Spatio-Temporal Hydro Forecasting of Multireservoir Inflows for Hydro-Thermal Scheduling

Timo Lohmann[a], Amanda S. Hering[b], Steffen Rebennack[a,*]

[a] *Division of Economics and Business, Colorado School of Mines, Golden, Colorado, USA.*
[b] *Department of Applied Mathematics and Statistics, Colorado School of Mines, Golden, Colorado, USA.*

Abstract

Hydro-thermal scheduling is the problem of finding an optimal dispatch of power plants in a system containing both hydro and thermal plants. Since hydro plants are able to store water over long time periods, and since future inflows are uncertain due to precipitation, the resulting multi-stage stochastic optimization problem becomes challenging to solve. Several solution methods have been developed over the past few decades to compute practically useful operation policies. One of these methods is stochastic dual dynamic programming (SDDP). SDDP poses strong restrictions on the forecasting method generating the necessary inflow scenarios. In this context, the current state-of-the-art in forecasting are periodic autoregressive (PAR) models. We present a new forecasting model for hydro inflows that incorporates spatial information, *i.e.*, inflow information from neighboring reservoirs of the system, and that also satisfies the restrictions posed by SDDP. We benchmark our model against a PAR model that is similar to the one currently used in Brazil. Three multi-reservoir basins in Brazil serve as a case study for the comparison. We show that our approach outperforms the benchmark PAR model and present the root mean squared error (RMSE) as well as the seasonally-adjusted coefficient of efficiency (SACE) for each reservoir modeled. The overall decrease in RMSE is 8.29% using our approach for one month-ahead forecasts. The decrease in RMSE is achieved without additional data collection while only adding 11.8% more state variables for the SDDP algorithm.

Keywords: Space-time models, forecasting, stochastic dual dynamic programming (SDDP), stochastic hydro-thermal scheduling, periodic autoregressive (PAR) model

1. Introduction

The optimization of power systems containing both hydro and thermal plants in a regulated market environment, also known as hydro-thermal scheduling, is a well-studied problem [65, 33]. In its basic form, a central dispatcher operates the power system and satisfies consumers' electricity demand while minimizing operating costs. The planning horizon of a mid-term hydro-thermal scheduling problem typically spans three to five years with a monthly time resolution and depends on the cumulative storage capacity of the hydro reservoirs in the system relative to total demand. The inflows into the reservoirs of the hydro plants of the system are not known and, therefore, are modeled as uncertain in order to hedge against extreme scenarios. Using too much water for power generation today may lead to high operating costs or, even worse, electricity rationing, in an unforeseen dry period in the future. Whereas, keeping too much water in the reservoirs today may lead to spillage in an unforeseen wet period in the future, again resulting in high operating costs. Similar concepts apply to the case of a deregulated electricity market environment with hydro-power plants in which individual power companies maximize their individual profit. In both cases, the resulting problem can be formulated as a large-scale multi-stage stochastic linear programming (SLP) problem.

Several solution methods have been developed over the past two decades, including scenario tree and various decomposition approaches. In particular, the stochastic dual dynamic programming (SDDP) method

*Corresponding author
Email address: srebenna@mines.edu (Steffen Rebennack)
URL: www.rebennack.net (Steffen Rebennack)

[33] is a well established solution method in this context. The Brazilian power system, for instance, is operated with the aid of a version of SDDP, *cf.* Section 4. The SDDP method is capable of handling a large hydro system, multiple time periods, and an extremely detailed representation of the uncertainty in inflows [25]. However, it poses the strong restriction on the forecasting method that if inflows are modeled as interstage dependent, they have to be sampled from a linear, additive model [19, 38]. Consequently, periodic autoregressive (PAR) type models are commonly used to forecast monthly inflows for a horizon of three to five years ahead [24]. Any improvements in the forecasting model will translate into superior optimization results. To date, spatial information has not been directly included in hydro-inflow forecasting models for use with SDDP. We propose and develop such a model; evaluate its performance using real data; and demonstrate how to incorporate such forecasts in the SDDP algorithm.

First, we propose a spatio-temporal approach to forecasting the hydro inflows and evaluate it using three basins from the Brazilian hydro system as a case study. More specifically, instead of only fitting a model to each location's inflows individually, we also use inflow series of neighbors to improve the forecasts at a given location. Selecting neighbors based on their upstream relationship along the river provides the most improvement. The Brazilian hydro system is among the largest in the world with about 140 hydro plants. Historic inflow data is publicly available through the system operator's website, *Operador Nacional do Sistema Elétrico (ONS)* [31].

We compare our spatial periodic autoregressive (SPAR) model to a benchmark PAR model that is similar to the one used in Brazil [24, 25]. The main difference between our benchmark PAR and the Brazilian PAR is that the latter is based on an equivalent reservoir scheme that aggregates reservoirs based on geographical location. In order to better compare the PAR to our SPAR, we apply the PAR on the reservoir level instead. We show that our SPAR is superior in terms of more accurate point forecasts while adding little complexity to the SDDP algorithm, *i.e.*, the number of state variables slightly increases or stays the same. Furthermore, our approach does not require additional data collection. For the mathematical programming practitioner who is familiar with hydro-thermal scheduling and the SDDP method, but not necessarily with forecasting inflow scenarios for it, we present a detailed description of our point forecasting approach as applied to a historical inflow data set and provide algorithms for generating forecast scenarios.

Second, we describe how the forecasted inflow scenarios are incorporated into the SDDP algorithm. We derive Benders optimality cuts for our spatio-temporal model and compare them to cuts associated with a PAR model.

The unique contributions of this paper are three-fold.

- First, we describe how the widely used PAR model in Brazil can be made more efficient without even considering spatial information. In particular, we discuss parameter reduction with regard to deseasonalization and model fitting.
- Second, we develop a spatio-temporal hydro-inflow forecasting model for a system of power plants with a monthly fidelity.
- Third, we show how inflow scenarios can be obtained from our model and that they can be incorporated into the SDDP method with little additional effort.

The remainder of this paper is organized as follows. Section 2 reviews the literature on hydro-thermal scheduling and hydro inflow forecasting. Section 3 presents the space-time model along with a description of how to incorporate it into an SDDP routine. Following this, we demonstrate and evaluate the model in Section 4 using the Brazilian hydro system as a case study. Finally, Section 5 offers concluding remarks.

2. Literature Review

The main goal of this paper is to introduce and demonstrate the performance of a new forecasting model for hydro inflows that incorporates spatial information. Ultimately, these forecasts would be used in a hydro-thermal scheduling model. Thus, we review the relevant literature in both hydro forecasting and hydro-thermal scheduling in order to motivate our hydro forecasting model choices.

2.1. Solution Methods for Hydro-Thermal Scheduling

Several approaches can be found in the literature on how to tackle mid-term to long-term hydro-thermal scheduling problems. Until the 1990s, the predominant method to solve reservoir management problems was the stochastic dynamic programming (SDP) algorithm, which essentially decomposes the multi-stage decision process into decisions for each stage following the principle of Bellman [2]. Different versions of the SDP algorithm emerged, each aimed at mitigating its computational drawbacks [15, 32]. Survey papers

[63, 65] present a review of hydro-thermal scheduling algorithms in this period. Then, the stochastic dual dynamic programming (SDDP) algorithm was proposed [33], which successfully overcomes the biggest drawback of the SDP related methods (at least computationally): the curse-of-dimensionality. The curse-of-dimensionality is the combinatorial "explosion" of the state or scenario space with an increase in i) reservoirs and reservoir level discretization, ii) scenarios, and iii) time stages. The SDDP algorithm has been extensively analyzed, especially with regard to statistical and convergence properties [35, 46].

Extensions on many fronts have been made for the SDDP method. On the inflow scenario front, cut sharing methods were developed to overcome the restriction of independent inflows and to allow interstage dependency [19, 38]. Discrete inflow residuals for each stage have also been proposed to overcome the discretization error exposed to by sampling from a continuous distribution [36]. The SDDP algorithm has been applied to answer irrigation and environmental questions [58] as well as to model emission caps [42, 40]. Furthermore, the SDDP algorithm, designed for a regulated market environment, has also been adopted to a deregulated electricity market, wherein instead of a single regulator operating all plants in the system, multiple players (electricity companies) own different parts of the system and submit bids into an electricity wholesale market. In this setting, prices of electricity are typically assumed to be uncertain, giving rise to hybrid SDP-SDDP methods [13, 12], bidding strategy analyses [45, 11, 55, 56], and incorporation of risk aversion [34, 47]. SDDP can also be applied to nonlinear models such as those that arise when incorporating nonlinear water head effects [5].

All of the above SDDP-type algorithms share the same restriction with respect to uncertainty in the inflows. If the uncertain inflows are to be modeled as interstage dependent, they must come from a linear additive forecasting model; our proposed model satisfies this requirement. Incorporating improved forecasts in an optimization problem, such as a power system economic dispatch model, has been shown to result in considerable cost savings [69].

SDDP is not the only approach to solve hydro-scheduling problems. In particular, most other methods do not require the hydro inflows to come from a linear additive model. One such method is the scenario tree based method [16, 4], which approximates the random space by a set of scenarios in order to construct a deterministic equivalent and, thus, can use any forecasting model. In addition, constructive dual dynamic programming [39, 64], a technique that is similar to SDDP, solves the dual of the dynamic programming formulation directly and can also be used with any forecasting model. However, the advantages to using SDDP are that (1) it can handle scenario trees that are several times larger in magnitude, and (2) it is currently used for the Brazilian power system.

2.2. Models for Hydro-Inflow Forecasting

The literature on hydro-inflow forecasting may be divided into two categories: process-driven methods and data-driven methods. Process-driven methods aim to model the underlying physical processes of the system in order to explain the resulting inflows, *e.g.*, in a rainfall-runoff hydrological model, which is governed by variables such as evaporation, soil moisture, temperature, and land use. Data-driven methods make use of empirical relationships observed in historical time series to calibrate model parameters. They are able to represent complex processes based on mathematical criteria without deeper knowledge of the underlying physical processes. A survey of both categories can be found in [61]. In this paper, we focus on data-driven methods in the literature review due to the linearity limitation imposed by the SDDP algorithm, *cf.* Section 3.3.

Different approaches exist to forecast hydro inflows using data-driven methods. Regression models are a commonly used method to forecast hydro inflows, and variations include principal component analysis [27], multiple regression [54], and nonparametric regression [52, 50], among others. Regression models allow the use of exogenous variables that describe precipitation, temperature, and large-scale climatological patterns, such as the El Niño Southern Oscillation and the North Atlantic Oscillation, in addition to the historical inflow time series to forecast hydro inflows. Semi-parametric methods that look for predictive analogues in the past whose predictors are similar to the current set of predictors have also been explored [10].

Time series models used for forecasting are ARIMA (autoregressive integrated moving average) models and its derivatives such as AR (autoregressive) and ARMA (autoregressive moving average) models. They assume stationarity in the time series and are, thus, typically used for annual inflow forecasts. For forecasts on higher time resolutions, seasonality aspects have to be incorporated. Models commonly used are seasonal ARIMA (SARIMA), periodic ARMA (PARMA), deseasonalized ARMA, and PAR models [18]. For monthly forecasts, the PAR model has been applied in India [28] and Brazil [24, 53]. The ARIMA model has been applied in Turkey [67], and the SARIMA model has been applied in Canada [3] as well as in a broad case study of

North and South America [30]. Dynamic Bayesian linear models [23] overcome the stationarity restrictions imposed by the classical time series approach.

Artificial neural networks (ANN) describe a third data-driven approach that have become popular for hydrological forecasting, see [26] and [7] for a review. A comparison of ANN and ARIMA models for monthly inflow forecasts for a reservoir in India finds that the ANN predicts higher inflows better than the ARIMA model [20]. However, the opposite is true for lower inflows. A comparison of ANN and PARMA models for monthly inflow forecasts for 37 reservoirs in Brazil shows that the ANN approach outperforms the PARMA model [59]. However, the ANN models cannot be used in combination with the SDDP algorithm since it is a nonlinear modeling approach.

No linear statistical models, to our knowledge, exist that directly incorporate spatial information for hydro-inflow forecasting at a monthly temporal scale. An ANN with spatial information for 1 to 7 hours ahead forecasts improved upon hydrological models [62]. Also, an ANN has shown the strong spatial dependence among 6 gaging stations in a river network in northeast Brazil [10]. Regression models incorporating spatial information have been shown to improve forecasting in a variety of applications [43, 60, 6, 9]. In this particular setting, we do not desire to make spatial predictions since the locations of the hydro plants are known and fixed. Instead, we want to use spatial information to improve the temporal forecasts, and since nearby hydro inflows tend to be more alike than those that are far apart, this information can be used in addition to the time series of inflows at the location of interest. Such methods have been used extensively in wind forecasting for wind farms, another area wherein the locations are sparse in space, and forecasts may be improved by including off-site information [8, 14, 17, 66]. All of these works use a spatially motivated algorithm for testing the inclusion of "upstream" locations in the forecasting model, and we adapt such an approach for forecasting inflows.

3. Methodology

In the following sections, we present our inflow forecasting model that can be used to generate inflow scenarios for the SDDP method. In particular, a detailed description is provided on how to obtain inflow scenarios from historical inflow data. Then, we discuss relevant error measures to compare different forecasting models and close with a description on how the SDDP method must be adjusted in order to function with scenarios from our point forecasting approach.

3.1. Forecasting Models

This section lays the foundation for our method by describing the properties of AR, and more specifically, PAR models. We then present our approach, a PAR model with a spatial component.

