the actual values" (Armstrong, 1985) due to the fact that the denominator is only based on the observed, real value. This means that when X is 60 and F is 90 the error is 30/60=0.5 which is higher than when X is 90, F is 60 and the error is 30/90=0.33. To cope with this, one can use a "Symmetric mean absolute percentage error" as (Makridakis, Hibon 2000 describes). Which is defined as where the denominator consists of the average of the forecasted and real value. This makes the errors in the two example cases described above where X and F is 60 or 90 result in the same error value for both situations.

Efficiency Coefficient

- NSE - Nash-Sutcliffe model efficiency coefficient, also called coefficient of determination, is a metric that is designed for assessing the predictive power of hydrological models and was presented by the two researchers Nash and Sutcliffe in the Journal of Hydrology in 1970 (Nash and Sutcliffe, 1970). The metric is defined as $1 - \frac{\sum_{t=1}^{T}(Q_m - Q_o)^2}{\sum_{t=1}^{T}(Q_o - \bar{Q_o})^2}$, where Q_o is observed values, $\bar{Q_o}$ is the observed values' mean, and Q_m is the model's predicted values.

The output ranges from 1 to $-\infty$ where 1 is the value for perfectly matching predictions to the observed values, where 0 is the value for when the predictions match the mean of the observed data and negative values for worse predictions where the residual variance between the predicted and observed values, which is in the numerator is greater than the variance of the data itself which is in the denominator.

NSE is a correlation and efficiency performance measure that preserves the data pattern, which neither of the above measures does, which rather consider each time step event as separate items. NSE on the other hand preserves these patterns by also considering how data points and their errors relate to each other (Bennett et al., 2013). This behaviour enables the models to be evaluated also according to if their predictions are following the patterns of the targets, but with lag or if they are preceding the target's patterns.

2.2 Structured Literature Review - SLR

The Structured Literature Review were carried out in the project preceding this thesis (Osberg, 2019) As it is not expected that the readers of this thesis will have read the pre-project report, it's contents have been re-stated here.

This section introduces the process and methods which is used to conduct the review. The SLR research objective (SLR RO) and SLR research questions (SLR RQ) that guide the review are presented. The query for the search is presented and the chosen inclusion and quality critera for filtering the search results is presented. From the basis of the research question- and objective, concepts and terms for categorizing the papers are described.

The use of a structured literature review (SLR) is done to support evidence based research with the goal of using and aggregating previous empirical research to answer a set of research questions. When comparing a structured review with other review types, narrative and thematic for instance, the systematic review utilises tools to minimise potential author bias and mistakes by rigorously cover all the papers that fulfil the inclusion criteria. However, a perceived weakness is that restricting inclusion criteria can possibly limit insights to effectiveness rather than seeking answers to more complex search questions (Grant and Booth, 2009). The perceived strengths definitely outweighs the perceived weaknesses still, as the authors also describe that this is the best known type of review.

2.2.1 SLR Objectives

The goal of the review is twofold. The first goal is to provide a systematic review of progress in the field of hydrological inflow modelling with the use of different types of neural networks. The second goal is to prepare the necessary evaluation metrics, benchmarks and data to conduct experiments of utilising neural networks to improve inflow predictions and modelling.

Table 2.1: Research objectives

ID	Research Objective
SLR RO1	Review research of water reservoir inflow modelling
SLR RO2	Define the necessary tools in order to conduct an experiment for measuring the effectiveness of BNN for water reservoir inflow modelling as future work

2.2.2 Concise & formal SLR research questions

Based upon the research objectives defined above, research questions have been made in order to aid the systematic review process. The question in focus on the review is what types of neural network architectures are used in research in the field of hydrological inflow modelling. As it is expected that multiple results to the chosen query will be multiclassifier models; either bagged or boosted in the case of ensembles, or stacked in the case of hybrid methods; another research question is which other method is used in conjunction to the machine learning methods.

Answering the first two research questions, SLR RQ1 and SLR RQ2, will also naturally answer question three, SLR RQ3, which is whether peer reviewed research has been conducted on the use of bayesian neural networks for hydrological inflow modelling, which is one of the described architectures in the background section of which the model outputs probability density functions, which enables the evaluation of the models certainty of its predictions. Question 4, SLR RQ4, is highly relevant for future work in the field and is about what is necessary for performing good experiments in the field of hydrological inflow modelling, in the form of tools, in the form of experiment setup, data foundation or other aspects discussed in the reviewed literature.

Table 2.2: SLR Research Questions

ID	**Research Objective**
SLR RQ1	What are the main neural network architectures used in research for reservoir inflow modelling
SLR RQ2	In the case of research on hybrid models for hydrological inflow modelling, which methods are used in conjunction with the machine leaning methods
SLR RQ3	Have peer reviewed research been done on the topic of using bayesian neural networks for hydrological inflow modelling
SLR RQ4	What are the necessary tools in order to conduct a good future experiment for measuring the effectiveness of a neural network architecture for water reservoir inflow modelling

2.2.3 Sources

The process of the structured literature review was done by searching among publications from the portals: IEEE Xplore, The ACM Guide to Computing literature, Science Direct and Scopus. The sources were selected by being portals for peer reviewed research, of which Scopus is the largest one of the four, hosting abstracts from 25300 journal titles (Data sources — The ISSN Portal, 2019). Scopus is also a multidisciplinary database enabling the search among titles both published under the discipline of computer science due to the focus on machine learning methods as well as in the discipline of hydrology under earth sciences due to the problem in focus.

2.2.4 Query construction

2.2.4.1 Query nr. 1

The first attempted query was:

$$Q1 : TITLE\_ABS\_KEY(hydrological\ AND\ inflow\ AND\ modelling)$$

This is a query that requires the three words “hydrological”, “inflow” and “modelling” to appear together in the set consisting of the title, abstract and keywords. It returned a vast number of articles, in the range of thousands. The returned documents was observed to in most cases to not be able to answer SLR RQ1 and SLR RQ3, as the documents did not use neural networks, or any machine learning method for that matter, to model the inflow. To account for this, query number 2 was created.

2.2.4.2 Query nr. 2

After iterating on how to narrow the search to focus solely on neural networks and multiclassifiers utilizing neural networks for predicting inflow, to match the thesis motivation, the following query was selected:

$$Q2 : TITLE\_ABS\_KEY(hydrological\ AND\ inflow\ AND\ modelling\ AND\ neural)$$

This query, which requires the words to appear in either the the title, abstract and keywords; narrowed the search substantially and specialized it to be able to answer SLR RQ1, SLR RQ2 easily. All returned papers used neural networks as one of usually multiple methods in focus in the article.

2.2.4.3 Selection Criteria

The following inclusion criteria (IC) and quality criteria (QC) have been used to select relevant literature in the search for hydrological time series forecasting

IC1 - The study's main focus should be inflow modelling by utilising measured or predicted weather data, such as precipitation and temperature.

IC2 - The type of publication of the literature is an original study presenting empirical results, and not a review or editorial.

IC3 - The study focuses on improving accuracy / decreasing errors in hydrological inflow modelling by the means of neural networks, either alone or in combination with other methods, in the form of a multiclassifier.

IC4 - The literature is reasonably recent, by being published in between and including the years 2014 and 2020

QC1 - The study must evaluate model performance by using objective measures, such as MAE and other evaluation methods described under "Evaluation Metrics"

QC2 - The study should have been peer-reviewed, which is fulfilled by choice of databases.

QC3 - The study is put into context of other studies and research

The reason for choosing IC3 as a criteria is due to the recent advancements in machine learning, such as described by Turing Award winners here (Lecun et al., 2015)

2.2.5 Results

This section describes the findings from the conducted structured literature review and relates them to the defined research questions and objectives from section 2.2.2.

2.2.5.1 Excluded Papers

The result from reading all returned abstracts to the query was that 13 articles, 42%, were marked as “Not relevant”, due to not meeting IC1 “The study’s main focus should be inflow modelling by utilising measured or predicted weather data, such as precipitation and temperature”. The most common reason for this was that the study did not attempt inflow modelling, but rather modelling of a different aspect of hydrology, by using inflow as input to the model. Therefore it is very understandable that these articles also were returned from the chosen query.

The full description of why each article did not meet the inclusion and/or quality criteria is described in the full spreadsheet file of the returned articles, which in the appendix of (Osberg, 2019). Other examples of focus for excluded articles was: attempting synthetic data generation rather than modelling, prediction of discharge within the inflow and not the inflow itself, creating a modelling framework for entire ecosystems, and prediction of where, when, and for how long algal blooms will occur in water body.

2.2.5.2 Neural Network Architecture

SLR RQ1 questions what kind of neural network architecture is used in research in this field. The following table, 2.3, shows the aggregated data of how many papers described each of the listed neural network architectures:

Table 2.3: Neural network architecture in focus

NN architecture	Corresponding papers
MLP - MultiLayer Perceptron	11
WANN - Wavelet Artificial Neural Network	5
ANFIS - Adaptive Neural-based Fuzzy Inference System	3
RBFNN - Radial Basis Function Neural Network	1
ENN - Elman Neural Network	1
RNN - Recurrent Neural Network	1
Total amount of relevant returned papers	19

2.2.5.3 Multiclassifier Models

In the same manner as the previous table, 2.3, answers SLR RQ1, the following table shows aggregated data in regards to SLR RQ2: "In the case of research on hybrid models for hydrological inflow modelling, which methods are used in conjunction with the machine leaning methods". The literature review and Database in the Appendix of (Osberg, 2019) shows that 9 of the 13 relevant articles, 69%, addressed multiclassifier models, and their structure is described in the tables 2.4 and 2.5.

Table 2.4: Multiclassifier individual methods in use, excluding methods in 4.1

Method	**Corresponding papers**
AR - Autoregressive modelling	2
ARX - Combination of AR and exogenous values	1
ARMAX - Autoregressive Moving average with exogenous input	1
WNARX - Wavelet-based Non-linear Autoregressive with Exogenous input	1
SARIMA - Seasonal Autoregressive Integrated Moving Average	1
CPMDE - Combined Pareto Multi-Objective Differential Equation)	1
VIC - Variable Infiltration Capacity	1
MLR - Multiple Linear Regression	1
WBMLR - Wavelet Based MLR	1
SVR - Support Vector Regression	1
TF - Thomas Fiering model	1
Total amount of multiclassifier articles	9

Table 2.5: Full multiclassifier architectures. Full names in table 2.3 and 2.4

Multiclassifier architecture	**Corresponding papers**
Ensemble of 3 MLP & Bayesian model averaging	1
Ensemble of WANN & WBMLR & MLP & MLR	1
AR & MLP stacked	1
CPMDE & MLP stacked	1
VIC & WANN stacked	1
VIC & AR stacked	1
VIC & ARMAX stacked	1
VIC & WNARX stacked	1
RBFNN & SVR stacked	1
WANN & TF stacked	1
ENN & ENN stacked	1
Total amount of multiclassifier articles	9

Among the hybrid structures most of the models are multiclassifiers in a hybrid form that utilized “stacking” which is a method that uses the output of one model as input to the next, and often then training one part of the multiclassifier to model and correct for the errors of a preliminary method in the hybrid system.

2.2.5.4 Comparison of chosen evaluation metrics

The reviewed articles have chosen somewhat different sets of evaluation metrics, which are summarized in table 2.6. In the table they are also categorised into “Deterministic, absolute”-, “Deterministic, relative”-, “Probabilistic”- and “Manual” metrics, which respectively output an error metric either deterministically as an absolute value, deterministically as a relative size of the predicted values size, as a histogram/aggregation of histogram, or as a description from a manually conducted case analysis often with the use of the other described metrics. Among the retrieved relevant articles 2 articles does not provide readily available information as one of the relevant papers is written in spanish, with an english abstract, and in another case the journal article is not available without purchase, even under the collaboration between the university and the journal publisher, Elsevier. The article in question has been ordered in the form of a physical copy via the university library, and its details will be added at the time of working on future steps.

Table 2.6: Used Evaluation Metrics

Category	**Evaluation metric**	**Corresponding papers**
Deterministic, Absolute	RMSE - Root Mean Square Error	10
	NSE - Nash Sutcliffe Efficiency Coefficient	10
	MAE - Mean Absolute Error	5
	NRMSE - Normalized RMSE	1
	Bias & Mean Bias error$_{(4)}$	1
Deterministic, Relative	CC - Correlation Coefficient	8
	Coefficient of Determination, square of CC	3
	RAE - Relative Absolute Error	3
	MAPE - Mean Absolute Percentage Error	2
	MARE - Mean Absolute Relative Error	1
	RRSE - Root Relative Square Error	1
	Peak Flow Criteria$_{(3)}$	1
	Scatter Index$_{(4)}$	1
	Willmott Index of Agreement$_{(4)}$	1
	Confidence Index$_{(4)}$	1
	Percentage Bias	1
Probabilistic	PIT - Percentage Integrated Transform Histogram	1
	CD - Calibration Deviation$_{(1)}$	1
	IGN - Ignorance Score$_{(2)}$	1
	CRPS - Continuous Ranked Probability Score	1
	Efficiency Index	1
Manual	Case Analysis	3
	Own Defined Metric$_{(5)}$	1
	Amount of articles with complete description of evaluation metrics	17

The following less frequent metrics are referenced directly for ease of access to documentation:
(1) The metric CD - Calibration Deviation is the degree of deviation from flat PIT histogram and is referenced from (Nipen and Stull, 2011)
(2) The metric IGN is referenced from (Roulston and Smith, 2002)
(3) Peak flow criteria is referenced from (Budu, 2014)
(4) The metric is referenced from (Allawi et al., 2019)
(5) The defined metric is found in (Wang and Lou, 2019)

2.3 SLR Discussion

2.3.1 Sensitivity Analysis

During the research process decisions are made, of which have the possibility of affecting the outcome of the research. This section addresses this issue by reflecting over possible different decisions.

2.3.1.1 Database Choices

The chosen databases of articles are described in 2.2.3, and contains multiple high regarded databases, of which Scopus is the definitively largest one in the set used in this study. According to (Elsevier's Webpage, 2019) in 2017, it hosted 75 million items from 16 million Author profiles and 70 000 institution profiles dating back to 1970.

There are however more large scale abstract databases than Scopus and Martín-Martín et al. (2018) compared systematically the citations of articles shown by Google Scholar, Web of Science, and Scopus. Their review concluded with amongs other things that Google Scholar finds significantly more citations than the Web of Science and Scopus across all subject areas. As such Google Scholar also finds a vastly higher amount of articles than the other two detabases in question, which could lead to more relevant retrievals to the query of this study.

However a higher article retrieval count is not the only factor in focus. Google Scholar was described by Martín-Martín et al. (2018) to find 48% to 65% from other materials than journals, and rather from theses/dissertations, books or book chapters, conference proceedings, unpublished materials, and other document types. They continue by concluding that "Google Scholar has reached a high level of comprehensiveness... However, at this point there is no reliable and scalable method to extract data from google Scholar, and the metadata offered by the platform is still very limited...".

It is uncertain whether search with other large scientific search engines such as Google Scholar would have increased the number of relevant retrieved articles by a large amount, and whether or not retrieved articles would to a larger extent not meet the inclusion criteria. Especially so as Google Scholar does not retrieve solely peer-reviewed articles, according to Martín-Martín et al. (2018), which is one of the Quality Criteria of this SLR.

2.3.2 Principal Findings

This study conducts a systematic review of 31 papers returned from the chosen query. 18 of the 31 returned papers, 58%, were marked as relevant as relevant or semi-relevant. Reasons for the semi-relevant classification was that the papers met IC1 and IC3 still, however inflow modelling was not the single most important focus of the article, but rather ranked to equal importance with other aims. Examples include the focus on modelling both inflow and outflow in the article.

2.3.2.1 Neural Network Architecture in Focus

As can be seen in table 2.3, a substantial majority of the relevant papers used multilayer perceptrons either as the full model, as a part of an ensemble or as a apart of a stacked hybrid model. The second most used architecture was the wavelet artificial neural network, then followed by the adaptive neural-based fuzzy inference system. This data provide good objective answers to SLR RQ1, in the context of the utilised query and databases.

It is clear to see that the neural network types used are rather homogeneous when compared to the joined methods used in the case of multiclassifiers. In the light of these articles, there has been more effort put into testing similar neural network architectures with many different supporting methods than to test the effectiveness of different neural network types. Different neural network types that are dominant in the field of time series modelling and that are either not found here or that is barely touched upon in the reviewed articles, are described in section 2.3.3.

2.3.2.2 Multiclassifier Methods in Focus

Table 4.2 and 4.3 shows that the architecture of multiclassifier models in this review is very heterogeneous, with almost no single architecture being used in two different returned articles. However when the scope is broadened slightly and methods are aggregated into categories, ARIMA, versions of ARIMA or submethods of ARIMA is the most common method to use in conjunction to neural networks in multiclassifiers.

2.3.2.3 Experiment Tools in Reviewed Papers

The tools that has been found to be used in the context of this review's query, databases, and inclusion criteria with extra emphasis on IC1, is: Historical weather data, which consists of precipitation, wind data, temperature data and in one case typhoon data; measured inflow data for both evaluation and training, objective evaluation metrics that in most cases are put into the context of relevant benchmarks by comparing the model to existing well tested models. Some were also tested on a public readily available dataset.

2.3.2.4 Untouched/barely Touched Topics According to SLR

The homogeneity of the neural network types used in the reviewed articles show that either advancements and improvements from the field of machine learning on neural networks has not made its way into the field of hydrology, or that the language and terms that are used are different depending on whether the researchers main field is hydrology, under Earth Sciences or if its machine learning, under Computer Science. Which explanation that are more likely is not certain, and is described in section 2.4.1.

Some examples of barely touched neural network types are: Recurrent Neural Networks and corresponding architectures such as LSTM(Long Short-Term Memory), GRU(Gated Recurrent Unit); Elman Neural Networks; Convolutional Neural Networks; M-Competition winning hybrid methods such as ES-RNN(Exponential Smoothing-Recurrent Neural Network) which is described in section 2.2.

2.3.3 Recurrent Neural Networks and LSTM's for Inflow Modelling

One of the untouched upon types of neural network for this application according to this structured review is the RNN architecture long short term memory networks(LSTM). As can be seen from the results in 2.2.5.2, none of the retrieved articles attempts using a LSTM's for inflow modelling. We will the following sections of the discussion look into what prerequisites are necessary to perform a good experiment with this architecture in this domain, in order to give more details to the answer of SLR RQ4.

2.3.3.1 Experiment Regarded Tools

In order to be able to evaluate the performance of bayesian neural networks or other methods, a benchmark for comparison is needed for context. The state of the art in commercial industry, for Trønderenergi Kraft AS, is utilizing Powel's "Powel Inflow" tool (Powel, 2019), which is a commercial implementation of the HBV-model. The tool is closed source software where state variables are hidden from the user. The variables that are available for the user are the hyperparameters for the given inflow field that experts/users select and the total inflow to a reservoir from all HBV-model buckets combined. The Powel Inflow model's output is an example of a contextualising model that can be used for comparison when evaluating an experiment.

A numerical HBV models usability does not only limit itelf to verification however. It will also enable building a stacked hybrid aswell, as the predicted bucket values can be treated as inputs to a neural network, or the predicted inflwo. Then the stacked networks can correct errors or get more data to use for inflow modelling rather than just using a standalone neural network model. Howeeer, because the commercial implementation from Powel is closed source, an open implementation of an HBV-model isnecessary in order to benefit from the stacked hybrid approach.

A suggested open source HBV-model is the numerical HBV-96 implementation based on the research of Amir Agha Kouchak from the University of California - Irvine (AghaKouchak, 2010), who have implemented a MATLAB version of the HVB-96 model in conjunction with his paper in 2010. This model has been converted to Python by water resource engineer John Robert Craven (Craven, 2016).

2.3.3.2 Model Comparison, Academical - Commercial

The commercial HBV model built by Powel, which is served to B2B customers as a purchaseable tool. This model is clearly more advanced than the academic HBV_96 implementation, as it has more than 6x the amount of hyperparameters than the academical version. Distinct differences is especially seen in how they model a snowpack, and altitude temperature differences. Here the commercial Powel Inflow tool utilises a distinct hyperparameter per temperature difference in 10 height regions and has over 10 parameters solely for adjust the behaviour of a snowpack. such as distinctions for quartile, 2*quartile, 3*quartile and fully covered land areas of snow.

Due to the increased complexity, and adaptability the commercial model is naturally expected to perform better than the open implementation for the Søa field, especially due to the difficult snowpack modelling features that are especially valuable in Norway, in a climate with yearly snowfall, and also due to the great expert help that has been provided when this models parameters has been tuned by professional experts.

2.4 SLR conclusion

As can be seen in the SLR discussion 2.3, this review has summarised research in the field of Hydrological inflow modelling by the use of neural networks either standalone or as ensembles or hybrids in conjunction with numerical or regression techniques. The research objectives of answering which NN architectures have been in focus in this field of research has been accomplished, and information from the review process to reach this objective has also provided solid and good information on how to prepare for proposed experiments that will be described further in section 2.4.1.

The most critical details to SLR RQ1 to SLR RQ4 can be summarised as follows:

- SLR RQ1, What are the main neural network architectures used in research for reservoir inflow modelling?
  - The main neural network architecture types use in research in the field is by a large margin MLP networks, followed by Wavelet neural networks, and other architectures described in table 2.3.
- SLR RQ2, In the case of research on hybrid models for hydrological inflow modelling, which methods are used in conjunction with the machine leaning methods?
  - The main methods used in conjunction to neural networks in the case of multiclassifiers are varied and the method that is the most used is autoregression. When the scope is broadened slightly and methods are aggregated into categories, ARIMA, versions of ARIMA or submethods of ARIMA is the most common method to use in conjunction to neural networks in multiclassifiers.
- SLR RQ3, Have peer reviewed research been done on the topic of using bayesian neural networks for hydrological inflow modelling?
  - Peer reviewed research on bayesian neural networks has not been done in this field in the search and database context of this study. How this can be achieved in the future is described in chapter 8, Future Work, and in the context of answering SLR RQ4.
- SLR RQ4, What are the necessary tools in order to conduct a good future experiment for measuring the effectiveness of a neural network architecture for water reservoir inflow modelling?
  - The tools that has been found to be necessary in the context of this review's query, databases, and inclusion criteria with extra emphasis on IC1, is: Enough historical weather data with as low uncertainty as obtainable, measured inflow data for both evaluation and training,

objective evaluation metrics put into the context of relevant benchmarks by comparing the model to existing well tested models also on a public readily available dataset.

2.4.1 Untested Architectures & Open Challenges

As described in section 2.3 and 2.4, multiple architectures have not been tested in this review's context. These architectures and open unsolved problems are e.g:

- LSTM networks have not been tested in this review's context, this is a suggested step for future work.
- The same applies to hybrid models that utilise the HBV bucket values as input to a neural network.
- In regards to the differences in the academical and commercial model for modelling a snowpack, research with special focus on modelling inflow when there is a snowpack on the landscape is highly interesting and a challenging problem to model accurately according to hydrologist Frode Vassenden (Domain expertise interview, October 24, 2019).
- Research with the focus of doing flood prediction is a very encouraged field of future work, as it is of high importance to residents safety near inflow fields.

2.5 Motivation

The underlying motivation for this thesis is to improve water inflow forecasting. This can however be specified and concretized by categorizing it into societal, ecological, economical, personal and technological motivation in the following manner:

- Societal motivation - Being able to accurately predict inflow to water reservoirs accurately can help predict floods, especially during spring, and ensure time for a safe temporary evacuation for citizens of a flood prone residential area.
- Environmental motivation - Being able to accurately predict inflow can help power plant operators ensure that enough water is available in the future to help ecosystems around water power plants downstream to remain stable (Suen and Eheart, 2006).
- Economic motivation - Being able to accurately predict inflow enables water power plant operators to maximize revenue per water available due to:

  - Lower risk of loosing water due to overflow caused by erroneous inflow predictions forecasting to low values, and when the reservoir is close to full and the inflow is of a higher value than the maximum possible consumption at the power plant;
  - Lower risk of water reservoir depletion as a consequence of lower inflow than predictions suggest or due to too higher evaporation than expected during dry seasons;
  - These lower risks helps water power plant operators's attempt at optimising the revenue per water used by creating more power when national consumption is high and less when consumption is lower but still avoid depletion and overflow.

  Which enables power plant operators to give more accurate bids for the amount of power they should create in the day-ahead market, described in chapter 2. Due to the structure of the Nordic Power market, also described in chapter 2, these economic benefits for operators can also affect prices beneficially for consumers which translates to a socioeconomic motivation as well.

- Personal motivation - It has been a personal motivation to learn more about neural networks through this thesis, because the amount of data available (155304 points) is a relatively large dataset capable of tuning neural networks.

- Technical motivation - Being able to more accurately predict inflow by using modern machine learning techniques in conjunction with domain specific models in order to verify if deep learning can provide value in the field of hydrology

  - To measure which architectures perform the best of a physical model, a deep machine learning model or a combination model.
  - To measure the effectiveness of stacked hybrid models with the commercially used HBV model and a neural network architecture
  - Measure whether and if so, the amount of positive impact on performance knowing all internal state variables of the numerical part of the hybrid has on predictions
  - Testing an open source HBV model that is free and a white box model with full access to it's inner state.

Chapter 3

Architecture/Model

The models that will be tested in these experiments are models that all use an LSTM network as a component. The differences between them are what data is used as input to the LSTM network, where they all use precipitation and temperature data. The hybrid models with an open source or commercial HBV model are tested in 3 stages from simpler to more complex:

- 1 - Only predicted inflow from HBV, precipitation and temperature
- 2 - All data from 1, and access to previous inflow, where datapoints up until the last 4 are smoothed and the last 4 points are unsmoothed.
- 3 - All data from 2, and access to the available state variables of the HBV models, which is complete access in the case of the open source model, and limited access in the case of the commercial model.

The commercial HBV model is far more complex than the open source model, however the limited state available mimics the full state of the open source model to some degree because the data available is the global bucket contents, which mimics the single bucket that the open source model consists of. The commercial model consists of many buckets divided into different reservoir heights with different air temperature, but this state data is closed source.

