PSR Technical Reports
Stochastic programming models

REPRESENTATION OF PARAMETER UNCERTAINTY IN SDDP's PROBABILISTIC INFLOW MODELS

Bernardo Bezerra

July 21st, 2016

Abstract The optimal stochastic scheduling of hydrothermal systems requires the analytical representation of uncertainties in the inflows to the hydro plants. Multistage stochastic optimization methods such as SDDP use a multivariate periodic autoregressive (PAR) inflow model in this task. Because the inflow model parameters are estimated from historical data, they have a natural estimation error. However, this error in the parameter values is not taken into account in the stochastic optimization, i.e. the model assumes that they are population values.

In this paper, we show that the consequence of ignoring parameter value uncertainty is an "optimistic" operation policy. We also show that, with the increased penetration of wind energy, whose historical records are typically much shorter than inflow records, this optimistic bias tends to get more severe. We also propose two approaches to eliminate this problem: (i) carry out a parameter estimation that takes into account its effect on expected operation cost, using a minimax regret or a CVaR-based criterion; (ii) incorporate multiple streamflow models in the SDDP recursion.

Keywords Power Systems · SDDP · Parameter Uncertainty

1 SDDP FORMULATION

1.1 Notation

Indices

$t = 1, \ldots, T$	time stages (typically weeks or months)
$\tau = 1, \ldots,$	intra-stage time blocks (e.g. peak, medium and low demand or 168 hours in a week)
$s = 1, \ldots, S$	scenarios for each stage t produced by the stochastic models (typically inflows and renewable generation; also loads, equipment availability and fuel costs)
$l = 1, \ldots, L$	set of scenarios for stage $t+1$ conditioned to scenario s in stage t
$i = 1, \ldots, I$	storage devices (typically hydro plants; also fuel storage, batteries, emission limits and some types of contracts)
$m \in M_i$	set of hydro plants immediately upstream of plant i

B. Bezerra
Centro Empresarial Rio Praia de Botafogo 228 / 1701-A Botafogo 22250-145 - Rio de Janeiro, RJ - Brasil
Tel.: +55 21 3906-2100
E-mail: bernardo@psr-inc.com

$j = 1, \ldots, J$	dispatchable devices (typically, thermal plants; also, some controllable renewables and price-responsive demand)
$n = 1, \ldots, N$	non-dispatchable devices (typically, wind, solar and biomass)
$p = 1, \ldots, \mathcal{P}$	number of hyperplanes (Benders cuts) in the future cost function

Decision variables for the operation problem in stage t, scenario s

$v_{t+1,i}$	stored volume of hydro i by the end of stage t
$u_{t,i}$	turbined volume of hydro i stage t
$s_{t,i}$	spilled volume of hydro i in stage t
$e_{t,\tau,i}$	energy produced by hydro i in block τ, stage t
$g_{t,\tau,j}$	energy produced by thermal plant j in block τ, stage t
α_{t+1}^{l}	present value of expected future cost from $t+1$ to T conditioned to scenario l in $t+1$

Note: for notational simplicity, we will not include the transmission network model in the formulations.

Known values for the operation problem in stage t, scenario s

$\hat{a}_{t,i}^{s}$	lateral inflow to hydro i in stage t, scenario s ($\hat{a}_t^s$ set of inflows for all hydro plants)
$\hat{v}_{t,i}^{s}$	stored volume of hydro i in the beginning of stage t, scenario s ($\hat{v}_t^s$ set of stored volumes for all hydro plants)
$\overline{v}_i$	maximum storage of hydro i
$\overline{u}_i$	maximum turbined outflow of hydro i
ρ_i	production coefficient (kWh/m^3) of hydro i
$\overline{g}_j$	maximum generation of thermal plant j
c_j	variable operating cost of thermal plant j
$\hat{r}_{t,\tau,n}^{s}$	energy produced by renewable plant n in stage t, block τ, scenario s
$\hat{d}_{t,\tau}$	demand of block τ, stage t

Multipliers

$\pi_{ht,i}$	multiplier of the storage balance equation of hydro i (see problem formulation)
$\pi_{\mathrm{at},i}$	multiplier of the conditioned inflow equation of hydro i (see problem formulation)

p^{th} Benders cut coefficients

$\hat{\varphi}_{ht+1,i}^{p}$	coefficient of hydro plant i's storage, $v_{t+1,i}$
$\hat{\varphi}_{\mathrm{at}+1,i}^{p}$	coefficient of hydro plant i's inflow, $a_{t+1,i}^{l}$
$\hat{\varphi}_{0t+1}^{p}$	constant term

Stochastic streamflow model coefficients

$\hat{\mu}_{t,i}$	mean of the lateral inflow to hydro i in stage t.
$\hat{\sigma}_{t,i}$	standard deviation of the lateral inflow to hydro i in stage t.
$\hat{\rho}_{t,i}$	serial correlation of the lateral inflow to hydro i in stage t.

1.2 Problem Formulation

Objective function (SDDP recursion)

$$\alpha_t\left(\hat{v}_t^s, \hat{a}_t^s\right) = Min \sum_j c_j \sum_\tau g_{t,\tau,j} + \frac{1}{L}\sum_l \alpha_{t+1}^l \tag{1a}$$

Storage balance for each stage

$$v_{t+1,i} = \hat{v}_{t,i}^s + \hat{a}_{t,i}^s - (u_{t,i}+s_{t,i}) + \sum_{m \in M_i} (u_{t,m}+s_{t,m}) \qquad \leftarrow \pi_{ht,i} \tag{1b}$$

Note: for notational simplicity, we will not represent real-life features of the storage balance equations such as evaporation, filtration, water diversion for irrigation and city supply, transposition and others.