3.1.1. AR and PAR Models

A time series, $Z(t)$, $t = 1, \ldots, T$, is an autoregressive process of order p if

$$\frac{Z(t) - \mu}{\sigma} = \sum_{k=1}^{p} \phi_k \frac{Z(t-k) - \mu}{\sigma} + \epsilon(t), \tag{1}$$

where ϕ_k are the model parameters, and μ and σ are the mean and standard deviation of the process. The $\epsilon(t)$ term is assumed to be independent and identically distributed (*iid*) with mean zero and variance $\sigma_\epsilon^2 < \infty$ for $t = 1, \ldots, T$. If the $\epsilon(t)$ also follow a normal distribution, then they are said to be a *Gaussian* white noise process. The model parameters can be estimated by minimizing the sum of squared errors. For a system of reservoirs, a model of type (1) is fitted for each location, and the set of locations is denoted by $\mathbb{I}$ in the remainder of this paper.

We introduce a monthly time index, m, a yearly time index, n, and an $|\mathbb{I}| \times 1$ vector of locations, $\boldsymbol{s}$. We index the inflow at a particular location, s_i, as $Z(s_i; t_{m,n})$. We say the time series $Z(s_i; t_{m,n})$, $i \in \mathbb{I}$, $m = 1, \ldots, 12$, $n = 1, \ldots, N$, is a periodic autoregressive process of orders $p_{i,m}$ if

$$\frac{Z(s_i; t_{m,n}) - \mu_{i,m,n}}{\sigma_{i,m}} = \sum_{k=1}^{p_{i,m}} \phi_{i,k,m} \frac{Z(s_i; t_{m-k,n}) - \mu_{i,m-k,n}}{\sigma_{i,m-k}} + \epsilon(s_i; t_{m,n}). \tag{2}$$

Given a location s_i, each of the twelve months is allowed to have different coefficients, $\phi_{i,k,m}$, and lags, $p_{i,m}$. The previous month's inflow for month $m = 1$, $Z(s_i; t_{1,n})$, is $Z(s_i; t_{12,n-1})$. The mean, $\mu_{i,m,n}$, has a month and a year time index to account for both monthly seasonality (*e.g.*, wet summer months, dry winter months) and yearly seasonality (*e.g.*, El Niño events that occur every 2-12 years). If there is no yearly seasonality present in the time series, the yearly index, n, can be omitted. In this case, $\mu_{i,m}$ and $\sigma_{i,m}$ can be estimated with

$$\hat{\mu}_{i,m} = \frac{1}{N} \sum_{n=1}^{N} Z(s_i; t_{m,n}), \tag{3}$$

and

$$\hat{\sigma}_{i,m} = \sqrt{\frac{1}{N-1} \sum_{n=1}^{N} \left(Z(s_i; t_{m,n}) - \hat{\mu}_{i,m} \right)^2}, \tag{4}$$

and 24 parameters per location would be necessary to deseasonalize the time series. However, a low-frequency persistence can be present in inflow time series [50], so standardizing on monthly data would ignore such year-to-year variability.

Pairs of harmonics can provide a deseasonalization that both uses fewer parameters and can model year-to-year variability by using the fitted values of the following models:

$$\mu_{i,m,n} = \beta^{\mu}_{i,0} + \sum_{l=1}^{L^{\mu}_i} \beta^{\mu}_{i,l,1} \cos(\gamma^{\mu}_l \alpha_{m,n}) + \beta^{\mu}_{i,l,2} \sin(\gamma^{\mu}_l \alpha_{m,n}), \quad (5)$$

and

$$\sigma_{i,m} = \beta^{\sigma}_{i,0} + \sum_{l=1}^{L^{\sigma}_i} \beta^{\sigma}_{i,l,1} \cos(\gamma^{\sigma}_l \alpha_{m}) + \beta^{\sigma}_{i,l,2} \sin(\gamma^{\sigma}_l \alpha_{m}), \quad (6)$$

where $\alpha_{m,n}$ is the angle of month m and year n observed in radians. For example, in a yearly harmonic, we divide the 12 months of the year across 2π radians, and the angles would be $\alpha_{1,n} = \frac{2\pi}{12}$, $\alpha_{2,n} = \frac{4\pi}{12}$, ..., $\alpha_{12,n} = \frac{24\pi}{12}$ for each year n. The regression parameters are denoted $\beta^{\mu}_{i,0}$, $\beta^{\mu}_{i,l,1}$, $\beta^{\mu}_{i,l,2}$, $\beta^{\sigma}_{i,0}$, $\beta^{\sigma}_{i,l,1}$, and $\beta^{\sigma}_{i,l,2}$; L^{μ}_i and L^{σ}_i denote the number of pairs of harmonics used; and γ^{μ}_l and γ^{σ}_l denote the corresponding multipliers. By using the periodogram [51], significant pairs of harmonics can be identified. The idea of the periodogram is to identify values of γ^{μ}_l and γ^{σ}_l that are highly correlated with the observed series. An example is provided in Section 3.1.3. The number of parameters needed for the deseasonalization using pairs of harmonics depends on the number of pairs added since each additional pair of harmonics leads to two additional parameters.

Lags, $p_{i,m}$, of the PAR models can either be determined by consulting partial autocorrelation function (PACF) plots of the corresponding month, or by minimizing an information criterion such as the Akaike or Bayesian information criteria, denoted AIC and BIC, respectively [1, 44]. For this work, we choose the lags of all models using BIC because it penalizes additional parameters in the model more heavily than AIC. This is especially important with regard to the SDDP algorithm discussed in Section 3.3.

3.1.2. Spatial PAR Model

The PAR models described in the previous section are widely used when forecasting inflows for hydrothermal applications because they have appealing properties with respect to the SDDP method. However, they do not take spatial information, such as distance between hydro plants or upstream/downstream relations, into account. Correlation between locations is typically added to the error terms $\epsilon(s_i; t_{m,n})$ at each location when forecasting different inflow scenarios, see Section 3.1.3. We extend the PAR model by directly incorporating a spatial component and, thus, present a SPAR model for each individual location. The idea is that inflows into "neighboring" plants are correlated. This correlation can be accounted for by using neighboring inflow series to explain the inflow at a given location. In order to do so, we extend the PAR framework of (2) following the work of [8] and model it as follows:

$$\frac{Z(s_i; t_{m,n}) - \mu_{i,m,n}}{\sigma_{i,m}} = \sum_{i' \in \mathbb{I}_{i,m}} \sum_{k=1}^{p_{i,i',m}} \phi_{i,i',k,m} \frac{Z(s_{i'}; t_{m-k,n}) - \mu_{i',m-k,n}}{\sigma_{i',m-k}} + \epsilon(s_i; t_{m,n}), \quad \forall i \in \mathbb{I} \quad (7)$$

for $m = 1, \ldots, 12$. Set $\mathbb{I}_{i,m} \subset \mathbb{I}$ describes the set of neighbors chosen in month m that includes location i itself. Neighbors can be selected with respect to distance, upstream/downstream relationship, or expert knowledge, and can be ranked accordingly. For example, the immediate upstream neighbor may be given a rank of 1, the neighbor immediately upstream of the neighbor with rank 1 a rank of 2, and so on (see Figures 4-6 for a visualization). Once neighbor candidates have been ranked, the selection strategy for inclusion in $\mathbb{I}_{i,m}$ is the following:

Step 0: Choose one of the reservoirs, s_i (the order does not matter).

Step 1: Identify $p_{i,i,m}$ by selecting the model whose BIC is the smallest from the set, where $p_{i,i,m} = 1, \ldots, 6$. Repeat for each month, $m = 1, \ldots, 12$.

Step 2: Choose the neighbor with rank 1 as s_{i_1}. Identify $p_{i,i_1,m}$ by selecting the model whose BIC is the smallest from the set where $p_{i,i_1,m} = 1, \ldots, (6 - p_{i,i,m})$, $m = 1, \ldots, 12$, while $p_{i,i,m}$ is fixed from the previous step. Do not add location s_{i_1} if BIC does not decrease.

Step 3: Repeat Step 2 for all neighbor candidates, in order of their ranks, starting with the lowest rank first.

Step 4: Repeat Steps 0 through 3 for each reservoir.

Other algorithms for building the SPAR are possible, but given the number of reservoirs and months, and, therefore, models to build, we require an automated process. The reason we choose the BIC is the same as in

Section 3.1.1. We further limit the number of total lags for a given location and month, *i.e.*, $\sum_{i' \in \mathbb{I}_{i,m}} p_{i,i',m}$, to a maximum of six, as it is crucial for the performance of the SDDP algorithm. For example, if $p_{i,i,m} = 2$ is chosen for location s_i and month m in step 1 of the algorithm above, only $p_{i,i_1,m} = 1, \ldots, 4$ are candidate lags for the first neighbor in step 2. If it is desired to keep the state variables required for the SPAR at the same level as the underlying PAR, a more restrictive neighbor selection scheme can be used. In this case, a lagged neighbor's time series would only be added if that neighbor's PAR also contains it.

3.1.3. From Data to Forecasted Inflow Scenarios

In this section, we explain how to build inflow scenarios that can be used as input for the SDDP algorithm, starting with the original inflow data set.

Training and Testing Data Sets: We assume the data set is in a space-time format, *i.e.*, historical inflow time series are given for a set of locations, $\mathbb{I}$. We further assume that the inflow time series have a monthly time resolution, are complete, *i.e.*, no entries are missing, and span a number of N years. The operator of the Brazilian power system, ONS, provides so-called natural and artificial time series [31]. Natural time series represent the observed inflows at fluviometer stations whereas artificial time series are calculated for locations in complicated cascade schemes in which multiple locations are very close to each other and are potentially connected by more than one flow path. We ignore the artificial inflows and only consider natural inflows in the following.

We create incremental inflows by subtracting the natural inflow time series of immediately upstream locations from each location's time series. The resulting time series only accounts for that part of the river upstream of the plant and not the entire river. Consider the cascade in Figure 5 as an example. In order to calculate location **(172)**'s incremental inflow, we subtract **(169)**'s natural inflow from **(172)**'s natural inflow. Then, to calculate **(169)**'s incremental inflow, we subtract both **(156)**'s and **(158)**'s natural inflows from **(169)**'s natural inflow. Similar calculations are performed to obtain incremental inflows at every location. The next step is to divide the data set into a training and a testing set. The training set is used to fit the forecasting models whereas the testing set is used to evaluate the quality of the fitted models. A rule of thumb is to use the first two thirds of the data set as the training, and the remaining third as the testing set. We denote the length in months of the training and testing set as T^{train} and T^{test}, respectively, and $T^{\text{train}}/12 + T^{\text{test}}/12 = N$.

Seasonal Detrending: Until we begin forecasting, we only work with the training data set. For convenience, we omit the term "training." Once the training set has been obtained, we need to account for seasonality, typically present in inflow time series. The output of this step is a detrended data set in which seasonality has been removed. We assume that the detrended data are stationary, meaning that the monthly means are constant, and the covariance between any two observations depends only on the length of time between them. In the following, we discuss two approaches to detrend the data.

The first approach using monthly means (3) and standard deviations (4) as introduced in Section 3.1.1 is the most popular and straight-forward one. These monthly summaries have to be determined for each of the locations in $\mathbb{I}$, leading to a total of $2 \cdot 12 \cdot |\mathbb{I}|$ parameters. Although this approach is very simple, it has proven to work well in practice [68]. The obvious disadvantages of these monthly summaries are that i) a large number of parameters is necessary, ii) only monthly but not yearly seasonality can be modeled, and iii) these statistics are not robust to outliers. One or two extremely high inflow values for a given month in the data set could lead to an overestimation of that month's mean and standard deviation.

The second approach, the use of pairs of harmonics, overcomes the aforementioned drawbacks of monthly summaries. Two parameters are required per pair of harmonics, and only a few pairs of harmonics are typically needed for each series. Furthermore, they can account for yearly seasonality and are not as sensitive to outliers. As an example, a seasonality occurring every eight years can be represented by adding a pair of harmonics spanning eight years. We can compare the two approaches. If two pairs of harmonics for the mean and one pair of harmonics for the standard deviation are sufficient, 8 parameters are needed in total using (5)-(6) compared to 24 parameters using (3)-(4). The clear disadvantage of this method is the challenge of selecting significant pairs of harmonics. A tool that can aid in this process is the aforementioned periodogram [51]. Figure 1 shows the periodogram for the incremental inflow at a given location. The peaks at 5, 1, and $\frac{1}{2}$ are indicators that the corresponding pairs of harmonics might be useful to detrend the time series. Once the pairs of harmonics for the mean have been selected, their coefficients, $\boldsymbol{\beta}$, can be determined through a least squares approach. The fitted harmonic model yields a vector of T^{train} monthly values that are subtracted from the original incremental inflows. The resulting detrended time series has a mean of approximately 0, but its monthly

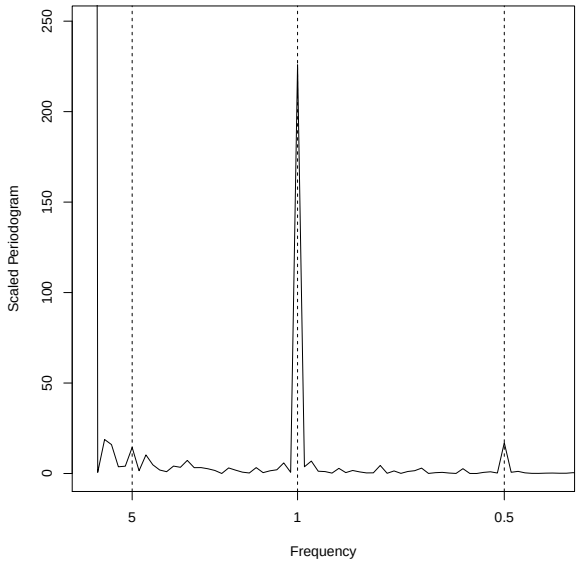

Figure 1: Periodogram showing frequencies for an incremental inflow time series. A yearly seasonality (=1) displays the highest peak. Other noteworthy frequencies represent a 5-yearly and a $\frac{1}{2}$-yearly seasonality.

standard deviations are not yet equal to 1. This can be achieved by either using the monthly standard deviation or using pairs of harmonics for the standard deviation as well. The harmonic model for the standard deviations is fitted based on 12 standard deviations, one computed from the observations for each month. In this case, we do not use the periodogram to choose potential pairs of harmonics. Instead, we simply plot the 12 values and determine the pairs to use accordingly. The BIC may be applied to decide between harmonics and monthly summaries. A mix of the two approaches is possible as well, *e.g.*, harmonics for the mean and monthly summaries for standard deviation.

