

IMPROVING ENERGY GENERATION FORECASTS OF SMALL HYDROPOWER PLANTS BY SPATIAL DEPENDENCE ANALYSIS

Margarete Afonso de Sousa

Industrial Engineering Department, PUC-Rio
Rua Marquês de São Vicente, 225, Gávea - Rio de Janeiro, RJ - Brasil
margarete.afonso.sousa@gmail.com

Reinaldo Castro Souza

Industrial Engineering Department, PUC-Rio
Rua Marquês de São Vicente, 225, Gávea - Rio de Janeiro, RJ - Brasil
reinaldo@puc-rio.br

ABSTRACT

Electric power generation forecasting has an important role for distribution utilities when planning their operations. The difficulty increases when involving Small Hydropower Plants, in general, run-of-river plants with a strong dependence on hydrological regimes. But that is not the only concern, these plants in many cases do not have inflow or precipitation data available. Joining inflow data as exogenous variable from neighboring basins seems to be an interesting approach to solve data unavailability. This study aims to present a spatial dependence analysis to identify the best inflow series to obtain good prediction closer to reality. In this study, SARIMA and Transfer Function models were used.

KEYWORDS: Time Series. Spatial Autocorrelation. Small Hydropower Plants. SARIMA. Transfer Function. Neighboring Inflow Series.

EN – PO na Área de Energia, Petróleo e Gás.

1. Introduction

Small Hydropower plants (SHP) are hydroelectric plants with power capacity between 5 and 30 megawatts (MW), they are considered as a renewable source with low impact in the environment because they do not need a reservoir to generate energy. In some cases, they have a reservoir that cannot exceed 13 km². The Brazilian legislation define also another category considered as small hydroelectric plant, called Power Generation Plants (PGP) and their generation capacity is lower than 5 MW [ANEEL, 2018]. For now on, for notation simplify, SHP and PGP denote both SHP.

Nowadays in Brazil, the hydraulic source represents 65,2% of all the electric energy generated as shown in Figure 1 [EPE, 2018]. There are 428 SHPs and 683 PGPs in operation representing 3,5% of the installed capacity. Those numbers may increase due to many programs of the Ministry of Mines and Energy that incentive projects of renewable sources as wind, solar and small hydropower plants [ANEEL, 2018; ABRAPCH, 2018].

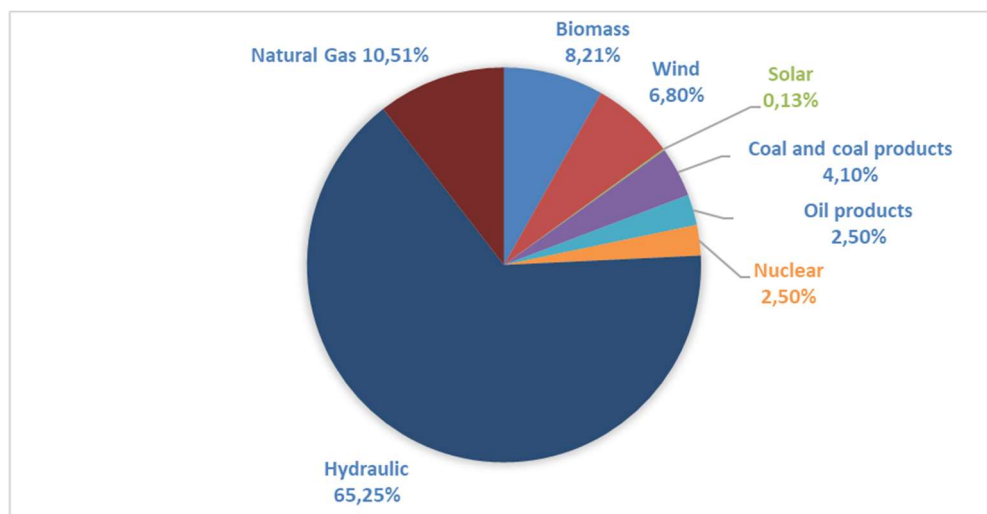


Figure 1: Brazilian Electrical Matrix. Source: EPE, 2018.