3.1 Models

The model diagrams show:

- Each hybrid's sub-models: Open source HBV model, commercial HBV model and the LSTM network; shown in coloured round-edged boxes
- Measured data: Inflow, temperature and precipitation; shown in rightward facing arrows
- Sub-model output and state data; shown in leftward facing arrows

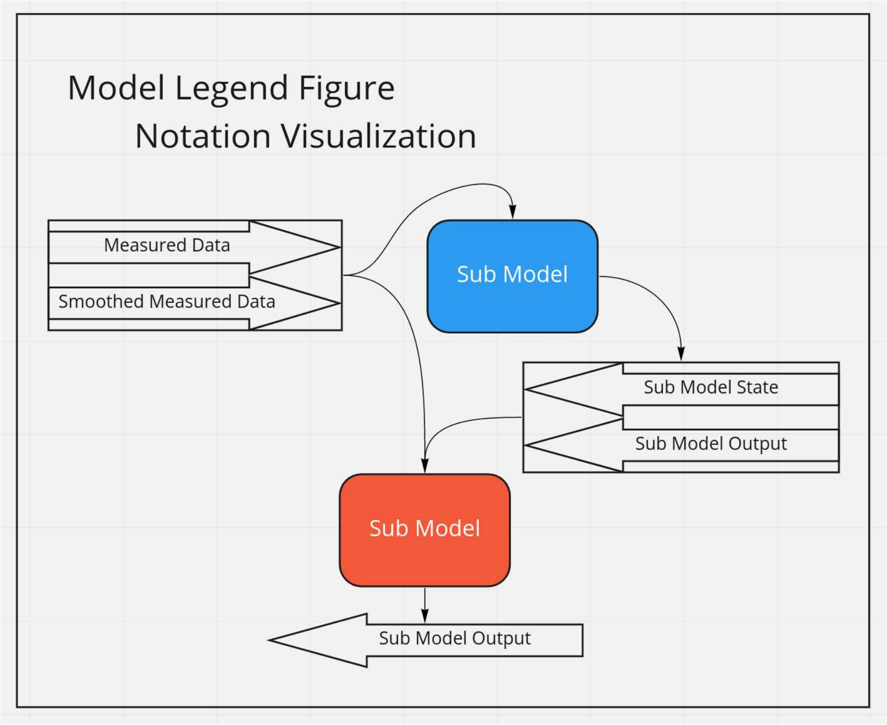

Figure 3.1: Model Notation Visualization, as described above

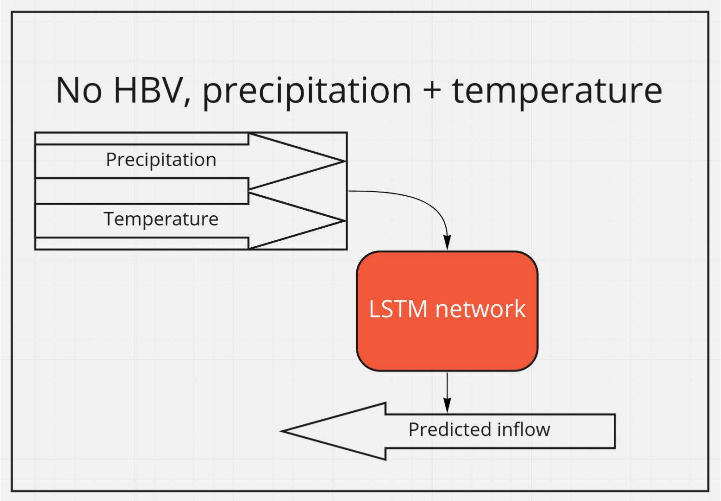

Figure 3.2: Model using only an LSTM net with precipitation and temperature as input

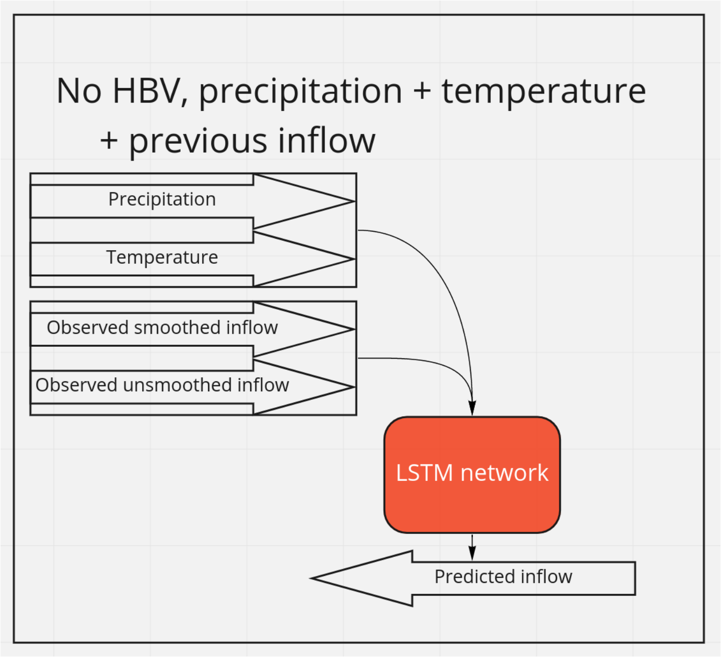
No HBV, precipitation + temperature + previous inflow
Precipitation
Temperature
Observed smoothed inflow
Observed unsmoothed inflow
LSTM network
Predicted inflow

Figure 3.3: Model using only an LSTM net with precipitation, temperature and previous inflow as input

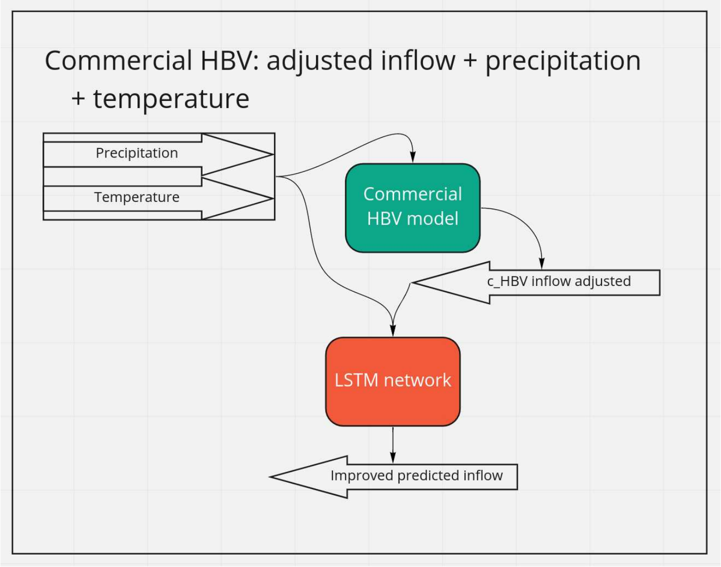
Commercial HBV: adjusted inflow + precipitation + temperature
Precipitation
Temperature
Commercial HBV model
c_HBV inflow adjusted
LSTM network
Improved predicted inflow

Figure 3.4: Model using the commercial HBV's predicted inflow together with precipitation and temperature as input to an LSTM net

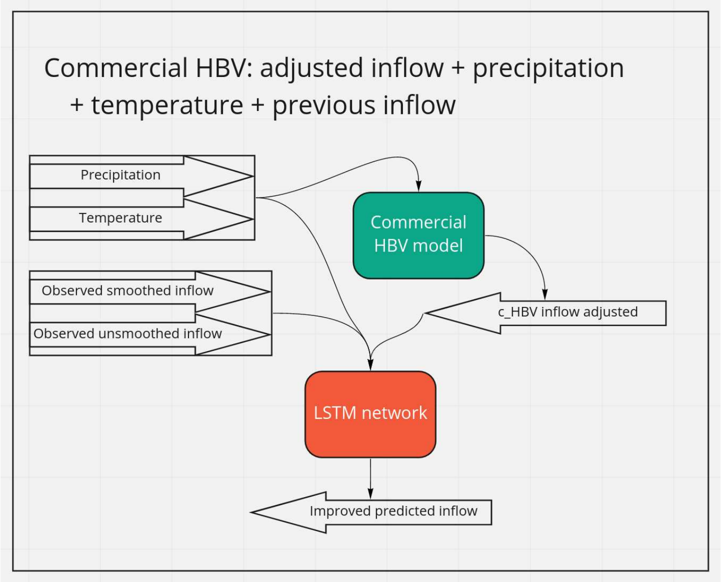
Commercial HBV: adjusted inflow + precipitation + temperature + previous inflow
Precipitation
Temperature
Commercial HBV model
Observed smoothed inflow
Observed unsmoothed inflow
c_HBV inflow adjusted
LSTM network
Improved predicted inflow

Figure 3.5: Model using the commercial HBV's predicted inflow together with precipitation, temperature and previous inflow as input to an LSTM net

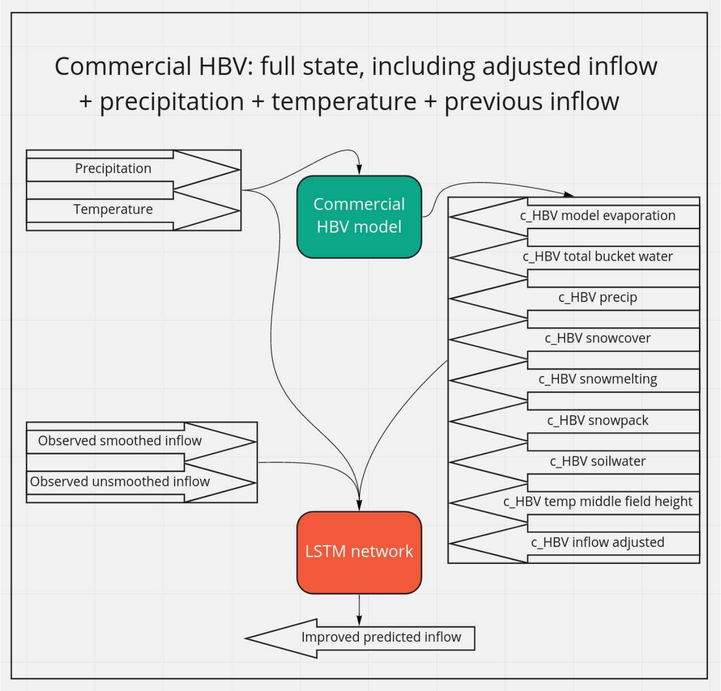

Figure 3.6: Model using the commercial HBV's limited internal state together with precipitation, temperature and previous inflow as input to an LSTM net

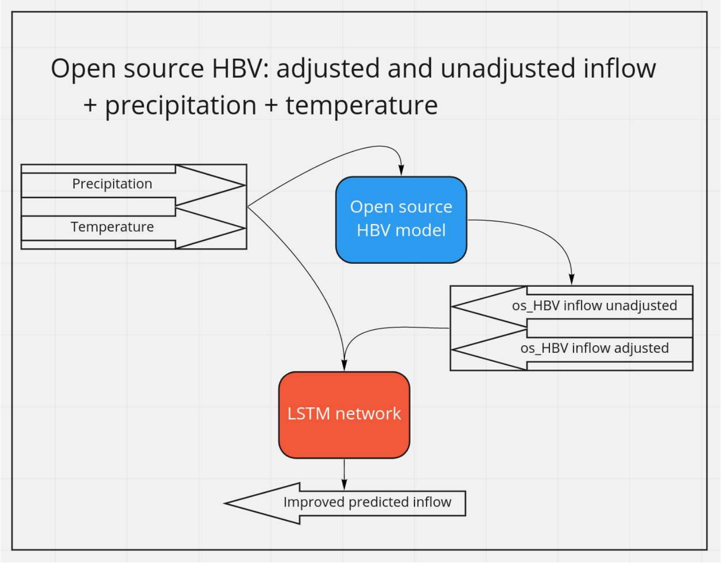

Figure 3.7: Model using the open source HBV's predicted inflow together with precipitation and temperature as input to an LSTM net

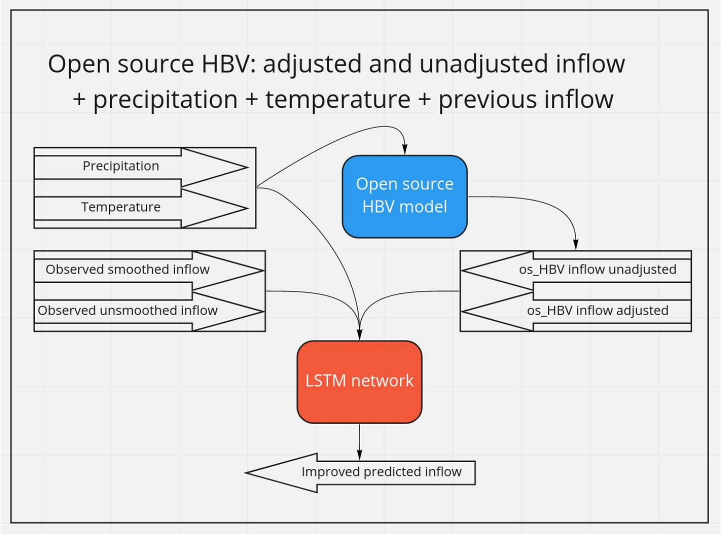
Open source HBV: adjusted and unadjusted inflow
+ precipitation + temperature + previous inflow
Precipitation
Temperature
Open source HBV model
Observed smoothed inflow
Observed unsmoothed inflow
os_HBV inflow unadjusted
os_HBV inflow adjusted
LSTM network
Improved predicted inflow

Figure 3.8: Model using the open source HBV's predicted inflow together with precipitation, temperature and previous inflow as input to an LSTM net

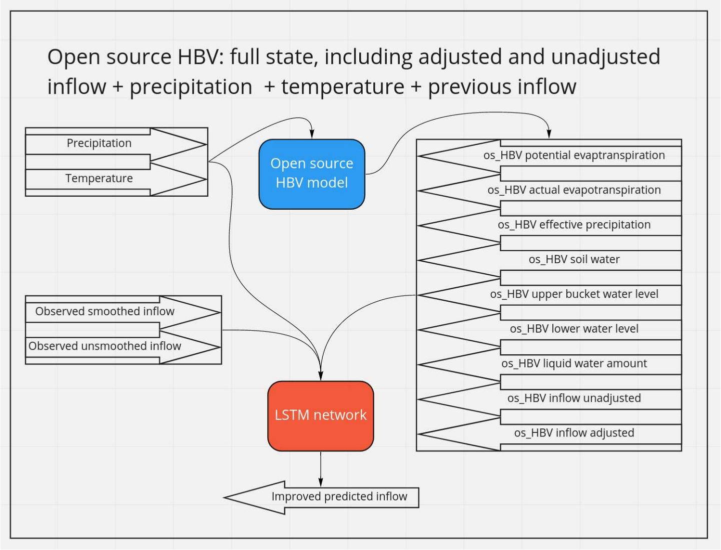
Open source HBV: full state, including adjusted and unadjusted inflow + precipitation + temperature + previous inflow
Precipitation
Temperature
Open source HBV model
os_HBV potential evaptranspiration
os_HBV actual evapotranspiration
os_HBV effective precipitation
os_HBV soil water
os_HBV upper bucket water level
os_HBV lower water level
os_HBV liquid water amount
os_HBV inflow unadjusted
os_HBV inflow adjusted
Observed smoothed inflow
Observed unsmoothed inflow
LSTM network
Improved predicted inflow

Figure 3.9: Model using the open source HBV's full internal state together with precipitation, temperature and previous inflow as input to an LSTM net

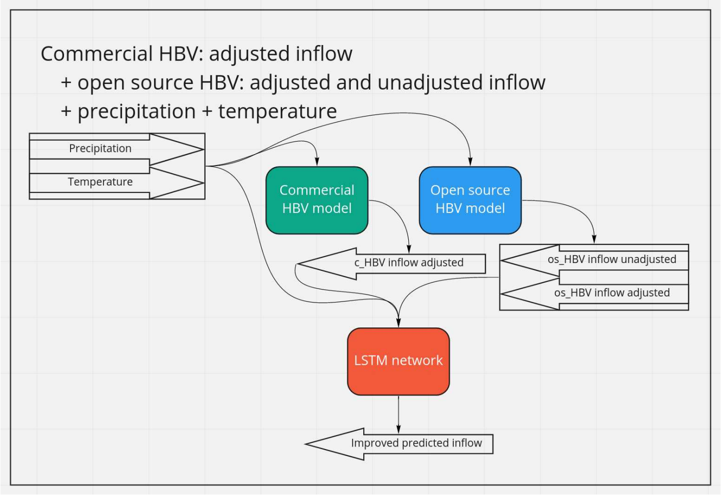

Figure 3.10: Model using both the commercial HBV's- and open source HBV's predicted inflow together with precipitation and temperature as input to an LSTM net

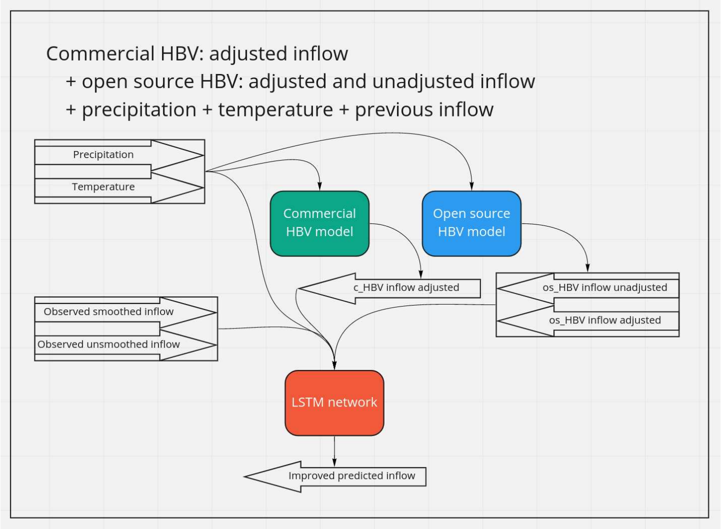
Commercial HBV: adjusted inflow
+ open source HBV: adjusted and unadjusted inflow
+ precipitation + temperature + previous inflow
Precipitation
Temperature
Commercial HBV model
Open source HBV model
c_HBV inflow adjusted
os_HBV inflow unadjusted
os_HBV inflow adjusted
Observed smoothed inflow
Observed unsmoothed inflow
LSTM network
Improved predicted inflow

Figure 3.11: Model using both the commercial HBV's- and open source HBV's predicted inflow together with precipitation, temperature and previous inflow as input to an LSTM net

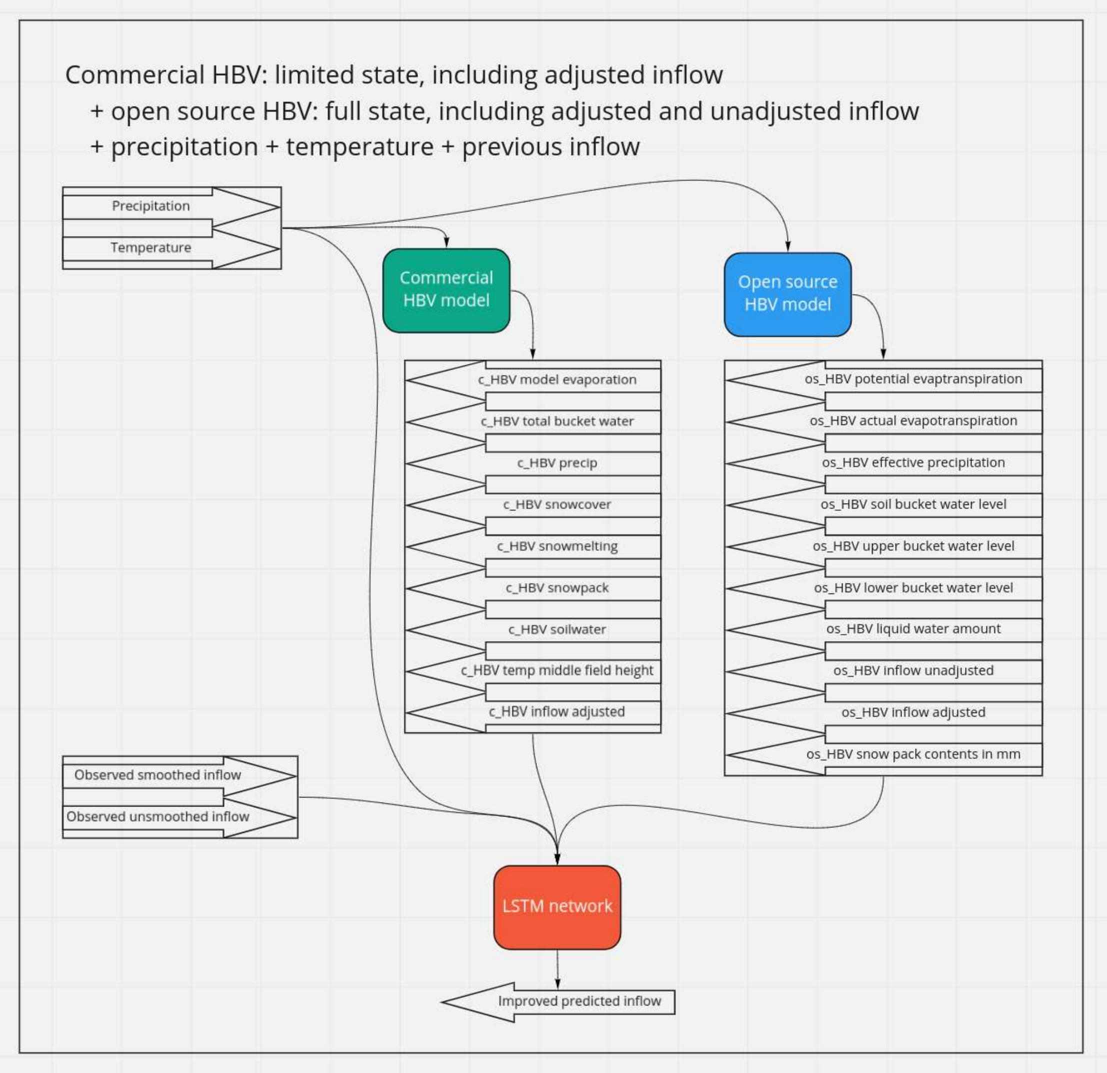
Commercial HBV: limited state, including adjusted inflow
+ open source HBV: full state, including adjusted and unadjusted inflow
+ precipitation + temperature + previous inflow
Precipitation
Temperature
Commercial HBV model
Open source HBV model
c_HBV model evaporation
c_HBV total bucket water
c_HBV precip
c_HBV snowcover
c_HBV snowmelting
c_HBV snowpack
c_HBV soilwater
c_HBV temp middle field height
c_HBV inflow adjusted
os_HBV potential evaptranspiration
os_HBV actual evapotranspiration
os_HBV effective precipitation
os_HBV soil bucket water level
os_HBV upper bucket water level
os_HBV lower bucket water level
os_HBV liquid water amount
os_HBV inflow unadjusted
os_HBV inflow adjusted
os_HBV snow pack contents in mm
Observed smoothed inflow
Observed unsmoothed inflow
LSTM network
Improved predicted inflow

Figure 3.12: Model using both the commercial HBV's limited internal state and open source HBV's full internal state together with precipitation, temperature and previous inflow as input to an LSTM net

3.2 LSTM Network Architecture

The LSTM network Visualization in 3.13 created by (StackOverflow nnnmmm, 2018) shows a general visualization of depth and width of LSTM networks, where a depth of 2 stacked LSTM layers is the case for $w = 1$, and the width of hidden dimensions decides the value of n. The effect of a higher or lower n value decides how many timesteps that will be evaluated at the same time to create a prediction

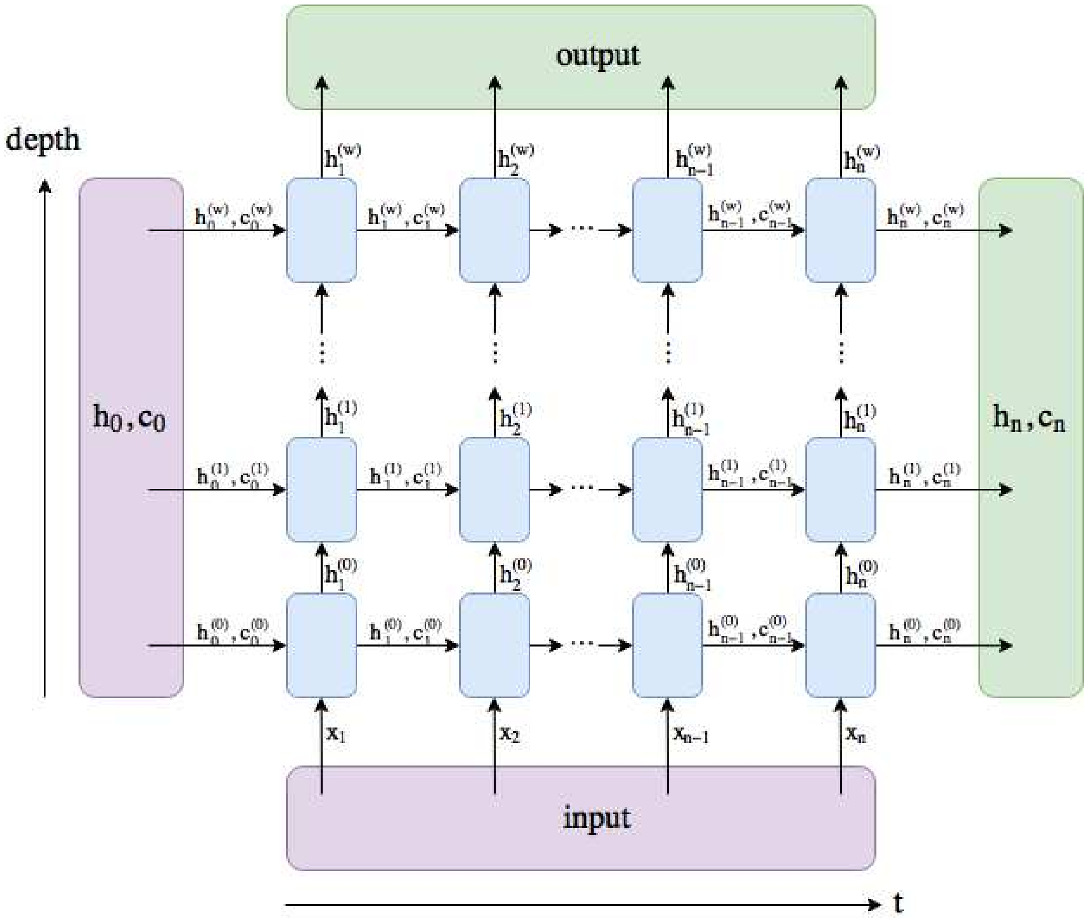

Figure 3.13: LSTM visualization of LSTM layers and hidden dimensions

In addition to tuning the depth and width of the LSTM network, the architecture and pipeline is also comprised of:

- The LSTM network is initialized with an input dimension according to the amount of data fields sent as input from precipitation, previous inflow and HBV models for the respective hybrid models.
- A hidden state is initialized with the dimensions(lstm layers, batch size, hidden dimension). The hidden state's weights are initialized as zero values

for reproducibility, and passed to CUDA on the GPU in order to achieve as reproducible results as we can on a GPU. See section 5.1 for more details about reproducibility.