PAR Model: The output of the previous step is a detrended time series that should show no seasonality. We can consult autocorrelation function (ACF) plots of the detrended time series to check for this. A desired ACF plot of a detrended series that follows an autocorrelated but stationary process can be seen in Figure 2 (right). The temporal dependence decreases with the lags, and there are no repeating patterns as there are for the original series (Figure 2, left). At this point an autoregressive model can be fitted to explain the remaining autocorrelation. In particular, PAR models have proven useful for inflow time series. Fitting a PAR model can be seen as fitting an AR model for each month separately. As an example, to fit an AR(p_1) model for month January, we discard the first year of data and collect all remaining detrended months that belong to a January. For our training set, the number of obtained "Januaries" is $T^{\text{train}}/12 - 1$. We then collect all "Decembers" preceding these "Januaries." Thus, we use the December of the first year and discard the last year's December. Again, the number of obtained "Decembers" is $T^{\text{train}}/12 - 1$. An AR($p_1 = 1$) model tries to explain each January using its preceding December. We continue in a similar fashion to fit AR models for $p_1 = 2$ through $p_1 = 6$ and compute the BIC for each of the six models. We then pick the model with the lowest BIC and repeat this process for all twelve months to fit PAR model (2) for each location separately.

Space Time Index (STI): At this point, most of the temporal dependency in the inflow series should be removed. We first removed seasonality by either using monthly summaries or harmonics, and then used a PAR model to remove autocorrelation in the detrended series, which leaves us with estimated error terms. These residuals from (2) are denoted by $\hat{\epsilon}(s_i; t_{m,n})$. The question is whether there is spatial dependence in these residuals across the set of locations. A statistical test can be carried out to check for significant spatial dependence. In particular, the following space-time index statistic,

$$\text{STI} := \frac{\sum_{t=2}^{T^{\text{train}}} \sum_{j\in\mathbb{I}} \sum_{i\in\mathbb{I}} \omega_{i,j,t-1}\hat{\epsilon}(s_i; t_{m,n})}{|\mathbb{I}|(T^{\text{train}}-1)\left(\sum_{t=2}^{T^{\text{train}}} \sum_{j\in\mathbb{I}} \sum_{i\in\mathbb{I}} \omega_{i,j,t-1}\right)} \cdot \frac{\hat{\epsilon}(s_j; t_{m,n})}{\left(\sum_{t=1}^{T^{\text{train}}} \sum_{i\in\mathbb{I}} \hat{\epsilon}(s_i; t_{m,n})^2\right)}, \tag{8}$$

can be used to test the null hypothesis that no space-time autocorrelation is present, where the $\omega_{i,j,t}$ are weights describing the neighbor relationship between locations i and j at time t [6]. We assume that the weights do not change over time, so $\omega_{i,j,t} = \omega_{i,j}\ \forall t \in 1, \ldots, T^{\text{train}}$. The permutation distribution of (8) under the null hypothesis has to be generated, which is computationally intractable for a realistic number of time periods and locations. However, it is sufficient to draw a large enough sample ($\approx$ 10,000) using a random number generator, permuting $\hat{\epsilon}(s_i; t_{m,n})$ across locations and times, and comparing the observed STI with those values obtained by random permutations. Figure 3 shows the permutation distribution and the observed STI. Assuming that most of the temporal dependence has been removed in prior steps, a remaining spatial dependence is evident, given the p-value of 0.0138.

SPAR Model: In the previous step, we have seen that there is an unaccounted spatial component in the residuals of the PAR model. There are two options.

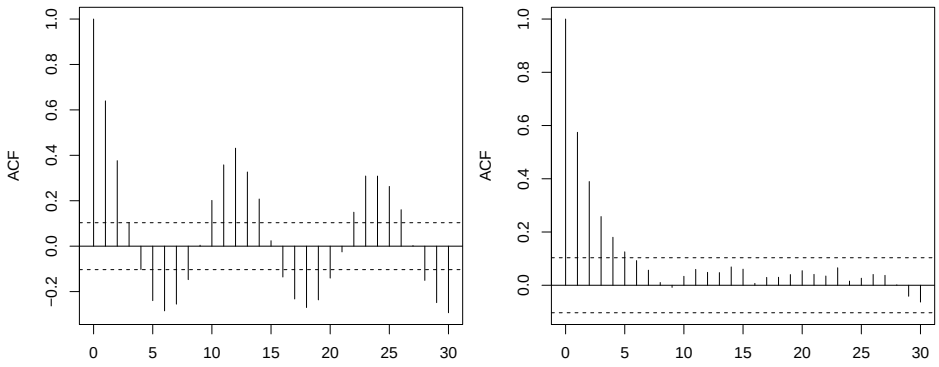

Figure 2: Example ACF plot before (left) and after (right) removal of temporal seasonality.

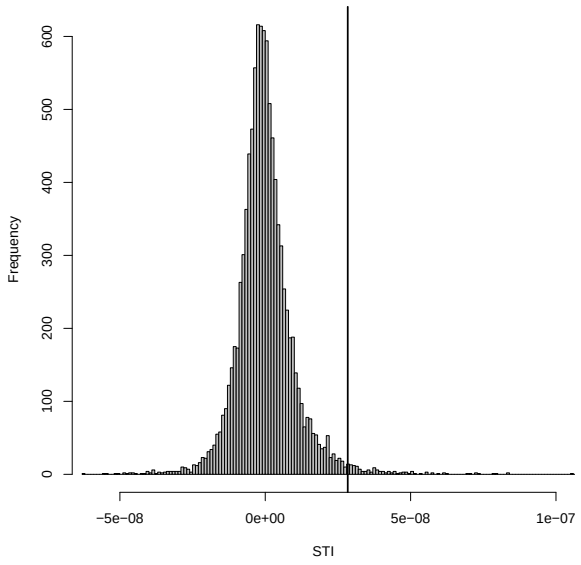

Figure 3: Histogram of the STI permutation distribution with the observed statistic STI marked by the vertical line. Number of simulations is 10,000, and the observed STI has a p-value of 0.0138.

We can ignore the spatial dependence, forecast the time series, and then impose correlation between locations later when generating scenarios for the SDDP. This may not necessarily improve the forecasts. In addition, the strength of the spatial correlations may differ depending on how the reservoirs are related to each other, not just their distance from one another, making it more difficult to find a realistic spatial covariance model. The second option is to treat this dependence explicitly in the model. Since forecasts are only desired for given locations (and not in between locations), constructing a full spatial covariance model is not necessary. The latter is the approach discussed here.

The "art" in setting up the SPAR model is the neighbor candidate selection mentioned in Section 3.1.2. We experimented with the following two options and ultimately decided to use the latter: neighbors based on i) distance between locations or ii) the upstream relation of locations, see Figures 4–6. Let us consider Figure 5 as an example. Starting with the most downstream location in the cascade, location **(172)**, we assign **(169)** as rank 1 neighbor, **(156)** as rank 2 neighbor, **(155)** as rank 3 neighbor, and **(158)** as rank 4 neighbor. Moving up the river, location **(169)** has three neighbors: **(156)** with rank 1, **(155)** with rank 2, **(158)** with rank 3. In our case, locations **(155)** and **(158)** have no upstream locations and, therefore, no potential neighbors in the SPAR. However, Brazil's hydro plants in total are divided into 15 basins, and the outflows of some of these basins serve as inflows into downstream basins. In this case, upstream neighbor candidate selection can extend into other basins. This is even more apparent when using distance to select neighbors because some of the basins are geographically close to each other. Once neighbor candidates have been chosen, we follow the five-step variable selection strategy for $\mathbb{I}_{i,m}$ presented in Section 3.1.2. The

process is similar to that for the PAR model. The difference is that instead of only considering "Decembers," "Novembers," etc. to explain the "Januaries" of location i, we consider "Decembers," "Novembers," etc. of selected neighbors as well.

Point Forecasting: At this point, the model has been fitted, and we are ready to forecast using the testing set. Note that we fitted one model of either type (2) or (7) and will use this to construct the entire forecast. We did not fit different models for different month-ahead forecasts. Conceptually, the SDDP algorithm would allow this, but it requires a significantly increased number of parameters to estimate, and our goal has been to decrease this number. Also, it typically requires a much larger amount of historical data than what we have available. Furthermore, a multi-model approach is very uncommon in mid-term, *i.e.*, three to five years, hydro inflow forecasting. On the other hand, wind forecasting is a very common application for a multi-model methodology [22]. However, most importantly, we focus on one-month ahead forecasts in this paper.

The general idea of forecasting with the same model is as follows: We are at the end of the training data period and wish to forecast the next unknown month. In our case, December is the last month in the training set, so we wish to forecast an inflow value for January of the following year. The first step is to use the SPAR (or PAR) model to forecast the detrended value for that January using the detrended values of prior months (which we have observed, *i.e.*, they are known). The error term in (2) or (7) can be ignored because it is assumed to have mean 0. Once we have forecast the detrended value for January, we multiply this by $\hat{\sigma}_{i,1}$, obtained either using (4) or (6). Then, we add $\hat{\mu}_{i,1}$, obtained either using (3) or (5). The resulting value is the forecast one-month ahead for January. In order to forecast two-months ahead, in this example February, we cannot use a known detrended value for January anymore, because January lies in the future as well. Instead, we use the one-month ahead value that has been forecast for January. In a similar fashion, all desired month-ahead forecasts can be obtained. Note that the forecasts of the detrended values go to 0 for higher month-ahead forecasts, so the forecasts converge to the monthly mean (3) or (5). This is due to the fact that less information is available the further in the future we wish to forecast.

Scenario Forecasting: The SDDP method requires not only one forecast, but a series of forecasts, which are called scenarios. While the SPAR model incorporates spatial dependence explicitly at the model level (7), spatial information in the PAR must be induced upon its errors in (2) when constructing forecasting scenarios. In both cases, the error term in (2) or (7) is used to build scenarios. Let us first consider the PAR model (2). We represent the dependence between locations using an $|\mathbb{I}| \times |\mathbb{I}|$ spatial covariance matrix, denoted by C. This matrix is typically estimated by computing the observed pairwise covariances [37], but it contains a large number of unknowns that must be estimated, namely $(|\mathbb{I}|^2 + |\mathbb{I}|)/2$, so it should be modeled with a spatial covariance based on distances, such as in [57]. This is a nontrivial task with such a complex and sparse network of reservoirs. Assuming that we have such an estimate of C, to correlate the error terms, we require a function of the spatial covariance, denoted by A. We denote the variance of a $|\mathbb{I}| \times 1$ sample, $\bar{\epsilon}$, from the processes in (2) with Λ, which is a diagonal matrix with diagonal entry $\hat{\sigma}^2_{\epsilon(s_i;m,n)}$. Thus, $Var(\bar{\epsilon}) = \Lambda = \Lambda^{1/2}\Lambda^{1/2}$, and we need to find A such that $Var(A\bar{\epsilon}) = A\Lambda A^\top = C$. This can be achieved with a spectral decomposition of C in which $C = PLP^\top$ for P the matrix of eigenvectors of C and L a matrix with the corresponding eigenvalues on the main diagonal [21]. Then, $A = PL^{1/2}\Lambda^{-1/2}$. One month-ahead forecast scenarios from PAR model (2) can then be obtained as follows:

1: **for** each scenario $r = 1, \ldots, R$ **do**
2: **for** each location $i \in \mathbb{I}$ **do**
3: Sample a residual, $\bar{\epsilon}_{i,r}$, from locations i's estimated noise process, $\hat{\epsilon}(s_i; t_{m,n})$, in (2).
4: **end for**
5: **for** each location $i \in \mathbb{I}$ **do**
6: Correlate the sample residual, $\epsilon_{i,r} = \sum_{i' \in \mathbb{I}} A_{ii'} \bar{\epsilon}_{i',r}$, using the matrix A.
7: Calculate a forecast scenario using residual $\epsilon_{i,r}$ and the estimated parameters of (2).
8: **end for**
9: **end for**

We obtain R realizations of ϵ for the first stage and, thus, R scenarios. We treat each of these R scenarios as plausible one-month ahead forecasts and calculate R second-stage forecast scenarios for each of them, leading to a total of R^2 inflow scenarios in the second stage. This can be done by using the same R realizations of ϵ in the second stage for each scenario. By continuing in this fashion, we could construct an exponentially growing scenario tree. However, the SDDP method discussed in Section 3.3 does not require construction or storage of the full scenario tree. Instead, it suffices to store the R realizations per stage for the case of interstage independence, discussed in Section 3.3.1. For the case of interstage dependence, discussed in Section 3.3.2, it suffices to store the R realizations of ϵ. A path in the scenario tree for the SDDP algorithm can then be con-

structed based on the appropriate history of the process and the realizations of ϵ.

For the SPAR model (7), this process is much simpler because spatial covariance between locations is explicitly accounted for by the model and does not have to be estimated with C, and the initially sampled error terms, $\bar{\epsilon}_{i,s}$, can remain independent. The algorithm then reads:

1: **for** each scenario $s = 1, \ldots, S$ **do**
2: **for** each reservoir $i \in \mathbb{I}$ **do**
3: Sample a residual, $\bar{\epsilon}_{i,s}$, from location i's estimated noise process, $\hat{\epsilon}(s_i; t_{m,n})$, in (7).
4: Calculate a forecast scenario using residual $\bar{\epsilon}_{i,s}$ and the estimated parameters of (7).
5: **end for**
6: **end for**

The required scenario tree can be constructed in the same way as for the PAR model above.