In general, SHPs do not have reservoirs to guarantee generation, they are considered as Run-of-the-river plants [Lopes and Borges, 2015]. So, their generation will be proportional to river inflow which in turn depends on the amount of rainfall in each region during the year. This dependence on hydrological regimes directly affects their future generation [Maçaira et al, 2017, Lima et al., 2014]. Therefore, the importance of taking into account in the forecasting models, not only the past generation, but also the inflow data as an important explanatory variable.

Searching in literature some references of spatial-temporal forecasting, there are many recent applications in forecasting of tourism demand. [Yang and Zhang, 2019; Liu et al., 2019]. This subject is also present in econometrics, Cabral et al. (2007) analyses spatially dependence in regional electricity consumption in Brazil and Gomez et al., (2013) considered spatial effects to predict the price and income elasticity of electricity demand; a Spanish residential electricity demand case study.

In many situations, SHP inflow data is not available and sometimes there is not enough observations to carry out the energy forecasts. Lohmann et al. (2016) suggested the use of neighboring inflow data in their study to improve forecasts considering the inflows at the location and also hydro inflows from nearby neighbors. Sousa et al. (2018) use inflow data from neighboring hydroelectric plants as exogenous variable in Regression models obtaining good results.

This paper aims to show an approach for the application of inflow series to solve problems of unavailable data to forecast generation of SHPs. This approach uses a spatial autocorrelation study in order to find the inflow time series that could reduce forecasting error

compared with a simple correlation study. The inflow data from nearby plants with a significant positive Local Indicator of Spatial Association – LISA [Anselin, 1995], will be used as explanatory variable in a causal model.

All the future energy generation forecasts obtained will be compared with those obtained by the SHP real inflow as explanatory variable and a univariate benchmark SARIMA model.

This work is organized in 4 sections. The first section is the introduction where the motivations of this study are presented and, in section 2, the methodology used to develop the study is described. Section 3 presents the results and discussions and section 4 the final considerations.

2. Methodology

There are many ways to produce forecasts and finding the right technique, but this is a difficult task. The specialized literature on the forecasting subject via univariate and/or casual models is rather lengthy. Maçaira et al. (2018) in their systematic review of the literature shows that causal models are widely used, mainly linear regressions, ARIMAX and neural networks. The use of exogenous variables can improve the assertiveness of models representing the impacts of one variable over another. The appropriate choice of variables is an important issue in the implementation of these models.

In this paper the authors opted for the use of popular models, particularly those from the Box and Jenkins family. In this study, the univariate time series used as benchmark model was the SARIMA (p,d,q)x(P,D,Q); i.e. the so called Seasonal Autoregressive Integrated Moving Average [Box and Jenkins, 1976; Box et al., 1994].

As mentioned above, the SARIMA model will be rated as the benchmark formulation in the comparison with the causal counterpart formulations, since the objective of the article is to compare the approaches using the inflow series as explanatory variable.

Considering the SHP time series as Y_t , with $S = 12$ periods (all monthly time series), X_t as the exogenous variable and a_t the white noise; i.e. a stationary stochastic process formed by independent and identical distributed random variables. The multiplicative ARIMA can be written as:

$$\Phi(B^S) * \phi(B) * \nabla_S^D * \nabla^d Y_t = \Theta(B^S) * \theta(B) * a_t$$

Where,

B is the backward shift operator $B^1 Y_t = Y_{t-1}$.

$\Phi(B^S) = (1 - \phi_1 B^S - \dots - \phi_P B^{PS})$ and $\phi_i, i = 1, 2, \dots, P$.

$\phi(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$ and $\phi_i, i = 1, 2, \dots, p$.

$\nabla_S^D = (1 - B^S)^D$ and $D = 0, 1$ or 2 in most of the times.

$\nabla^d = (1 - B)^d$ and $d = 0, 1$ or 2 .

$\Theta(B^S) = (1 - \theta_1 B^S - \dots - \theta_Q B^{QS})$ and $\theta_i, i = 1, 2, \dots, Q$.

$\theta(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$ and $\theta_i, i = 1, 2, \dots, q$.

As mentioned before, some models produced good results with exogenous variables and the SARIMAX (Box & Jenkins Family with exogenous variable) is among these models.