- The input data is passed into the network together with the hidden state by the training algorithm and passed through the lstm layers and hidden dimensions, with a dropout function after each LSTM layer, also when there is only 1 layer.
- The dropout function prevents overfitting by randomly setting output state values to zero, according to the dropout probability tuned as a hyperparameter.
- After the data is passed through the dropout function after the final lstm layer, it is passed to an MLP fully connected layer which changes the dimensionality of our data to the output dimension.
- The output dimension of the fully connected layer is set to (1 x forecast horizon) where out forecast horizon is 40 time steps
- After the output is received from the network, backward propagation is done which accumulates gradients based the loss function.
- Then gradient clipping is performed on these accumulated gradients. This prevents exploding gradients by clipping the derivative loss so that it is kept within a certain range defined by the gradient clipping hyperparameter.
- After the gradient clipping the optimizer updates the net parameters based on the currently stored gradients
- Then finally the networks accumulated gradients are set to zero toprepare for another iteration.

3.2.1 Loss Function and Evaluation Metric

The experiments in this thesis use MSE as the chosen loss function because minimizing fewer larger errors are of higher importance than minimizing smaller errors often. The reasons behind this are both due to the cost of reservoir overflow, the way the Nordic power market's day-ahead market operates and also because of the environmental and societal cost of floods. A water power plant that registers that it will produce X amount of power in the day-ahead market but fails to do so because of a lack of water in the reservoir and is not able to trade or hedge their errors in the intra-day market, described in paragraph 2.1.4.5, is penalized heavily because of the increased market price per megawatt in the balance market supplied by the contry's operating TSO, described in paragraph 2.1.4.6. On the other side of the spectrum the cost for overfilling a water reservoir is also vast. The cost here consists of the alternative cost for the plant operators of lost water, and also the cost of damage from floods to infrastructure during the spring flood period, as well as societal and environmental cost.

MSPE (Mean Squared Percentage Error) or other relative metrics will not be used as a loss function because the errors in focus is not always in relation to the expected inflow, but rather in relation to the reservoir water amount. A negative deviation, where the observed inflow is lower than the predicted inflow is also a more severe error when the reservoir water levels are low, and likewise a positive deviation, where the observed inflow is higher than the predicted inflow is a more severe error when the reservoir water levels are high. As such, a relevant problem for future work is to design and test a relative error metric that measures inflow error relative to reservoir water contents, as is described in section 5.5.

After the training and tuning of the models has been completed with the use of MSE, NSE will be used for the primary evaluation of the models because of its ability to consider data patters, as described in section 2.1.14. However, NSE is calculated as sums over the forecast horizon so that the output value will be a scalar that represents the entire forecast of 40 hours, meaning that it will not be possible to use NSE to visualize performance per hour when the data has the same granularity. Because of this, MSE, MAE and Smooth L1 Loss will be used to present the models' hourly performance.

Chapter 4

Experiments and Results

4.1 Experimental Plan

The experimental plan consists the following steps:

1. Decide upon suitable inflow field for experiments, together with hydrologist Frode Vassenden
2. Join available data into dataset for the suitable inflow field
3. Analyse outliers and data quality for the data set
4. Clarify what operations or processing, if any, have been performed on the dataset
5. Implement tunable LSTM network that will aim at forecasting inflow in the case of non-hybrid models with only an LSTM network, and aim at improving inflow forecasts in the case of hybrid models with an HBV model component
6. Tune the open source HBV model on the training partition of the data
7. Clarify data acquisition latency from time of measurement and the time it is available to the prediction models, and implement that latency as a gap between the input data and the target labels.
8. Implement custom dataset logic to be able to combine smoothed and unsmoothed data as model input and target lables.

9. Implement hybrid model structure for three stages of simple to complex models for the case of using the open source HBV, commercial HBV and both HBV models at the same time:
   - Simple: LSTM has access to measured weather data and HBV inflow predictions
   - Medium: LSTM has access to the data for the simple model, and in addition access to previous inflow data
   - Complex: LSTM has access to the data for the medium model, and in addition access to all available data ablit the HBV model(s) inner state
10. Select suitable loss function/evaluation metric for the problem definition
    - Implement custom evaluation metric for the pyTorch framework for NSE, Nash Sutcliffe Efficiency
11. Select suitable forecast horizon length for the problem definition
12. Perform initial upper boundary search for the models' hyperparameters
13. Perform hyperparameter search using bayesian optimization for efficient search, by evaluating performance on a validation partition of the data
14. Create model performance metrics by using the tuned parameters to train the model on the train partition, validating and selecting the model that performs best on the validation partition, and creating the performance metrics based on error on the test partition of the data.
15. Evaluate performance by comparing performance between the models on the test partition
16. Measure performance by total mean error and standard deviation and also by hourly mean error and standard deviation during the selected forecast horizon.

As the data is collected before the conduction of these experiments this is a retrospective study, as the goal in this study is predicting temporal events it is prognostic, and as the forecasts will predict continuous variables it's a regression problem.

4.2 Experimental Setup

4.2.1 Data

The data available are time series dating back to 1st of September 2002 and up to 20th of may 2020 for Søa water power plant which is operated by TrønderEnergi Kraft AS. The datasets used during the current study are not publicly available due to them being owned by TrønderEnergi Kraft AS, and is considered confidential by the company. The data is still available from the author with approval from TrønderEnergi on reasonable request as is recommended by SpringerLink (SpringerLink, 2020). The data has a granularity of 1 hour, meaning that there are 155304 data points over the 6471 days, with the following dimensions:

- Measured data:
  - Measured precipitation in mm per hour
  - Measured temperature in degrees Celsius
  - Measured inflow, smoothed in m^3 per second The smoothing operation evens out each data point according to 4 neighbouring values(a smooth width of 5) (Domain expertise interview, May 20, 2020), two preceding and two succeeding data points. This is the timeseries used by TrønderEnergi to tune and evaluate the commercial HBV model's performance.
  - Measured inflow, unsmoothed in m^3 per second
- Commercial HBV limited state data:
  - Evaporation in mm
  - Total HBV model bucket contents in mm
  - Predicted inflow in m^3 per second
  - Snowpack-coverage fraction, in percentage
  - HBV predicted snowpack-melting in mm
  - HBV total snowpack in mm
  - Soil water level in mm
  - HBV predicted field-temperature at mid-height in unit of degrees Celsius
- Open-source HBV model complete state data:
  - Soil water level in mm
  - Upper reservoir water level in mm

- Lower reservoir water level in mm
- Predicted inflow Unadjusted in m^3 per second
- Predicted inlfow adjusted reservoir field size in m^3 per second
- Snowpack contents in mm
- Liquid water amount in mm
- Effective precipitation(precipitation that contributes to runoff)
- Potential evapotranspiration
- Actual evapotranspiration

4.2.2 Preprocessing

4.2.2.1 Outlier analysis

An outlier analysis was performed on all of the measured and smoothed data, which is: temperature, precipitation, smoothed inflow and unsmoothed inflow. These timeseries was analyzed to find values deviating both from the global mean, but more importantly the values that were deviating from the surrounding local values. The amount of neighbouring values tocalculate the local mean was set to the same amount as was used by the domain expert who performed the smoothing operation on the unsmoothed inflow data where the neighbouring values were 4 preceding and 4 succeeding values (Domain expertise interview, May 20, 2020). The table 4.2.2.1 shows the amount of outliers more than 1-5 and 1-3 standard deviations away from the global and local mean.

Table 4.1: Outlier analysis

global/local	**std**	**Temp.**	**Precip.**	**Inflow not smoothed**	**Inflow Smoothed**
Deviation from global mean	1 std	44321	11360	14098	16432
	2 std	8487	5732	4394	7266
	3 std	794	3155	1582	3478
	4 std	45	1832	776	1802
	5 std	0	1108	375	1123
Deviation from mean of 8 neighbours	1 std	32280	17811	44470	42922
	2 std	933	7838	6774	1133
	3 std	9	3130	727	0

As is expected, the smoothed inflow values minimize the local deviations, as is the effect of smoothing. It is however interesting to see how much variation it is for the unsmoothed inflow, temperature and precipitation, where the local deviations provide a sense of quality measurement of the data. It is also interesting to see that the smoothing operation pull the data points away from the global mean, so that the smoothed data have more outliers when comparing points to a global mean, while having fewer outliers when comparing to the local mean.

It can be seen that the dataset contains values deviating vastly from the local mean, it is however difficult to without bias decide a definite threshold for correct or faulty values, and it is also difficult to decide upon a suitable replacement value of which to substitute deviating values without introducing bias to the research. Because of these reasons, the time series has been used with all original values in the experiments following below.

4.2.2.2 Negative Values

The smoothed and unsmoothed inflow time series was also analysed for negative inflow values, shown below. The inflow is expected to not be negative, however it is a common occurrence due to inaccuracies in water level readings, as inflow is calculated as the sum of reservoir water height changes and water spent in electricity generating turbines. A few centimeters of deviance as a result of wind will result in massive changes in reported water content when the surface area spans square kilometers.

It is a potential great bias factor introduced when choosing what values the negative readings can be replaced by. If replacing them by 0, the accumulated total water inflow per week, month, year and so forth will be greatly affected and altered, which could alter the accumulated state to a higher inflow value than is accurate by removing only negative values. Due to this and the fact that smoothing the inflow improved the negative readings by lowering them by 76%, negative reading has been unaltered, and smoothed inflow is used in conjunction with unsmoothed inflow as the LSTM-models input.

	Inflow not smoothed	**Inflow smoothed**
Negative inflow values	34160	8343

Table 4.2: Negative values

4.2.2.3 Data scaling

The input data is scaled using a min-max scaler which both fits and tranforms the data by the formula below so that it is in a range from 0 to 1.

All input data is transformed for the models which do not use previous inflow data as input, and for the models that use previous inflow as input, all data except for the previous inflow is scaled. This is due to the fact that we fit and transform input according to the target values, and when previous inflow also is used as input it still remains as target values as well. The reason the labels and output isn't scaled is because we want the model output to be a meaningful metric, directly applicable to power plant operation.

Min-max scaling uses the following formula to scale data and normalize the values, making all values fit a range between 0 and 1. x_{new} is the new value to substitute, X is the previous value, x_{min} is the lowest recorded value in the series and x_{max} is the highest recorded value:

$$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

The same scaling has been applied to all partitions of data: train, validation and test, as the model is trained and adapted to input data for all non-inflow related dimensions to be in the range from 0 to 1 to improve convergence during training.

4.2.3 Open source HBV model Tuning

The HBV model requires tuning to perform well, and the open source HBV model was implemented with the PEST framework (Doherty et al., 2018) and its method SCEUA_P, an acronym for "Shuffled Complex Evolution method developed at The University of Arizona" and implemented by the PEST framework. The framework optimizes the 10 tunable parameters that are described in 4.2.3.1 and does so by using the SCE(Shuffled Complex Evolution)-algorithm which is a generative evolution algorithm described by (Duan and Publisher, 1991), (Duan et al., 1992), (Duan et al., 1993) and (Duan et al., 1994) together with their runoff model, where both the model and optimization algorithm is designed specifically for inflow models.

It creates an initial set of parameters randomly by drawing samples from the defined parameter ranges and densities, uniform in our case, that are used to run the model and measures an error metric, which in our case is NSE which is modified to range from 0 to inf where smaller values are better. The parameter combinations, called complexes, are ranked by their error score where all are evolved by another algorithm CCE(Complex Competitive Evolution), described by the same authors.

CCE evolves the complexes by evaluating if they are within a boundary from their centroid value, a value that averages all other complexes than the one being altered by evolution. If the complex is outside thos boundary, a it is changed to a randomly generated point in the parameter space. If this complex performs better than the average of the other complexes, it is kept, otherwise the performance of parameters averaging the new random complex and the centroid is evaluated. If this peforms better, it is kept, otherwise it is randomly changed again and iteratively evaluated aswell.

After the completion of each iteration's CCE-algorithm the parent complexes are replaced by the evolved offspring. The evolved complexes are then shuffled and measured against a convergence criteria which is a user specified metric for how small the difference between each the order of shuffled complex and iteratively ordered and evolved again if the criteria is not met.

The method requires you to define an initial number of complexes to evaluate, which again affects default values for parameter sets per complex, parameter set per sub-complex, evolution steps before shuffling, minimum number of complexes and a random number seed. According to documentation the initial number of complexes should be between 2 and 20, and in most cases be set to 5 (Doherty, 2015). The remaining parameters were set to their default values, affected by the value of initial complexes, such that 20 sub complexes also were created when choosing 20 initial complexes. Both initial values of 5 and 20 were tested to perform the most exhaustive search available, though 20 could and did lead to slight overfitting when tuned on only the training partition of the data and evaluated on the validation partition. The parameter combination that was used was therefore with 5 initial parameter complexes.

The 11th configurable parameter, called "ca" or "Watershed Area" is manually set to the total area of the power plants inflow field.

4.2.3.1 Open source HBV model parameters

The open source HBV model has the following tunable parameters

Table 4.3: Open source HBV tunable parameters

Parameter	**Description**
d	tuning air temps effect on liquid water amount when air temp is over snow threshold(falls as rain)
fc	Tuning effective precip(dq) by dividing yesterday's soil water by fc and exponentiate this base to the power of beta
beta	Tuning effective precip(dq) by exponentiating the base of yesterday's soil water by fc to the power of beta
c	Tuning potential ET(evapotranspiration)
k0	Tuning inflow factor from S1, fast flow when water contents are above l0
l0	Minimum water amount in S1 cutoff for fast inflow from S1
k1	Tuning inflow factor from S1, slow continuous flow
k2	Tuning inflow factor from S2
kp	Tuning factor for flow down from S1 to S2
pwp	Permanent wilting point: the minimum amount of water content in the soil for water to be available to plants and agriculture, where water contents below the pwp will lead to the plants wilting (Kirham, 2005) (Rai et al., 2017)

The parameter search requires us to define an upper and lower band of parameter values. For the permanent wilting point the upper and lower band was set to cover the range that is described to apply for both sand, clay and loam of up to 250 (Brouwer et al., 1985). The remaining parameters had both the upper and lower bound widened from the ranges provided by water resources engineer (Craven, 2016) in order to increase the chance of finding an optimal configuration. Where the tunable parameters were used in both the open source and commercial HBV model, the parameter ranges for the open source was made sure to cover the values used for the commercial model, such as was the case for fc and $beta$.

To accommodate for the larger parameter space and the high amount of parameter combinations to test, the tuning was given 3 days of processing time on an Intel i5-6200U 2.30GHz cpu combined for the two parameter searches.

The parameter ranges that was used during tuning, and the discovered tuned values were:

Table 4.4: Open source HBV tunable parameter ranges and tuned values

Parameter	**Lower bound**	**Upper bound**	**Tuned value**
d	1.000	40.00	38.909
fc	50.00	500.0	53.07
beta	3.000E-1	7.600	0.3080
c	5000E-3	1.750	2.374E-02
k0	1.000E-2	2.000E-1	2.381E-02
l0	2.000	10.00	5.479
k1	1.000E-2	4.000E-1	0.3991
k2	1.000E-2	0.2000E-1	1.759E-02
kp	1.000E-3	5.000E-2	1.611E-03
pwp	25.00	250.0	249.8

4.2.4 LSTM Hyperparameter Constraint Analysis and Parameter Distribution

The possible hyperparameter space for a machine learning model is vast and in order to minimize the time needed for the hyperparameter search, an initial search was conducted for each model with hyperparameters that provided an upper bound for the parameter space that still overfitted the model.

Because we initially have little information about the parameter space and how the model will perform within it, the initial parameter distributions were set to be uniform distributions for all parameters. While the Tree-Parzen estimator(see below) improves the loss, the distributions are iteratively changed for each run, which contributes to faster convergence.

4.2.5 Hyperparameter Search

4.2.5.1 TPE - Tree-structured Parzen Estimator

Tree-structured Parzen Estimator(TPE) optimization is a technique that uses a surrogate model for tuning parameters and falls under the category "Sequential Model-Based Optimization"(SMBO). The original paper referenced by the HyperOpt library described below was written in 2011 by Bergstra et. al. (Bergstra et al., 2011)

The algorithm begins by sampling the response surface by random search to initialize the algorithm (Bergstra et al., 2011). Rather than modelling p(y—x) directly, TPE models p(x—y) and p(y). Then the observations are split into groups according to the best performing quantile, y^*, according to the defined loss, and the remaining worse performing quantiles. p(x—y) is as such defined by these two densities: (Bergstra et al., 2011)

$$p(x|y) = \begin{cases} l(x), & \text{if } y < y^* \\ g(x), & \text{if } y \geq y^* \end{cases} \tag{4.1}$$

The density l(x) is formed by observations x_i from past evaluations such that the loss is less than the threshold y^*. g(x) is the remaining worse performing quantile.

$$EI_{y^*} := \int_{-\infty}^{\infty} max(y^* - y, 0)p(y|x)dy \tag{4.2}$$

When using the TPE algorithm, it can be shown that maximizing EI(Expected Improvement, above) amounts to choosing x values that minimizes $g(x)/l(x)$. i.e. we would like points with a high likelihood of being in $l(x)$ and a low likelihood of being in $g(x)$. At each iteration, the algorithm draws several samples from $l(x)$, the wanted density, evaluates them in terms of $g(x)/l(x)$ and returns the candidate with highest EI^1 (Bergstra et al., 2011) before the next iteration begins.

A benefit of using TPE for Bayesian Optimization is the fact that TPE easily can optimize both mixed continuous and discrete spaces (Falkner et al., 2018), which is applicable for this model as e.g. the amount of LSTM layers is a discrete space while dropout probability on the other hand is a continuous space.

4.2.5.2 Early stopping

In order to be able to test a higher amount of hyperparameter settings, early stopping was implemented and set to stop the training if the validation loss was increasing for 2 consecutive epochs. Then after the hyperparameter search, the best hyperparameter settings were applied and the network was trained for a higher amount of epochs, before selecting the network weights that provided the lowest validation loss.

4.2.5.3 Hyperparameter Space

The hyperparameters that are being tuned are the following:

- Learning rate
- In window, the amount of timesteps that are being used as input to the model
- Batch size

- Dropout probability
- Clipping coefficient, for gradient clipping, described in section 3.2
- Hidden dimensions, the width of the network
- Amount of stacked LSTM layers

Frozen parameters, description of the parameter and the reason behind the freeze:

- Out window: the amount of timesteps forward to predict the reservoir inflow
  - Frozen at 40 hours due to the specific time step requirement for power plant operators in the joint Nordic power market of how many hours ahead they need to report their power production, in the Intra day market described in chapter 2.
- Input-output latency: The delay associated with receiving the measurement data and HBV-prediction data, resulting in the necessity to use input data from x amount of timesteps backwards to predict the future out window of inflow.
  - Frozen at 1 hour due to the specific time delay of up to 1 hour to receive the HBV model predictions from measurements are being made. This ensures that no data leakage is experienced due to using data that would not be available in the moment of prediction.

4.2.5.4 HyperOpt - Hyperparameter Optimization Library & Parameter Configuration Space

The algorithm and configurations described above were implemented for the experiment using the optimization library HyperOpt (Bergstra et al., 2013). The library does in addition to using the TPE algorithm described in 2.1 also allow and require the definition of a parameter configuration space, which is the prior probability distribution used for the defined parameter.

As the parameter space is complex and it is difficult to predict the models behaviour prior to optimization, all continuous parameters were initialized with a uniform prior distribution to not introduce any bias.

4.2.6 Reproducibility

Reproducibility is of high importance in order to provide results that are trustworthy and reliable. In order to enhance the reproducibility of the results in this thesis there added some critical configurations to the implemented pipeline

4.2.6.1 Random Seed

In order to make computation more reproducible, the random seed both for pytorch and cuda calculations have been manually fixed. In order to remove bias even more, the seed was frozen to a value selected by the random integer generator of random.org within 0 and 1 000 000 000.

The same value was also used for Numpy's random seed as any libraries that rely on Numpy are dependent on a fixed random seed for Numpy to obtain reproducibility. An example of a library that uses numpys random seed is Hyperopt, that is used for the hyperparameter search.

4.2.6.2 CuDNN backend

The experiments were runned on an NVIDIA Tesla P100 GPU with 16gb of memory. This gpu uses cuda for network tuning and in order to ensure more reproducible results, two configuration flags has been set according to pytorch documentation (Torch Contributors 2019, 2019):

- Torch.backends.cudnn.deterministic = True
- Torch.backends.cudnn.benchmark = False

Where the CuDNN benchmark configuration according to pytorch moderators allows you to enable the inbuilt auto-tuner to find the best algorithm to use for your hardware. It usually leads to faster runtime (Massa and AlbanD, 2017). By locking the random seed in accordance with configuring the CuDNN backend to deterministic, we are able to ensure deterministic computation within the same pytorch releases on the same GPU.

4.2.6.3 Limitations to the effect of random seed

The random seed in pytorch, does not guarantee reproducible results across software releases, individual commits or different platforms, (Torch Contributors 2019, 2019). Running the same pytorch code on different GPU's will also be able to yield different results, confirmed by pytorch's own moderators (AlbanD, 2019). This case would be the same if Tensorflow was used rather than pytorch.

4.2.6.4 Weight initialization

Weight initialization can affect the outcome of the trained model heavily. In order to ensure that different weight initialization is not the reason behind different model performance the cell state and hidden state of the LSTM layers are therefore initialized as zero in all experiments.

4.3 Experimental Results

4.3.1 NSE - Nash-Sutcliffe Efficiency for Entire Forecast

The figures show the mean of each forecast's NSE and their standard deviation, with the range $(\infty, 1]$.