3.2. Error Measures

There are many error measures in use to evaluate the quality of a forecasted inflow time series compared to the actual, observed time series [61]. In this paper, we focus on the following two: the root mean squared error (RMSE) and the seasonally-adjusted coefficient of efficiency (SACE). The $RMSE^{(k)}$ of the k month-ahead forecast for all reservoirs is computed as follows

$$RMSE^{(k)} := \sqrt{\frac{1}{|\mathbb{I}| \cdot M \cdot N} \sum_{i,m,n} \left(Z(s_i; t_{m,n}) - \hat{Z}^{(k)}(s_i; t_{m,n})\right)^2}, \quad (9)$$

where $\hat{Z}^{(k)}(s_i; t_{m,n})$ denotes the k month-ahead forecast at location i in month m of year n. Similarly, the $RMSE^{(i,k)}$ of the k month-ahead forecast for location i can be obtained as

$$RMSE^{(i,k)} := \sqrt{\frac{1}{M \cdot N} \sum_{m,n} \left(Z(s_i; t_{m,n}) - \hat{Z}^{(k)}(s_i; t_{m,n})\right)^2}. \quad (10)$$

The RMSE penalizes overestimating and underestimating forecasts equally, and its unit is the same as that of the time series. A forecast that is close to the actual observed time series is represented by a small RMSE relative to the magnitude of the time series.

Another forecast evaluation tool is the SACE, which compares model forecasts to the monthly mean. The $SACE^{(i,k)}$ of the k-month ahead forecast for location i is computed as follows:

$$SACE^{(i,k)} := 1 - \frac{\sum_{m,n} \left(Z(s_i; t_{m,n}) - \hat{Z}^{(k)}(s_i; t_{m,n})\right)^2}{\sum_{m,n} \left(Z(s_i; t_{m,n}) - \mu_{i,m})\right)^2}. \quad (11)$$

It is unit-less and, therefore, well suited to compare model performances. The best achievable SACE is 1, meaning that forecasted and observed inflow series are identical. An SACE of 0 means that the forecasted series on average performs as well as simply using monthly means. It follows that a negative SACE implies that the model performs worse than the monthly means. An overall SACE for all locations can be calculated as

$$SACE^{(k)} := \frac{1}{|\mathbb{I}|} \sum_{i \in \mathbb{I}} SACE^{(i,k)}.$$

3.3. Stochastic Dual Dynamic Programming and Cut Derivation

In this section, we describe how the SDDP algorithm must be adjusted to use forecasted inflow scenarios from SPAR model (7). First, we present an SLP model for least-cost hydro-thermal scheduling problems with uncertain hydro inflows. Second, we briefly describe the SDDP method for that problem. Third, we derive the necessary modifications to the Benders optimality cuts used in SDDP in order to work with our SPAR.

We assume that the uncertain inflows can be characterized by a discrete distribution with a finite number of outcomes. We also assume that we have full information on the distribution. This allows us to collect all possible inflow scenarios s in set $\mathbb{S}$ over T stages, *i.e.*, the inflow into reservoir $i \in \mathbb{I}$ at stage $t \in \mathbb{T}$ for scenario $s \in \mathbb{S}$ is $a_{i,s,t}$ (or $\boldsymbol{a}_{s,t}$ in vector notation); we measure the inflows in MWh. The probability of the occurrence of scenario s is denoted by p_s. Typically, the collection of possible inflow scenarios build a scenario tree; to that end, we denote by set $\mathbb{S}_{s,t}$ the collection of all scenarios $\tilde{s} \in \mathbb{S}$ that are identical to scenario s in stage t. Instead of using $Z(s_i; t_{m,n})$ to describe the inflow time series, we use $|\mathbb{I}| \times 1$ column vectors $\boldsymbol{a}_{s,t}$ to be consistent with the hydro-thermal scheduling literature.

A T-stage SLP for the least-cost stochastic hydro-thermal scheduling problem in the so-called extensive

form can be stated as follows

$$z^* := \min \sum_{s\in\mathbb{S}} p_s \sum_{t\in\mathbb{T}} \boldsymbol{c}_t^\top \boldsymbol{g}_{s,t} \tag{12}$$

$$\text{s.t. } \mathbf{1}^\top \boldsymbol{g}_{s,t} + \boldsymbol{\rho}^\top \boldsymbol{u}_{s,t} = d_t \quad \forall s \in \mathbb{S},\ t \in \mathbb{T} \tag{13}$$

$$v_{i,s,t} = v_{i,s,t-1} - u_{i,s,t} - w_{i,s,t} + \sum_{i'\in\mathbb{U}_i} (u_{i',s,t} + w_{i',s,t}) + a_{i,s,t} \quad \forall i \in \mathbb{I},\ s \in \mathbb{S},\ t \in \mathbb{T} \tag{14}$$

$$g_{s,t} = g_{\tilde{s},t}, \quad w_{s,t} = w_{\tilde{s},t}, \quad u_{s,t} = u_{\tilde{s},t}, \quad v_{s,t} = v_{\tilde{s},t} \quad \forall s \in \mathbb{S},\ \tilde{s} \in \mathbb{S}_{s,t} \tag{15}$$

$$\underline{\boldsymbol{G}}_t \le \boldsymbol{g}_{s,t} \le \overline{\boldsymbol{G}}_t, \quad \underline{\boldsymbol{W}}_t \le \boldsymbol{w}_{s,t} \le \overline{\boldsymbol{W}}_t, \quad \underline{\boldsymbol{U}}_t \le \boldsymbol{u}_{s,t} \le \overline{\boldsymbol{U}}_t, \quad \underline{\boldsymbol{V}}_t \le \boldsymbol{v}_{s,t} \le \overline{\boldsymbol{V}}_t \quad \forall t \in \mathbb{T},\ s \in \mathbb{S}, \tag{16}$$

where

$\mathbb{U}_i$: (set) set of reservoirs immediately upstream of reservoir i

$\mathbf{1}$: (data) vector of ones [-]

$\boldsymbol{c}_t$: (data) marginal cost vector for thermal plants in stage t [\$/MWh]

d_t : (data) electricity demand in stage t [MWh]

$\boldsymbol{\rho}$: (data) efficiency coefficient vector for hydro plants [100%]

$\underline{\boldsymbol{G}}_t, \overline{\boldsymbol{G}}_t$: (data) lower and upper thermal generation limit vectors in stage t [MWh]

$\underline{\boldsymbol{W}}_t, \overline{\boldsymbol{W}}_t$: (data) lower and upper spillage limit vectors in stage t [MWh]

$\underline{\boldsymbol{U}}_t, \overline{\boldsymbol{U}}_t$: (data) lower and upper hydro generation limit vectors in stage t [MWh]

$\underline{\boldsymbol{V}}_t, \overline{\boldsymbol{V}}_t$: (data) lower and upper volume storage limit vectors in stage t [MWh]

$\boldsymbol{g}_{s,t}$: (decision variable) thermal generation vector in stage t for scenario s [MWh]

$\boldsymbol{w}_{s,t}$: (decision variable) spillage vector in stage t for scenario s [MWh]

$\boldsymbol{u}_{s,t}$: (decision variable) hydro generation vector in stage t for scenario s [MWh]

$\boldsymbol{v}_{s,t}$: (decision variable) water level vector for hydro plants at the end of stage t for scenario s [MWh].

The objective (12) minimizes (expected) power generation cost of the thermal plants. Demand constraints (13) ensure that monthly electricity demand is satisfied for all inflow scenarios. Constraints (14) describe the water balance in a given month t for scenario s: the water level in reservoir i at the end of month t is the water level at the end of the previous month minus water used for generation and spillage, plus water used for generation and spillage from the reservoirs immediately upstream, plus inflows (for that scenario s). Constraints (15) are the so-called nonanticipativity constraints which ensure that decisions for non-distinct scenarios (in the scenario tree) are identical. Constraints (16) describe lower and upper bounds for all variables. Typically, first-stage decisions are assumed to be deterministic, *i.e.*, there is only one distinct scenario s for stage 1, and we denote $\boldsymbol{g}_1 \equiv \boldsymbol{g}_{s,1}$, $\boldsymbol{w}_1 \equiv \boldsymbol{w}_{s,1}$, $\boldsymbol{u}_1 \equiv \boldsymbol{u}_{s,1}$, and $\boldsymbol{v}_1 \equiv \boldsymbol{v}_{s,1}$.

Since LP problems of type (12)-(16) tend to be very large, decomposition methods are used to solve these problems – SDDP is such a method. The general idea of the SDDP method is to decompose the original T-stage SLP (12)-(16) into multiple one-stage subproblems that are "connected" via the water balance equations (14). Therefore, we define the *t-stage value function* ($t \in \mathbb{T},\ s \in \mathbb{S}$) as

$$Q_{s,t}(\boldsymbol{v}_{t-1}) := \min \ \boldsymbol{c}_t^\top \boldsymbol{g}_t + \sum_{\tilde{s}\in\tilde{\mathbb{S}}_{s,t}} p_{s,\tilde{s},t+1} Q_{\tilde{s},t+1}(\boldsymbol{v}_t) \tag{17}$$

$$\text{s.t. } \mathbf{1}^\top \boldsymbol{g}_t + \boldsymbol{\rho}^\top \boldsymbol{u}_t = d_t \tag{18}$$

$$v_{i,t} = v_{i,t-1} - u_{i,t} - w_{i,t} + \sum_{i'\in\mathbb{U}_i} (u_{i',t} + w_{i',t}) + a_{i,s,t} \quad \forall i \in \mathbb{I} \tag{19}$$

$$\underline{\boldsymbol{G}}_t \le \boldsymbol{g}_t \le \overline{\boldsymbol{G}}_t, \quad \underline{\boldsymbol{W}}_t \le \boldsymbol{w}_t \le \overline{\boldsymbol{W}}_t, \quad \underline{\boldsymbol{U}}_t \le \boldsymbol{u}_t \le \overline{\boldsymbol{U}}_t, \quad \underline{\boldsymbol{V}}_t \le \boldsymbol{v}_t \le \overline{\boldsymbol{V}}_t \tag{20}$$

where we have dropped the scenario index, "s", for the decision variables and where $p_{s,\tilde{s},t}$ denotes the conditional probability of the occurrence of scenario $\tilde{s}$ at stage t, given scenario s (with $\sum_{\tilde{s}\in\tilde{\mathbb{S}}_{s,t}} p_{s,\tilde{s},t} = 1$). Set $\tilde{\mathbb{S}}_{s,t}$ contains those scenarios which are duplicates of scenario s in stage t but differ in stage $t+1$; $\tilde{\mathbb{S}}_{s,t} \subset \mathbb{S}_{s,t}$. For the final stage, T, we use the convention $Q_{s,T+1}(\cdot) \equiv 0$ for all $s \in \mathbb{S}$. For initial water reservoir levels $\boldsymbol{v}_0$, we call the first stage problem the *Master Problem* and $Q_1(\boldsymbol{v}_0) \equiv Q_{s,1}(\boldsymbol{v}_0) = z^*$ holds.

Because $\boldsymbol{v}_{s,t-1}$ is a constant for the t-stage value function that appears only in the right-hand-side of the water balance constraints (19), functions $Q_{s,t}(\cdot)$ are convex in $\boldsymbol{v}_{s,t-1}$. The SDDP algorithm underestimates this function by constructing a piecewise-linear function using hyperplanes, typically referred to as Benders (optimality) cuts, in an iterative fashion. Therefore, the SDDP is a nested Benders decomposition (NBD) type algorithm and relies on the convexity of $Q_{s,t}(\cdot)$.

In general NBD algorithms (we point out the difference to SDDP in the next paragraph), the piece-wise

linear underestimation of function $Q_{s,t}(\cdot)$ is constructed in the backward step in which problems (17)-(20) are solved for $t = T, \dots, 2$ and $s \in \mathbb{S}$. In the forward step, (17)-(20) are solved for $t \in \mathbb{T}$ and $s \in \mathbb{S}$, constructing a feasible solution to the original problem (12)-(16). The objective function value of such a feasible solution is an upper bound on z^*. Because of the underestimation of the functions $Q_{s,t}(\cdot)$, the first stage problem is the so-called *relaxed* Master Problem. It provides a lower bound on z^*. NBD converges once the difference in lower and upper bounds reaches a user-defined tolerance.

SDDP differs from the NBD described above in that SDDP considers only a subset of all possible scenarios $\mathbb{S}$ in the forward step. The chosen scenarios are the so-called sample paths. Consequently, the computed upper bound is no longer exact, and a statistical convergence criterion might be employed. We refer to the literature for a more detailed description of NBD and the SDDP algorithm specifically [33, 41].

Besides the forward step, the backward step in the SDDP algorithms also differs from the NBD algorithm described above. SDDP is tailored to problems where the scenarios, and with that also the scenario tree, have a special structure. This leads us to the interstage independent case next. After that, we finally present the interstage dependent case which allows the incorporation of our SPAR model into SDDP algorithms.