To check the effect of the exogenous variable in the SHP energy generation it was chosen a simple transfer function model as shown below in the next equation, which is easily implemented in the software R with the “forecast” package. [Hyndman and Athanasopoulos, 2013].

$$Y_t = \beta(B)X_t + \frac{\theta(B)}{\phi(B)}a_t$$

Where,

B is the backward shift operator $B^1 Y_t = Y_{t-1}$.

$$\begin{aligned}\beta(B) &= 1 - \beta_1 B^1 - \dots - \beta_w B^w \\ \theta(B) &= 1 - \theta_1 B^1 - \dots - \theta_q B^q \\ \phi(B) &= 1 - \phi_1 B^1 - \dots - \phi_p B^p \\ a_t &\text{ is the white noise process defined above.}\end{aligned}$$

The first challenge in the autocorrelation analysis is the specification of the spatial weighting matrix W . According to Cliff et al. (1975) distance-based matrix (d_{ij}) and contingency-based matrix are some of the spatial weighting matrix methods.

The HPPs are distributed in points with specific coordinates, so the distance-based matrix, is the method used to define which HPPs will be considered as neighbors. In this case, distance (d_{ij}) could be the matrix element w_{ij} , or $w_{ij} = 1/d_{ij}^2$ [Downey, 2006]. Anselin (1995) calls attention to the fact that when the distance between the points is very large, it can generate a matrix with many zeros. Transforming the matrix into a binary one can be interesting. So, after the distance was calculated, the choice of neighbor followed some assumptions: Considering only HPPs distance inferior of 200 km and in the same basin or river.

The spatial weighting matrix is a $n \times n$ matrix with zeros in the diagonal elements (w_{ii}). In this case, the element w_{ij} will be 1 when the HPPs meet the above assumption and zero otherwise [Cliff et al. , 1975].

As indicated by Anselin (1995) the Local Moran I is chosen to verify the similarity of the nearby regions, a significant and positive result indicate similarity between neighbors. The Local Moran I statistics can be written as:

$$I_i = \frac{(y_i - \bar{y})}{m_0} \sum_{j=1}^n w_{ij} (y_j - \bar{y})$$

$$m_0 = \sum_{i=1}^n (y_i - \bar{y})^2 / (n)$$

Where

i is the interest location.

j is the neighbor location.

y_i is the variable of interest, in this study the inflow.

$\bar{y} = \sum_i y_i / n$ is the y average in n .

w_{ij} is the spatial weight between i and j .

The software R is chosen to test, fit and forecast the models used in this study [R Core Team, 2015]. The Root Mean Square Error (RMSE) and the Mean Absolute Percentage Error (MAPE) are the metrics used to validate the out-of-sample results. Their equations are shown below:

$$RMSE = \sqrt{\frac{\sum_{t=1}^N (\hat{Y}_t - Y_t)^2}{N}}$$

$$MAPE = 100 \frac{1}{N} \sum_{t=1}^N \left| \frac{\hat{Y}_t - Y_t}{Y_t} \right|$$

Where Y_t is the observed value of period t , $\hat{Y}_t$ is the predicted value for period t and N is number of observations.

Summarizing, the proposed methodology considering the neighboring inflow series is:

- Fit and forecast with SARIMA model playing the role of benchmarking (case 1).
 - Carry out a spatial autocorrelation study to identify neighboring inflow series that could reduce the forecast error.
 - Carry out a correlation study between Generation and Inflow considering the situation the strongest correlation with the inflow series.
 - Fit SARIMAX with the Inflow as exogenous variable.
 - Using SHP Real Inflow (case 2).
 - Inflow from neighbors' plants chosen after spatial autocorrelation analysis (case 3).
 - Inflow with the stronger correlation with SHP Generation. (case 4).
- Forecast.
 - Compare and analyze the results.

3. Results and discussions

It was first conducted a graphical analysis of the time series to identify basic components such as: trends, seasonality and level. The electric power generation of 4 time series located at *Pinho River* and *Pomba River*, *Minas Gerais* state are shown in Figure 2. *Ivan Botelho 1* SHP shows seasonality and a slight decreasing trend. These characteristics can be observed in detail with the time series decomposition (Figure 3).