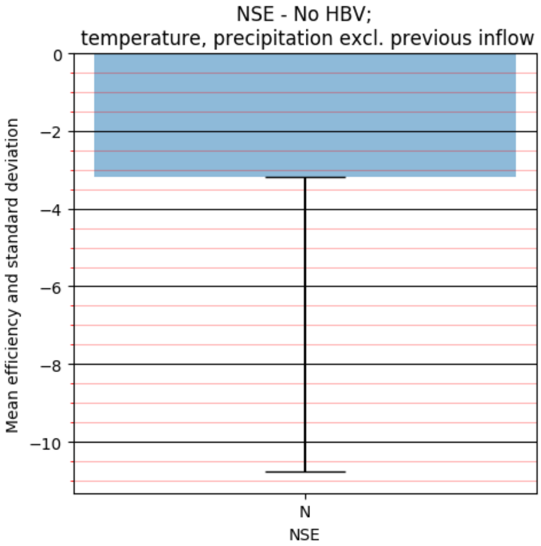

Figure 4.1: Model using only an LSTM net with precipitation and temperature as input

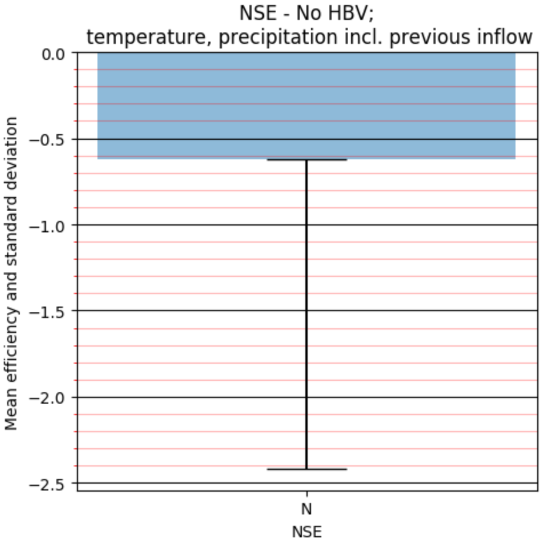

Figure 4.2: Model using only an LSTM net with precipitation, temperature and previous inflow as input

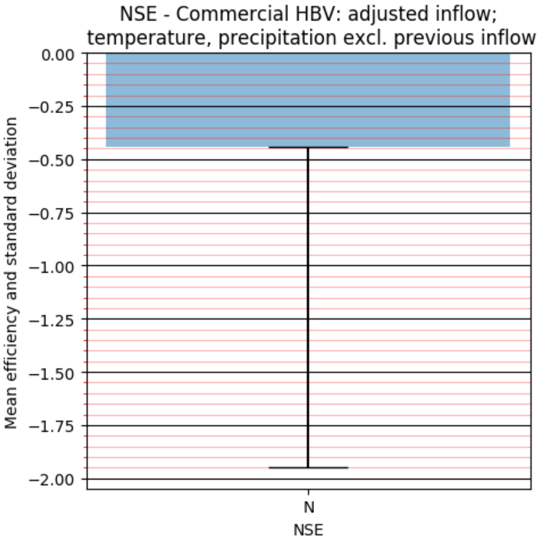

Figure 4.3: NSE for model using the commercial HBV's predicted inflow together with precipitation and temperature as input to an LSTM net

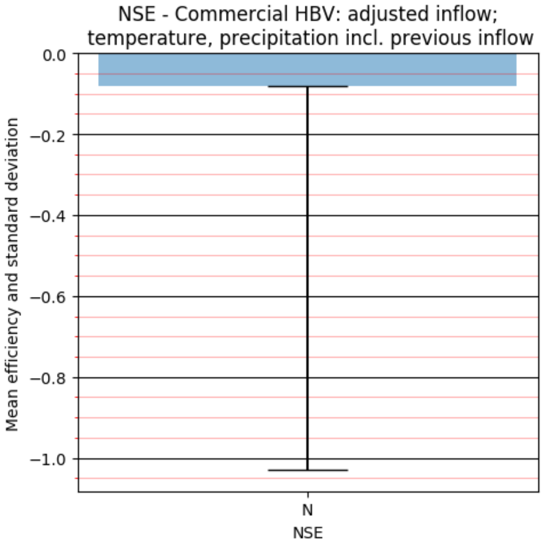

Figure 4.4: NSE for model using the commercial HBV's predicted inflow together with precipitation, temperature and previous inflow as input to an LSTM net

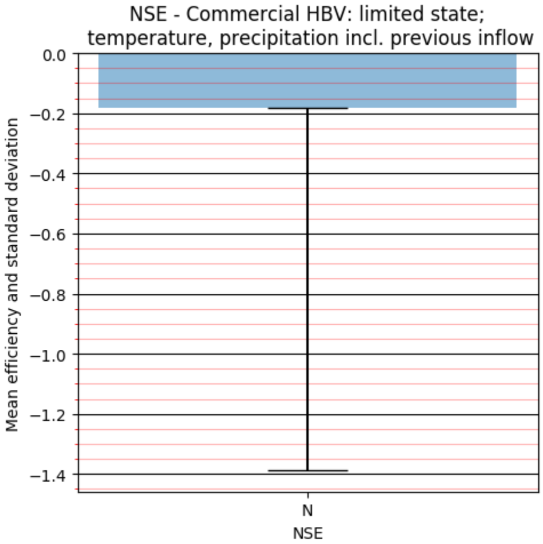

Figure 4.5: NSE for model using the commercial HBV's limited internal state together with precipitation, temperature and previous inflow as input to an LSTM net

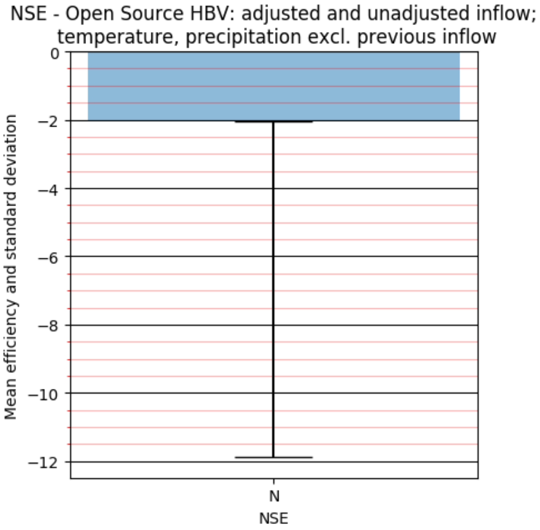

Figure 4.6: NSE for model using the open source HBV's predicted inflow together with precipitation and temperature as input to an LSTM net

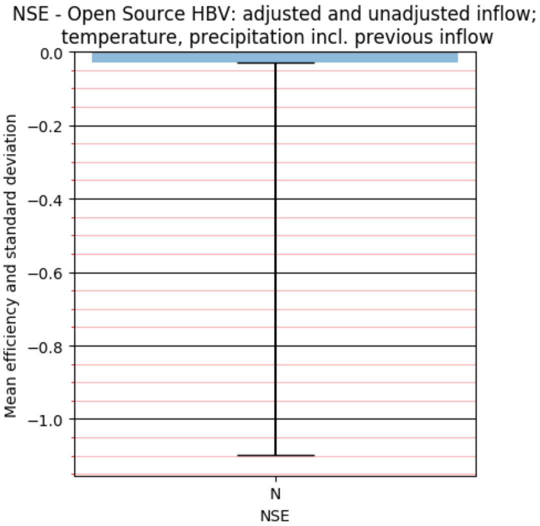

Figure 4.7: NSE for model using the open source HBV's predicted inflow together with precipitation, temperature and previous inflow as input to an LSTM net

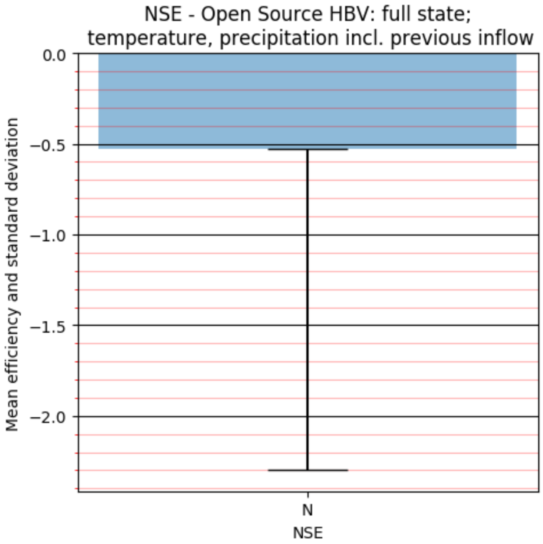

Figure 4.8: NSE for model using the open source HBV's full internal state together with precipitation, temperature and previous inflow as input to an LSTM net

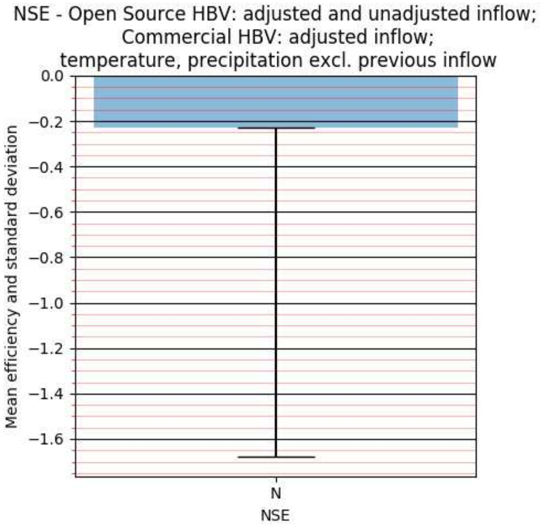

Figure 4.9: NSE for model using both the commercial HBV's- and open source HBV's predicted inflow together with precipitation and temperature as input to an LSTM net

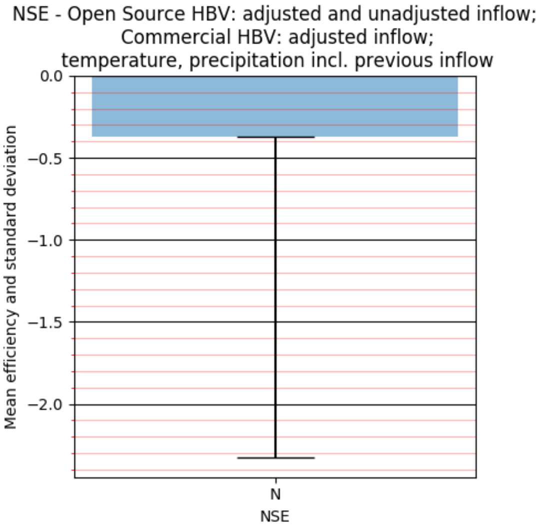

Figure 4.10: NSE for model using both the commercial HBV's- and open source HBV's predicted inflow together with precipitation, temperature and previous inflow as input to an LSTM net

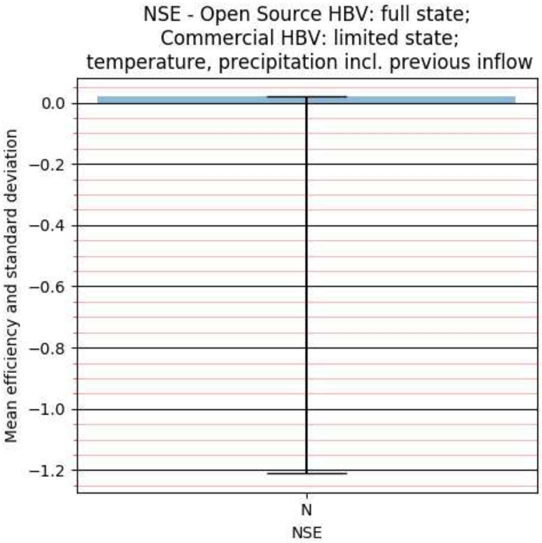

Figure 4.11: NSE for model using both the commercial HBV's limited internal state and open source HBV's full internal state together with precipitation, temperature and previous inflow as input to an LSTM net

4.3.2 MSE - Models' Mean Square Error and Standard Deviation per Hour

The graphs here show mean MSE per hour as blue bars, and standard deviation as black capped lines.

General trends to notice:

- For all but one model, both the mean error and standard deviation is strictly increasing with time in the forecast horizon
- There is a large bump in mean error from hour 36 to hour 37. Remember the input and target data structure where due to the smoothing, the last 4 data points of both the input and targets are unsmoothed.
- All eleven models, hybrids and LSTM models perform better than the HBV models, which are presented below in this section.

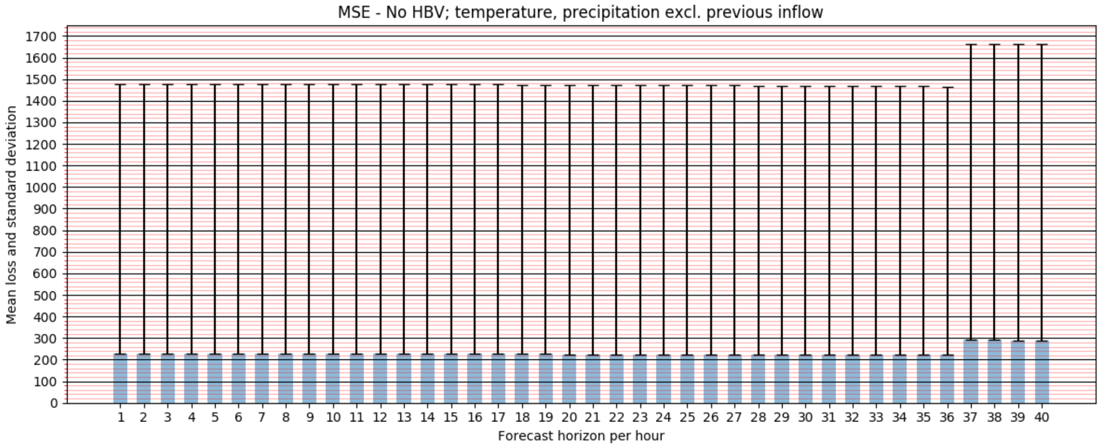

Figure 4.12: MSE for model using only an LSTM net with precipitation and temperature as input

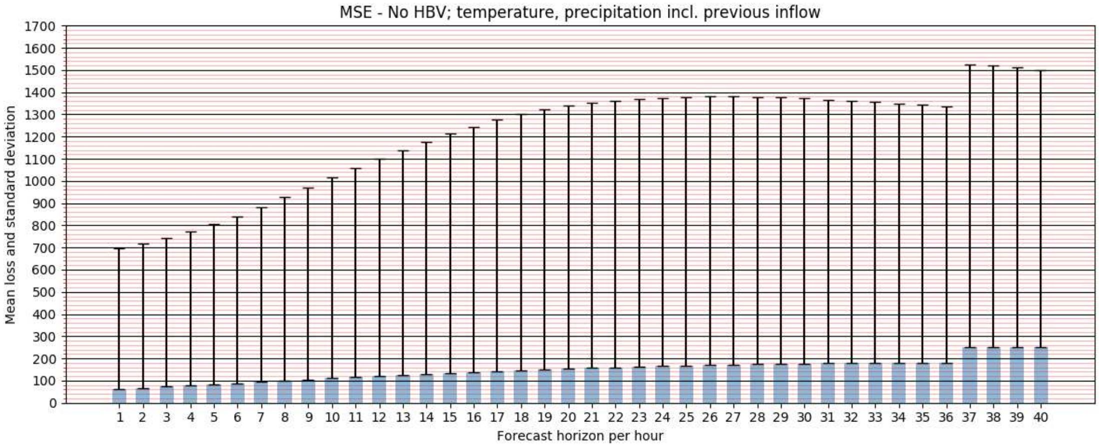

Figure 4.13: MSE for model using only an LSTM net with precipitation, temperature and previous inflow as input

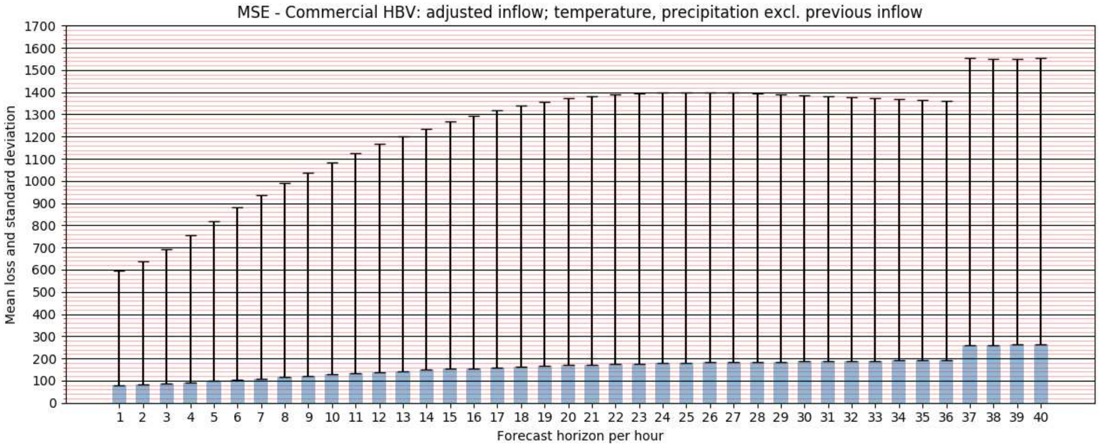

Figure 4.14: MSE for model using the commercial HBV's predicted inflow together with precipitation and temperature as input to an LSTM net

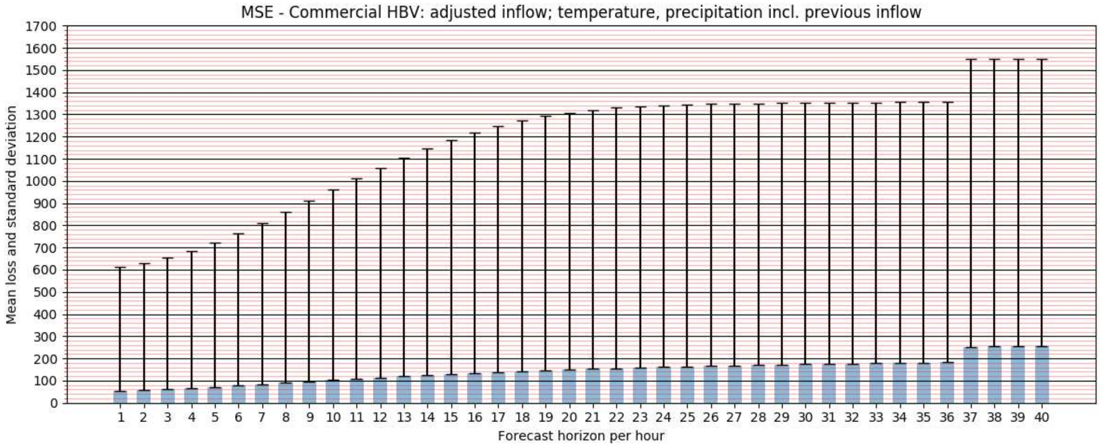

Figure 4.15: MSE for model using the commercial HBV's predicted inflow together with precipitation, temperature and previous inflow as input to an LSTM net

Notice the fact that the standard deviation decays from hour 26 to 36. The mean error does however strictly increase with time.

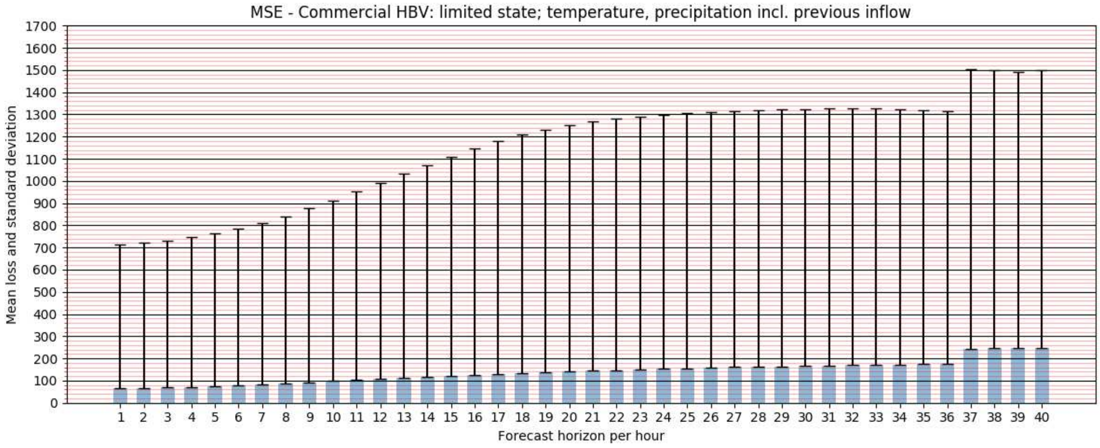

Figure 4.16: MSE for model using the commercial HBV's limited internal state together with precipitation, temperature and previous inflow as input to an LSTM net

Notice how much longer into the forecast horizon the standard deviation is kept to a low value compared to the model without access to previous inflow as well as how the mean error decreased. However, it can also be seen that for the first 4 hours, the standard deviation is higher for this model with access to previous inflow compared to the one without.

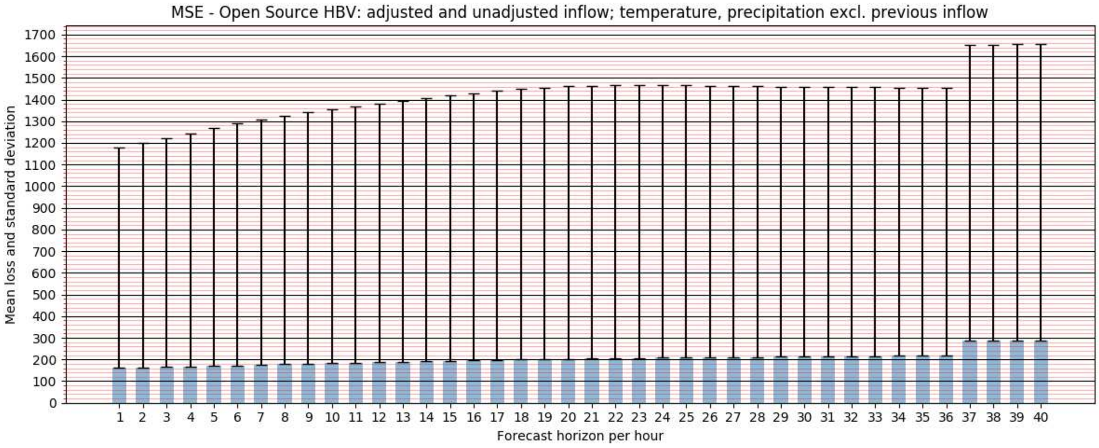

Figure 4.17: MSE for model using the open source HBV's predicted inflow together with precipitation and temperature as input to an LSTM net

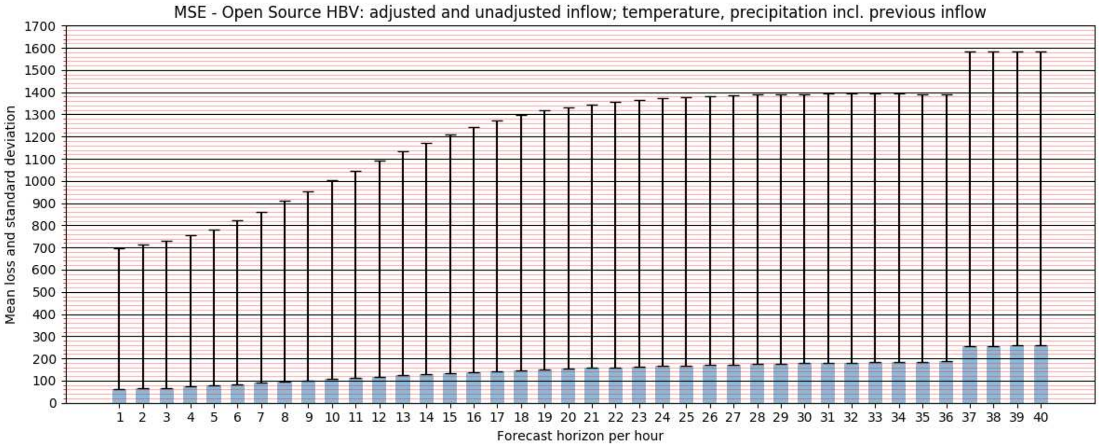

Figure 4.18: MSE for model using the open source HBV's predicted inflow together with precipitation, temperature and previous inflow as input to an LSTM net

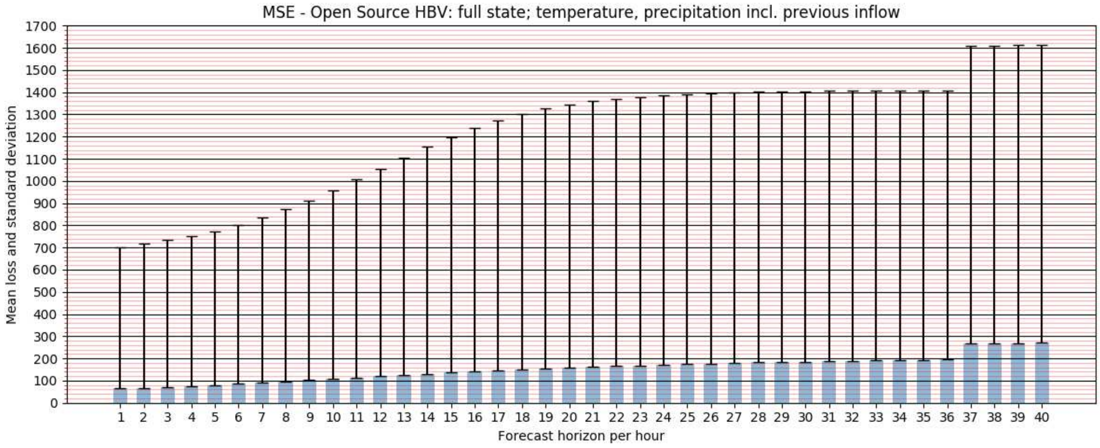

Figure 4.19: MSE for model using the open source HBV's full internal state together with precipitation, temperature and previous inflow as input to an LSTM net

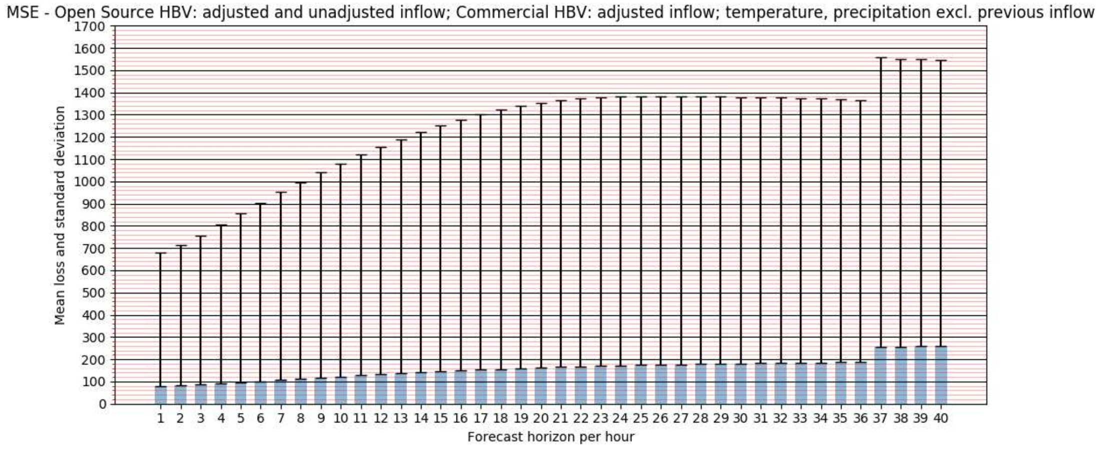

Figure 4.20: MSE for model using both the commercial HBV's- and open source HBV's predicted inflow together with precipitation and temperature as input to an LSTM net

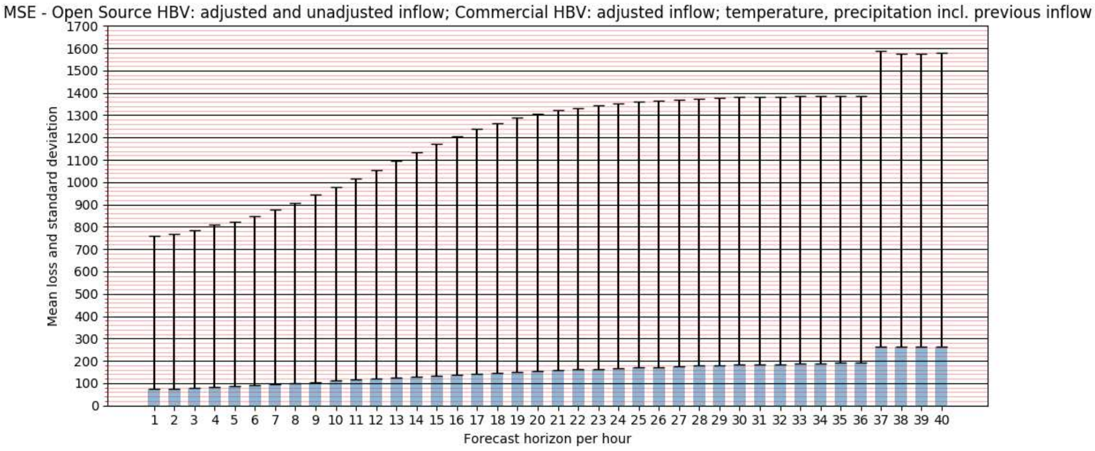

Figure 4.21: MSE for model using both the commercial HBV's- and open source HBV's predicted inflow together with precipitation, temperature and previous inflow as input to an LSTM net

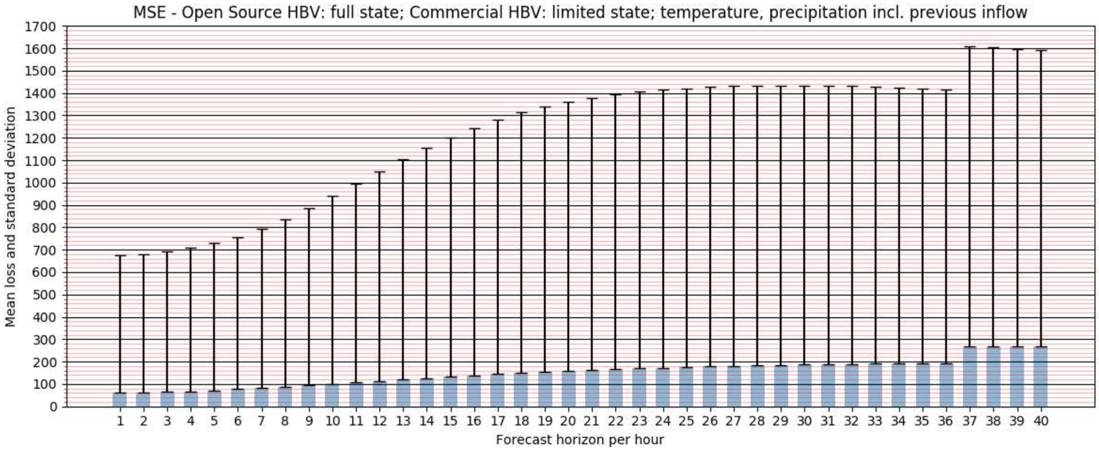

Figure 4.22: MSE for model using both the commercial HBV's limited internal state and open source HBV's full internal state together with precipitation, temperature and previous inflow as input to an LSTM net

4.3.2.1 MAE and Smooth L1 Loss

Error metrics for each model is also available measured in MAE and Smooth L1 Loss, and is located in the appendix in A.1. These metrics show the error for the exact same models, that are all trained using MSE as their loss function, and provide more perspective about the model's performances.