Storage and turbined outflow limits

$$v_{t+1,i} \leq \overline{v}_i \tag{1c}$$

$$u_{t,i} \leq \overline{u}_i \tag{1d}$$

Generation and demand balance for each block

$$e_{t,i} = \rho_i u_{t,i} \tag{1e}$$

$$\sum_\tau e_{t,\tau,i} = e_{t,i} \tag{1f}$$

$$e_{t,\tau,i} \leq \overline{e}_i \tag{1g}$$

$$g_{t,\tau,j} \leq \overline{g}_j \tag{1h}$$

$$\sum_i e_{t,\tau,i} + \sum_j g_{t,\tau,j} = \hat{d}_{t,\tau} - \sum_n \hat{r}_{t,\tau,n}^s \tag{1i}$$

Conditioned inflow scenarios for t+1

For simplicity of presentation, we show a multivariate $AR(1)$ model (in practice, SDDP uses a multivariate periodic autoregressive (PAR)) model with up to six past time stages):

$$\frac{\left(a_{t+1,i}^l - \hat{\mu}_{t+1,i}\right)}{\hat{\sigma}_{t+1,i}} = \hat{\rho}_{t,i} \times \frac{\left(\hat{a}_{t,i}^s - \hat{\mu}_{t,i}\right)}{\hat{\sigma}_{t,i}} + \sqrt{1-\hat{\rho}_{t,i}^2} \times \hat{\xi}_{t,i}^l \qquad \leftarrow \pi_{at,i} \tag{1j}$$

where the parameters $\{\hat{\mu}_{t,i}, \hat{\sigma}_{t,i}, \hat{\rho}_{t,i}\}$ are respectively the mean, standard deviation and serial correlation of the lateral inflow to hydro i in stage t. Spatial dependence is represented through a correlation matrix in the sampling the innovation values $\hat{\xi}^l_{t,i}$ for all hydro plants.

Note: For clarity of presentation, the stochastic streamflow models are shown explicitly. In the actual SDDP implementation, they are represented implicitly.

Future cost functions

As it is well known, the FCFs in SDDP are represented by a set of hyperplanes

$$\alpha^l_{t+1} \geq \sum_i \hat{\varphi}^p_{ht+1,i} \times v_{t+1,i} + \sum_i \hat{\varphi}^p_{at+1,i} \times a^l_{t+1,i} + \ \hat{\varphi}^{0p}_{0t+1} \qquad \forall p = 1,\ \ldots,\ \mathcal{P};\ l = 1,\ \ldots,\ L \tag{1k}$$

1.3 Backward recursion step

After the one-stage dispatch problem 1.2 is solved, we can generate a Benders cut to improve the future cost function approximation of the previous stage. Note that Benders cuts are now generated in both the forward and backward steps.

Assuming that the FCF for the previous stage already has $\mathcal{P}$ hyperplanes, the Benders cut will correspond to the $(\mathcal{P}+1)^{\text{th}}$ FCF constraint:

$$\alpha^l_t \geq \sum_i \hat{\varphi}^{\mathcal{P}+1}_{ht,i} \times v_{t,i} + \sum_i \hat{\varphi}^{\mathcal{P}+1}_{at,i} \times a^l_{t,i} + \ \hat{\varphi}^{\mathcal{P}+1}_{0t} \tag{2a}$$

The Benders cut coefficients $\hat{\varphi}^{\mathcal{P}+1}_{ht,i}$, $\hat{\varphi}^{\mathcal{P}+1}_{\text{at},i}$ and $\hat{\varphi}^{\mathcal{P}+1}_{0t}$ are calculated from a linear expansion of the optimal solution α^*_t of the one-stage dispatch problem 2.2

$$\alpha_t\left(v_t, a_t\right) \approx \alpha^*_t + \sum_i \frac{\partial \alpha_t}{\partial v_{t,i}} \times \left(v_{t,i} - \hat{v}^s_{t,i}\right) + \sum_i \frac{\partial \alpha_t}{\partial a_{t,i}} \times \left(a_{t,i} - \hat{a}^s_{t,i}\right) \tag{2b}$$

The coefficient $\hat{\varphi}^{\mathcal{P}+1}_{ht,i}$ corresponds to $\frac{\partial \alpha_t}{\partial v_{t,i}}$, which is the simplex multiplier $\pi_{ht,i}$. In turn, $\hat{\varphi}^{\mathcal{P}+1}_{\text{at},i}$ corresponds to $\frac{\partial \alpha_t}{\partial a_{t,i}}$, calculated as: $\pi_{ht,i} + \left(\frac{\hat{\rho}_{t,i}}{\hat{\sigma}_{t,i}}\right) \times \pi_{\text{at},i}$. Finally, the constant term is obtained by adding all the constants of the linear expansion:

$$\hat{\varphi}^{\mathcal{P}+1}_{0t} = \alpha^*_t - \sum_i \hat{\varphi}^{\mathcal{P}+1}_{ht,i} \hat{v}^s_{t,i} - \sum_i \hat{\varphi}^{\mathcal{P}+1}_{\text{at},i} \hat{a}^s_{t,i} \tag{2c}$$

1.4 Forward simulation step

Upper bound calculation

In stage t, scenario s of the forward simulation step, we calculate the immediate operation cost associated to optimal solution (indicated by the superscript "*").

$$z^s_t = \sum_j c_j \sum_\tau g^*_{t,\tau,j} \tag{3a}$$

As in the traditional SDDP formulation, the upper bound is calculated as:

$$\overline{z} = \frac{1}{S} \sum_t \sum_s z_t^s \tag{3b}$$

Inflow scenario for stage $\mathbf{t+1}$

We also calculate in the forward simulation step the inflow scenario vector for the next stage $t+1$: $\left\{\hat{a}_{t+1,i}^s,\ i=1,\ \ldots,\ I\right\}$. This is done by sampling from the expression for the conditioned inflows of section 1.2.5:

$$\frac{\left(a_{t+1,i}^l - \hat{\mu}_{t+1,i}\right)}{\hat{\sigma}_{t+1,i}} = \hat{\rho}_{t,i} \times \frac{\left(\hat{a}_{t,i}^s - \hat{\mu}_{t,i}\right)}{\hat{\sigma}_{t,i}} + \sqrt{1-\hat{\rho}_{t,i}^2} \times \xi_{t,i}^l \quad \forall l = 1, \ldots, L;\ \forall i = 1, \ldots, I \tag{3c}$$