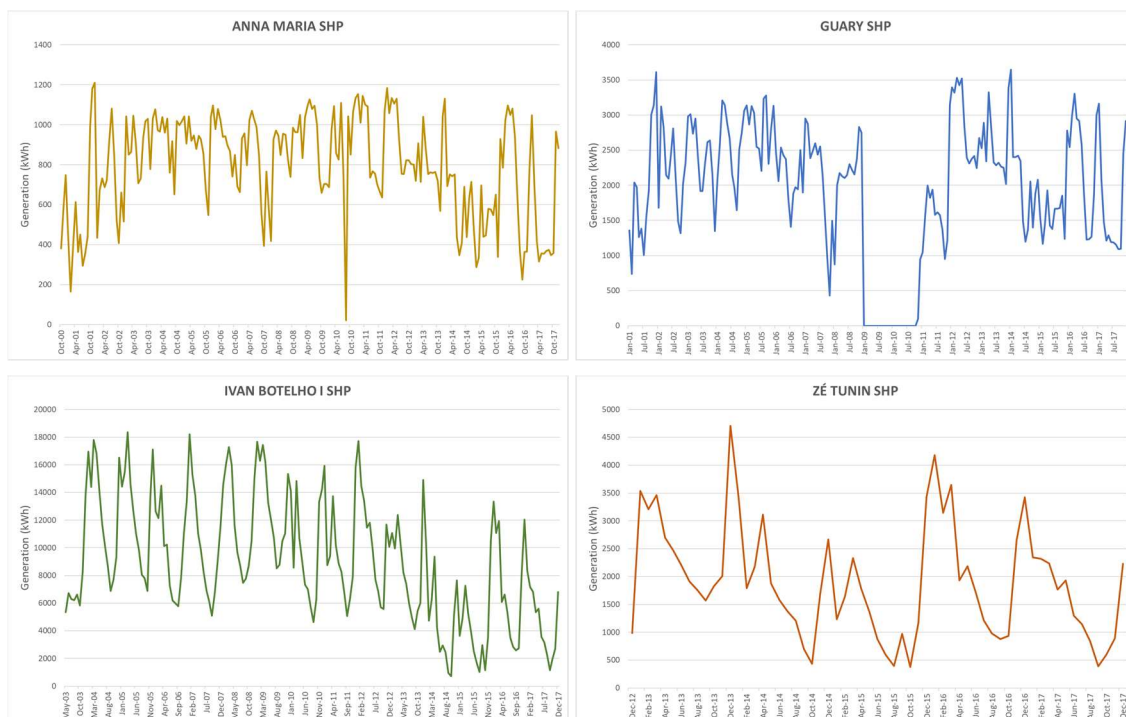


Figure 2 – Energy Generation Time Series. Source: The author, 2019

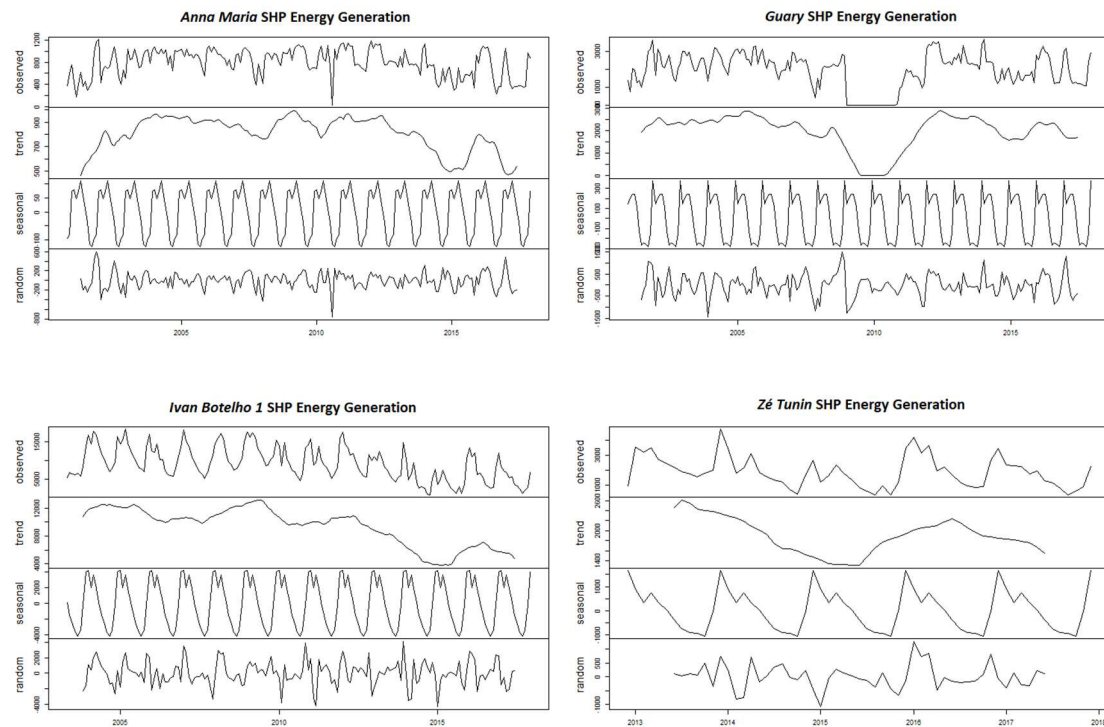


Figure 3 – Time series decomposition in trend, seasonality and residuals. Source: The author, 2019.

Continuing the analysis of the time series presented in Figure 2 and 3, it is necessary to verify if the series is stationary to allow the application of Box & Jenkins models. For all time series, the Augmented Dickey-Fuller test was performed finding $p\text{-value} < 0.02$, rejecting, therefore, the null hypothesis and accepting the stationarity of the series.

The SHP Energy Generation time series was provided by a Brazilian Utility located at the Southern Region. The time series energy generation analyzed is described in Table 1.

Table 1 – SHP considered.

SHP	River	Operation start date	Power (MW)	Observations
<i>Anna Maria</i>	Pinho	oct-00	1.5	207
<i>Guary</i>	Pinho	jan-01	5.3	204
<i>Ivan Botelho 1</i>	Pomba	mai-03	24.4	176
<i>Zé Tunin</i>	Pomba	dec-12	8.0	61

The series was divided into training (from the operation start date to December 2016) and validation (January 2017 to December 2017). This strategy allows the check of the fitted model by comparing the forecast with the actual data.

As the inflow data (exogenous variable) from these SHPs is available from January 2015 to December 2017, when testing case 2, the energy generation training series (response variable) will have the same number of observations, it means that it will start in January 2015.

The next step consists of the spatial autocorrelation analysis. To do so, 38 Hydroelectric Power Plants (HPP) with inflow data, the blue points in the map shown in Figure 4, located in *Rio de Janeiro*, *Minas Gerais* and *Espírito Santo* states are candidates to participate in this study. These data were provided by the National Electric System Operator [ONS, 2018].

As mentioned before, sometimes there is no inflow data available, so it is why spatial dependence approach could help to identify one or more neighbor's inflow time series which approximate or reduce forecasting error.