3.3.1. Interstage Independent Case

Interstage independence of the scenarios $s \in \mathbb{S}$ means that inflows $a_{i,s,t}$ are independent from the inflows of previous stages. If we assume that each stage ($t = 2, \dots, T$) has R different realizations $a_{i,r,t}$, then there are only R different t-stage value functions per stage $t = 2, \dots, T$. Note, however, the scenario tree consists of R^{T-1} scenarios. For $r = 1, \dots, R$, the *approximated* t-stage value function then reads

$$\begin{aligned}
\underline{Q}_{r,t}(\boldsymbol{v}_{t-1}) := \min \quad & \boldsymbol{c}_t^\top \boldsymbol{g}_t + \eta_t \\
\text{s.t. } & v_{i,t} = v_{i,t-1} - u_{i,t} - w_{i,t} + \sum_{i' \in \mathbb{U}_i} (u_{i',t} + w_{i',t}) \\
& \qquad + a_{i,r,t} \quad \forall i \in \mathbb{I} && (21) \\
& -(\boldsymbol{\gamma}^v_{f,t})^\top \boldsymbol{v}_t + \eta_t \geq \gamma^c_{f,t} \quad \forall f \in \mathbb{F} && (22) \\
& (18), (20) && (23)
\end{aligned}$$

where constraints (22) represent the Benders optimality cuts (which we describe in detail below). We denote the dual variable vectors corresponding to constraints (21), (22), and (23) as $\boldsymbol{\pi}^a_t$, $\boldsymbol{\pi}^\gamma_t$, and $\boldsymbol{\pi}^b_t$. Further, we define

$$\boldsymbol{b}_t := (d_t, \underline{\boldsymbol{G}}_t, \overline{\boldsymbol{G}}_t, \underline{\boldsymbol{U}}_t, \overline{\boldsymbol{U}}_t, \underline{\boldsymbol{V}}_t, \overline{\boldsymbol{V}}_t, \underline{\boldsymbol{W}}_t, \overline{\boldsymbol{W}}_t)^\top$$

and collect all Benders optimality cuts in set $\mathbb{F}$. The cut coefficient vector, $\boldsymbol{\gamma}^v_{f,t}$, and cut constant, $\gamma^c_{f,t}$, $t = 1, \dots, T-1$, for Benders cut $f \in \mathbb{F}$ are calculated as follows:

$$\begin{aligned}
\boldsymbol{\gamma}^v_{f,t} &= \sum_{r=1}^{R} p_{r,t+1} \boldsymbol{\pi}^a_{r,t+1} \\
\gamma^c_{f,t} &= \sum_{r=1}^{R} p_{r,t+1} \big[(\boldsymbol{\pi}^a_{r,t+1})^\top \boldsymbol{a}_{r,t+1} + (\boldsymbol{\pi}^b_{r,t+1})^\top \boldsymbol{b}_{t+1} \\
&\qquad + (\boldsymbol{\pi}^\gamma_{r,t+1})^\top \boldsymbol{\gamma}^c_{t+1} \big], \qquad (24)
\end{aligned}$$

where $p_{r,t+1}$ denotes the probability of realization r for stage $t+1$ with the convention that $(\boldsymbol{\pi}^\gamma_{r,T})^\top \boldsymbol{\gamma}^c_T \equiv 0$. Equivalently, constant $\gamma^c_{f,t}$ can be calculated through

$$\gamma^c_{f,t} = -(\boldsymbol{\gamma}^v_{f,t})^\top \boldsymbol{v}_t + \sum_{r=1}^{R} p_{r,t+1} \underline{Q}_{r,t+1}(\boldsymbol{v}_t).$$

3.3.2. Interstage Dependent Case

If the hydro inflows are not stage-wise independent, function $Q_{s,t}(\boldsymbol{v_{t-1}})$ depends on the previous inflows. To show this dependency, we denote the t-stage value function as $Q_{s,t}(\boldsymbol{v}_{t-1}, \boldsymbol{a}_{t-1}, \boldsymbol{a}_{t-2}, \dots, \boldsymbol{a}_{t-p})$ for some $p \geq 1$. To guarantee the convexity of this t-stage value function jointly in all arguments, strong assumptions are necessary on the underlying inflow models. Currently, only linear additive models, such as an AR or PAR model, are known to conserve convexity. The cut constant (24) has to be adjusted accordingly.

We follow the approach presented in [19] to handle stage-wise dependency. Let us assume that the stage-wise dependency of inflows is modeled with a PAR model as described in (2). We omit the deseasonalization parameters and state

$$\boldsymbol{a}_t = \sum_{k=1}^{p} \operatorname{diag}(\boldsymbol{\phi}_k) \boldsymbol{a}_{t-k} + \boldsymbol{\epsilon}_t, \qquad (25)$$

where $\operatorname{diag}(\boldsymbol{\phi}_k)$, $k = 1, \dots, p$, are $|\mathbb{I}| \times |\mathbb{I}|$ matrices with model coefficients $\phi_{i,k,m}$ on the main diagonal. Note that some of these entries are allowed to be zero. In fact, (25) depends on the month that is associated with t, thus, we would have twelve models of type (25).

The adjusted Benders cut then reads as follows:

$$-(\boldsymbol{\gamma}^v_{f,t})^\top \boldsymbol{v}_t - \sum_{k=1}^{p} (\boldsymbol{\gamma}^a_{k,f,t})^\top \boldsymbol{a}_{t-k+1} + \eta_t \geq \tilde{\gamma}^c_{f,t}. \qquad (26)$$

In remains to show how cut coefficients $\boldsymbol{\gamma}^a_{k,f,t}$ and cut constant $\tilde{\gamma}^c_{f,t}$ are calculated.

Therefore, for scenario $s \in \mathbb{S}$ and stage $t = 2, \ldots, T-1$, cut (26) approximates (more precisely, underestimates) the multivariate function

$$\mathbb{Q}_{s,t+1}(\boldsymbol{v}_t, \boldsymbol{a}_t, \ldots, \boldsymbol{a}_{t-p+1}) := \sum_{\tilde{s} \in \tilde{\mathbb{S}}_{s,t}} p_{s,\tilde{s},t+1} Q_{\tilde{s},t+1}(\boldsymbol{v}_t, \boldsymbol{a}_t, \ldots, \boldsymbol{a}_{t-p+1}). \quad (27)$$

To derive the cut coefficients in (26), we expand the objective function associated with (27) in dual form (of the approximated problem, *e.g.*, we replace $\mathbb{Q}_{s,t+1}(\cdot)$ by $\underline{\mathbb{Q}}_{s,t+1}(\cdot)$ and $Q_{\tilde{s},t+1}(\cdot)$ by $\underline{Q}_{\tilde{s},t+1}(\cdot)$), by recognizing the zero duality gap of LPs, to obtain

$$\begin{aligned}
&\underline{\mathbb{Q}}_{s,t+1}(\boldsymbol{v}_t, \boldsymbol{a}_t, \ldots, \boldsymbol{a}_{t-p+1}) \\
&= \sum_{\tilde{s} \in \tilde{\mathbb{S}}_{s,t}} p_{s,\tilde{s},t+1} \Bigg[(\boldsymbol{\pi}^a_{\tilde{s},t+1})^\top (\boldsymbol{a}_{\tilde{s},t+1} + \boldsymbol{v}_t) + \sum_{f \in \mathbb{F}} \pi^\gamma_{f,\tilde{s},t+1} \Bigg(\\
&\quad (\boldsymbol{\gamma}^a_{1,f,t+1})^\top \boldsymbol{a}_{\tilde{s},t+1} + \sum_{k=2}^{p} (\boldsymbol{\gamma}^a_{k,f,t+1})^\top \boldsymbol{a}_{t-k+2} + \tilde{\gamma}^c_{f,t+1} \Bigg) \\
&\quad + (\boldsymbol{\pi}^b_{\tilde{s},t+1})^\top \boldsymbol{b}_{t+1} \Bigg] \\
&= \sum_{\tilde{s} \in \tilde{\mathbb{S}}_{s,t}} p_{s,\tilde{s},t+1} \Bigg[\Bigg(\boldsymbol{\pi}^a_{\tilde{s},t+1} + \sum_{f \in \mathbb{F}} \pi^\gamma_{f,\tilde{s},t+1} \boldsymbol{\gamma}^a_{1,f,t+1} \Bigg)^\top \Bigg(\\
&\quad \sum_{k=1}^{p} \mathrm{diag}(\boldsymbol{\phi}_k) \boldsymbol{a}_{t-k+1} + \boldsymbol{\epsilon}_{\tilde{s},t+1} \Bigg) \qquad (28) \\
&\quad + \sum_{k=1}^{p-1} \sum_{f \in \mathbb{F}} \pi^\gamma_{f,\tilde{s},t+1} (\boldsymbol{\gamma}^a_{k+1,f,t+1})^\top \boldsymbol{a}_{t-k+1} + (\boldsymbol{\pi}^a_{\tilde{s},t+1})^\top \boldsymbol{v}_t \\
&\quad + \sum_{f \in \mathbb{F}} \pi^\gamma_{f,\tilde{s},t+1} \tilde{\gamma}^c_{f,t+1} + (\boldsymbol{\pi}^b_{\tilde{s},t+1})^\top \boldsymbol{b}_{t+1} \Bigg]
\end{aligned}$$

$$\begin{aligned}
&= \sum_{k=1}^{p-1} \Bigg[\sum_{\tilde{s} \in \tilde{\mathbb{S}}_{s,t}} p_{s,\tilde{s},t+1} \Bigg(\Bigg(\boldsymbol{\pi}^a_{\tilde{s},t+1} + \sum_{f \in \mathbb{F}} \pi^\gamma_{f,\tilde{s},t+1} \boldsymbol{\gamma}^a_{1,f,t+1} \Bigg) \mathrm{diag}(\boldsymbol{\phi}_k) \\
&\quad + \sum_{f \in \mathbb{F}} \pi^\gamma_{f,\tilde{s},t+1} \boldsymbol{\gamma}^a_{k+1,f,t+1} \Bigg) \Bigg]^\top \boldsymbol{a}_{t-k+1} \\
&\quad + \Bigg[\sum_{\tilde{s} \in \tilde{\mathbb{S}}_{s,t}} p_{s,\tilde{s},t+1} \Bigg(\boldsymbol{\pi}^a_{\tilde{s},t+1} + \sum_{f \in \mathbb{F}} \pi^\gamma_{f,\tilde{s},t+1} \boldsymbol{\gamma}^a_{1,f,t+1} \Bigg) \mathrm{diag}(\boldsymbol{\phi}_p) \Bigg]^\top \\
&\quad \boldsymbol{a}_{t-p+1} + \Bigg[\sum_{\tilde{s} \in \tilde{\mathbb{S}}_{s,t}} p_{s,\tilde{s},t+1} \boldsymbol{\pi}^a_{\tilde{s},t+1} \Bigg]^\top \boldsymbol{v}_t + const
\end{aligned}$$

with constant value *const*, independent of $\boldsymbol{v}_t$ and $\boldsymbol{a}_t, \ldots, \boldsymbol{a}_{t-p+1}$. The expressions in front of state variables $\boldsymbol{v}_t$ and $\boldsymbol{a}_t, \ldots, \boldsymbol{a}_{t-p+1}$ are the t-stage cut coefficients, $\boldsymbol{\gamma}^v_{f,t}$ and $\boldsymbol{\gamma}^a_{1,f,t}, \ldots, \boldsymbol{\gamma}^a_{p,f,t}$. The cut constant is given by

$$\tilde{\gamma}^c_{f,t} = \underline{\mathbb{Q}}_{s,t+1}(\boldsymbol{v}_t, \boldsymbol{a}_t, \ldots, \boldsymbol{a}_{t-p+1}) - (\boldsymbol{\gamma}^v_{f,t})^\top \boldsymbol{v}_t - \sum_{k=1}^{p} (\boldsymbol{\gamma}^a_{k,f,t})^\top \boldsymbol{a}_{t-k+1}.$$

Let us now assume that the inflows are coming from a SPAR model of type (7). Again, we omit the deseasonalization parameters and state

$$\boldsymbol{a}_t = \sum_{k=1}^{p} \mathscr{P}_k \boldsymbol{a}_{t-k} + \boldsymbol{\epsilon}_t,$$

where $\mathscr{P}_k$, $k = 1, \ldots, p$, are $|\mathbb{I}| \times |\mathbb{I}|$ matrices with model coefficients not only on the main diagonal as it is the case for PAR, but each row i represents the neighbors i' selected and their associated coefficients $\phi_{i,i',k,m}$. In the cut coefficient calculations above, $\mathrm{diag}(\boldsymbol{\phi}_k)$ in (28) is replaced by $\mathscr{P}_k$, and all other terms remain unchanged.

For the special case that the matrices $\mathscr{P}_k$, $k = 1, \ldots, p$, do not reference any other state variable than those referenced by $\mathrm{diag}(\boldsymbol{\phi}_k)$, the total number of state variables required for the PAR and SPAR models is the same. This is discussed in Section 4.3.

4. Case Study

The Brazilian hydro-power system is one of the largest systems in the world with about 140 installed hydro-power plants (reservoirs and run-of-the-river plants), and their cumulative installed capacity equals about 88% of Brazil's total installed capacity of power plants. The hydro inflow time series of these plants are publicly available from the operator of the system, ONS [31]. They span 80 years, from 1931 through 2010, of monthly inflow data. Additional information can be found on the ONS website related to grouping the hydro plants into basins, the cascade scheme of the hydro plants, and the types of time series. In particular, some of the time series do not describe observed data but were artificially constructed by ONS; we omitted those time series from our analysis. We calculate incremental inflows from the natural inflow data, see Section 3.1.3 for a detailed description. We obtain the locations of the hydro plants separately from the Brazilian Ministry of Mines and Energy website and identify each of the plants manually.

We focus our analysis on 3 of 15 basins described by ONS that do not contain any artificially constructed

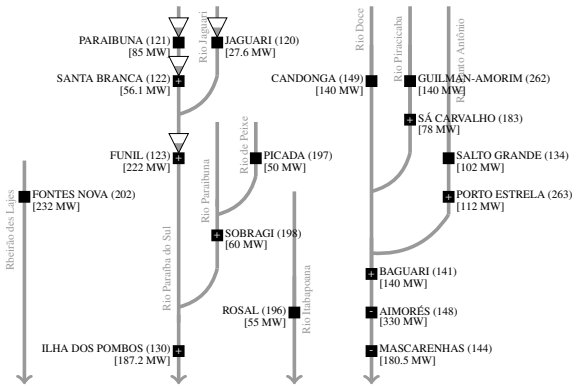

Figure 4: Basin map of Southeast Atlantic with power plant capacities. A plus, minus, or empty square corresponds to column SPAR-F in Table 3 and indicates a deterioration, improvement, or no change to the forecasts, respectively, relative to the PAR-F.