Basically, $\hat{s}$ is randomly sampled from the set $\{1, \ldots,\ L\}$, and the inflows $\left\{\hat{a}_{t+1,i}^s,\ i=1,\ \ldots,\ I\right\}$ are calculated for the corresponding innovation vector $\left\{\hat{\xi}_{t,i}^{\hat{s}},\ i=1,\ \ldots,\ I\right\}$. Note that the entire innovation vector($i=1,\ldots I$) has to be used in order to preserve the spatial correlation.

$$\frac{\left(a_{t+1,i}^s - \hat{\mu}_{t+1,i}\right)}{\hat{\sigma}_{t+1,i}} = \hat{\rho}_{t,i} \times \frac{\left(\hat{a}_{t,i}^s - \hat{\mu}_{t,i}\right)}{\hat{\sigma}_{t,i}} + \sqrt{1-\hat{\rho}_{t,i}^2} \times \hat{\xi}_{t,i}^{\hat{s}} \quad \forall i = 1, \ldots, I \tag{3d}$$

At this point we have two possibilities, either we pre calculate $a_{t,i}$ for all t before even starting the sddp recursion or we repeat the above described procedure in each iteration so that we have higher probability of generating new cuts.

2 Uncertainty on the inflow model parameters

For convenience of presentation, we reproduce below the AR(1) stochastic model of equation (3d).

$$\frac{\left(a_{t+1,i}^l - \hat{\mu}_{w(t+1),i}\right)}{\hat{\sigma}_{w(t+1),i}} = \hat{\rho}_{w(t),i} \times \frac{\left(\hat{a}_{t,i}^s - \hat{\mu}_{w(t),i}\right)}{\hat{\sigma}_{w(t),i}} + \sqrt{1-\hat{\rho}_{w(t),i}^2} \times \hat{\xi}_{t,i}^l \quad \forall i = 1, \ldots, I \tag{4}$$

As it is well known, the values of the mean, standard deviation and serial correlation parameters $\left\{\hat{\mu}_{w(t)}, \hat{\sigma}_{w(t)}, \hat{\rho}_{w(t)}\right\}$ are *estimated* from the historical records:

$$\hat{\mu}_{w,i} = \frac{1}{Y_H} \sum_{t \in T_w} \hat{a}_{t,i}^H \qquad w = 1,\ \ldots,\ 52\ (or\ 12) \tag{5a}$$

$$\hat{\sigma}_{w,i} = \sqrt{\frac{1}{Y_H - 1} \sum_{t \in T_w} \left(\hat{a}_{t,i}^H - \hat{\mu}_{w,i}\right)^2} \tag{5b}$$

$$\hat{\rho}_{w,i} = \frac{\frac{1}{Y_H - 1} \sum_{t \in T_w} \left(\hat{a}_{t,i}^H - \hat{\mu}_{w,i}\right)\left(\hat{a}_{t+1,i}^H - \hat{\mu}_{w(t+1),i}\right)}{\hat{\sigma}_{w,i}\hat{\sigma}_{w(t+1),i}} \tag{5c}$$

where:

- w is a scalar variable that indexes the weeks (or months)
- $w(t)$ is a function that indicates the week (or month) associated to stage $tt \in T_w$ set of time stages associated to week (or month) w
- Y_H is the number of years in the historical record

It is also well known that there is an *uncertainty* around these estimated values, which depends on the number of years Y_H of the historical record. For example, the estimator of the mean $\hat{\mu}_{w,i}$ is a Normal random variable with a standard deviation of $\frac{\hat{\sigma}_{\mathrm{wt},i}}{\sqrt{Y_H}}$.

This means that, when we simulate the system operation with a set of S inflow scenarios produced by the AR(1) model above with *fixed values* for $\{\hat{\mu}_t, \hat{\sigma}_t, \hat{\rho}_t\}$, we are ignoring the uncertainty on the parameter values. This uncertainty may be relevant when the historical records are fairly short, which is the case of many modern renewable sources such as wind.

2.1 Modeling inflow scenarios with parameter uncertainty

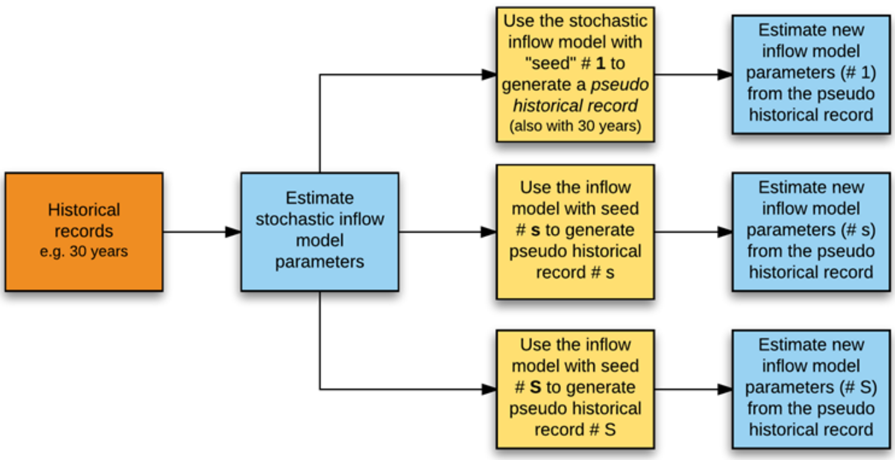
Initially, we show how to produce a set of inflow scenarios $\{\tilde{a}_t^s,\ s = 1, \ldots S; t = 1,\ \ldots,\ T\}$ which incorporate parameter uncertainty.