4.3.3 Benchmark Models' Efficiency and Theoretical Hourly Error

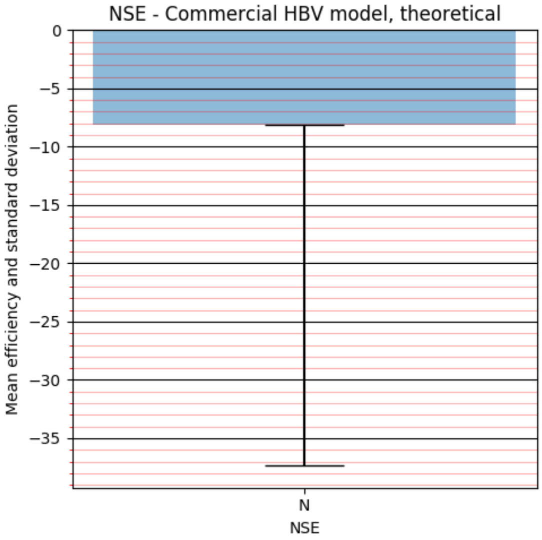

Figure 4.23: NSE for the commercial HBV model in a theoretical scenario where it makes 40 hour forecasts based on measured data

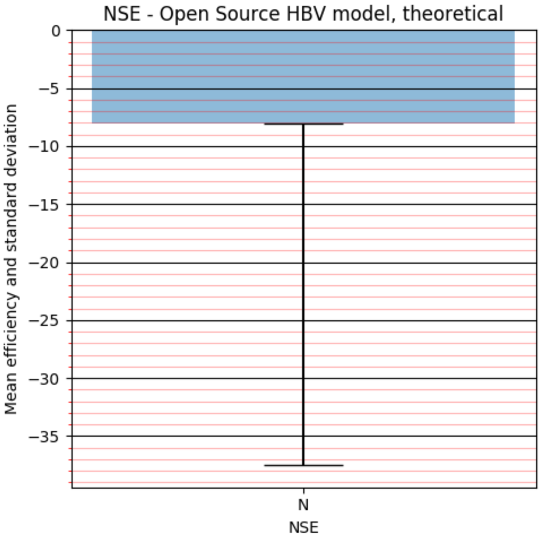

Figure 4.24: NSE for the open source HBV model in a theoretical scenario where it makes 40 hour forecasts based on measured data

The graphs below show what the theoretical error would be if the HBV models were to make 40 hour forecasts based on measured data. As they don't provide such forecasts, one would need to use weather forecasts as input to the models, which has a higher uncertainty and error rate than measured data. Because of this, the errors reported here are best case scenarios for the HBV models where 1 hour forecasts only based on measured data are stacked, while in reality, the error would on average be higher for all but the first timestep. As the forecast here is 40 stacked 1 hour forecasts, the error per hour is approximately uniform between the 36 smoothed values and between the 4 unsmoothed values. The only difference between teh approximate uniform values and true uniform values are the 40 first and last data points as the timeseries' are not padded.

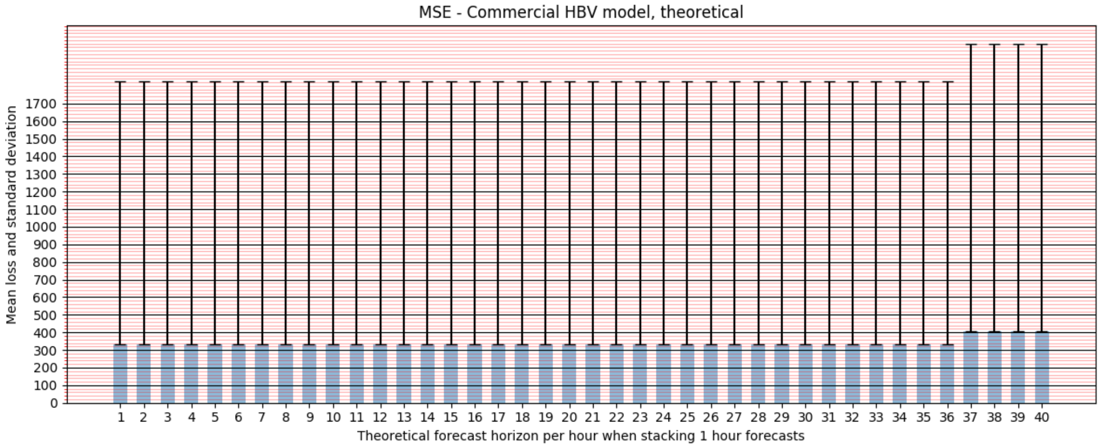

Figure 4.25: MSE for the commercial HBV model in a theoretical scenario where it makes 40 hour forecasts based on measured data

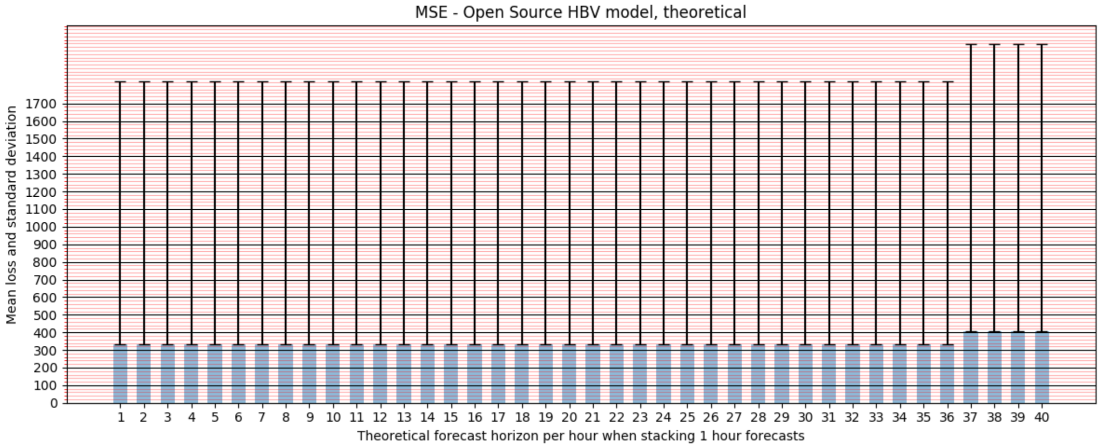

Figure 4.26: MSE for the commercial HBV model in a theoretical scenario where it makes 40 hour forecasts based on measured data

4.3.4 Model Mean and Standard Deviation

Abbreviations:

os HVB = open source HBV

c HVB = commercial HBV

no HVB = only LSTM net

both HVB = both os HBV and c HBV

Table 4.5: Benchmark Model Performance, NSE efficiency and MSE error

Metric	**Model Name**	**Mean**	**Std**
NSE	c HBV	-8.161738	29.219992
NSE	os HBV	-8.093324	29.447031
MSE	c HBV	341.472595	1507.546875
MSE	os HBV	339.355225	1508.660400

Notice in table 4.3.4 that the two HBV models perform almost identically, even though they are of vastly different complexity.

Table 4.6: Model Mean Efficiency and Standard Deviation by NSE

Metric	**Model Name**	**Access to Previous Inflow?**	**Access to state?**	**Mean**	**Std**
NSE	no HBV	N	N	-2.277737	1.948010
NSE	no HBV	Y	N	-0.621070	0.586458
NSE	os HBV	N	N	-2.355785	2.960371
NSE	os HBV	Y	N	-0.015438	0.164231
NSE	os HBV	Y	Y	-0.5276523	0.408788
NSE	c HBV	N	N	-0.368335	0.363243
NSE	c HBV	Y	N	-0.173352	0.196634
NSE	c HBV	Y	Y	-0.049486	0.127185
NSE	both HBV	N	N	-0.429513	0.370954
NSE	both HBV	Y	N	-0.252391	0.232111
NSE	both HBV	Y	Y	0.028954	0.096195

Table 4.7: Model Ranking Based on Mean NSE

Ranking	Model Name	Access to Previous Inflow?	Access to state?	Mean
#1	both HBV	Y	Y	0.028954
#2	os HBV	Y	N	-0.015438
#3	c HBV	Y	Y	-0.049486
#4	c HBV	Y	N	-0.173352
#5	both HBV	Y	N	-0.252391
#6	c HBV	N	N	-0.368335
#7	both HBV	N	N	-0.429513
#8	os HBV	Y	Y	-0.5276523
#9	no HBV	Y	N	-0.621070
#10	no HBV	N	N	-2.277737
#11	os HBV	N	N	-2.355785

Table 4.8: Model Mean Error and Standard Deviation by MSE

Metric	Model Name	Access to Previous Inflow?	Access to state?	Mean	Std
MSE	no HBV	N	N	228.416988	626.435014
MSE	no HBV	Y	N	155.350956	541.777600
MSE	os HBV	N	N	228.152865	626.574387
MSE	os HBV	Y	N	145.194504	519.628831
MSE	os HBV	Y	Y	150.131950	526.282740
MSE	c HBV	N	N	154.310330	525.421053
MSE	c HBV	Y	N	139.346881	511.562233
MSE	c HBV	Y	Y	139.712514	515.380863
MSE	both HBV	N	N	165.523761	542.951148
MSE	both HBV	Y	N	148.771959	519.132510
MSE	both HBV	Y	Y	168.525514	564.390794

Table 4.9: Model Ranking Based on Mean MSE

Ranking	**Model Name**	**Access to Previous Inflow?**	**Access to state?**	**Mean**
#1	c HBV	Y	N	139.346881
#2	c HBV	Y	Y	139.712514
#3	os HBV	Y	N	145.194504
#4	both HBV	Y	N	148.771959
#5	os HBV	Y	Y	150.131950
#6	c HBV	N	N	154.310330
#7	no HBV	Y	N	155.350956
#8	both HBV	N	N	165.523761
#9	both HBV	Y	Y	168.525514
#10	os HBV	N	N	228.152865
#11	no HBV	N	N	228.416988

Table 4.10: Model Mean MSE Ranking for Hour 1

Ranking	**Model Name**	**Access to Previous Inflow?**	**Access to state?**	**Mean**
#1	c HBV	Y	N	53.95747
#2	both HBV	Y	Y	59.860035
#3	os HBV	Y	N	60.18853
#4	no HBV	Y	N	63.598618
#5	os HBV	Y	Y	64.541534
#6	c HBV	Y	Y	64.815865
#7	both HBV	Y	N	72.46247
#8	both HBV	N	N	79.609604
#9	c HBV	N	N	79.71544
#10	os HBV	N	N	161.90985
#11	no HBV	N	N	227.43687

The results in table 4.3.4 are relevant for a TSO in the balance market. Notice that access to inner state is beneficial for the hybrid with both HBV models, but not for the commercial or open source hybrids. Notice also the impressive performance of the LSTM net with no HBV model, but with access to previous inlfow.

Table 4.11: Model Mean Error and standard deviation for hours 2-5

Ranking	**Model Name**	**Access to Previous Inflow?**	**Access to state?**	**Mean**
#1	c HBV	Y	N	64.25658375
#2	both HBV	Y	Y	66.598165
#3	c HBV	Y	Y	70.57612875
#4	os HBV	Y	N	70.715311
#5	os HBV	Y	Y	74.00548225
#6	no HBV	Y	N	74.9794625
#7	both HBV	Y	N	80.56453
#8	both HBV	N	N	88.8856565
#9	c HBV	N	N	90.20724625
#10	os HBV	N	N	167.01682
#11	no HBV	N	N	227.1588125

The results in table 4.3.4 are especially relevant for the intra-day market. Notice that access to inner state is beneficial for the hybrid with both HBV models, but not for the commercial or open source hybrids.

Table 4.12: Model Mean Error and standard deviation for hours 12-36

Ranking	**Model Name**	**Access to Previous Inflow?**	**Access to state?**	**Mean**
#1	c HBV	Y	Y	150.68379
#2	c HBV	Y	N	158.2374521
#3	no HBV	Y	N	162.1238992
#4	os HBV	Y	N	162.8817623
#5	both HBV	Y	N	165.1701258
#6	both HBV	Y	Y	168.0984185
#7	os HBV	Y	Y	168.4151928
#8	both HBV	N	N	169.0070413
#9	c HBV	N	N	175.1553508
#10	os HBV	N	N	206.7631413
#11	no HBV	N	N	224.8984004

These results in table 4.3.4 are especially relevant for the day-ahead market. Notice that state access is beneficial for the commercial HBV hybrid, but not the open source HBV hybrid or the hybrid with both HBV models. Notice also the impressive performance of the LSTM net with no HBV model, but with access to previous inlfow.

Chapter 5

Evaluation and Conclusion

5.1 Evaluation

5.1.1 Hyperparameter Tuning

The dimensionality of the input data to the LSTM is high for the most complex models in these experiments, such as using both the limited state of the commercial HBV, the full state of the open source HBV and the measured temperature, precipitation and previous inflow. Because of this factor, the more complex models with access to full state- and limited state variables was seen to converge at a slower rate during the Bayesian hyperparameter tuning. It is likely that the more complex models would improve more relative to the simpler models with no access to state variables and potentially no access to previous inflow data either. The amount of hyperparameter combinations that were tested was 100 per model. When training the models with the selected hyperparameters, the max amount of epochs was increased from 25 to 100 and the number of epochs with a non-increasing validation error to stop the training was increased from 2 to 20, in order to increase the chance of good performance.

An additional test run was performed on the most complex model, with access to the states of both HBV models, where 300 different hyperparameters were tested and compared to the one tuned only 100 parameter combinations and seen to improve as described in the result tables.

When working with complex models it is however important to remember the findings of Makridakis & Hibon when reviewing the M-3 competition in year

2000 (Makridakis and Hibon, 2000) they presented four conclusions of the M-competition, where one of them was: “Statistically sophisticated or complex methods do not necessarily produce more accurate forecasts than simpler ones” (Makridakis and Hibon, 2000) Such that a model with a higher time till convergence does not always perform better than a simpler model that takes a shorter time to train.

This can also be seen in these experiments where the bayesian hyperparameter tuning for all models tuned the LSTM layer depth of the network down to low numbers around 2. This is likely due to the massive amount of data needed to train really deep machine learning archictures, and due to and vanishing gradients for the deeper configurations. It is however natural that shallower nets prevailed as when comparing to very deep machine learning models 155304 data points is not a large dataset.

5.1.2 Hardware and Time Limitations

The thesis' experiments are characterized by being restrained by hardware and time constraints, specifically about the time available for hyperparameter tuning for each model. This means that conclusions drawn must be seen in conjunction with these limitations. The performance differences between each model could have been different with a longer tuning period. However, we know that each model's performance is lower bound to the performance seen in these experiments, and can perform the same or better with further tuning.

5.1.3 Forecast Horizon

The forecast horizon decides the models' priorities as they during training are measured against an error metric across the entire horizon. Therefore the results for 1 hour, 2-5 hours and 12-36 hours are using models that are not specifically tuned and trained for the best performance for that horizon. The results are however, interesting in order to evaluate subsets of the behaviour of the models.

5.2 Discussion

Interpreting Results When interpreting the results, they are discussed in a matter where if the data is correct and the tuning is acceptable, the discussion topics and conclusions following can be drawn.

5.2.1 Complex Model's Tuning Time Required for Convergence

The amount of hyperparameter combinations attempted before new best validation losses were recorded more rarely was very different from model to model, where the simpler models were seen to converge much more rapidly than the models with access to state and previous inflow. This suggests that the complex models, especially the models with access to the full and limited state variables would improve more than the models without access to this data if given more time for tuning parameters.

5.2.2 Time Step Loss Differences

As can be seen for all the models, there is a significant difference between the loss of the 36 first hours and the last 4. This big difference is likely caused by the label-structure where the smoothing operation done on the inflow where 8 neighbouring points, 4 earlier and 4 later timesteps, are used to smooth out the inflow values. Because we in the event of predicting 40 hours into the future do not have 4 later timesteps to use for smoothing for the 4 last hours, the labels created therefore consists of 36 smoothed values and 4 unsmoothed values where the smoothed values have a lower variance and the unsmoothed values have a higher variance. Both its higher variance and the fact that these datapoints only consists of 5% of the forecast horizon, making the model adapt much less to these datapoints than the smoothed ones makes it natural that the difference between the 36 first and 4 last time steps are larger.

5.2.3 Performance comparison of the commercial and open source HBV model

The comparably good performance of the open source HBV model, when both have access to previous inflow, which is free and simpler than it's commercial counterpart is likely to partly be a result of the tuning of the models. The models need thorough and regular tuning in order to keep performance high according to hydrologist Frode Vassenden (Domain expertise interview, October 24, 2019). The open source HBV was tuned with the entire train partition of the data, while the commercial model according is not tuned as recently. Both models also require data about monthly average evaporation for the field, where the two models used the exact same values. Another reason is the highly beneficial effect the access to data about previous inflow has on the models, so that the dependence on the HBV models weakens slightly. This is also made very apparent by the great gap in performance for the LSTM net without HBV models with and without access to previous inflow.

However, as can be seen for the case of when the models don't have access to previous inflow, the commercial model outperforms the open source model drastically for the first few hours both in mean loss and standard deviation. With longer forecast horizons especially the standard deviation evens out, though the error mean is still much better for the commercial model.

5.2.4 Effect of having access to previous inflow data

The models that utilizes input from the open source HBV model and the commercial HBV model on their own were tested both with and without access to previous inflow data and its effect can clearly be seen on the graphs for hourly prediction error, where the first hours have a significantly lower error than for the models without access to previous inflow. The error for the first 10 hours for the models without access to previous inflow was 3 times as large while the error for the 10 last hours was only 25% as large. All hours in the forecast horizon benefited from access to previous inflow and we see that the improvement from access to previous inflow was much greater than the improvement from also having access to state variables in addition to inflow. The only part of the prediction that did not improve with access to previous inflow was that the relative error increase from hours 30-36 to 37-40 with the unsmoothed inflow values was larger for the models with access to previous inflow.

A possible explanation for this was that the inflow input data consisted of smoothed values up until the last 4 values, such as for the target values, so the model did only have access to 0.009% of the input window size of 222 as unsmoothed values per time step in the case for the open source HBV hybrid.

This result emphasises the importance of having low latency systems for inflow measurements, such as for this inflow field where access to reservoir data is available within one hour. Although it also seems that the effect of access to previous inflow diminishes severely around 24 to 26 hours into the future in these experiments which means 25 to 27 hours between the latest data input to the forecasted values due to the 1 hour latency. This makes it likely that the relative difference between these two models would be much lower for the case where spring floods should be predicted, as the forecast horizon increases from hours to days.

5.2.5 Relevance for Future Behaviour

An important question when evaluating the performance for a model using historical data to predict the future behaviour of a system is whether it will perform well in the future. This is a difficult question to answer as the real life system can and most likely will change over time. However, by structuring the experiments so that the test partition is the 5% most recent data and only the first 95% of the timeseries was used for training and validation, it is likely that the performance experienced in these experiments is comparable to future behaviour as it was tested on the most recent data.

5.2.6 Effect of Access to HBV Model's Inner State

The effect of having access to the state of the HBV models had a positive impact on mean error of predictions, where the improvement for the commercial HBV model improved the most of the two from access to this data. Possible reasons for this are the fact that access was given to the complete state in comparison to only a limited part of the state for the commercial, which then would behave less predictably than the open source one. Another possible reason is that the commercial model is more advanced and that improving the models become exponentially more difficult the better they perform. The effect of access to subsets of the inner states of the models is a relevant research problem for future work, described in section 5.5.2.

5.2.7 Answering the Research Questions

As we see in these experiments, and discussed in sections 5.2.6 and 5.2.4 we can answer the research questions presented in section 1.3 in the light of the limitations presented in the evaluation section, 5.1. Once again, the research questions and their answers are set into the context of a climate with yearly snowfall, which makes the prediction problem more complex to solve.

Research question 1 *Will a machine learning model that has access to previous inflow data and the state of an HBV-model be able to predict inflow more accurately than the HBV-model?*

- Yes it will, for both HBV models: Although, in conjunction with access to data about previous inflow, inner state access was not beneficial for the models performance for forecast horizons under 12 hours with this restricted hyperparameter search, likely due to longer convergence time for the more complex models.

Research question 1.1 *Will a machine learning model that does not have access to the state of an HBV-model be able to predict inflow more accurately than the HBV-model?*

- Yes it will, for both HBV models. This is especially true for short forecast horizons of up to 12 hours, and also for horizons of 25-27 hours from inflow observation when having access to data about previous inflow.

Research question 2 *Will a machine learning model that has access to previous inflow be able to predict inflow more accurately than the HBV-model?*

- Yes it will, as is shown for the full 40 hour forecast horizon in these experiments, even for the LSTM network with no HBV model input, and especially for short term predictions of 25-27 hours from the time of measurement or less

5.3 Conclusion

According to the results shown in this thesis, the best performing model overall during the forecast horizon is the hybrid model that has access to previous inflow, the full state of the open source HBV and the limited state of the commercial HBV. The model received a positive mean Nash-Sutcliffe efficiency and the lowest standard deviation of NSE as well.

The other well performing models according to MSE are:

Open source HBV hybrid with access to previous inflow, but no state This was the second best performing model by NSE over the full forecast horizon, and even outperformed the commercial HBV model both with and without access to limited state.

- A conclusion that should be drawn from these results is that the commercial HBV model should be re-tuned and re-tested. If the performance does not improve significantly, the open source HBV model should be used in stead of- or in conjunction with the commercial HBV model.

Commercial HBV hybrid with access to previous inflow, but no state This model was the best performing model according to MSE for both hour 1/balance market, and for hour 2-5/intra-day market.

Commercial HBV hybrid with access to previous inflow and state This model was the best performing model according to MSE for the 12-36/day-ahead market.

It has also been clearly shown that the performance of all the hybrid models was greatly improved when compared to their respective benchmarks, the commercial HBV and the open source HBV, and also when compared to a standalone LSTM network without an HBV model and no acces to previous inflow.

The most significant change for model performance was having access to previous inflow, especially for short forecasting horizons, but also across the entire horizon.

5.4 Contributions

The contributions of this thesis are:

- Creation and evaluation of sequence to sequence models that uses measured data as input in order to forecast inflow
- Creation and evaluation of 9 Hybrid models that forecast inflow.
- Evaluation of the importance of access to previous inflow and low latency for inflow measurements.
- Creation of- and running experiments with target values from two different dimensions of the dataset to be able to smooth data with a large smoothing width and still avoid leakage of the targets.
- Evaluating LSTM networks performance for inflow forecasting alone and as sub-models in a hybrid architecture.
- Tuning and evaluation of a free and open source alternative to a commercial, complex and costly HBV model.

5.5 Future Work

5.5.1 Error metric and reservoir state

Error metric that analyses deviations in accordance with reservoir water level. This is interesting because

- When the reservoir level is low and the model predicts a higher inflow than what happens, the outcome is more negative than when the model predicts a too low inflow
- When the reservoir level is high and the model predicts a lower inflow than what happens, the outcome is more negative than when the model predicts a too high inflow

5.5.2 Effect of access to specific subsets of the HBV states

A relevant research problem for future work is evaluating the effect on performance of access to subsets of the HBV model's states. This is interesting as it could provide details to what kind of behaviour is extra difficult to predict for machine learning models, compared to numerical models created by domain experts.

5.5.2.1 Effect on performance of access to snowpack data for flood predictions

Of all the variables accessible in the full and limited states of the HBV models, one state variable is especially connected to a difficult subsection of inflow prediction: Flood prediction during the spring. According to hydrologist Frode Vassenden (Domain expertise interview, October 24, 2019), it is difficult to tune the models to maximise the resemblance of real life snow pack contents and also the snowpack's effect on inflow.

5.5.3 Deep LSTM-layered hybrid networks

As the best performing networks were shallow networks with few hidden dimensions and with few LSTM layers, it should be experimented with techniques that could improve deep RNN's performance such as minimizing the vanishing gradient problem in recurrent neural networks even more than the LSTM architecture does on its own.