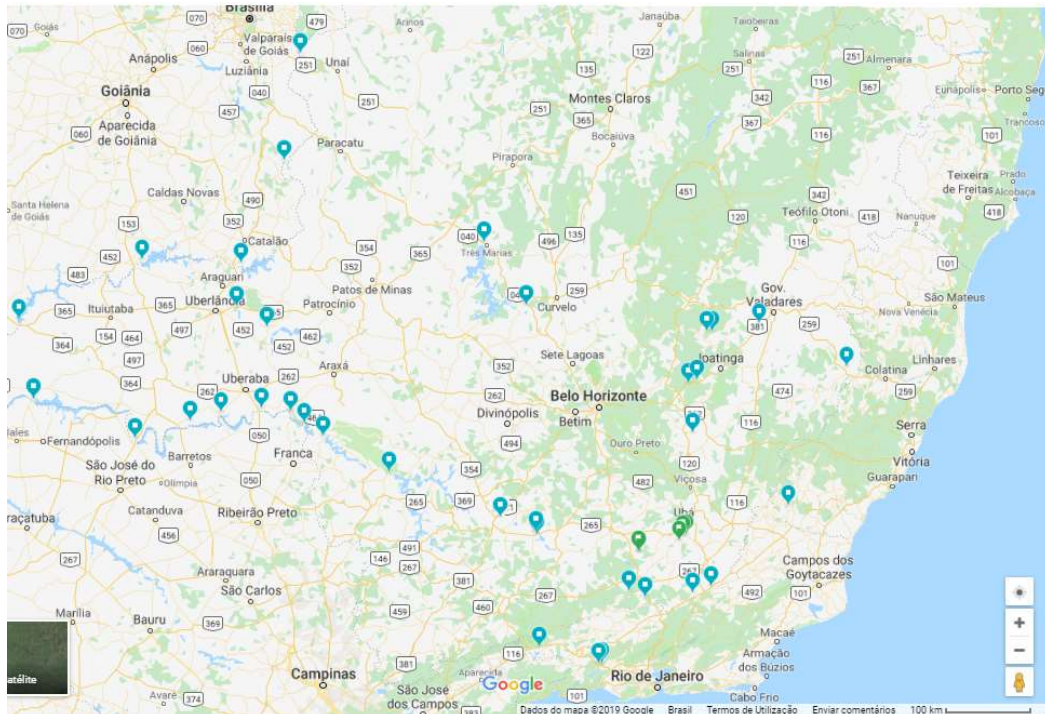


Figure 4 - HPP and SHP geographical distribution. Source: The authors, 2019.

To perform the spatial autocorrelation taking into account the plants inflow data, it is necessary to construct the spatial weighting matrix related to the location of each plant analyzed.

According to Tobler's First Law, "everything is related to everything else, but things are more related than distant things." [Tobler, 1970, p. 236], it means, the nearest points are more spatially correlated than those that are far away. For that reason, the study started the analysis only with the hydropower plants that are geographically located in the states of *Minas Gerais*, *Rio de Janeiro* and *Espírito Santo* were considered. But not only the distance between HP plants are considered in this study. It was also considered if plants are located in the same river and/or basin.

The neighboring hydroelectric plants inflow that will be considered in the study of the spatial correlation belongs to *Paraná*, Southeast Atlantic and *São Francisco* hydrographic basins (Figure 5). However, after calculating the distances in km between the SHPs and the HPPs, it was found that the plants located geographically in the same sub-basin as the SHPs are strong candidates for neighbors. Going further, the nearest HPPs from the *Paraíba do Sul* sub-basin chosen as neighbors are *Picada*, *Sobrangi*, *Simplicio* and *Ilha dos Pombos*, and they are located at less than 120 km from SHPs.

The Local Moran I (0.6720) calculated with these four HPPs selected as neighbors presents a significant and positive result, what indicates that a local spatial association with nearby regions.



Figure 5 – Brazil Hydrographic basins. Source: Adapted from ONS, 2019.

The strongest correlation between SHP Generation (kWh) and HPP Inflow (m³/s) was displayed in Table 2. *Ivan Botelho 1* and *Zé Tunin* show strong correlation with *Sobragi*, one of the HPP considered as neighbor in the spatial dependence analysis with approximately 100 km distant. On the other hand, in the case of *Anna Maria*, Pearson's correlation coefficient indicates a moderate correlation with *Batalha* HPP which reinforce the Tobler's first law. *Batalha* HPP is 800 km far from *Anna Maria* SHP.

Table 2 – Correlation between neighboring Inflow and Generation

SHP Generation	HPP Inflow	Correlation
<i>Anna Maria</i>	<i>Batalha</i>	0,48
<i>Guary</i>	<i>Simplicio</i>	0,59
<i>Ivan Botelho 1</i>	<i>Sobragi</i>	0,86
<i>Zé Tunin</i>	<i>Sobragi</i>	0,91

In total 4 models were tested and forecast 12-steps-ahead were produced. The case 1 (benchmark model) is a SARIMA model. The other 3 models are those that used explanatory variables. It was applied, in case 2, the real SHP inflow as explanatory variable directly in SARIMAX. Case 2 is expected to be the best result, but for *Zé Tunin* the case 4 shows a better MAPE. The third case considers all inflow neighbors (*Picada*, *Sobragi*, *Simplicio* and *Ilha dos Pombos*) as explanatory variables and shows a satisfactory result compared with case 2. The Case 4 that uses the inflow data from HPP stronger correlated with SHP Generation has results close to the univariate SARIMA model. Table 3 shows the full results of all the models tested and Figure 6 presents the 12-steps-ahead out-of-sample forecasts.