1. Estimate the AR^H model parameters $\{\hat{\mu}_t, \hat{\sigma}_t, \hat{\rho}_t\}$ from the historical record.
2. Repeat for $s = 1,\ \ldots,\ S$
   Apply the AR^H model (with parameters $\{\hat{\mu}_t, \hat{\sigma}_t, \hat{\rho}_t\}$) to generate a *pseudo-historical inflow record* $\mathcal{H}^s$,that is, an inflow scenario with the same number of years N as the historical record.
   Estimate the AR^s model parameters $\{\tilde{\mu}_t^s, \tilde{\sigma}_t^s, \tilde{\rho}_t^s\}$ from $\mathcal{H}^s$.
3. Repeat for $s = 1,\ \ldots,\ S$

   Apply the AR^s model with parameters $\{\tilde{\mu}_t^s, \tilde{\sigma}_t^s, \tilde{\rho}_t^s\}$ to generate *one* scenario of length T, $\{\tilde{a}_t^s,\ t = 1,\ \ldots,\ T\}$

   Figure (1) illustrates the scenario generation scheme.

Fig. 1 Scenario Generation scheme



2.2 Operating cost impact of ignoring parameter uncertainty

The effect on expected operating costs of ignoring parameter uncertainty can be assessed as follows:

1. Run a SDDP stochastic policy/probabilistic simulation *without* parameter uncertainty, i.e. where both the inflow scenarios $\{\hat{a}_t^s,\ s = 1, \ldots S;\ t = 1, \ldots, T\}$ and the conditioned inflows $\{a_{t+1}^l,\ l = 1, \ldots, L\}$ are produced with the parameters $\{\hat{\mu}_t, \hat{\sigma}_t, \hat{\rho}_t\}$ estimated from the historical sample:

2. Let $\overline{z}$ be the expected operation cost in the final probabilistic simulation of step (a). As discussed, $\overline{z}$ will be correct only if the estimates of parameter values are equal (or very close) to their "true" (population) values.
3. Run another probabilistic simulation using the set of FCFs from the run of step (a), but with inflow scenarios $\{\tilde{a}_t^s,\ s = 1,\ \ldots S; t = 1,\ \ldots,\ T\}$ produced by the two-phase procedure above[1]. Let $\tilde{z}$ be the expected operation cost of this simulation. From the previous discussions, we know that $\tilde{z} \geq \overline{z}$ (for the confidence interval).

The difference $\frac{(\tilde{z}-\overline{z})}{\overline{z}}$ is a measure of the cost of ignoring parameter uncertainty in the stochastic optimization. If this cost is significant, we need to represent parameter uncertainty in the SDDP stochastic optimization scheme. This topic is discussed next.

2.3 Calculation of inflow model clusters

The first step is to group the S different sets of parameters (remember that S can be as high as 1,200 scenarios) into a smaller number $\mathcal{M}$ of clusters, for example, $\mathcal{M} = 5$. This is done as follows:

1. Repeat for each pseudo-historical record $\mathcal{H}^s,\ s = 1,\ \ldots,\ S$
   Calculate the total annual energy inflow e_y^s for each year $y = 1,\ \ldots,\ Y_H$ as:

$$e_y^s = \sum_{t \in T_y} \sum_i \overline{\rho}_i a_{t,i}^s \tag{6}$$

   where:
   - $t \in T_y$ is the set of stages (months or weeks) of year y.
   - $\overline{\rho}_i$ is the sum of production coefficients of hydro i plus all plants downstream of i.

   Calculate the mean and standard deviation of the yearly annual inflows:

$$\hat{\mu}_Y^s = \frac{1}{Y_H} \sum_y e_y^s \tag{7}$$

$$\hat{\sigma}_Y^s = \sqrt{\frac{1}{Y_H - 1} \sum_y \left(e_y^s - \hat{\mu}_Y^s\right)^2} \tag{8}$$

   Define a scalar *severity index* χ^s based on a severe, but plausible, drought (e.g. 2.0 standard deviations from the mean):

$$\chi^s = \hat{\mu}_Y^s - 2.0\hat{\sigma}_Y^s \tag{9}$$

2. Use a clustering method over the set $\{\chi^s,\ s = 1,\ \ldots,\ S\}$ to create $\mathcal{M}$ sets of parameter values $\{\tilde{\mu}_t^m, \tilde{\sigma}_t^m, \tilde{\rho}_t^m\},\ m = 1,\ \ldots,\ \mathcal{M}$. Each of these sets will be associated with a different stochastic streamflow model.
3. Estimate the *probability* p_m of each inflow model as $p_m = \frac{S_m}{S}$, $m = 1,\ \ldots,\ \mathcal{M}$, where S_m is the number of scenarios in cluster m.

Next, we will show how to represent different stochastic inflow models in the SDDP policy calculation.

[1] Note that the conditioned inflows for the next stage, $\{a_{t+1}^l\ ,\ l = 1,\ \ldots,\ L\}$, used in the operating problem in each stage and scenario are still calculated with the parameters $\{\hat{\mu}_t, \hat{\sigma}_t, \hat{\rho}_t\}$ estimated from the historical sample.

2.4 Selection of inflow model based on risk-adjusted operation costs

1. Repeat for each inflow model $m = 1, \ldots, \mathcal{M}$
   Run a SDDP stochastic policy calculation assuming that an "oracle" indicated that inflow model m is correct, i.e. where both the inflow scenarios $\{\tilde{a}_t^{ms}, \ s = 1, \ldots S; t = 1, \ldots, T\}$ and the conditioned inflows $\left\{a_{t+1}^{ml}, \ l = 1, \ldots, L\right\}$ are produced with the parameters $\{\tilde{\mu}_t^m, \tilde{\sigma}_t^m, \tilde{\rho}_t^m\}$ of inflow model m.
2. Repeat for each inflow model $n = 1, \ldots, \mathcal{M}$

   Run a probabilistic *simulation* using the set of FCFs from the SDDP run of step (1), but using a different set of inflow scenarios $\{\tilde{a}_t^{\mathrm{ns}}, \ s = 1, \ldots S; t = 1, \ldots, T\}$ produced by inflow model n. Let $\tilde{z}_{mn}$ be the expected operation cost of this simulation. From the previous discussions, we know that $\tilde{z} \geq \overline{z}$ (for the confidence interval).