The dimensionality of the input data to the LSTM is also high for the most complex models in these experiments, such as using both the limited state of the c_HBV, the full state of the os_HBV and the measured temperature, precipitation and previous inflow. These models could potentially benefit from performing a tree search to weight the input data according to importance

Bibliography

Abdelrahim, M., Merlosy, C., Wang, T., 2016. Hybrid Machine Learning Approaches: A Method to Improve Expected Output of Semi-structured Sequential Data, in: Proceedings - 2016 IEEE 10th International Conference on Semantic Computing, ICSC 2016, Institute of Electrical and Electronics Engineers Inc.. pp. 342–345. doi:10.1109/ICSC.2016.72.

Abunama, T., Othman, F., Ansari, M., El-Shafie, A., 2019. Leachate generation rate modeling using artificial intelligence algorithms aided by input optimization method for an MSW landfill. Environmental Science and Pollution Research 26, 3368–3381. doi:10.1007/s11356-018-3749-5.

AghaKouchak, A., 2010. Application of a Conceptual Hydrologic Model in Teaching Hydrologic Processes. International Journal of Engineering Education URL: http://amir.eng.uci.edu/software.php.

AlbanD, 2019. Pytorch reproducibility over Different Machines. URL: https://discuss.pytorch.org/t/reproducibility-over-different-machines/63047/2.

Ali, A., 2015. Multi-objective operations of multi-wetland ecosystem: iModel applied to the everglades restoration. Journal of Water Resources Planning and Management 141. doi:10.1061/(ASCE)WR.1943-5452.0000511.

Allawi, M.F., Jaafar, O., Mohamad Hamzah, F., Koting, S.B., Mohd, N.S.B., El-Shafie, A., 2019. Forecasting hydrological parameters for reservoir system utilizing artificial intelligent models and exploring their influence on operation performance. Knowledge-Based Systems 163, 907–926. doi:10.1016/j.knosys.2018.10.013.

Amnatsan, S., Yoshikawa, S., Kanae, S., 2018. Improved forecasting of extreme monthly reservoir inflow using an analogue-based forecasting method: A case study of the Sirikit Dam in Thailand. Water (Switzerland) 10. doi:10.3390/w10111614.

Aravena, I., Gil, E., 2015. Hydrological scenario reduction for stochastic optimization in hydrothermal power systems. Applied Stochastic Models in Business and Industry 31, 231–240. doi:10.1002/asmb.2027.

Armstrong, J., 1985. Long-Range Forecasting URL: https://repository.upenn.edu/cgi/viewcontent.cgi?article=1227&context=marketing_papers.

Armstrong, J.S., 2005. A Commentary on Error Measures URL: https://papers.ssrn.com/abstract=663642.

Baesens, B., Viaene, S., Van Den Poel, D., Vanthienen, J., Dedene, G., 2002. Bayesian neural network learning for repeat purchase modelling in direct marketing. European Journal of Operational Research 138, 191–211. doi:10.1016/S0377-2217(01)00129-1.

Bai, Y., Wang, P., Xie, J., Li, J., Li, C., 2015. Additive model for monthly reservoir inflow forecast. Journal of Hydrologic Engineering 20. doi:10.1061/(ASCE)HE.1943-5584.0001101.

Bennett, N.D., Croke, B.F., Guariso, G., Guillaume, J.H., Hamilton, S.H., Jakeman, A.J., Marsili-Libelli, S., Newham, L.T., Norton, J.P., Perrin, C., Pierce, S.A., Robson, B., Seppelt, R., Voinov, A.A., Fath, B.D., Andreassian, V., 2013. Characterising performance of environmental models. Environmental Modelling and Software 40, 1–20. doi:10.1016/j.envsoft.2012.09.011.

Bergstra, J., Bardenet, R., Bengio, Y., Kégl, B., 2011. Algorithms for Hyper-Parameter Optimization .

Bergstra, J., Yamins, D., Cox, D.D., 2013. Hyperopt: A Python Library for Optimizing the Hyperparameters of Machine Learning Algorithms. Technical Report. URL: http://www.youtube.com/watch?v=Mp1xnPfE4PY!

Bergström, S., 1976. Development and Application of a Conceptual Runoff Model for Scandinavian Catchments. URL: https://www.researchgate.net/publication/255274162_Development_and_Application_of_a_Conceptual_Runoff_Model_for_Scandinavian_Catchments.

Bermúdez, J.D., Segura, J.V., Vercher, E., 2010. Bayesian forecasting with the Holt-Winters model. Journal of the Operational Research Society 61, 164–171. doi:10.1057/jors.2008.152.

de Boodt, M., Verdonck, O., 1972. The Physical Properties of the Substrates in Horticulture. Acta Horticulturae , 37–44doi:10.17660/actahortic.1972.26.5.

Bradshaw, J., Matthews, A.G.d.G., Ghahramani, Z., 2017. Adversarial Examples, Uncertainty, and Transfer Testing Robustness in Gaussian Process Hybrid Deep Networks URL: http://arxiv.org/abs/1707.02476.

Brockwell, P.J., Davis, R.A., Fienberg, S.E., 1998. Time Series: Theory and Methods: Theory and Methods. URL: https://books.google.no/books?hl=en&lr=&id=ZW_ThhYQiXIC&oi=fnd&pg=PR7&dq=time+series+definition&ots=g-fViSHGha&sig=GsWrdeMkpy4TlWqraMK1W8Uztl8&redir_esc=y#v=onepage&q=time%5C%20series%5C%20definition&f=false.

Brouwer, C., Goeffeau, A., Heiblom, M., 1985. Irrigation Water Management: Training Manual No. 1 - Introduction to Irrigation. URL: http://www.fao.org/3/r4082e/r4082e03.htm.

Budu, K., 2014. Comparison of wavelet-based ANN and regression models for reservoir inflow forecasting. Journal of Hydrologic Engineering 19, 1385–1400. doi:10.1061/(ASCE)HE.1943-5584.0000892.

Chatfield, C., 2000. Time-series forecasting. URL: https://content.taylorfrancis.com/books/download?dac=C2006-0-13501-8&isbn=9781420036206&format=googlePreviewPdf.

Chengzhao, Z., Heiping, P., Ke, Z., 2015. Comparison of back propagation neural networks and EMD-Based neural networks in forecasting the three major Asian stock markets. Journal of Applied Sciences 15, 90–99. doi:10.3923/jas.2015.90.99.

Chevalier, G., 2018. LARNN: Linear Attention Recurrent Neural Network URL: http://arxiv.org/abs/1808.05578.

Cho, S., Lim, B., Jung, J., Kim, S., Chae, H., Park, J., Park, S., Park, J.K., 2014. Factors affecting algal blooms in a man-made lake and prediction using an artificial neural network. Measurement: Journal of the International Measurement Confederation 53, 224–233. doi:10.1016/j.measurement.2014.03.044.

Coerver, H.M., Rutten, M.M., Van De Giesen, N.C., 2018. Deduction of reservoir operating rules for application in global hydrological models. Hydrology and Earth System Sciences 22, 831–851. doi:10.5194/hess-22-831-2018.

Costabile, P., Macchione, F., 2015. Enhancing river model set-up for 2-D dynamic flood modelling. Environmental Modelling and Software 67, 89–107. doi:10.1016/j.envsoft.2015.01.009.

Craven, J.R., 2016. Open source HBV model adapted by johnrobertcraven (John Craven) - Github. URL: https://github.com/johnrobertcraven/hbv_hydromodel.

Cui, Q., Wang, X., Li, C., Cai, Y., Liang, P., 2016. Improved Thomas–Fiering and wavelet neural network models for cumulative errors reduction in reservoir inflow forecast. Journal of Hydro-Environment Research 13, 134–143. doi:10.1016/j.jher.2015.05.003.

Data sources — The ISSN Portal, 2019. Data sources — The ISSN Portal. URL: https://portal.issn.org/data-sources.

Dehghani, M., Riahi-Madvar, H., Hooshyaripor, F., Mosavi, A., Shamshirband, S., Zavadskas, E.K., Chau, K.w., 2019. Prediction of hydropower generation using Grey wolf optimization adaptive neuro-fuzzy inference system. Energies 12. doi:10.3390/en12020289.

Doherty, J., 2015. Calibration and Uncertainty Analysis for Complex Environmental Models. Watermark Numerical Computing, Brisbane, Australia.

Doherty, J., Muffels, C., Rumbaugh, J., Tonkin, M., 2018. PEST - Model-Independent Parameter Estimation &amp; Uncertainty Analysis. URL: http://www.pesthomepage.org/Downloads.php.

Doña, C., Chang, N.B., Caselles, V., Sánchez, J.M., Pérez-Planells, L., Bisquert, M.d.M., García-Santos, V., Imen, S., Camacho, A., 2016. Monitoring hydrological patterns of temporary lakes using remote sensing and machine learning models: Case study of La Mancha Húmeda Biosphere Reserve in Central Spain. Remote Sensing 8. doi:10.3390/rs8080618.

Duan, A., Publisher, Q., 1991. A global optimization strategy for efficient and effective calibration of hydrologic models. Item Type text; Dissertation-Reproduction (electronic). Technical Report. URL: http://hdl.handle.net/10150/185655.

Duan, Q., Sorooshian, S., Gupta, V., 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. Water Resources Research 28, 1015–1031. doi:10.1029/91WR02985.

Duan, Q., Sorooshian, S., Gupta, V.K., 1994. Optimal use of the SCE-UA global optimization method for calibrating watershed models. Journal of Hydrology 158, 265–284. doi:10.1016/0022-1694(94)90057-4.

Duan, Q.Y., Gupta, V.K., Sorooshian, S., 1993. Shuffled complex evolution approach for effective and efficient global minimization. Journal of Optimization Theory and Applications 76, 501–521. doi:10.1007/BF00939380.

Elsevier's Webpage, 2019. Getting the most out of published research - Scopus — Elsevier Solutions. URL: `https://www.elsevier.com/solutions/scopus/how-scopus-works`.

Energifakta Norge - Ministry of Petroleum and Energy, 2019. Regulering av nettvirksomheten - Energifakta Norge. URL: `https://energifaktanorge.no/regulering-av-energisektoren/regulering-av-nettvirksomhet/`.

Falkner, S., Klein, A., Hutter, F., 2018. BOHB: Robust and Efficient Hyperparameter Optimization at Scale. Technical Report.

Flores, G., Ferreira, V.H., 2018. A rain-streamflow model for prediction of limnimetric behavior of reservoirs using artificial neural networks, in: SBSE 2018 - 7th Brazilian Electrical Systems Symposium, Institute of Electrical and Electronics Engineers Inc.. pp. 1–6. doi:`10.1109/SBSE.2018.8395879`.

Fragoso, T.M., Neto, F.L., 2014. Bayesian model averaging: A systematic review and conceptual classification *. Technical Report.

Gardner, E.S., 1985. Exponential smoothing: The state of the art. Journal of Forecasting 4, 1–28. doi:`10.1002/for.3980040103`.

Grant, M.J., Booth, A., 2009. A typology of reviews: An analysis of 14 review types and associated methodologies. doi:`10.1111/j.1471-1842.2009.00848.x`.

Graves, A., Schmidhuber, J., 2005. Framewise phoneme classification with bidirectional LSTM and other neural network architectures, in: Neural Networks, Pergamon. pp. 602–610. doi:`10.1016/j.neunet.2005.06.042`.

Josue de las Heras Torres, R., 2018. 7 Ways Time Series Forecasting Differs from Machine Learning — Oracle Data Science. URL: `https://blogs.oracle.com/datascience/7-ways-time-series-forecasting-differs-from-machine-learning`.

Hochreiter, S., Schmidhuber, J., 1997. LSTM can solve hard long time lag problems. Technical Report.

Holt, C.C., 2004. Forecasting seasonals and trends by exponentially weighted moving averages. International Journal of Forecasting 20, 5–10. doi:`10.1016/j.ijforecast.2003.09.015`.

House, C., Nicholas, S., Sutton, W., 2017. Iinternational Hydropower Assiciation: 2017 report. Technical Report. URL: `www.hydropower.org`.

Hutter, F., Kotthoff, L., Vanschoren, J., 2019. The Springer Series on Challenges in Machine Learning Automated Machine Learning Methods, Systems, Challenges. URL: http://www.springer.com/series/15602.

Ji, C., Yu, H., Wu, J., Yan, X., Li, R., 2018. Research on cascade reservoirs' short-term optimal operation under the effect of reverse regulation. Water (Switzerland) 10. doi:10.3390/w10060808.

Joyce, J., Chang, N.B., Harji, R., Ruppert, T., Imen, S., 2017. Developing a multi-scale modeling system for resilience assessment of green-grey drainage infrastructures under climate change and sea level rise impact. Environmental Modelling and Software 90, 1–26. doi:10.1016/j.envsoft.2016.11.026.

Kalekar, P.S., 2004. Time series Forecasting using Holt-Winters Exponential Smoothing. Technical Report.

Kang, B., Ku, Y.H., Kim, Y.D., 2015. A case study for ANN-based rainfall–runoff model considering antecedent soil moisture conditions in Imha Dam watershed, Korea. Environmental Earth Sciences 74, 1261–1272. doi:10.1007/s12665-015-4117-0.

Kim, T., Shin, J.Y., Kim, H., Kim, S., Heo, J.H., 2019. The use of large-scale climate indices in monthly reservoir inflow forecasting and its application on time series and artificial intelligence models. Water (Switzerland) 11. doi:10.3390/w11020374.

Kirham, M., 2005. Field Capacity, Wilting Point, Available Water, and the Non-Limiting Water Range, in: Principles of Soil and Plant Water Relations. Academic Press, pp. 101–115. URL: https://www.researchgate.net/publication/286420023_Field_Capacity_Wilting_Point_Available_Water_and_the_Nonlimiting_Water_Range, doi:10.1016/B978-012409751-3/50008-6.

Kleiven, A., Steinsland, I., 2019. Inflow Forecasting for Hydropower Operations: Bayesian Model Averaging for Postprocessing Hydrological Ensembles, in: Proceedings of the 6th International Workshop on Hydro Scheduling in Competitive Electricity Markets. Springer International Publishing, pp. 33–40. doi:10.1007/978-3-030-03311-8{_}5.

Kristvik, E., Riisnes, B.K., 2015. Hydrological Assessment of Water Resources in Bergen Climate Change Impacts. Technical Report.

Kumar, S., Tiwari, M.K., Chatterjee, C., Mishra, A., 2015. Reservoir Inflow Forecasting Using Ensemble Models Based on Neural Networks, Wavelet Analysis and Bootstrap Method. Water Resources Management 29, 4863–4883. doi:10.1007/s11269-015-1095-7.

Langsholt, E., Beldring, S., 2020. HBV-modellen - NVE. URL: https://www.nve.no/hydrologi/analysemetoder-og-modeller/hbv-modellen/.

Lauret, P., Fock, E., Randrianarivony, R.N., Manicom-Ramsamy, J.F., 2008. Bayesian neural network approach to short time load forecasting. Energy Conversion and Management 49, 1156–1166. doi:10.1016/j.enconman.2007.09.009.

Lecun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. doi:10.1038/nature14539.

Lindström, G., Johansson, B., Persson, M., Gardelin, M., Bergström, S., 1997. Development and test of the distributed HBV-96 hydrological model. Journal of Hydrology 201, 272–288. doi:10.1016/S0022-1694(97)00041-3.

Lindström, G., Rodhe, A., 1992. Transit times of water in soil lysimeters from modeling of oxygen-18. Water, Air, & Soil Pollution 65, 83–100. doi:10.1007/BF00482751.

Londhe, S.N., Sonawane, K.P., 2018. Modelling stage-discharge relationship using artificial neural networks, in: Proceedings - International Association for Hydro-Environment Engineering and Research (IAHR)-Asia Pacific Division (APD) Congress: Multi-Perspective Water for Sustainable Development, IAHR-APD 2018, Department of Civil and Environmental Engineering, Faculty of Engineering, Universitas Gadjah Mada. pp. 931–938.

Louizos, C., Welling, M., 2017. Multiplicative Normalizing Flows for Variational Bayesian Neural Networks URL: http://arxiv.org/abs/1703.01961.

Makridakis, S., Hibon, M., 2000. The M3-competition: Results, conclusions and implications. International Journal of Forecasting 16, 451–476. doi:10.1016/S0169-2070(00)00057-1.

Makridakis, S., Spiliotis, E., Assimakopoulos, V., 2020. The M4 Competition: 100,000 time series and 61 forecasting methods. International Journal of Forecasting 36, 54–74. doi:10.1016/j.ijforecast.2019.04.014.

Martín-Martín, A., Orduna-Malea, E., Thelwall, M., López-Cózar, E.D., 2018. Google Scholar, Web of Science, and Scopus: a systematic comparison of citations in 252 subject categories URL: http://arxiv.org/abs/1808.05053http://dx.doi.org/10.1016/j.joi.2018.09.002, doi:10.1016/j.joi.2018.09.002.

Mason, S.J., Goddard, L., Graham, N.E., Yulaeva, E., Sun, L., Arkin, P.A., 1999. The IRI Seasonal Climate Prediction System and the 1997/98 El Niño Event.

Bulletin of the American Meteorological Society 80, 1853–1873. doi:10.1175/1520-0477(1999)080<1853:TISCPS>2.0.CO;2.

Massa, F., AlbanD, 2017. Pytorch torch.backends.cudnn.benchmark. URL: https://discuss.pytorch.org/t/what-does-torch-backends-cudnn-benchmark-do/5936.

Mastrantonio, L.J., 1990. Lysimeters. Forest Research West URL: https://www.fs.fed.us/psw/publications/Popular/Lysimeters.html.

McDermott, P., Wikle, C., 2019. Bayesian Recurrent Neural Network Models for Forecasting and Quantifying Uncertainty in Spatial-Temporal Data. Entropy 21, 184. URL: http://www.mdpi.com/1099-4300/21/2/184, doi:10.3390/e21020184.

Moeeni, H., Bonakdari, H., 2017. Forecasting monthly inflow with extreme seasonal variation using the hybrid SARIMA-ANN model. Stochastic Environmental Research and Risk Assessment 31, 1997–2010. doi:10.1007/s00477-016-1273-z.

Moon, S., Kang, B., 2016. Terrestrial sediment yield projection under the bias-corrected nonstationary scenarios with hydrologic extremes. Water (Switzerland) 8. doi:10.3390/w8100433.

Mosavi, A., Ozturk, P., Chau, K.W., 2018. Flood prediction using machine learning models: Literature review. doi:10.3390/w10111536.

Mukheibir, P., Cole, C., Drinkwater, K., Abeysuriya, K., 2015. Consultative multi-criteria decision making process for drought security. Water Practice and Technology 10, 725–738. doi:10.2166/wpt.2015.089.

Nanda, T., Sahoo, B., Chatterjee, C., 2019. Enhancing real-time streamflow forecasts with wavelet-neural network based error-updating schemes and ECMWF meteorological predictions in Variable Infiltration Capacity model. Journal of Hydrology 575, 890–910. doi:10.1016/j.jhydrol.2019.05.051.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I - A discussion of principles. Journal of Hydrology 10, 282–290. doi:10.1016/0022-1694(70)90255-6.

Niggemann, O., Stein, B., Maier, A., Vodencarevic, A., Kleine, H., 2012. Learning Behavior Models for Hybrid Timed Systems. Technical Report. URL: www.aaai.org.

Nipen, T., Stull, R., 2011. Calibrating probabilistic forecasts from an NWP ensemble. Tellus, Series A: Dynamic Meteorology and Oceanography 63, 858–875. doi:10.1111/j.1600-0870.2011.00535.x.

Norwegian Ministry of Petroleum and Energy, 2019a. The electricity grid - Energifakta Norge. URL: https://energifaktanorge.no/en/norsk-energiforsyning/kraftnett/.

Norwegian Ministry of Petroleum and Energy, 2019b. The Power Market - Energifakta Norge. URL: https://energifaktanorge.no/norsk-energiforsyning/kraftmarkedet/.

Olofintoye, O., Otieno, F., Adeyemo, J., 2016a. Real-time optimal water allocation for daily hydropower generation from the Vanderkloof dam, South Africa. Applied Soft Computing Journal 47, 119–129. doi:10.1016/j.asoc.2016.05.018.

Olofintoye, O., Otieno, F., Adeyemo, J., 2016b. Real-time optimal water allocation for daily hydropower generation from the Vanderkloof dam, South Africa. Applied Soft Computing Journal 47, 119–129. doi:10.1016/j.asoc.2016.05.018.

Ord, J.K., 1988. Future developments in forecasting. The time series connexion. International Journal of Forecasting 4, 389–401. doi:10.1016/0169-2070(88)90106-9.

Osberg, M.M., 2019. Inflow modelling with probability distributions. Technical Report. Department of Computer Science Technology, NTNU – Norwegian University of Science and Technology.

Palit, A.K., Popovic, D., 2016. Computational Intelligence in Time Series Forecasting. URL: https://books.google.no/books?hl=en&lr=&id=fcv9Z7uiFasC&oi=fnd&pg=PA3&dq=time+series+forecasting+definition&ots=qdiBfssH5v&sig=3J8t9alj9ZUD92Y0lptV0Mif0eo&redir_esc=y#v=onepage&q=time%5C%20series%5C%20forecasting%5C%20definition&f=false.

Powel, 2019. Powel Inflow Tool. URL: https://www.powel.com/energy-trading-optimisation/forecasting/powel-inflow.

Rai, R., Singh, V., Upadhyay, A., 2017. Planning and Evaluation of Irrigation Projects Methods and Implementation. Academic Press. URL: https://www.elsevier.com/books-and-journals, doi:10.1016/B978-0-12-811748-4.00017-0.

Roulston, M.S., Smith, L.A., 2002. Evaluating Probabilistic Forecasts Using Information Theory. Monthly Weather Review 130, 1653–1660. URL: http://journals.ametsoc.org/doi/abs/10.1175/1520-0493%282002%29130%3C1653%3AEPFUIT%3E2.0.CO%3B2, doi:10.1175/1520-0493(2002)130<1653:EPFUIT>2.0.CO;2.

Rushdi, M., Perera, A., 2019. K-Medoids Clustering Based Approach to Predict the Future Water Height of a Reservoir, Institute of Electrical and Electronics Engineers (IEEE). pp. 279–286. doi:10.1109/icter.2018.8615461.

Santos, C.D., Girling, B., Rogner, M., Samuel, D., Troja, N., Ubierna, M., Costa, J., Faraday, F., Henley, W., Scorza, L., 2018. 2018 Hydropower Status Report — International Hydropower Association. Technical Report. URL: www.hydropower.org.

Seim, T.O., Thorsnes, O.R., 2007. Analyzing the price-and inflow relationships in hydroelectric scheduling. Technical Report.

da Silva Filho, J.A., de Farias, C.A.S., 2018. Stochastic modeling of monthly river flows by self-organizing maps. Journal of Urban and Environmental Engineering 12, 219–230. doi:10.4090/juee.2018.v12n2.219230.

Smyl, S., 2018. M4 Forecasting Competition: Introducing a New Hybrid ES-RNN Model. URL: https://eng.uber.com/m4-forecasting-competition/.

Spiliotis, E., Kouloumos, A., Assimakopoulos, V., Makridakis, S., 2020. Are forecasting competitions data representative of the reality? International Journal of Forecasting 36, 37–53. doi:10.1016/j.ijforecast.2018.12.007.

Springenberg, J.T., Klein, A., Falkner, S., Hutter, F., 2016. Bayesian Optimization with Robust Bayesian Neural Networks. Technical Report. URL: https://github.

SpringerLink, 2020. Data Availability Statements — Authors — Springer Nature. URL: https://www.springernature.com/gp/authors/research-data-policy/data-availability-statements/12330880.

StackOverflow nnnmmm, 2018. LSTM Visualization. URL: https://stackoverflow.com/revisions/48305882/1.

Suen, J.P., Eheart, J.W., 2006. Reservoir management to balance ecosystem and human needs: Incorporating the paradigm of the ecological flow regime. Water Resources Research 42. URL: https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2005WR004314https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005WR004314https:

//agupubs.onlinelibrary.wiley.com/doi/10.1029/2005WR004314, doi:10.1029/2005WR004314.

Tan, Q., Wang, X., Cai, S., Lei, X., 2016. Daily runoff time-series prediction based on the adaptive neural fuzzy inference system, in: 2015 12th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2015, Institute of Electrical and Electronics Engineers Inc.. pp. 506–512. doi:10.1109/FSKD.2015.7381994.

Taylor, J.W., 2003. Short-Term Electricity Demand Forecasting Using Double Seasonal Exponential Smoothing. Technical Report.

The Norwegian Water Resources and Energy Directorate, 2019. Horing - forskriftsbestemmelser om selskapsmessig og funksjonelt skille, samt merkevare og kommunikasjon - NVE. URL: https://www.nve.no/reguleringsmyndigheten/nytt-fra-rme/nyheter-reguleringsmyndigheten-for-energi/horing-forskriftsbestemmelser-om-selskapsmessig-og-funksjonelt-skille-samt-merkevare-og-kommunikasjon/.

Torch Contributors 2019, 2019. Reproducibility — PyTorch 1.5.0 documentation. URL: https://pytorch.org/docs/stable/notes/randomness.html.

Vodenčarević, A., Büning, H.K., Niggemann, O., Maier, A., 2011. Using behavior models for anomaly detection in hybrid systems, in: 2011 23rd International Symposium on Information, Communication and Automation Technologies, ICAT 2011. doi:10.1109/ICAT.2011.6102093.

Wan, X., Yang, Q., Jiang, P., Zhong, P., 2019. A Hybrid Model for Real-Time Probabilistic Flood Forecasting Using Elman Neural Network with Heterogeneity of Error Distributions. Water Resources Management 33, 4027–4050. doi:10.1007/s11269-019-02351-3.

Wang, Z., Lou, Y., 2019. Hydrological time series forecast model based on wavelet de-noising and ARIMA-LSTM, in: Proceedings of 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference, ITNEC 2019, Institute of Electrical and Electronics Engineers Inc.. pp. 1697–1701. doi:10.1109/ITNEC.2019.8729441.

Wei, C.C., 2016. Comparing single- and two-segment statistical models with a conceptual rainfall-runoff model for river streamflow prediction during typhoons. Environmental Modelling and Software 85, 112–128. doi:10.1016/j.envsoft.2016.08.013.