Figure 6 - 12-steps-ahead out-of-sample forecasts.

In addition to the MAPE, the residuals were tested to verify dependence. The Ljung-Box test presented a p-value greater than 0.05, indicating the absence of any kind of dependence on the residuals series.

Table 3: Out-of-sample comparison: MAPE and RMSE

SHP	Metrics	Case 1	Neighbor Inflow		
			Real SHP Inflow Case 2	Case 3	Case 4
Anna Maria	MAPE (%)	64.3	28.3	29.8	53.9
	RMSE (kWh)	291.04	293.84	273.91	260.52
Guary	MAPE (%)	67.9	10.5	49.5	54.6
	RMSE (kWh)	958.22	205.03	749.96	816.29
Ivan Botelho 1	MAPE (%)	41.2	24.8	34.9	48.6
	RMSE (kWh)	2162.48	861.28	1263.58	1403.63
Zé Tunin	MAPE (%)	73.8	32.7	30.9	27.1
	RMSE (kWh)	914.65	321.09	1739.29	358.67

4. Final Considerations

The main objective of this paper was the application of spatial autocorrelation to identify neighboring inflow data from Hydroelectric Plants near SHP plants and compare with Real SHP inflow and others approaches to analyses if the results are satisfactory.

Utility companies need to plan their operation and small hydropower plants has a stronger dependence with the hydrological regimes due to its run-of-river characteristic, making the energy generation forecast a difficult task. Added to it, in order to produce accurate forecasts, it is important to join inflow data as exogenous variable and the use of such data were also included in this study. The inflow data from SHP are not always available and; even when they exist, the number of observations may be not enough to improve on the accuracy of the forecasts. Real case was used to test the proposed methodology considering univariate and causal forecasting models.

An important contribution of this research is the use of inflow data from hydroelectric plants spatial located near a river basin when the inflow data from a SHP are not available. By the

spatial autocorrelation study it is possible to identify useful data to increase accuracy of the predictions.

The approach that considered the correlation study between HPP inflow and SHP Generation (case 4) shows that sometimes it could find good time series, as the case of *Zé Tunin* that shows a stronger correlation (0.91) with *Sobragi* and the best out-of-sample MAPE. But in the case of *Anna Maria* SHP, *Batalha* HPP shows the stronger correlation (0.48) and is 800 km far away what can be a spurious correlation.

However, other models could be implemented with different SHP plants and different number of observation and also with different windows of validation.

Another important issue not considered in this research has to do with an analysis of extremes values that appear in the energy generation time series. These values may interfere in the results if not treated and explained the reason for their existence. For instance, they can be related to maintenance periods of the plant, or even errors caused by wrong measurements.

For further studies, it is also possible to test dynamic regression, a more complex Transfer Function model, as well as techniques based on artificial intelligent methods and hybrid approaches.

As a final word, this is a going on research. The basic idea of this modelling study is to arrive at a kind of space-time formulation, where the location and spatial weighting matrix will be included in the final forecasting model. So far, it was only considered the time dependence of the structural dependency of these HPP.

Acknowledgements

The author thanks the R&D program of the Brazilian Electricity Regulatory Agency (ANEEL) for the financial support (P&D 06585-1802/2018). This study was financed in part by the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil* (CAPES) - Finance Code 001.