The above procedure results in the $\mathcal{M} \times \mathcal{M}$ matrix shown below. Each element $\tilde{z}_{mn}$ in the matrix is the expected operation cost when we calculate the stochastic policy assuming that the inflow model is m, but the actual model is n. The values $\{p_m\}$ in the first row the probabilities of each inflow model.

$$\begin{vmatrix} p_1 & \cdots & p_m & \cdots & p_{\mathcal{M}} \\ \tilde{z}_{11} & \ldots & \tilde{z}_{1m} & \ldots & \tilde{z}_{1\mathcal{M}} \\ \vdots & \ddots & & & \vdots \\ \tilde{z}_{m1} & & \tilde{z}_{\mathrm{mm}} & & \tilde{z}_{\mathrm{mM}} \\ \vdots & & & \ddots & \vdots \\ \tilde{z}_{\mathcal{M}1} & \ldots & \tilde{z}_{\mathrm{Mm}} & \ldots & \tilde{z}_{\mathrm{MM}} \end{vmatrix}$$

The objective is now to select the "best" inflow model based on the policy calculation / simulation results. The decision criteria include:

a) choose the model that minimizes the expected operation cost over all possible models:

$$m^* = \underset{m}{\operatorname{argmin}} \sum_n p_n \tilde{z}_{mn} \tag{10}$$

b) minimax regret:

$$m^* = \underset{m}{\operatorname{argmin}} \max_n \{\tilde{z}_{mn} - \tilde{z}_{mm}\} \tag{11}$$

c) convex combination of (a) and (b):

$$m^* = \left[\lambda \sum_n p_n \tilde{z}_{\mathrm{mn}} + (1-\lambda) \ \ \{\tilde{z}_{\mathrm{mn}} - \tilde{z}_{\mathrm{mm}}\}\right] \tag{12}$$

d) convex combination of (a) and CVaR of expected operating costs:

$$m^* = \left[\lambda \sum_n p_n \tilde{z}_{\mathrm{mn}} + (1-\lambda) \mathrm{CVaR}_q \left\{\tilde{z}_{\mathrm{mn}}\right\}\right] \tag{13}$$

The aforementioned procedure assumes that we can only have one inflow model in the SDDP policy calculation. We show next how to represent several inflow models simultaneously in the SDDP recursion.

2.5 Case studies

The proposed approach was applied to the Brazilian Power System, which has an installed capacity of 125 GW, 160 hydro (85 with storage), 140 thermal plants (gas, coal, oil and), 8 GW wind, 5 GW biomass and 1 GW solar.

From the parameters of the historical record (Parent Level), 200 historical records with 85 years size (1931-2014) were generated (Child Level) and the corresponding sample parameters were clustered in 5 clusters $\{\hat{\mu}_t^m,\ \hat{\sigma}_i^m, \hat{\rho}_t^m\}$. Figure (2.5) depicts the standard deviation and average of each Child Sequence and for each cluster.

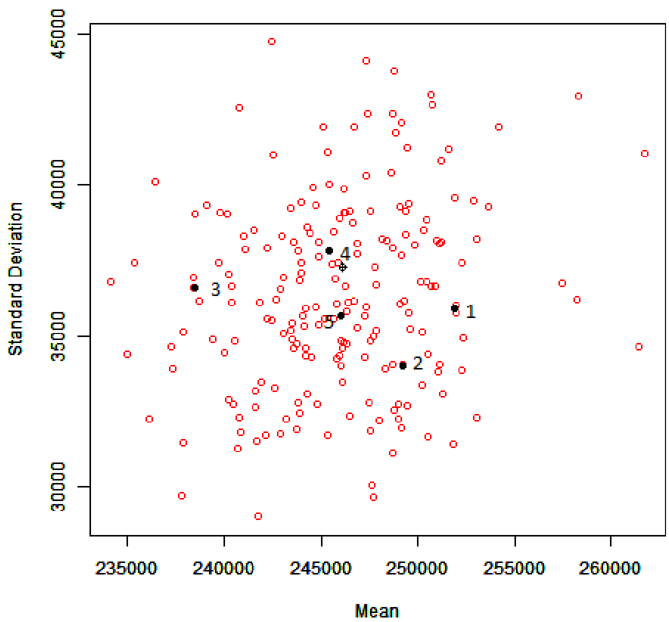


Fig. 2 Average and standard deviation of parent level, child level and centroids

For each cluster we have generated sequences for the study horizon (Grandchild Level). In order to calculate the best inflow model, we calculated an operating policy for each cluster and held a final simulation for each Grandchild sequence m. Figure (2.5) shows the matrix with the difference to the optimal policy, in percentage.

The simulations carried out with the “taylor made” policy are the best ones, as expected. We have calculated the different criteria discussed before. Figure (2.5) depicts the best policy for each criterion.

We can observe that the “best” inflow model based on the policy calculation/simulation depends on the decision criteria.