Winters, P.R., 1960. Forecasting Sales by Exponentially Weighted Moving Averages. Management Science 6, 324–342. doi:10.1287/mnsc.6.3.324.

Yang, S., Yang, D., Chen, J., Zhao, B., 2019. Real-time reservoir operation using recurrent neural networks and inflow forecast from a distributed hydrological model. Journal of Hydrology 579. doi:10.1016/j.jhydrol.2019.124229.

Zhong, Y., Guo, S., Ba, H., Xiong, F., Chang, F.J., Lin, K., 2018. Evaluation of the BMA probabilistic inflow forecasts using TIGGE numeric precipitation predictions based on artificial neural network. Hydrology Research 49, 1417–1433. doi:10.2166/nh.2018.177.

Appendix

A.1 Other Error/Efficiency Metrics

A.1.1 MAE - Mean Absolute Error

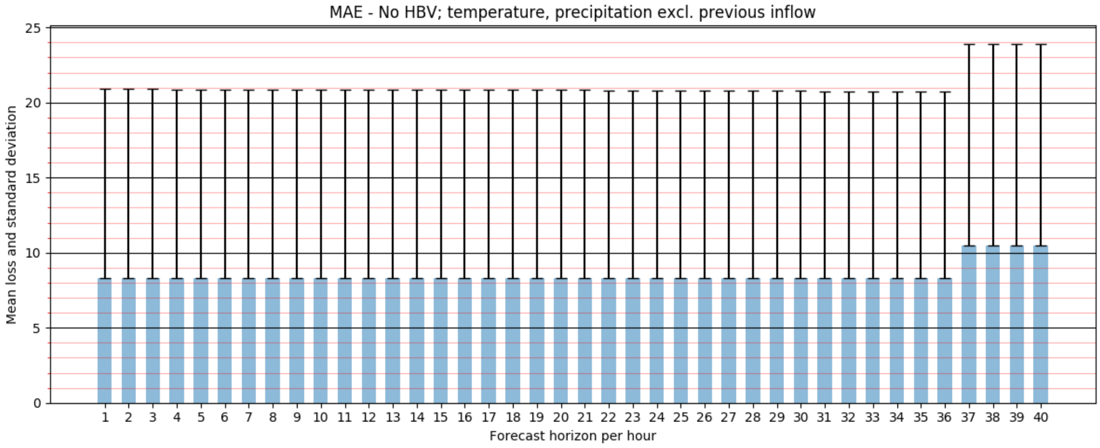

Figure 1: MAE for model using only an LSTM net with precipitation and temperature as input

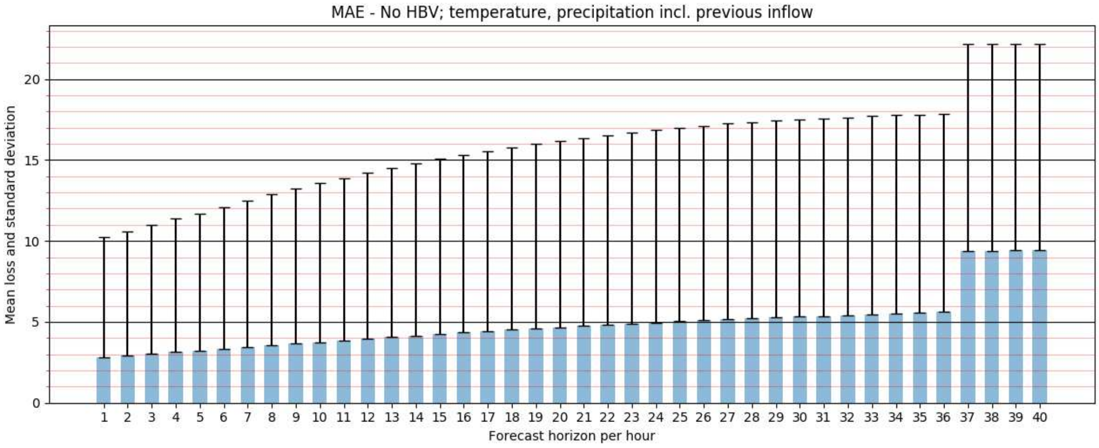

Figure 2: MAE for model using only an LSTM net with precipitation, temperature and previous inflow as input

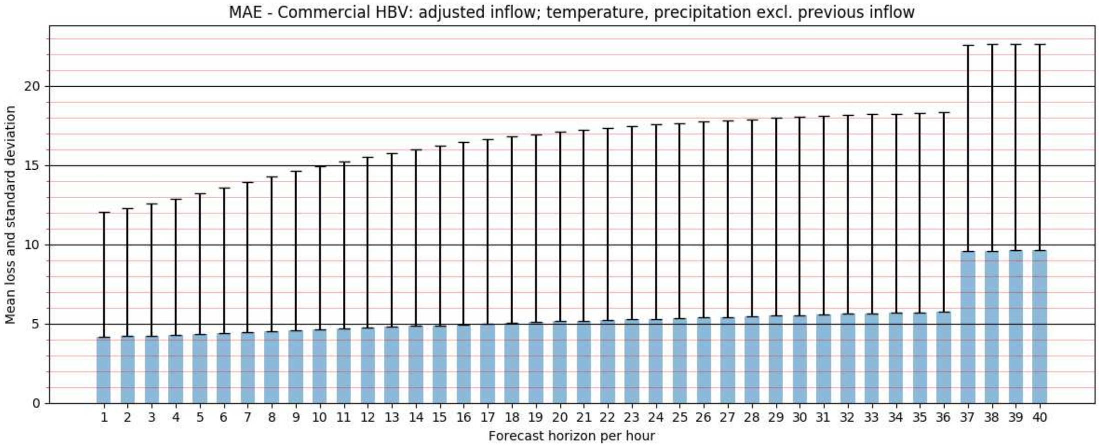

Figure 3: MAE for model using the commercial HBV's predicted inflow together with precipitation and temperature as input to an LSTM net

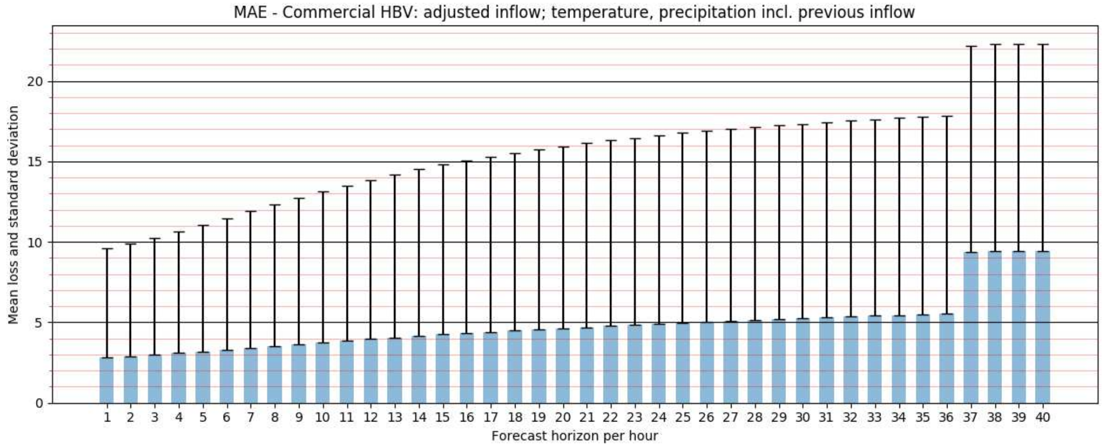

Figure 4: MAE for model using the commercial HBV's predicted inflow together with precipitation, temperature and previous inflow as input to an LSTM net

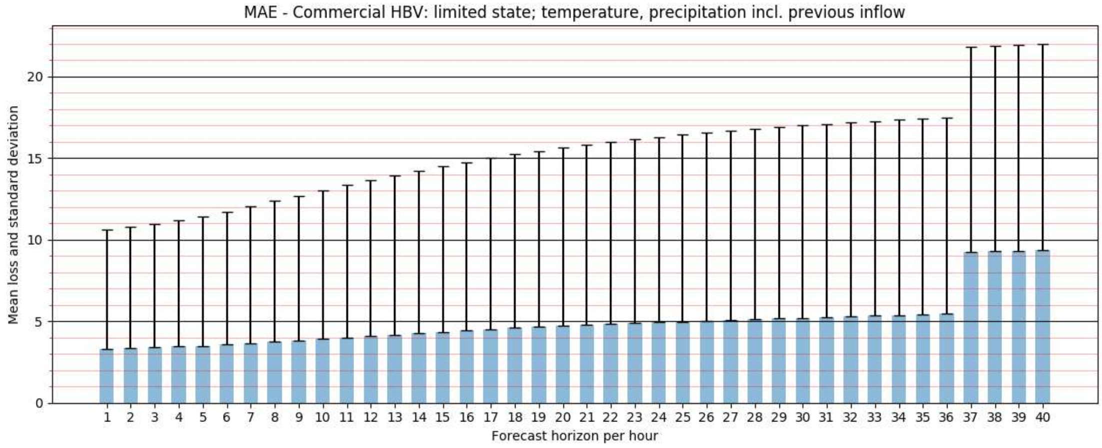

Figure 5: MAE for model using the commercial HBV's limited internal state together with precipitation, temperature and previous inflow as input to an LSTM net

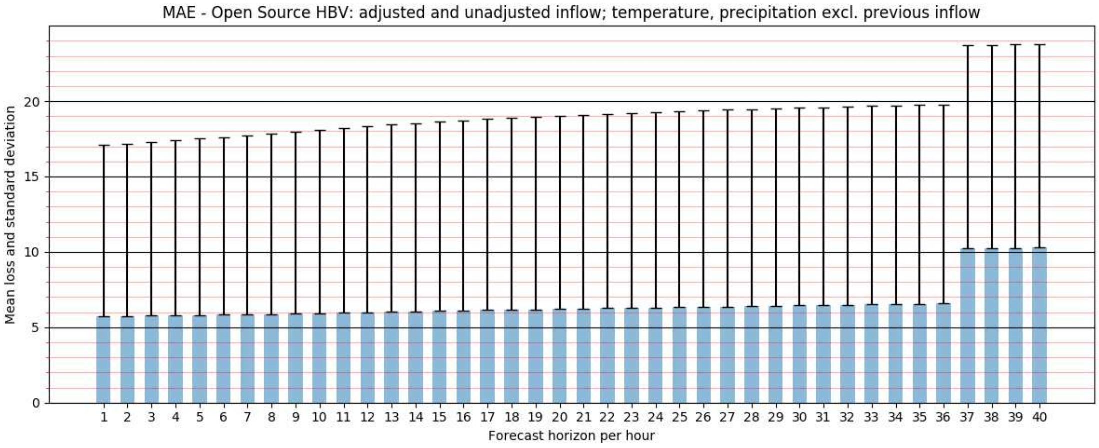

Figure 6: MAE for model using the open source HBV's predicted inflow together with precipitation and temperature as input to an LSTM net

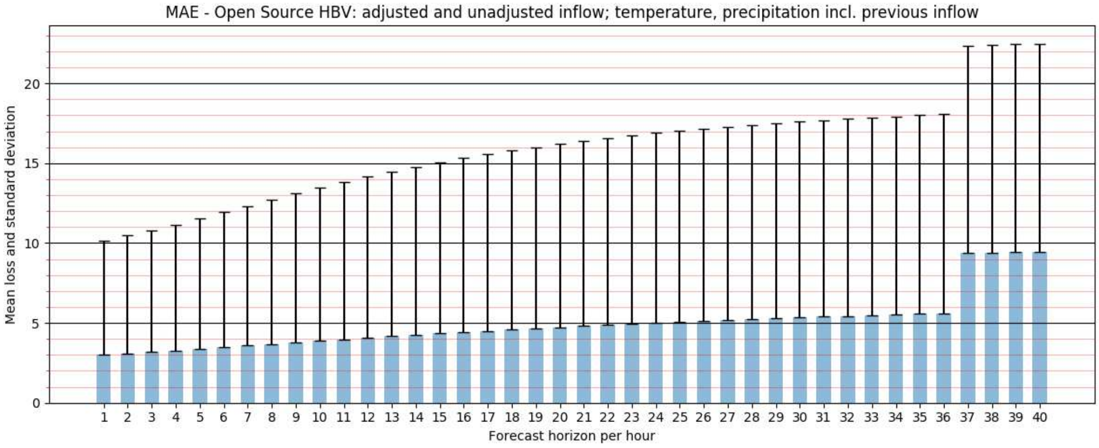

Figure 7: MAE for model using the open source HBV's predicted inflow together with precipitation, temperature and previous inflow as input to an LSTM net

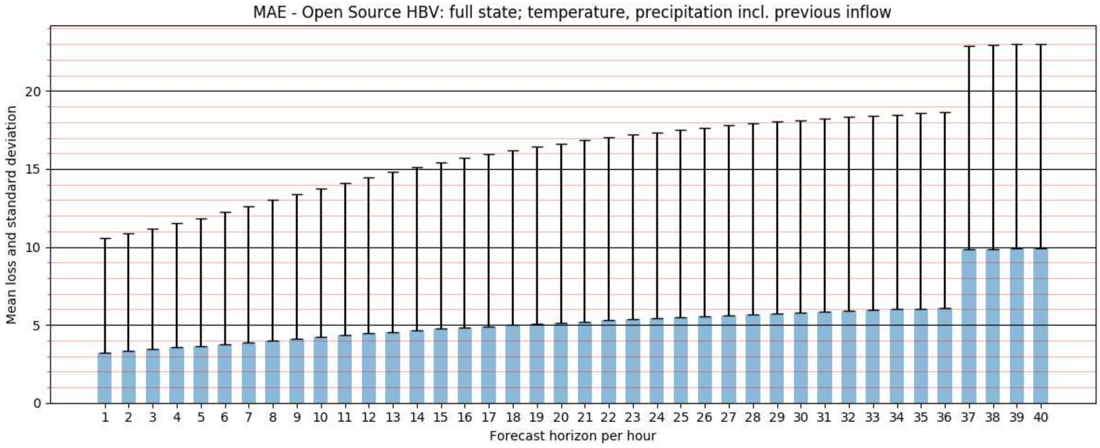

Figure 8: MAE for model using the open source HBV's full internal state together with precipitation, temperature and previous inflow as input to an LSTM net

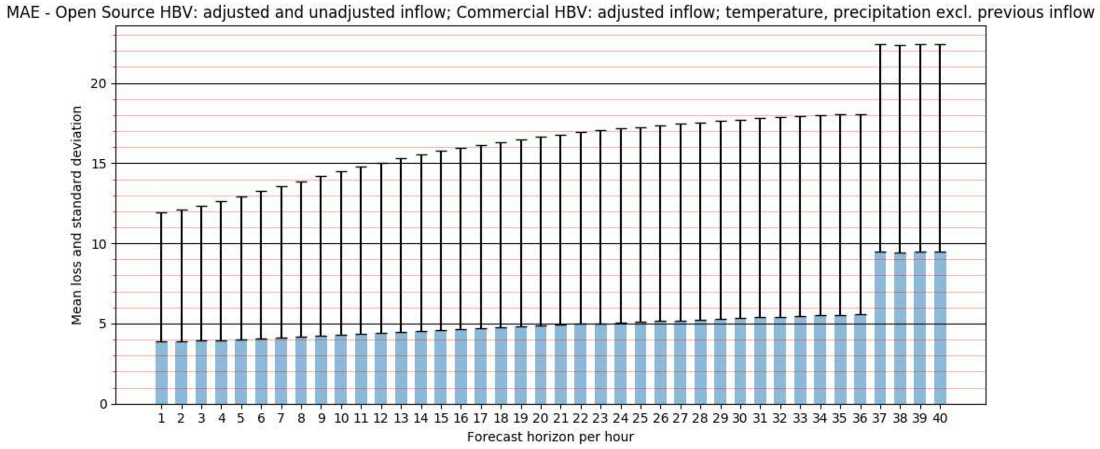

Figure 9: MAE for model using both the commercial HBV's- and open source HBV's predicted inflow together with precipitation and temperature as input to an LSTM net

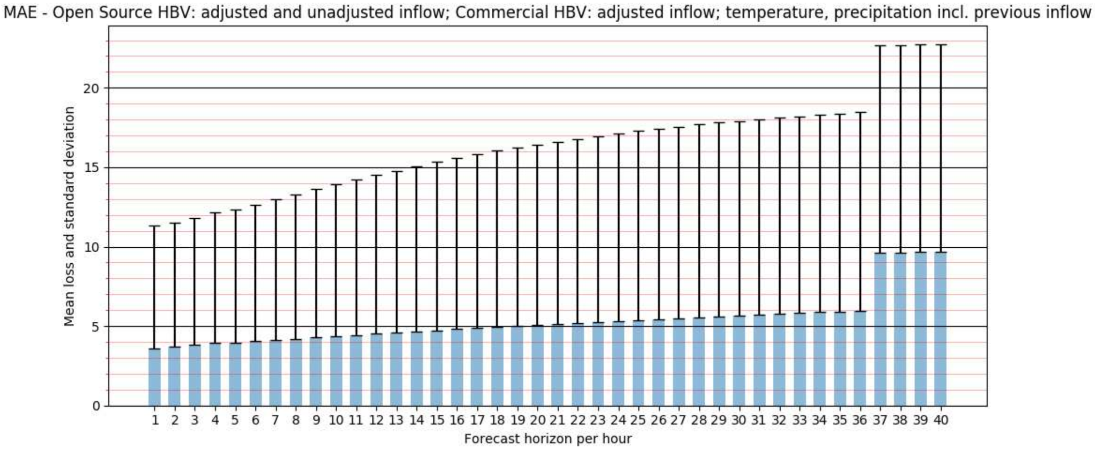

Figure 10: MAE for model using both the commercial HBV's- and open source HBV's predicted inflow together with precipitation, temperature and previous inflow as input to an LSTM net

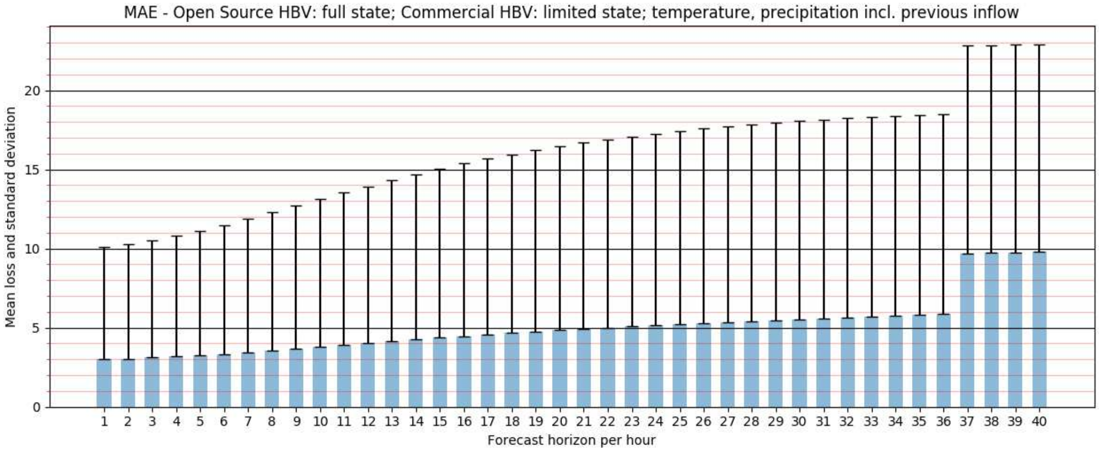

Figure 11: MAE for model using both the commercial HBV's limited internal state and open source HBV's full internal state together with precipitation, temperature and previous inflow as input to an LSTM net

Table 1: Model Mean Error and Standard Deviation by MAE

Metric	Model Name	Access to Previous Inflow?	Access to state?	Mean	Std
MAE	No HBV	N	N	8.441065	6.263072
MAE	No HBV	Y	N	4.811425	5.660452
MAE	os HBV	N	N	7.414157	6.555375
MAE	os HBV	Y	N	4.843859	5.397728
MAE	os HBV	Y	Y	5.086882	5.451234
MAE	c HBV	N	N	5.066901	5.592323
MAE	c HBV	Y	N	4.900718	5.289573
MAE	c HBV	Y	Y	4.586271	5.343112
MAE	both HBV	N	N	5.661468	5.698024
MAE	both HBV	Y	N	5.109941	5.445168
MAE	both HBV	Y	Y	5.213559	5.850499

A.1.2 Smooth L1 Loss

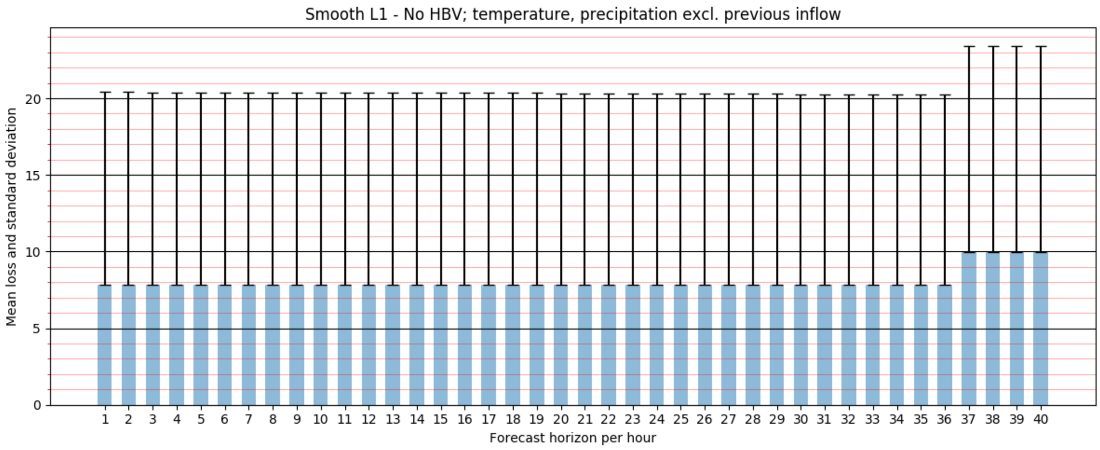

Figure 12: Smooth L1 Loss for model using only an LSTM net with precipitation and temperature as input

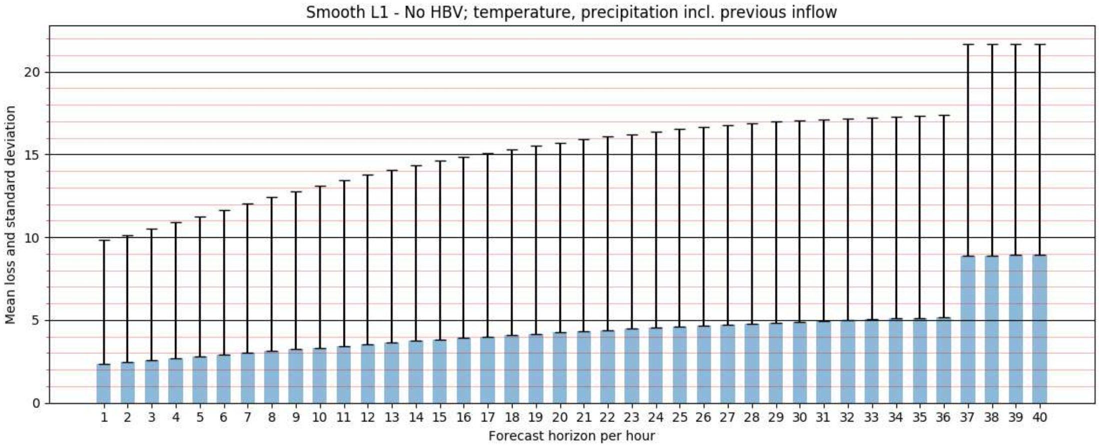

Figure 13: Smooth L1 Loss for model using only an LSTM net with precipitation, temperature and previous inflow as input

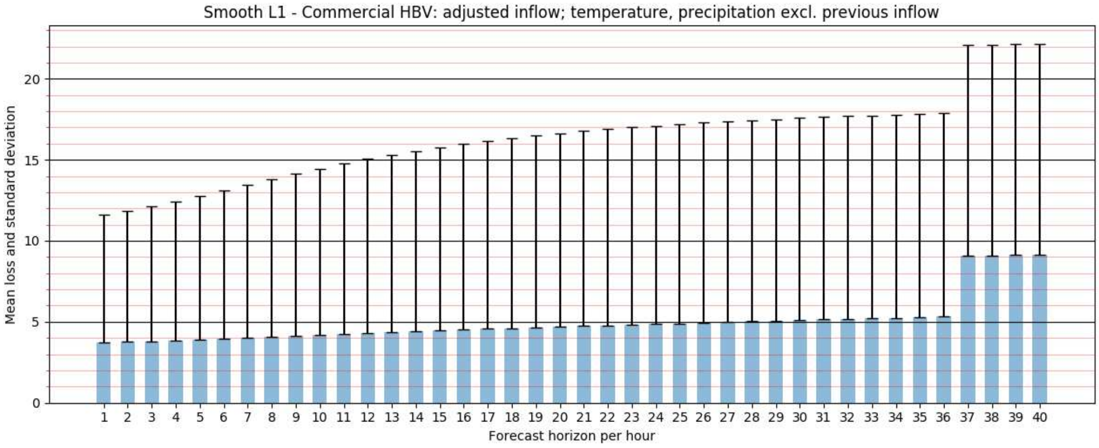

Figure 14: Smooth L1 Loss for model using the commercial HBV's predicted inflow together with precipitation and temperature as input to an LSTM net