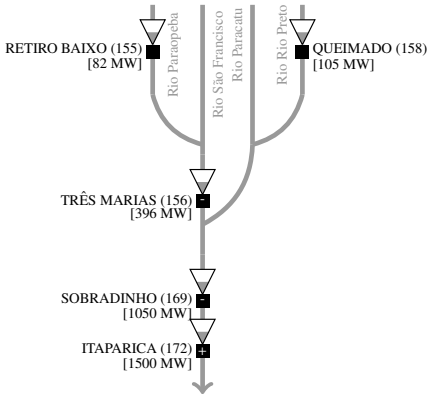

Figure 5: Basin map of Sao Francisco with power plant capacities. A plus, minus, or empty square corresponds to column SPAR-F in Table 3 and indicates a deterioration, improvement, or no change to the forecasts, respectively, relative to the PAR-F.

series: Southeast Atlantic (17 hydro plants), Sao Francisco (5 hydro plants), and Tocantins (7 hydro plants). Figures 4–6 show maps of these three basins. We divide the dataset into training and testing datasets as follows: The 30 years 1946-1975 are used for training, and the 10 years 1976-1985 are used for evaluating the models. With both the PAR and SPAR models, 30 years of monthly inflows equates to only 30 observations to fit the forecasting model for each month, so we need as many years as possible to train the models. In addition, change points or structural changes in the time series between the training and testing sets will cause the models to perform poorly. In 1986, a major change point occurs that affects nearly all locations, so this is why our testing set stops in 1985.

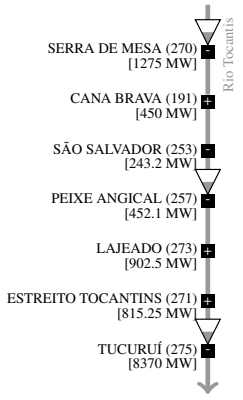

Figure 6: Basin map of Tocantins with power plant capacities. A plus, minus, or empty square corresponds to column SPAR-F in Table 3 and indicates a deterioration, improvement, or no change to the forecasts, respectively, relative to the PAR-F.

4.1. Benchmark Model: Periodic Autoregressive Model

We set up a PAR model (2) for each individual location as a benchmark model for our spatial PAR approach. We choose the PAR model because it is the most widely used type of model to forecast inflows in combination with the SDDP method. We remove seasonality using monthly means and standard deviations, (3) and (4), respectively. We compare this approach to removing seasonality with pairs of harmonics for the mean (5) and the monthly standard deviation (6). The advantage of using harmonics is the ability to model yearly trends, which are evident for some locations. Note that the training data set spans 30 years, so a 5-yearly pattern can only be observed 6 times. The reason for these patterns are events such as El Niño and La Niña, which have a strong influence on the coastal regions of Brazil. We construct the periodogram for each reservoir inflow series to obtain initial candidates for the pairs of harmonics. We investigate the frequencies 40/8, 80/19, 40/12, 80/29, 40/40, and 40/80, which roughly translate to 5-yearly, 4-yearly, $3\frac{1}{3}$-yearly, $2\frac{1}{3}$-yearly, yearly, and $\frac{1}{2}$-yearly harmonics, and use the periodogram to determine which of the frequencies are represented by harmonics at each location. We further refine the chosen pairs by comparing the BIC to a deseasonalization with more or fewer pairs chosen. We observe that locations close to each other show similar yearly seasonality.

Figure 7 depicts a comparison between fitted monthly means and pairs of harmonics for a given location. The improvement of the trend with the inclusion of the 5-yearly harmonic pair is especially noticeable. Table 1 shows a comparison between monthly means and pairs of harmonics for each location. As all point forecasts

Table 1: Comparison of fitted trends using means and pairs of harmonics. Values shown are $\text{RMSE}^{(i,1)}$ (10) of fitted means compared to observed inflows. The fifth column shows the difference in percent. The second column gives the average monthly inflow of each reservoir for reference.

Location	Avg. Monthly Inflow	RMSE Means	RMSE Harmonics	Percent Diff
JAGUARI (120)	28.6	10.6	10.3	-2.77
PARAIBUNA (121)	69.4	23.9	23.2	-2.79
SANTA BRANCA (122)	12.2	6.8	6.8	-0.01
FUNIL (123)	121.6	59.8	58.8	-1.59
ILHA DOS POMBOS (130)	268.9	133.5	133.7	0.13
SALTO GRANDE (134)	150.1	85.9	83.4	-2.85
BAGUARI (141)	163.2	92.1	91.3	-0.85
MASCARENHAS (144)	124.2	75.8	76.2	0.57
AIMORS (148)	233.8	193	193.1	0.02
CANDONGA (149)	146.1	68.1	68.3	0.25
RETIRO BAIXO (155)	143.5	96.8	96.9	0.03
TRS MARIAS (156)	515.8	357.6	357.4	-0.05
QUEIMADO (158)	54.5	26.6	26.7	0.56
SOBRADINHO (169)	1986.5	1127.8	1123.3	-0.40
ITAPARICA (172)	71.8	224.7	226.3	0.71
ANTNIO DIAS + S CARVALHO (183)	7.0	3.8	3.7	-1.51
CANA BRAVA (191)	78.9	71.3	71.0	-0.37
ROSAL (196)	31.0	16.8	16.9	0.94
PICADA (197)	35.7	17.0	17.0	0.48
SOBRAGI (198)	43.5	20.5	20.2	-1.65
LAJES/P.PASSOS/FONTES NOVA (202)	5.9	3.8	3.8	0.40
SO SALVADOR (253)	82.9	71.2	71.5	0.42
PEIXE ANGICAL (257)	655.9	512.0	498.6	-2.61
GUILMAN-AMORIM (262)	75.4	40.5	39.9	-1.55
PORTO ESTRELA (263)	6.6	3.8	3.7	-3.00
SERRA DA MESA (270)	707.9	498.5	496.5	-0.41
ESTREITO TOCANTINS (271)	1815.1	760.0	757.6	-0.32
LAJEADO (273)	722.8	496.6	491.1	-1.09
TUCURUI (275)	5783.2	3291.1	3340.0	1.49
Overall	-	687.9	695.2	1.05

eventually converge towards the fitted mean values for high month-ahead forecasts, using harmonics for deseasonalization increases the performance for some models in the long term. The PAR models (2) for each location are then built using the detrended time series, and parameters are estimated using least squares.

4.2. Our Model: Spatial PAR Model

We remove the seasonality via the approach discussed in Section 4.1 and use the upstream relationship (see Figures 4-6) of locations to find candidate neighbors for our spatial model. The immediate upstream neighbor is considered first, then the one upstream of that neighbor. In total, we identify up to four neighbors as candidates for each model. We experimented with more neighbors, but neighbors higher than the fourth could almost never be selected, either due to the maximum of six variables allowed in the SPAR model, or because of insignificant correlation. We use the identification strategy described in Section 3.1.2 to build the SPAR model (7) for each location.

4.3. Comparison of Results

The models that we fit and evaluate are defined with the acronyms below:

PAR-F: PAR model (2) using monthly summaries (3) and (4)

PAR-H: PAR model (2) using harmonics (5) and (6)

SPAR-F: SPAR model (7) using monthly summaries (3) and (4)

SPAR-H: SPAR model (7) using harmonics (5) and (6)

PAR-F is the benchmark model, which is widely used in Brazil for hydro-thermal scheduling problems solved with SDDP. In fact, PAR-F differs from the PAR used in Brazil in that it is fitted at the reservoir level instead of an aggregated regional level, see Section 1. This allows a better comparison of the PAR and SPAR, including an analysis of additional model lags required in the SPAR.

Table 2 presents the number of model lags for PAR-F and SPAR-F, respectively. The lags of the PAR are of particular interest with respect to the SDDP algorithm because they directly translate into the number of required state variables. The theoretical complexity of the algorithm increases exponentially in the number of state variables. Although SPAR-F uses additional lags compared to the PAR, most of these are already represented by state variables of upstream locations. It is possible that the SPAR does not increase the number of state variables at all. The overall number of lags across all months and locations shows that even though the SPAR has 220 more lags, only 68 new state variables need to be introduced. Locations **(156)** and **(273)** are examples for which no additional state variables are required using the SPAR.

The performance of all four models on the location level is presented in Tables 3 and 4, featuring RMSE and SACE, respectively. Replacing the monthly means with harmonics improved the performance of the PAR by 2% and typically does not add much complexity to the SDDP algorithm. In our case, the total number of model lags across all months and locations was reduced from 576 to 563. The SPAR-F provides the highest overall decrease in RMSE with 8.29%. Figures 4–6 show the improvements on a system map. Squares with a "-" correspond to locations for which SPAR-F performs better than PAR-F, and squares with a "+" the opposite. Empty squares represent locations that either have no upstream neighbors and/or forecasts that remain unchanged when going from PAR-F to SPAR-F.

The SACEs reported in Table 4 show improvements for some locations, but the overall differences between the models are marginal. Recall that an SACE close to one is desired, and an SACE of zero means the model does not perform better than simply using the monthly means (3). The overall SACE is the highest for PAR-H, but the other models are within close range. The SACE values on the location level follow the trend of the individual RMSE. The location with the least improvement for SPAR-F in the RMSE is location **(271)**. The

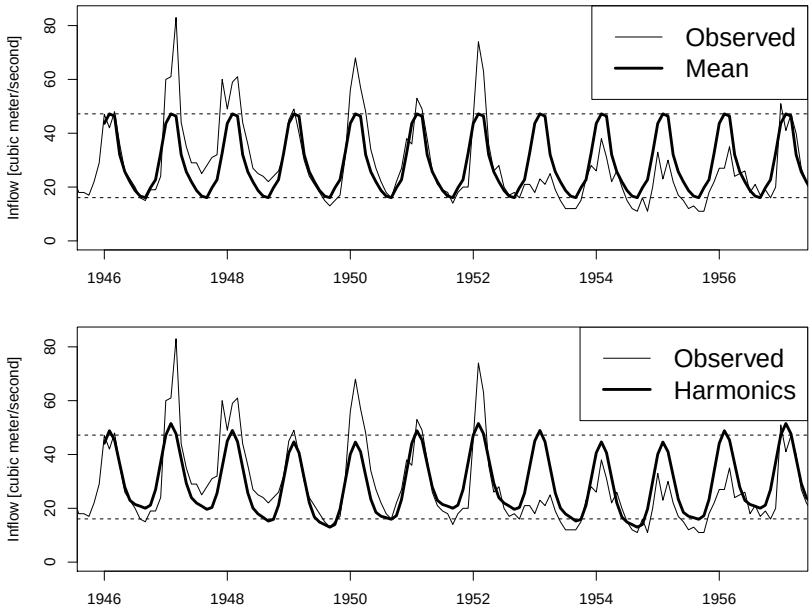

Figure 7: Fitted trend using mean summaries (top) versus pairs of harmonics (bottom). Pairs of harmonics selected for this series are 5-yearly, yearly, and $\frac{1}{2}$-yearly. The dashed vertical lines represent the minimum and maximum of the mean summaries, respectively.

Table 2: Number of lags per location for models PAR-F/SPAR-F and added state variables in SPAR-F compared to PAR-F.