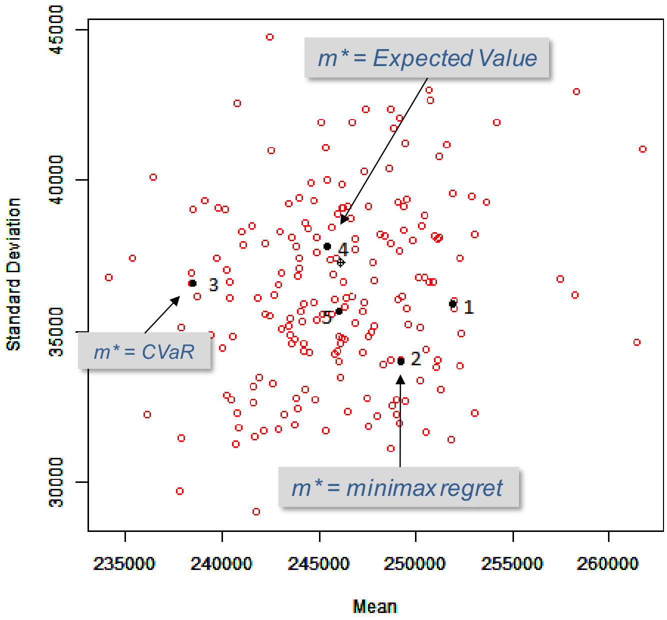


Fig. 3 Operating cost matrix (% optimal value)

		Policies				
		P1	P2	P3	P4	P5
Simulations	S1	0%	2%	6%	2%	18%
	S2	0%	0%	5%	0%	16%
	S3	6%	8%	0%	2%	20%
	S4	4%	6%	2%	0%	23%
	S5	18%	16%	18%	21%	0%

Fig. 4 Operating cost matrix (% optimal value)

3 SDDP policy calculation with different inflow models

In this section, we describe our current research on the representation of different inflow models in the SDDP recursion.

3.1 Basic formulation

The SDDP formulation below calculates the optimal stochastic policy (minimization of expected operation costs) when there are $\mathcal{M}$ alternative inflow models with probabilities $\{p_m, m = 1, \ldots, \mathcal{M}\}$.

$$\alpha_t\left(\hat{v}_t^s, \tilde{a}_t^s\right) = \text{Min} \sum_j c_j \sum_\tau g_{t,\tau,j} + \sum_m p_m \left[\frac{1}{L}\sum_l \alpha_{t+1}^{ml}\right] \tag{14a}$$

$$v_{t+1,i} = \hat{v}_{t,i}^s + \tilde{a}_{t,i}^s - (u_{t,i}+s_{t,i}) + \sum_{\eta \in M_i} (u_{t,\eta}+s_{t,\eta}) \;\leftarrow \pi_{ht,i} \tag{14b}$$

$$\ldots$$

$$\sum_i e_{t,\tau,i} + \sum_j g_{t,\tau,j} = \hat{d}_{t,\tau} - \sum_n \hat{r}_{t,\tau,n}^s \tag{14c}$$

$$\frac{\left(a^{ml}_{t+1,i} - \hat{\mu}^{m}_{w(t+1),i}\right)}{\hat{\sigma}^{m}_{w(t+1),i}} = \hat{\rho}^{m}_{w(t),i} \times \frac{\left(\tilde{a}^{s}_{t,i} - \hat{\mu}^{m}_{w(t),i}\right)}{\hat{\sigma}^{m}_{w(t+1),i}} + \sqrt{1 - \left[\hat{\rho}^{m}_{w(t),i}\right]^{2}} \times \hat{\xi}^{l}_{t,i} \quad \forall i, l, m \leftarrow \pi^{ml}_{at,i} \tag{14d}$$

$$\alpha^{ml}_{t+1} \geq \sum_{i} \hat{\varphi}^{p}_{ht+1,i} \times v_{t+1,i} + \sum_{i} \hat{\varphi}^{p}_{at+1,i} \times a^{ml}_{t+1,i} + \hat{\varphi}^{p}_{0t+1} \qquad \forall p, m, l \tag{14e}$$

Note that in this formulation there are $\mathcal{M} \times L$ piecewise linear future cost functions $\left\{\alpha^{ml}_{t+1}\right\}$. Therefore, the computational effort is higher than in the standard SDDP recursion.

Benders cut calculation

The storage coefficients $\left\{\hat{\varphi}^{\mathcal{P}+1}_{ht,i}\right\}$ are calculated in the same way as in the standard SDDP formulation. The inflow coefficients $\left\{\hat{\varphi}^{\mathcal{P}+1}_{\mathrm{at},i}\right\}$ are calculated as:

$$\hat{\varphi}^{\mathcal{P}+1}_{\mathrm{at},i} = \pi_{ht,i} + \sum_{m} p_m \left[\left(\frac{\hat{\rho}^{m}_{w(t),i}}{\hat{\sigma}^{m}_{w(t+1),i}}\right) \pi^{ml}_{at,i}\right] \tag{14f}$$

4 Case studies

The proposed approach was applied to a system with a single hydro plant and 3 thermal power plants, resulting in a power system with 80% hydro predominance. Although simple, this static configuration captures the essence of hydrothermal scheduling issues and was chosen for didactic reasons. The thermal plants have installed capacity of 80, 110 and 90 MW, and unit variable cost of 80, 30 and 115 R$/MWh, respectively. Table 1 presents the characteristics of the hydro plant.

A dummy load shedding unit was considered with a variable cost of 1000 $/MWh. The discount rate used was 8% per year, measured in inflation-free currency. The monthly energy load is 496 GWh. The monthly load was considered constant along the study period and flat during each month, i.e., seasonality and peak demand modeling was not considered.

	Hydro
Capacity (MW)	1147
Max. turbining outflow ($\mathbf{m^3/s}$)	1479
Mean production coefficient ($\mathbf{MW/m^3/s}$)	0.783
Min. storage ($\mathbf{hm^3}$)	20922
Max. storage ($\mathbf{hm^3}$)	22950

Table 1: Hydro Plant's Characteristics

From the parameters of the historical record (Parent Level), 4096 historical records with 40 years size were generated (Child Level) and the corresponding sample parameters were clustered in 5 clusters $\{\hat{\mu}^{m}_{t},\ \hat{\sigma}^{m}_{i}, \hat{\rho}^{m}_{t}\}$. Because the hydrothermal simulation has 12 stages, the clustering technique was applied to the average inflow of the period. Fig. 1 presents μ and σ of the average streamflows from March to February (study period) of the parent sequence level, child sequence level and the cluster centroids. The weights p of each cluster are, in order, 23.1%, 20.7%, 27.4%, 18.5% and 10.3%.