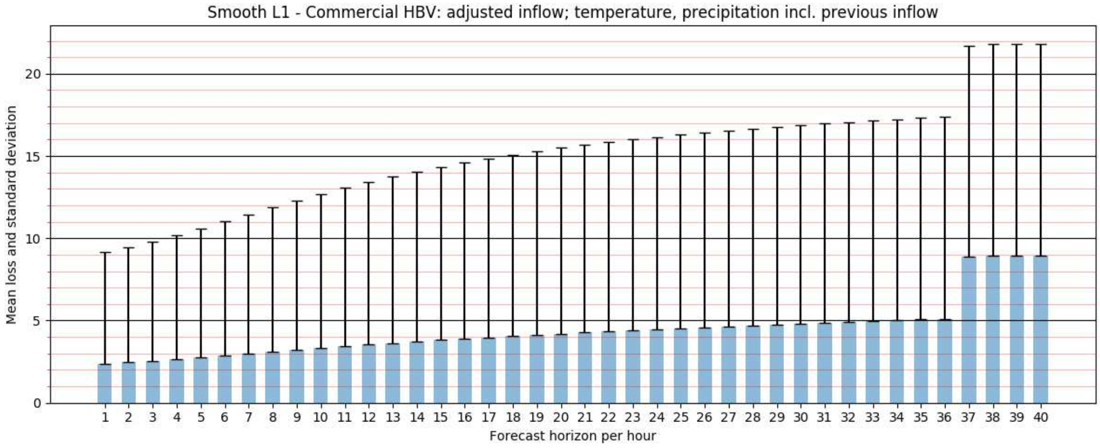

Figure 15: Smooth L1 Loss for model using the commercial HBV's predicted inflow together with precipitation, temperature and previous inflow as input to an LSTM net

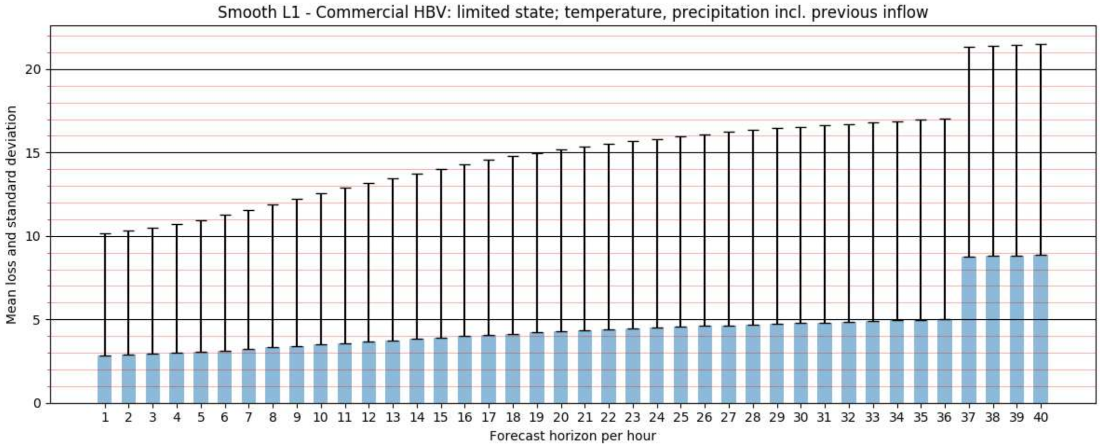

Figure 16: Smooth L1 Loss for model using the commercial HBV's limited internal state together with precipitation, temperature and previous inflow as input to an LSTM net

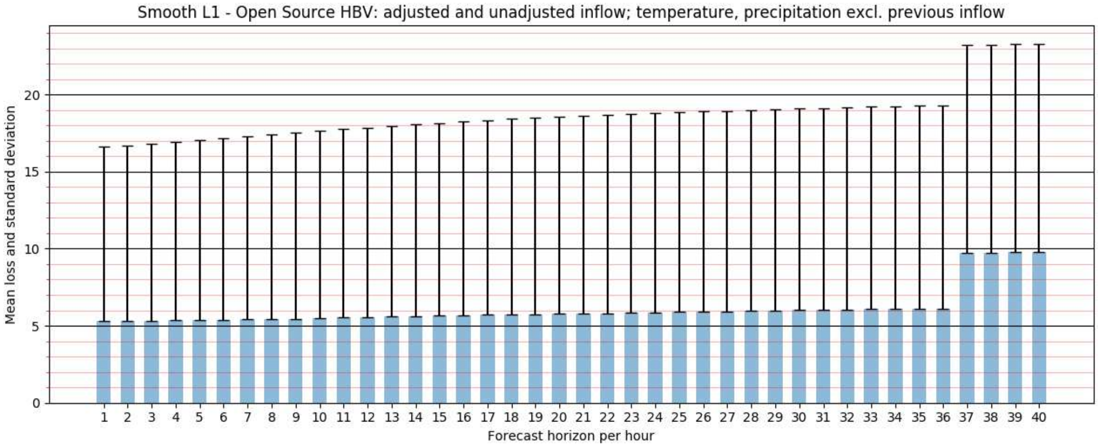

Figure 17: Smooth L1 Loss for model using the open source HBV's predicted inflow together with precipitation and temperature as input to an LSTM net

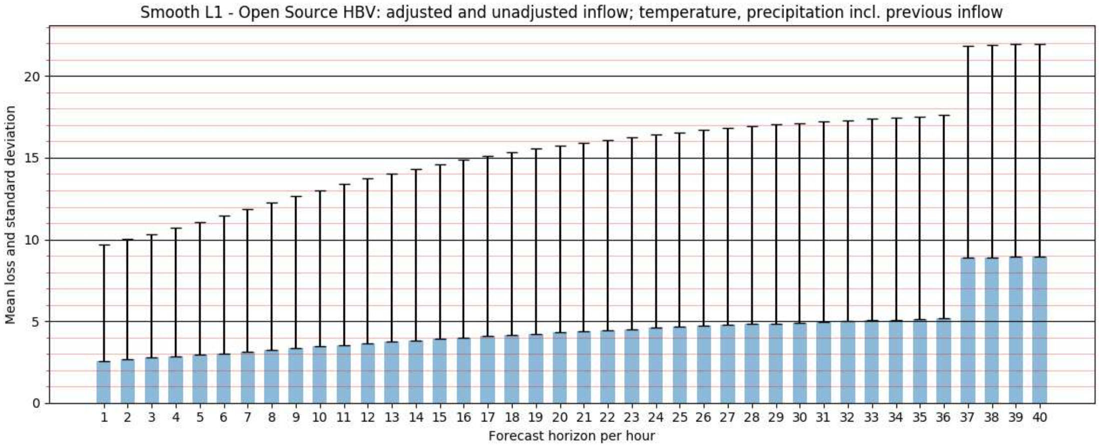

Figure 18: Smooth L1 Loss for model using the open source HBV's predicted inflow together with precipitation, temperature and previous inflow as input to an LSTM net

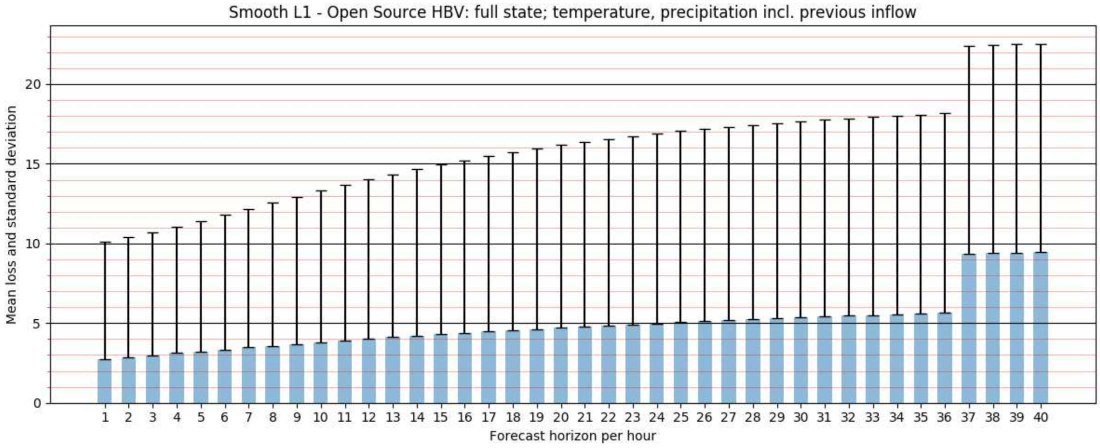

Figure 19: Smooth L1 Loss for model using the open source HBV's full internal state together with precipitation, temperature and previous inflow as input to an LSTM net

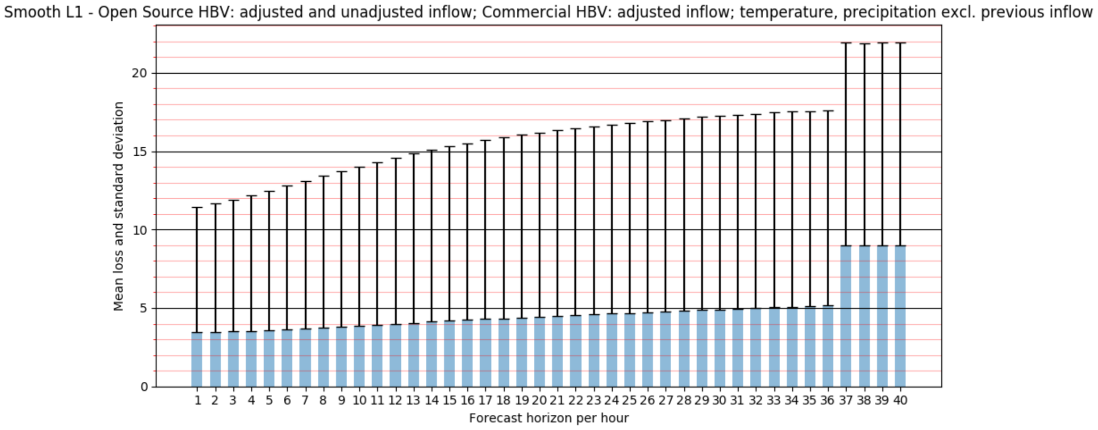

Figure 20: Smooth L1 Loss for model using both the commercial HBV's- and open source HBV's predicted inflow together with precipitation and temperature as input to an LSTM net

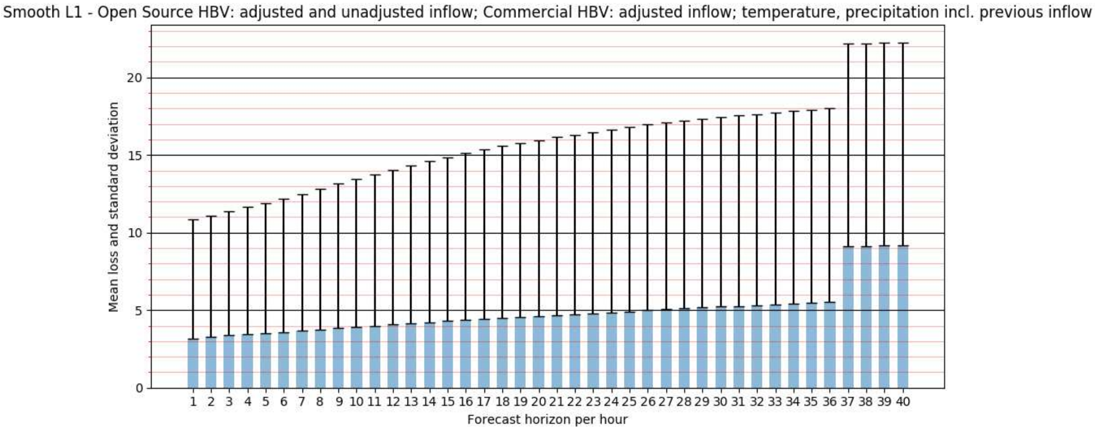

Figure 21: Smooth L1 Loss for model using both the commercial HBV's- and open source HBV's predicted inflow together with precipitation, temperature and previous inflow as input to an LSTM net

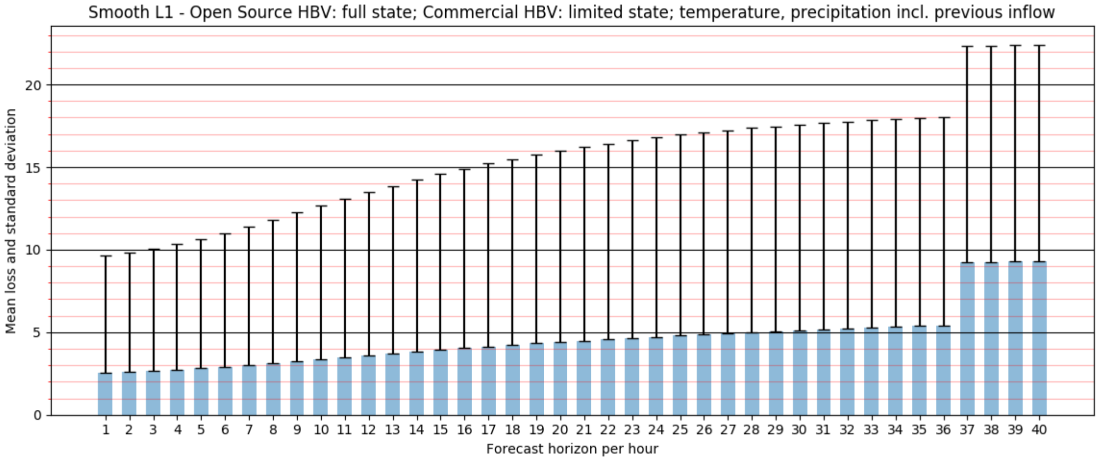

Figure 22: Smooth L1 Loss for model using both the commercial HBV's limited internal state and open source HBV's full internal state together with precipitation, temperature and previous inflow as input to an LSTM net

Table 2: Model Mean Error and Standard Deviation by Smooth L1 Loss

Metric	**Model Name**	**Access to Previous Inflow?**	**Access to state?**	**Mean**	**Std**
S L1	No HBV	N	N	7.953812	6.258918
S L1	No HBV	Y	N	4.382490	5.647306
S L1	os HBV	N	N	6.958296	6.543710
S L1	os HBV	Y	N	4.408518	5.384838
S L1	os HBV	Y	Y	4.649657	5.438251
S L1	c HBV	N	N	4.635481	5.578340
S L1	c HBV	Y	N	4.457947	5.277793
S L1	c HBV	Y	Y	4.158230	5.329552
S L1	both HBV	N	N	5.204155	5.688353
S L1	both HBV	Y	N	4.663914	5.433761
S L1	both HBV	Y	Y	4.780924	5.837101

A.2 Benchmark models' other error/efficiency metrics

A.2.1 Commercial HBV: MAE, Smooth L1 loss

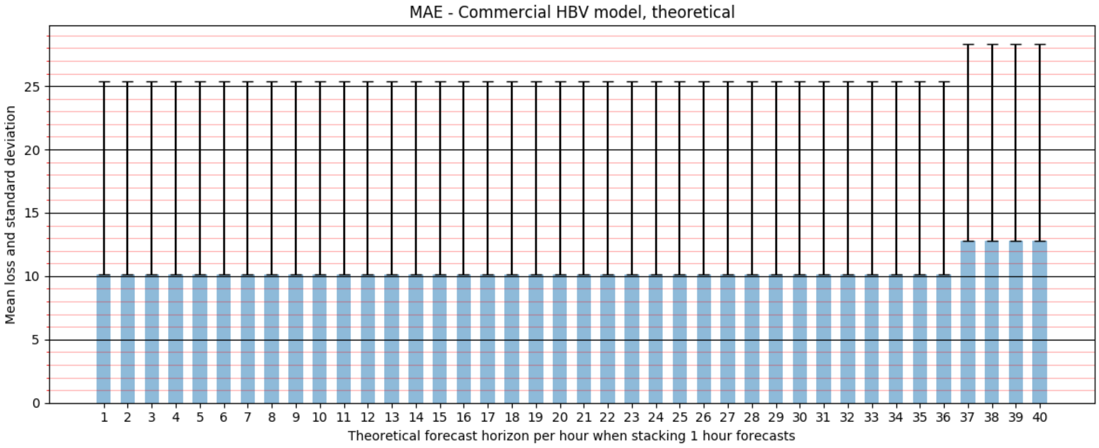

Figure 23: MAE for the commercial HBV model in a theoretical scenario where it makes 40 hour forecasts based on measured data

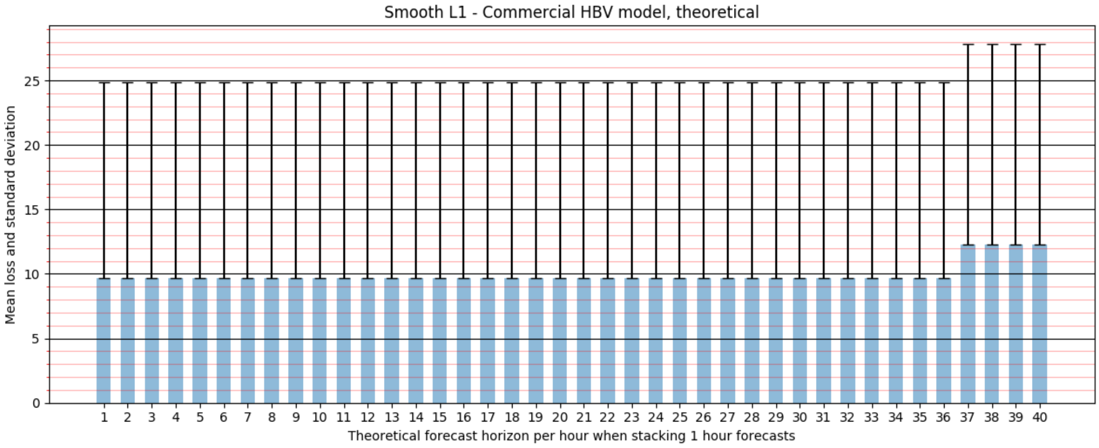

Figure 24: Smooth L1 Loss for the commercial HBV model in a theoretical scenario where it makes 40 hour forecasts based on measured data

A.2.2 Open Source HBV: MAE, Smooth L1 loss, NSE

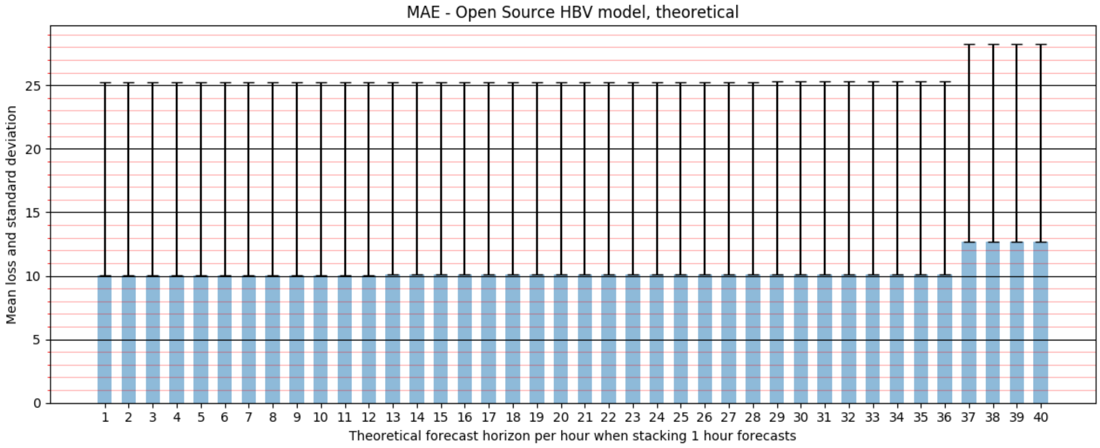

Figure 25: MAE for the open source HBV model in a theoretical scenario where it makes 40 hour forecasts based on measured data

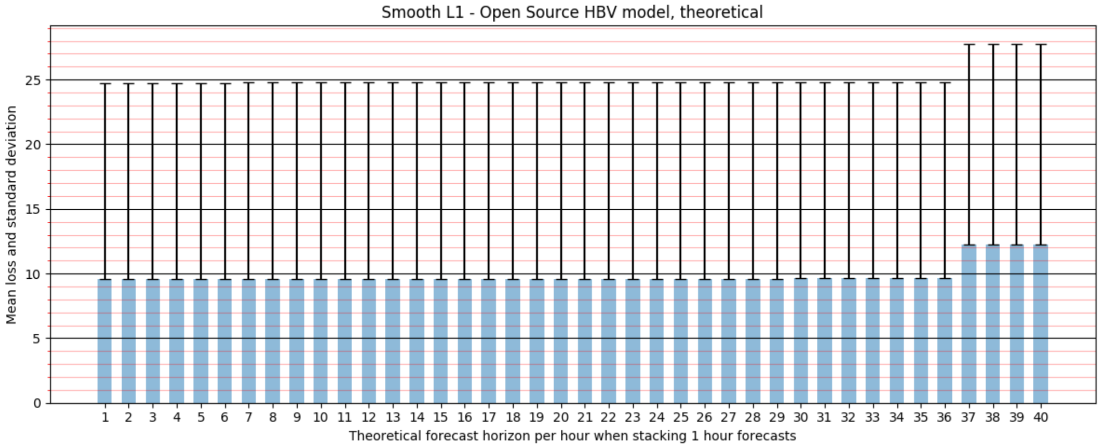

Figure 26: Smooth L1 Loss for the open source HBV model in a theoretical scenario where it makes 40 hour forecasts based on measured data

A.3 Tuned Hyperparameters for Each Model

Model using only an LSTM net with precipitation and temperature as input
{’batch_size’: 318.0, ’dropout_probability’: 0.09168090818037575, ’gradient_clipping’:
0.1258678646136906, ’hidden_dim’: 50.0, ’in_window’: 5.0,
’learning_rate’: 0.005358957245286223, ’lstm_layers’: 3.0}

Model using only an LSTM net with precipitation, temperature and previous inflow as input
{’batch_size’: 39.0, ’dropout_probability’: 0.2582174008849078, ’gradient_clipping’:
0.49896174269733257, ’hidden_dim’: 27.0, ’in_window’: 425.0,
’learning_rate’: 0.006618471494254368, ’lstm_layers’: 1.0}

Model using the commercial HBV’s predicted inflow together with precipitation and temperature as input to an LSTM net
{’batch_size’: 112.0, ’dropout_probability’: 0.055972153852199, ’gradient_clipping’:
0.43658919926509676, ’hidden_dim’: 15.0, ’in_window’: 118.0,
’learning_rate’: 0.00662901177440361, ’lstm_layers’: 2.0}

Model using the commercial HBV’s predicted inflow together with precipitation, temperature and previous inflow as input to an LSTM net
{’batch_size’: 227.0, ’dropout_probability’: 0.23446272129078952, ’gradient_clipping’:
0.6796383403042253, ’hidden_dim’: 18.0, ’in_window’: 234.0,
’learning_rate’: 0.04191621389991214, ’lstm_layers’: 1.0}

Model using the commercial HBV’s limited internal state together with precipitation, temperature and previous inflow as input to an LSTM net
{’batch_size’: 274.0, ’dropout_probability’: 0.3496138957680823, ’gradient_clipping’:
0.7274912860999065, ’hidden_dim’: 48.0, ’in_window’: 307.0,
’learning_rate’: 0.003197126320401476 4, ’lstm_layers’: 1.0}

Model using the open source HBV’s predicted inflow together with precipitation and temperature as input to an LSTM net
{’batch_size’: 84.0, ’dropout_probability’: 0.5400905317288727, ’gradient_clipping’:
0.321704976403499, ’hidden_dim’: 33.0, ’in_window’: 222.0,
’learning_rate’: 0.002208554017192015, ’lstm_layers’: 2.0}

Model using the open source HBV’s predicted inflow together with

precipitation, temperature and previous inflow as input to an LSTM net
{’batch_size’: 357.0, ’dropout_probability’: 0.12643475846392224, ’gradient_clipping’: 0.08871140522951365, ’hidden_dim’: 28.0, ’in_window’: 316.0, ’learning_rate’: 0.010157334192686128, ’lstm_layers’: 1.0}

Model using the open source HBV’s full internal state together with precipitation, temperature and previous inflow as input to an LSTM net
{’batch_size’: 73.0, ’dropout_probability’: 0.48076317042871436, ’gradient_clipping’: 0.4584051911951125, ’hidden_dim’: 48.0, ’in_window’: 998.0, ’learning_rate’: 0.0019632955333275955, ’lstm_layers’: 3.0}

Model using both the commercial HBV’s- and open source HBV’s predicted inflow together with precipitation and temperature as input to an LSTM net
{’batch_size’: 279.0, ’dropout_probability’: 0.39913642117439174, ’gradient_clipping’: 0.388739688450507, ’hidden_dim’: 35.0, ’in_window’: 172.0, ’learning_rate’: 0.015467203339116966, ’lstm_layers’: 2.0}

Model using both the commercial HBV’s- and open source HBV’s predicted inflow together with precipitation, temperature and previous inflow as input to an LSTM net
{’batch_size’: 188.0, ’dropout_probability’: 0.093306635161192, ’gradient_clipping’: 0.33827390746609365, ’hidden_dim’: 36.0, ’in_window’: 305.0, ’learning_rate’: 0.05924978100093053, ’lstm_layers’: 1.0}

Model using both the commercial HBV’s limited internal state and open source HBV’s full internal state together with precipitation, temperature and previous inflow as input to an LSTM net
{’batch_size’: 334.0, ’dropout_probability’: 0.18790784325908566, ’gradient_clipping’: 0.694783218334632, ’hidden_dim’: 10.0, ’in_window’: 995.0, ’learning_rate’: 0.007169395593965098, ’lstm_layers’: 4.0}

A.4 Tuned Hyperparameters for Selected Models when tuned for Longer Than 100 Combinations

Model using both the commercial HBV’s limited internal state and open source HBV’s full internal state together with precipitation, tem-

perature and previous inflow as input to an LSTM net

{’batch_size’: 209.0, ’dropout_probability’: 0.1501815850398151, ’gradient_clipping’: 0.7866505582273241, ’hidden_dim’: 20.0, ’in_window’: 67.0, ’learning_rate’: 0.004963359770016575, ’lstm_layers’: 2.0} After 164 combinations

Model using the commercial HBV’s limited internal state together with precipitation, temperature and previous inflow as input to an LSTM net

{’batch_size’: 345.0, ’dropout_probability’: 0.09544273297259243, ’gradient_clipping’: 0.48923633376708975, ’hidden_dim’: 28.0, ’in_window’: 20.0, ’learning_rate’: 0.004332006051395278, ’lstm_layers’: 1.0} After 272 combinations

Magnus Myrmo Osberg

LSTM Hybrid Model for Water Reservoir Inflow Forecasting

NTNU

Norwegian University of
Science and Technology