Location	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dev	Total
(120)	1/1/+0	1/1/+0	2/2/+0	1/1/+0	1/1/+0	1/1/+0	1/1/+0	1/1/+0	1/1/+0	4/4/+0	1/1/+0	1/1/+0	16/16/+0
(121)	1/1/+0	2/2/+0	1/1/+0	1/1/+0	1/1/+0	6/6/+0	4/4/+0	1/1/+0	2/2/+0	1/1/+0	1/1/+0	2/2/+0	23/23/+0
(122)	1/1/+0	1/1/+0	1/2/+0	1/3/+1	1/4/+2	1/1/+0	1/1/+0	1/1/+0	2/2/+0	1/1/+0	4/4/+0	2/2/+0	17/23/+3
(123)	1/3/+0	1/1/+0	2/2/+0	1/2/+0	1/2/+0	1/2/+0	4/6/+0	2/2/+0	2/3/+0	1/5/+3	1/2/+0	1/2/+0	18/32/+3
(130)	1/3/+0	1/1/+0	1/4/+1	1/1/+0	3/3/+0	3/4/+0	1/4/+1	1/1/+0	1/5/+2	2/4/+0	1/2/+0	1/2/+0	17/34/+4
(134)	1/1/+0	1/1/+0	1/1/+0	2/2/+0	3/3/+0	2/2/+0	5/5/+0	1/1/+0	1/1/+0	1/1/+0	2/2/+0	1/1/+0	21/21/+0
(141)	1/1/+0	3/5/+0	1/6/+4	3/6/+0	1/1/+0	1/1/+0	3/5/+0	1/1/+0	1/1/+0	1/3/+1	1/3/+1	1/2/+0	18/35/+6
(144)	2/3/+0	1/2/+0	1/1/+0	4/5/+0	6/6/+0	1/3/+0	2/2/+0	1/1/+0	2/5/+0	1/5/+3	1/3/+1	1/1/+0	23/37/+4
(148)	2/2/+0	1/2/+0	1/1/+0	4/4/+0	1/2/+0	2/2/+0	2/2/+0	1/1/+0	1/3/+1	1/4/+1	1/3/+1	1/1/+0	18/27/+3
(149)	5/5/+0	1/1/+0	1/1/+0	1/1/+0	3/3/+0	1/1/+0	2/2/+0	3/3/+0	1/1/+0	1/1/+0	1/1/+0	1/1/+0	21/21/+0
(155)	6/6/+0	1/1/+0	1/1/+0	2/2/+0	5/5/+0	1/1/+0	2/2/+0	2/2/+0	1/1/+0	1/1/+0	2/2/+0	1/1/+0	25/25/+0
(156)	1/1/+0	1/2/+0	1/1/+0	1/1/+0	3/3/+0	1/1/+0	1/1/+0	1/1/+0	1/1/+0	1/1/+0	1/1/+0	1/1/+0	14/15/+0
(158)	1/1/+0	2/2/+0	4/4/+0	2/2/+0	3/3/+0	3/3/+0	1/1/+0	1/1/+0	6/6/+0	1/1/+0	1/1/+0	1/1/+0	26/26/+0
(169)	1/3/+0	1/6/+3	2/4/+0	1/4/+1	1/6/+3	3/3/+0	2/2/+0	1/2/+0	3/5/+0	2/5/+1	1/3/+0	1/4/+0	19/47/+8
(172)	1/2/+0	1/1/+0	1/1/+0	3/5/+0	1/5/+2	1/2/+0	4/4/+0	2/2/+0	2/4/+0	3/3/+0	1/5/+1	1/3/+1	21/37/+4
(183)	1/1/+0	1/1/+0	2/2/+0	2/3/+0	3/3/+0	2/4/+0	2/2/+0	2/3/+0	1/3/+1	1/1/+0	1/2/+0	1/1/+0	19/26/+1
(191)	1/1/+0	2/2/+0	6/6/+0	1/1/+0	6/6/+0	2/6/+2	1/1/+0	3/3/+0	1/1/+0	4/4/+0	1/1/+0	2/2/+0	30/34/+2
(196)	5/5/+0	1/1/+0	1/1/+0	4/4/+0	1/1/+0	2/2/+0	1/1/+0	1/1/+0	1/1/+0	1/1/+0	4/4/+0	1/1/+0	23/23/+0
(197)	3/3/+0	1/1/+0	1/1/+0	1/1/+0	3/3/+0	1/1/+0	2/2/+0	1/1/+0	1/1/+0	1/1/+0	1/1/+0	1/1/+0	17/17/+0
(198)	1/1/+0	1/1/+0	5/5/+0	1/1/+0	3/3/+0	2/5/+1	1/4/+2	4/6/+0	2/2/+0	2/2/+0	1/1/+0	1/1/+0	24/32/+3
(202)	1/1/+0	1/1/+0	1/1/+0	1/1/+0	1/1/+0	1/1/+0	1/1/+0	4/4/+0	1/1/+0	2/2/+0	1/1/+0	3/3/+0	18/18/+0
(253)	1/1/+0	1/2/+0	1/1/+0	1/1/+0	1/2/+0	1/3/+1	1/6/+4	5/6/+0	1/4/+2	1/1/+0	2/2/+0	2/3/+0	18/32/+7
(257)	1/1/+0	1/1/+0	1/1/+0	1/1/+0	2/4/+0	3/6/+0	3/4/+0	6/6/+0	1/1/+0	1/5/+3	1/1/+0	1/1/+0	22/32/+3
(262)	1/1/+0	1/1/+0	2/2/+0	2/2/+0	2/2/+0	1/1/+0	2/2/+0	3/3/+0	2/2/+0	1/1/+0	1/1/+0	1/1/+0	19/19/+0
(263)	1/1/+0	1/1/+0	1/1/+0	2/2/+0	3/4/+0	1/1/+0	2/2/+0	2/2/+0	1/3/+1	1/2/+0	1/1/+0	1/1/+0	17/21/+1
(270)	1/1/+0	1/1/+0	1/1/+0	1/1/+0	1/1/+0	3/3/+0	2/2/+0	1/1/+0	2/2/+0	5/5/+0	1/1/+0	1/1/+0	20/20/+0
(271)	1/3/+1	1/2/+0	1/6/+4	1/2/+0	1/1/+0	1/2/+0	1/6/+4	1/3/+0	1/1/+0	1/2/+0	1/1/+0	1/4/+0	12/33/+9
(273)	1/2/+0	1/2/+0	1/1/+0	1/2/+0	1/1/+0	3/3/+0	2/5/+0	5/6/+0	3/3/+0	1/2/+0	1/1/+0	1/2/+0	21/30/+0
(275)	1/1/+0	1/6/+3	1/3/+0	1/3/+1	1/3/+1	5/5/+0	3/5/+0	1/1/+0	1/3/+0	2/2/+0	1/3/+0	1/5/+2	19/40/+7
Overall	46/57/+1	34/52/+6	46/64/+9	48/65/+3	63/83/+8	56/76/+4	59/85/+11	59/67/+0	46/69/+7	46/71/+12	38/55/+4	35/52/+3	576/796/+68

Table 3: The first column presents the $\text{RMSE}^{(i,1)}$ (10) of benchmark model, PAR-F, and the remaining columns report percent changes compared to the first column. The last row shows the overall $\text{RMSE}^{(1)}$ (9).

Location	PAR-F	PAR-H	SPAR-F	SPAR-H
(120)	9.36	-0.92 %	0.00 %	-0.92 %
(121)	24.01	-8.00 %	0.00 %	-8.00 %
(122)	5.51	1.13 %	1.43 %	2.82 %
(123)	51.28	0.10 %	2.00 %	-4.89 %
(130)	118.82	0.07 %	1.87 %	3.34 %
(134)	102.67	3.02 %	0.00 %	3.02 %
(141)	95.99	-4.43 %	3.66 %	5.28 %
(144)	96.75	-1.07 %	-1.19 %	15.18 %
(148)	252.27	-1.37 %	-2.15 %	14.32 %
(149)	77.39	0.56 %	0.00 %	0.56 %
(155)	120.39	-0.57 %	0.00 %	-0.57 %
(156)	397.12	-1.20 %	-2.72 %	-3.39 %
(158)	26.15	-3.40 %	0.00 %	-3.40 %
(169)	1276.31	-0.88 %	-36.54 %	-35.92 %
(172)	274.90	-0.45 %	5.42 %	-5.44 %
(183)	4.82	10.23 %	0.49 %	10.85 %
(191)	92.92	-1.81 %	0.07 %	0.35 %
(196)	24.11	2.79 %	0.00 %	2.79 %
(197)	15.33	-0.73 %	0.00 %	-0.73 %
(198)	21.09	-0.08 %	2.23 %	2.15 %
(202)	2.18	1.79 %	0.00 %	1.79 %
(253)	94.28	-1.33 %	-10.26 %	-8.29 %
(257)	665.09	-5.73 %	0.02 %	-5.46 %
(262)	51.53	8.18 %	0.00 %	8.18 %
(263)	4.48	1.20 %	0.44 %	1.62 %
(270)	639.61	-2.62 %	0.00 %	-2.62 %
(271)	871.26	-7.48 %	33.15 %	6.72 %
(273)	568.84	-3.66 %	4.89 %	2.83 %
(275)	3081.11	-1.69 %	-9.86 %	-4.53 %
Overall	680.83	-2.09 %	-8.29 %	-6.69 %

Table 4: Individual $\text{SACE}^{(i,1)}$ for 1 month ahead forecasts. The first column represents the benchmark model. All models use monthly mean summaries (3) as reference in (11).

Location	PAR-F	PAR-H	SPAR-F	SPAR-H
(120)	0.42	0.43	0.42	0.43
(121)	0.21	0.33	0.21	0.33
(122)	0.44	0.43	0.42	0.41
(123)	0.26	0.26	0.23	0.33
(130)	0.27	0.27	0.24	0.22
(134)	0.19	0.14	0.19	0.14
(141)	0.26	0.32	0.20	0.18
(144)	0.20	0.22	0.22	-0.06
(148)	0.17	0.20	0.21	-0.08
(149)	0.27	0.26	0.27	0.26
(155)	0.26	0.27	0.26	0.27
(156)	0.39	0.40	0.42	0.43
(158)	0.40	0.44	0.40	0.44
(169)	0.27	0.28	0.71	0.70
(172)	0.09	0.10	-0.01	0.18
(183)	0.13	-0.06	0.12	-0.07
(191)	0.30	0.33	0.30	0.30
(196)	-0.16	-0.22	-0.16	-0.22
(197)	0.35	0.36	0.35	0.36
(198)	0.14	0.14	0.10	0.10
(202)	0.22	0.19	0.22	0.19
(253)	0.27	0.29	0.41	0.38
(257)	0.15	0.25	0.15	0.24
(262)	0.13	-0.02	0.13	-0.02
(263)	0.19	0.17	0.19	0.17
(270)	0.25	0.28	0.25	0.28
(271)	-0.05	0.10	-0.86	-0.19
(273)	0.19	0.25	0.11	0.14
(275)	0.49	0.51	0.59	0.54
Overall	0.23	0.24	0.22	0.22

SACE for its PAR-F is already negative, meaning that it performs worse than the mean. Using harmonics for this location seems to be beneficial, improving both the PAR and the SPAR models. Figure 8 compares PAR-F with SPAR-F for location **(169)**, the one with the greatest improvement for the SPAR, for the ten year testing set. The SPAR model captures the peaks in inflows much better. It also picks up subtle sub-peaks such as those in 1982.

4.4. Model Fitting Discussion

In this work, we intentionally do not include any exogenous predictors, such as precipitation forecasts or climate indices such as NINO3.4 [29], which describes sea surface temperature in the region bounded by 5° N $\times$ 5° S and 170° W $\times$ 120° W. First, including more predictors increases the number of state variables required for the SDDP, which is undesirable because the number of major iterations in the SDDP algorithm depends directly on the number of state variables. In addition, to make forecasts more than one month ahead, the values of exogenous predictors must also be forecast, which introduces even more uncertainty into the forecasts. In [23], it is shown that incorporating the NINO3.4 index did not improve inflow predictions but incorporating precipitation did. Forecasting precipitation at even one location is nontrivial as in [49], [48], and the references therein, and it is even more challenging to forecast in both space and time. In [57], the space-time challenges in this context are described, and a summary of the relevant precipitation literature is provided in addition to an evaluation of the Matérn space-time covariance in modeling precipitation at various time scales. Furthermore, the model proposed in [23] would not extend naturally to incorporate spatial information, so we believe that the model that we propose here is an important advance in the field. Nevertheless, improvement of the models during the wet months (November through April) would be very important from the reservoir operation point of view, and we noted that when the SPAR model improves upon the PAR, it is typically during the wet months. However, we leave the formal incorporation of space-time precipitation forecasts into the SPAR model for future work.

We observed that the SPAR models, in general, can be sensitive to "bad neighbors," *i.e.*, neighbors that suddenly change their inflow behavior going from training set to testing set while the location to forecast remains consistent. The problem is that these "bad neighbors" have the potential to not only impair their own forecast, but the forecast of all their respective downstream

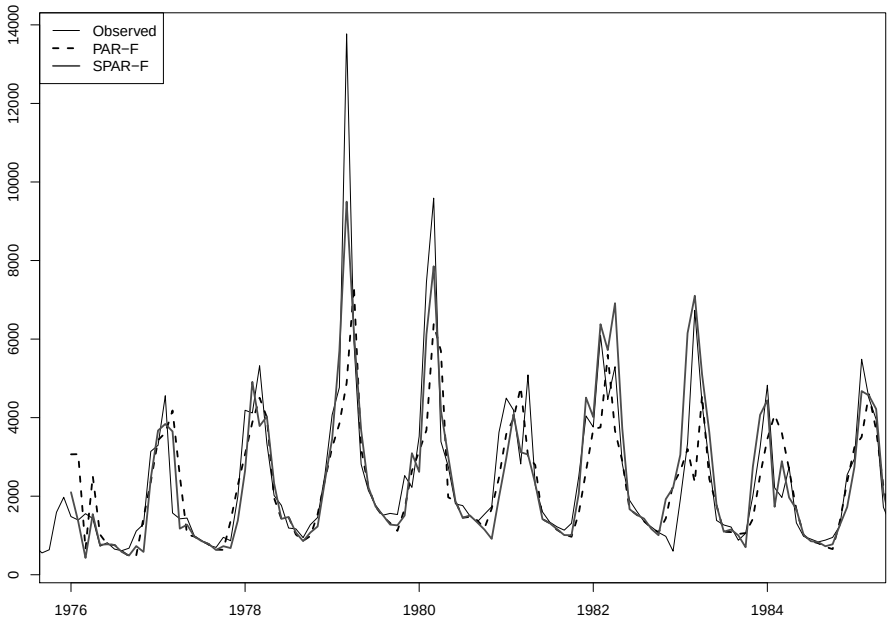

Figure 8: PAR-F and SPAR-F comparison for location **(169)** for the ten year testing set.

neighbors as their forecast inflow series may be used in those models. This problem will not occur with a PAR model to the same extent since errors in forecasts at one location do not propagate to neighbors.

Therefore, it is important in fitting a SPAR model (even more so than a PAR model) that the process used to obtain historic data is consistent. Regarding the Brazilian data set, we doubt that this is true across all locations and years for two reasons. First, it is unlikely that measuring techniques have remained consistent over the past 80 years. Second, and more importantly, the system of hydro plants has tremendously changed in those years, and its operation is also likely to have changed. Still, ONS offers inflow data for all locations in those years. This means that multiple inflow series must have been artificially constructed until the plant started operating, possibly affecting both existing locations and new locations.

The problem of "bad neighbors" can also have a strong impact on forecasts beyond one-month ahead, but in spite of this, the SPAR still showed improvement across multiple horizons. In particular, the SPAR reduced the RMSE relative to the PAR across all locations by 8.29%, 3.26%, and 1.59% for one, two, and three month ahead forecasts, respectively. We found random variation in four and five month ahead forecasts, and starting in month six, the point forecasts of both PAR and SPAR converge to each location's respective monthly mean.