5. References

- ANEEL. (2018). Metodologia de Cálculo Tarifário da Distribuição. Agência Nacional de Energia Elétrica. Available at http://www.aneel.gov.br/metodologia-distribuicao/-/asset_publisher/e2INtBH4EC4e/content/perdas/654800?inheritRedirect=false. Accessed date January 24, 2019.
- ABRAPCH. (2018). O que são PCHs e CGHs. Associação Brasileira de Pequenas Centrais Hidrelétricas e Centrais Geradoras Hidrelétricas. Available at <http://www.abrapch.org.br/pchs/o-que-sao-pchs-e-cghs>. Accessed date June 15, 2018.
- Anselin, L. (1995). Local Indicators of Spatial Association—LISA. *Geographical Analysis*, 27: 93-115.
- Box, G. E. P., Jenkins, G. M. (1976). Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco.
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C. (1994). Time Series Analysis: Forecasting and Control. Prentice Hall.
- Cabral, J.A., Legey L.F. L., Cabral M. V. F. (2017). Electricity consumption forecasting in Brazil: A spatial econometrics approach, *Energy*, Volume 126, 2017, p. 124-131.
- Cliff, A.D., Haggett, P., Ord, J.K., Bassett, K., Davies, R. (1975). Elements of Spatial Structure – A Quantitative Approach. Cambridge University Press.

Downey, L. (2006). Using Geographic Information Systems to Reconceptualize Spatial Relationships and Ecological Context. *American Journal of Sociology*, 112(2), 567-612.

EMPRESA DE PESQUISA ENERGETICA (BRASIL). (2018) Balanço Energético Nacional 2018: Ano base 2017 / Empresa de Pesquisa Energética. – Rio de Janeiro: EPE, 2018. Available at http://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-303/topico-419/BEN2018__Int.pdf. Accessed date January 21, 2018.

Gomez L. M. B., Heimsch M. F. F. (2013). Regional impact of changes in disposable income on Spanish electricity demand: A spatial econometric analysis. *Energy Economics*, Volume 40, Supplement 1, p. S58-S66.

Hyndman, R., Athanasopoulos, G. (2018). Forecasting: principles and practice. Available at <http://otexts.com/fpp2/>. Accessed date December, 2018.

Lima, L. M. M.; Popova, E.; Damien, P. (2014). Modeling and forecasting of Brazilian reservoir inflows via dynamic linear models. *International Journal of Forecasting*, 30(3):464–476.

Liu, P.; Zhang, H.; Zhang, J.; Sun, Y.; Qiu, M. (2019). Spatial-temporal response patterns of tourist flow under impulse pre-trip information search: From online to arrival. *Tourism Management*, Volume 73, p. 105-114.

Lohmann, T.; Hering, A. S.; Rebennack, S. (2016). Spatio-temporal hydro forecasting of multireservoir inflows for hydro-thermal scheduling. *European Journal of Operational Research*, 255(1):243–258.

Lopes, V. S.; Borges, C. L. T. (2015). Impact of the Combined Integration of Wind Generation and Small Hydropower Plants on the System Reliability. *IEEE TRANSACTIONS ON SUSTAINABLE ENERGY*, Vol. 6, no. 3.

Maçaira, P. M.; Thome, A. M. T.; Cyrino Oliveira, F. L.; Carvalho Ferrer, A. L. (2018). Time series analysis with explanatory variables: A systematic literature review. *Environmental Modelling & Software*, v. 107, p. 199-209.

ONS. 2018. Resultados da Operação – Histórico da Operação - Dados Hidrológicos/Vazões. Available at http://ons.org.br/Paginas/resultados-da-operacao/historico-da-operacao/dados_hidrologicos_vazoes.aspx. Accessed date July 10, 2018.

R Core Team. (2015). The R Project for Statistical Computing. Available at <https://www.r-project.org/>. Accessed date March 20, 2018.

Sousa, M. A.; Maçaira, P. M. ; Souza, R. C. ; Oliveira, F. L. C. (2018). Forecasting electricity generation of small hydro-power plants. In: *2018 POMS International Conference in Rio*, Rio de Janeiro. Proceedings of the 2018 POMS International Conference in Rio.

Tobler W. R. A. (1970). Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, Vol. 46, Supplement: *Proceedings. International Geographical Union. Commission on Quantitative Methods*, p. 234-240.

Yang, Y; Zhang, H. (2019). Spatial-temporal forecasting of tourism demand. *Annals of Tourism Research*, Vol. 75, p. 106-119.