First, the operating policy $\Omega_0$was calculated for the sample parameters $\{\hat{\mu}_t,\ \hat{\sigma}_t, \hat{\rho}_t\}$, which corresponds to the current hydrothermal dispatch methodology. Then, this policy was simulated for the sequences generated with the parameters of each cluster (Grandchild Sequence Level). These sequences would be possible candidates for the population values.

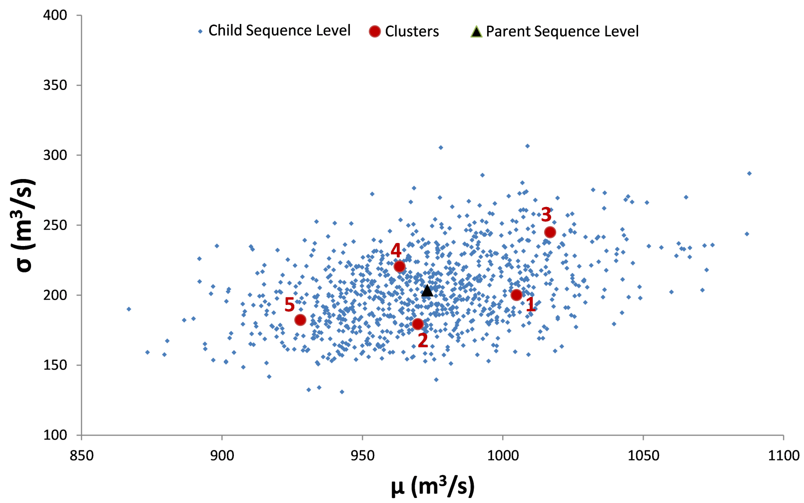


Fig. 5 Average and standard deviation of parent level, child level and centroids

Figure (4) compares the present value of the total expected operating cost (thermal cost plus energy deficit cost). We observe that, if the population values are close to cluster 5, the real operating costs are 41% higher than the initial expectation.

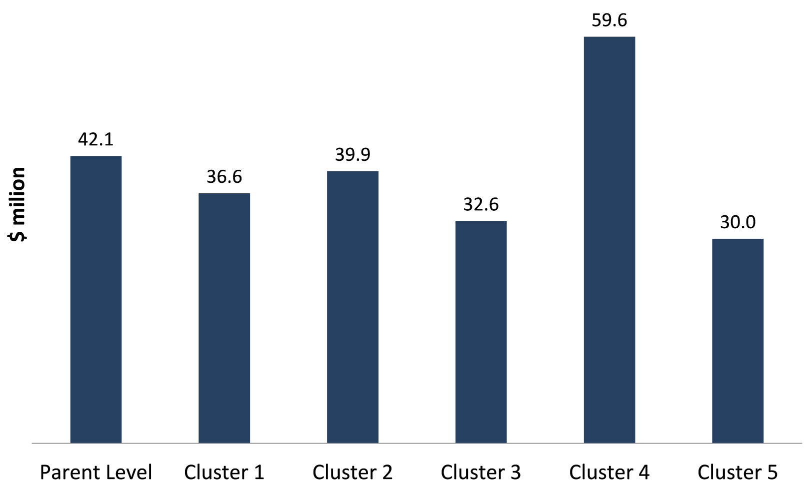


Fig. 6 Present value of the total expected operating cost – without parameter uncertainty

The average cost resulting from the weighting of the operating cost of each cluster in Figure 2, weighted by their respective weight p_m, is the expected cost of the hydrothermal dispatch simulated with parameter uncertainty, as this simulation represents jointly parameter and streamflow uncertainties. This results in a weighted cost \$40.0 million, 5.1% less than the simulation performed with series generated without parameter uncertainty. However, this does not mean that the policy $\Omega_0$calculated without parameter uncertainty is optimal for the system. To illustrate this point, the Ω_m operational policies were calculated for each cluster m, in order to identify the one that minimizes the operating cost of the final simulation with parameter uncertainty. The proposed policy Ω_6, which considers parameter uncertainty, was also calculated. Figure (4) compares the present value of the expected cost for each policy, all simulated with uncertainty in the parameters. The Ω_4 policy presented operating costs 30.4% higher. The policy strategy with parameter uncertainty reduced 7.7% the operating costs.

Fig. 7 Present value of the total expected operating cost – with parameter uncertainty

In order to analyze the optimality of proposed operating policy in this paper, we calculated an operating policy for each cluster and held a final simulation for each Grandchild sequence m. Figure (4) shows the matrix with the difference to the optimal policy, in percentage.

		Final Simulation				
		Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Policy	Ω_0 Parent Level	0.5%	0.2%	0.6%	0.5%	1.5%
	Ω_1 Cluster 1	0.0%	0.2%	2.3%	1.2%	2.1%
	Ω_2 Cluster 2	0.6%	0.0%	2.2%	1.2%	2.1%
	Ω_3 Cluster 3	1.1%	0.8%	0.0%	0.5%	0.8%
	Ω_4 Cluster 4	34.5%	29.1%	50.8%	0.0%	81.4%
	Ω_5 Cluster 5	0.9%	1.1%	0.4%	0.4%	0.0%
	Ω_6 Parameter Uncertainty	0.6%	0.4%	0.0%	0.2%	0.5%

Fig. 8 Operating cost matrix (k$)

Analyzing the table columns, we observed that the policy with parameter uncertainty Ω_6 is between the second and the fourth best of the seven tested policies. Analyzing the table rows, we note that the maximum regret (greater deviation from the optimum) of the policy Ω_6 was 0.6%, while for other policies this value is greater than 1.1% reaching 81.4 % in Ω_4. Therefore, proposed policy Ω_6 is the one that minimizes the expected value of the operating cost taking into account parameter uncertainty and that minimizes the maximum regret, considering as candidates for the population values the parameters of the 5 clusters.