5. Conclusion

We have built a new approach to forecasting hydro inflows that can be used as input for the SDDP algorithm. The approach uses spatial information, in our case upstream neighbors of locations, to improve the forecast of an underlying PAR model. This is possible because residuals of the PAR models show spatial correlation (*cf.* Figure 3), and upstream time series can therefore be incorporated as additional explanatory variables. Even though spatial dependence can be incorporated when forecasting scenarios with the PAR, we have shown in this work how this process can be handled more elegantly through a spatial model, avoiding the challenging task of estimating a large spatial covariance matrix.

Using three basins with 29 locations total from the Brazilian hydro system as a case study, we show that our SPAR model performs better for one month ahead forecasts than our benchmark PAR model while only increasing the number of state variables by 11.8%. Compared to the PAR model which is similar to the one currently used in Brazil, both models require significantly fewer state variables. We further describe how detrending with pairs of harmonics can both decrease the number of parameters to estimate and results in better long-term forecasts without adding complexity to the SDDP algorithm. The generation of scenario trees for use in the SDDP algorithm is also simplified with the SPAR

model because correlation between locations is already captured by the model.

However, for the SPAR model to perform well, we must have stationarity between the training and testing data sets. With hydrological inflows, the number of reservoirs in the system may change over longer time scales, which can have a dramatic impact on the inflows of all associated run-of-the-river plants. In future work, the effect that such decisions have on longer-term forecasts and, subsequently, on SDDP method outputs should be studied. We also focused on point forecasts in this paper, and future work could investigate the quality of scenario forecasting with regard to the representation of spatial dependence. Furthermore, as mentioned in Section 4.4, we excluded any exogenous predictors in our model. Their impact on the forecasting results and the additional complexity they introduce in the SDDP algorithm should be studied in future work.

Acknowledgements

The authors would like to thank two anonymous reviewers for valuable comments and suggestions that have improved this article's content and presentation. Further, the authors thank Professor Marc G. Genton of King Abdullah University of Science and Technology for his comments on an earlier version of this work.

[1] Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19, 716–723.

[2] Bellman, R., 1957. Dynamic Programming. Princeton University Press.

[3] Bender, M., Simonovic, S., 1994. Time-series modeling for long-range stream-flow forecasting. Journal of Water Resources Planning and Management 120 (6), 857–870.

[4] Boomsma, T., Juul, N., Fleten, S.-E., 2014. Bidding in sequential electricity markets: The Nordic case. European Journal of Operational Research 238 (3), 797–809.

[5] Cerisola, S., Latorre, J., Ramos, A., 2012. Stochastic dual dynamic programming applied to nonconvex hydrothermal models. European Journal of Operational Research 218, 687–697.

[6] Cressie, N., Wikle, C. K., 2011. Statistics for Spatio-Temporal Data. Wiley.

[7] Dawson, C., Wilby, R., 2001. Hydrological modelling using artificial neural networks. Progress in Physical Geography 25 (1), 80–108.

[8] de Luna, X., Genton, M., 2005. Predictive spatio-temporal models for spatially sparse environmental data. Statistica Sinica 15, 547–568.

[9] Elberg, C., Hagspiel, S., 2015. Spatial dependencies of wind power and interrelations with spot price dynamics. European Journal of Operational Research 241 (1), 260–272.

[10] Filho, F. A. S., Lall, U., 2003. Seasonal to inter annual ensemble streamflow forecasts for Ceara, Brazil: Applications of a multivariate, semi parametric algorithm. Water Resources Research 39 (11), 1–1–1–13.

[11] Gjelsvik, A., Belsnes, M., Haugstad, A., 1999. An algorithm for stochastic medium-term hydrothermal scheduling under spot price uncertainty. In: 13th Power Systems Computation Conference: Proceedings, Vol. 2. pp. 1079–1085.

[12] Gjelsvik, A., Mo, B., Haugstad, A., 2010. Long- and medium-term operations planning and stochastic modelling in hydro-dominated power systems based on stochastic dual dynamic programming. In: Rebennack, S., Pardalos, P., Pereira, M., Iliadis, N. (Eds.), Handbook of Power Systems. Energy Systems. Springer-Verlag, pp. 377–391.

[13] Gjelsvik, A., Wallace, S., 1996. Methods for stochastic medium-term scheduling in hydro-dominated power systems. Tech. rep., Norwegian Electric Power Research Institute, Trondheim.

[14] Gneiting, T., Larson, K., Westrick, K., Genton, M., Aldrich, E., 2006. Calibrated probabilistic forecasting at the Stateline Wind Energy Center: The regime-switching space-time method. Journal of the American Statistical Association 101, 968–979.

[15] Grygier, J., Stedinger, J., 1985. Algorithms for optimizing hydropower system operation. Water Resources Research 21 (1), 1–10.

[16] Helseth, A., Mo, B., Warland, G., 2010. Long-term scheduling of hydro-thermal power systems using scenario fans. Energy Systems 1 (4), 377–391.

[17] Hering, A., Genton, M., 2010. Powering up with space-time wind forecasting. Journal of the American Statistical Association 105, 92–104.

[18] Hipel, K., McLeod, A., 1994. Time Series Modeling of Water Resources and Environmental Systems. Elsevier.

[19] Infanger, G., Morton, D., 1996. Cut sharing for multistage stochastic linear programs with interstage dependency. Mathematical Programming 75 (2), 241–256.

[20] Jain, S., Das, A., Srivastava, D., 1999. Application of ANN for reservoir inflow prediction and operation. Journal of Water Resources Planning and Management 125 (5), 263–271.

[21] Johnson, R., Wichern, D., 2007. Applied Multivariate Statistical Analysis. Prentice Hall, Upper Saddle River, NJ.

[22] Kazor, K., Hering, A., 2015. The role of regimes in short-term wind speed forecasting at multiple wind farms. Stat 4, 271–290.

[23] Lima, L. M., Popova, E., Damien, P., 2014. Modeling and forecasting of Brazilian reservoir inflows via dynamic linear models. International Journal of Forecasting 30, 464–476.

[24] Maceira, M., Damázio, J., 2006. Use of the PAR(p) model in the stochastic dual dynamic programming optimization scheme used in the operation planning of the Brazilian hydropower system. Probability in the Engineering and Informational Sciences 20, 143–156.

[25] Maceira, M., Duarte, V., Penna, D., Moraes, L., Melo, A., 2008. Ten Years of Application of Stochastic Dual Dynamic Programming in Official and Agent Studies in Brazil - Description of the NEWAVE Program. In: 16th PSCC, Glasgow, Scotland, July 14-18, 2008.

[26] Maier, H., Dandy, G., 2000. Neural networks for the prediction and forecasting of water resources variables: A review of modelling issues and applications. Environmental Modelling & Software 15 (1), 101–124.

[27] Marsden, M., Davis, R., 1968. Regression on principle components as a tool in water supply forecasting. In: Proceedings of the 36th Western Snow Conference. pp. 33–40.

[28] Mondal, M., Shahjahan, M., Wasimi, S., 2006. Generating and forecasting monthly flows of the Ganges river with PAR model. Journal of Hydrology 323 (1–4), 41–56.

[29] NINO3.4 Index, 2015.
URL `http://www.cpc.ncep.noaa.gov/data/indices/`

[30] Noakes, D., McLeod, A., Hipel, K., 1985. Forecasting monthly riverflow time series. International Journal of Forecasting 1, 179–190.

[31] Operador Nacional do Sistema Elétrico, 2015.

URL http://www.ons.org.br/operacao/vazoes_naturais.aspx

[32] Pereira, M., Pinto, L., 1985. Stochastic optimization of a multireservoir hydroelectric system: A decomposition approach. Water Resources Research 21 (6), 779–792.

[33] Pereira, M., Pinto, L., 1991. Multi-stage stochastic optimization applied to energy planning. Mathematical Programming 52, 359–375.

[34] Philpott, A., de Matos, V., 2012. Dynamic sampling algorithms for multi-stage stochastic programs with risk aversion. European Journal of Operational Research 218 (2), 470–483.

[35] Philpott, A., Guan, Z., 2008. On the convergence of stochastic dual dynamic programming and related methods. Operations Research Letters 36, 450–455.

[36] Pritchard, G., 2015. Stochastic inflow modeling for hydropower scheduling problems. European Journal of Operational Research 246 (2), 496–504.

[37] PSRI, December 2008. SDDP Methodology Manual. Version 9.2.

[38] Queiroz, A., Morton, D., 2013. Sharing cuts under aggregated forecast when decomposing multi-stage stochastic programs. Operations Research Letters 41, 311–316.

[39] Read, E., 1990. A dual approach to stochastic dynamic programming for reservoir release scheduling. In: Esogbue, A. (Ed.), Dynamic Programming for Optimal Water Resources System Management. Prentice Hall, NJ, pp. 361–372.

[40] Rebennack, S., 2014. Generation expansion planning under uncertainty with emissions quotas. Electric Power Systems Research 114, 78–85.

[41] Rebennack, S., 2016. Combining sampling-based and scenario-based nested Benders decomposition methods: application to stochastic dual dynamic programming. Mathematical Programming 156 (1), 343–389.

[42] Rebennack, S., Flach, B., Pereira, M., Pardalos, P., 2012. Stochastic hydro-thermal scheduling under CO2 emission constraints. IEEE Transactions on Power Systems 27 (1), 58–68.

[43] Schabenberger, O., Gotway, C. A., 2000. Statistical Methods for Spatial Data Analysis. Chapman and Hall.

[44] Schwarz, G., 1978. Estimating the dimension of a model. Annals of Statistics 6, 461–464.

[45] Scott, T., Read, E., 1996. Modelling hydro reservoir operation in a deregulated electricity sector. International Transactions in Operational Research 3, 209–221.

[46] Shapiro, A., 2011. Analysis of stochastic dual dynamic programming method. European Journal of Operational Research 209 (1), 63–72.

[47] Shapiro, A., Tekaya, W., da Costa, J. P., Soares, M. P., 2013. Risk neutral and risk averse stochastic dual dynamic programming method. European Journal of Operational Research 224 (2), 375–391.

[48] Sharma, A., 2000. Seasonal to inter annual rainfall probabilistic forecasts for improved water supply management: Part 3—a nonparametric probabilistic forecast model. Journal of Hydrology 239, 249–258.

[49] Sharma, A., Luk, K. C., Cordery, I., Lall, U., 2000. Seasonal to inter annual rainfall probabilistic forecasts for improved water supply management: Part 2—predictor identification of quarterly rainfall using ocean-atmosphere information. Journal of Hydrology 239, 240–248.

[50] Sharma, A., O'Neill, R., 2002. A nonparametric approach for representing interannual dependence in monthly streamflow sequences. Water Resources Research 38 (7), 5–1–5–10.

[51] Shumway, R., Stoffer, D., 2011. Time Series Analysis and Its Applications. Springer.

[52] Smith, J., 1991. Long-range streamflow forecasting using non-parametric regression. Water Resources Bulletin 27 (1), 39–46.

[53] Souza, R., Marcato, A., Dias, B., Oliveira, F., 2012. Optimal operation of hydrothermal systems with hydrological scenario generation through bootstrap and periodic autoregressive models. European Journal of Operational Research 222 (3), 606–615.

[54] Stedinger, J., Grygier, J., Yin, H., 1989. Seasonal streamflow forecasts based upon regression. In: Computerized Decision Support Systems for Water Managers. ASCE, pp. 266–279.

[55] Steeger, G., Barroso, L., Rebennack, S., 2014. Optimal bidding strategies for hydro-electric producers: A literature survey. IEEE Transactions on Power Systems 29 (4), 1758–1766.

[56] Steeger, G., Rebennack, S., 2015. Strategic bidding for multiple price-maker hydroelectric producers. IIE Transactions 47 (9), 1013–1031.

[57] Sun, Y., Bowman, K. P., Genton, M. G., Tokay, A., 2015. A Matérn model of the spatial covariance structure of point rain rates. Stochastic Environmental Research and Risk Assessment 29, 411–416.

[58] Tilmant, A., Kelman, R., 2007. A stochastic approach to analyze trade-offs and risks associated with large-scale water resources systems. Water Resources Research 43, 1–11.

[59] Valença, M., Ludermir, T., 2002. NeuroInflow: The new model to forecast average monthly inflow. In: Proceedings of the VII Brazilian Symposium on Neural Networks. pp. 74–79.

[60] Waller, L., Gotway, C., 2004. Applied Spatial Statistics for Public Health Data. Wiley.

[61] Wang, W., 2006. Stochasticity, Nonlinearity and Forecasting of Streamflow Processes. IOS Press.

[62] Xu, Z., Li, J., 2002. Short-term inflow forecasting using an artificial neural network model. Hydrological Processes 16, 2423–2439.

[63] Yakowitz, S., 1982. Dynamic programming applications in water resources. Water Resources Research 18 (4), 673–696.

[64] Yang, M., Read, E., 1999. A constructive dual dynamic programming for a reservoir model with correlation. Water Resources Research 35 (7), 2247–2257.

[65] Yeh, W.-G., 1985. Reservoir management and operations models: A state of the art review. Water Resources Research 21 (12), 1797–1818.

[66] Yoder, M., Hering, A. S., Navidi, W. C., Larson, K., 2014. Short-term forecasting of categorical changes in wind power with Markov chain models. Wind Energy 17, 1425–1439.

[67] Yurekli, K., Kurunc, A., Ozturk, F., 2005. Application of linear stochastic models to monthly inflow data of Kelkit Stream. Ecological Modelling 183, 67–75.

[68] Zhu, X., Bowman, K., Genton, M., 2014. Incorporating geostrophic wind information for improved space-time short-term wind speed forecasting. Annals of Applied Statistics 8, 1782–1799.

[69] Zhu, X., Genton, M., Gu, Y., Xie, L., 2014. Space-time wind speed forecasting for improved power system dispatch. TEST 23, 1–25.