5 Improved SDDP policy calculation with different inflow models

The previous SDDP formulation uses a single probability value p_m for each inflow model $m = 1, \ldots, \mathcal{M}$. It is possible to refine the policy calculation by using the probability of each model *conditioned* to the current inflow value $\tilde{a}^s_{t,i}$. This is analogous to using the transition probabilities of a Markov chain instead of the steady-state marginal distribution probabilities.

Calculation of the conditioned model probabilities

The first step is to transform the S inflow vectors $\left\{\tilde{a}^s_{t,i}\right\}$ of stage t into S scalar values corresponding to the hydro energy inflow $\{\tilde{e}^s_t\}$ (analogous to the procedure used to cluster stochastic models). Next, we group the energy inflows into $\mathcal{K}$ clusters $\left\{\mathcal{C}^k_t,\ k = 1, \ldots, \mathcal{K}\right\}$. For example, $\mathcal{K} = 3$ could correspond to high, medium and low energy inflows. Finally, for each cluster $\mathcal{C}^k_t$, we estimate the conditioned probability of each model m, p^k_{tm}, as the ratio $\frac{S^k_{tm}}{S^k_t}$, where S^k_t is the number of scenarios in cluster $\mathcal{C}^k_t$, and S^k_{tm} is the number of scenarios produced by model m in that cluster.

The second step is to estimate the transition probabilities $\left\{p^{k\nu}_t\right\}$ from each cluster $\mathcal{C}^k_t$ in stage t to each of the clusters in stage $t+1$, $\left\{\mathcal{C}^\nu_{t+1}, \nu = 1, \ldots, \mathcal{K}\right\}$. This is similar to the procedure used in the Markov chain formulations of the previous chapter: $p^{k\nu}_t$ is the *fraction* of scenarios $s \in \mathcal{C}^k_t$ that also belong to $\mathcal{C}^\nu_{t+1}$. In summary:

(i) for each cluster $\mathcal{C}^k_t$ there is an associated vector of model probabilities $\left\{p^k_{tm}\right\}$; and (ii) the transition probability from cluster k in stage t to cluster ν in stage $t+1$ is $p^{k\nu}_t$.

SDDP recursion with different inflow models and conditioned probabilities

This means that the energy inflow cluster $\mathcal{C}^m_t$ will also be a state variable in the SDDP recursion:

$$\alpha_t\left(\hat{v}^s_t, \tilde{a}^s_t, \mathcal{C}^{k(s)}_t\right) = \text{Min} \sum_j c_j \sum_\tau g_{t,\tau,j} + \sum_\nu p^{k\nu}_t \left[\sum_m p^k_{tm}\left(\frac{1}{L}\sum_l \alpha^{ml\nu}_{t+1}\right)\right] \tag{15a}$$

$$v_{t+1,i} = \hat{v}^s_{t,i} + \hat{a}^s_{t,i} - (u_{t,i}+s_{t,i}) + \sum_{m\in M_i}(u_{t,m}+s_{t,m}) \tag{15b}$$

$$\ldots$$

$$\sum_i e_{t,\tau,i} + \sum_j g_{t,\tau,j} = \hat{d}_{t,\tau} - \sum_n \hat{r}^s_{t,\tau,n} \tag{15c}$$

$$\frac{\left(a^{ml}_{t+1,i} - \hat{\mu}^m_{w(t+1),i}\right)}{\hat{\sigma}^m_{w(t+1),i}} = \hat{\rho}^m_{w(t),i} \times \frac{\left(\tilde{a}^s_{t,i} - \hat{\mu}^m_{w(t),i}\right)}{\hat{\sigma}^m_{w(t+1),i}} + \sqrt{1 - \left[\hat{\rho}^m_{w(t),i}\right]^2} \times \hat{\xi}^l_{t,i} \quad \forall i, l, m \leftarrow \pi^{ml}_{at,i} \tag{15d}$$

$$\alpha^{ml\nu}_{t+1} \geq \sum_i \hat{\varphi}^{\nu p}_{ht+1,i} \times v_{t+1,i} + \sum_i \hat{\varphi}^{\nu p}_{at+1,i} \times a^{ml}_{t+1,i} + \hat{\varphi}^{p\nu}_{0t+1} \qquad \forall p = 1, \ldots, \mathcal{P}^\nu, m, l \tag{15e}$$

Benders cut calculation

The storage coefficients $\left\{\hat{\varphi}^{\mathcal{P}+1}_{ht,i}\right\}$ are calculated in the same way as in the standard SDDP formulation. In turn, the inflow coefficients $\left\{\hat{\varphi}^{\mathcal{P}+1,k}_{\text{at},i}\right\}$ are calculated as

$$\hat{\varphi}_{\text{at},i}^{\mathcal{P}+1,k} = \sum_{m} p_{tm}^{k} \left[\left(\frac{\hat{\rho}_{w(t),i}^{m}}{\hat{\sigma}_{w(t+1),i}^{m}} \right) \pi_{\text{at},i}^{ml} \right] \tag{16}$$

6 Conclusion

The impact of the uncertainty of the PAR(p) model parameters was quantified, in the context of stochastic hydrothermal dispatch, and it was proposed an improvement in the policy calculation in order to take into account this phenomenon. It was shown that parameter uncertainty has significant impact on the system's operating costs and its incorporation in the policy calculation allows a reduction of 7% in operating costs. The proposed operating policy had the lower expected cost minimized the maximum regret.

Although by computational limitation the application of the proposed methodology is not feasible for policy calculation in large systems, such as the Brazilian Power System, the work serves as starting point for discussions on the improvement of stochastic optimization models for hydrothermal dispatch. However, this does not prevent final simulations of operation and expansion planning studies to be carried out considering this uncertainty, in order to diagnose more accurately the situation of supply security.

Finally, the objective of the paper was to calculate an operating policy which is better than the policy without considering parameter uncertainty, which is the current approach used in hydro scheduling. The quality of the proposed policy can be improved by modeling the inflows as a Markov Chain, with transition probabilities between each cluster.