



Margarete Afonso de Sousa

**Desenvolvimento de Modelos de Previsão de
Geração de Energia Elétrica aplicados às Pequenas
Centrais Hidrelétricas**

Dissertação de Mestrado

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Produção da PUC-Rio como requisito parcial para obtenção do grau de Mestre em Engenharia de Produção.

Orientador: Prof. Reinaldo Castro Souza

Rio de Janeiro
Setembro de 2019



Margarete Afonso de Sousa

**Desenvolvimento de Modelos de Previsão de
Geração de Energia Elétrica aplicados às Pequenas
Centrais Hidrelétricas**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-Graduação em Engenharia de Produção da PUC-Rio. Aprovada pela Comissão Examinadora abaixo.

Prof. Reinaldo Castro Souza

Orientador

Departamento de Engenharia Industrial – PUC-Rio

Prof. André Luís Marques Marcato

Universidade Federal de Juiz de Fora - UFJF

Prof. José Francisco Moreira Pessanha

Centro de Pesquisa de Energia Elétrica - CEPEL

Rio de Janeiro, 13 de setembro de 2019

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, da autora e do orientador.

Margarete Afonso de Sousa

Graduou-se em Engenharia Elétrica com Ênfase em Eletrônica pelo Centro Federal de Educação Tecnológica – Celso Suckow da Fonseca no Rio de Janeiro (CEFET/RJ) em 1992. Concluiu a pós-graduação em Análise de Sistemas pela Pontifícia Universidade Católica do Rio de Janeiro em 2002, o MBA em Gestão de Negócios pelo IBMEC-RJ em 2004 e a pós-graduação em Economia Empresarial pela Universidade Cândido Mendes em 2009. Atuou em projetos de Engenharia e Tecnologia da Informação nas áreas de Petróleo e Gás, Energia Elétrica e empresas de Tecnologia da Informação e como professor/tutor em cursos de MBA à distância.

Ficha Catalográfica

Sousa, Margarete Afonso de

Desenvolvimento de modelos de previsão de geração de energia elétrica aplicados às pequenas centrais hidrelétricas / Margarete Afonso de Sousa ; orientador: Reinaldo Castro Souza. – 2019.

73 f. : il. color. ; 30 cm

Dissertação (mestrado)–Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Industrial, 2019.

Inclui bibliografia

1. Engenharia Industrial – Teses. 2. Pequenas centrais hidrelétricas. 3. Previsão de geração de energia elétrica. 4. Modelos causais. 5. Séries temporais. 6. Autocorrelação espacial. I. Souza, Reinaldo Castro. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Industrial. III. Título.

CDD: 658.5

Agradecimentos

Nesses dois anos vivi momentos muito importantes na minha formação e gostaria de agradecer:

A Deus por me conduzir com sabedoria e determinação em todos os momentos de minha vida.

À minha filha, Catarina, pela paciência nesses dois anos abrindo mão, em muitos momentos, da minha presença, em função da rotina em dias de aulas, noites de estudo, viagens para congressos e muitos finais de semana na frente do computador. É por você que me esforço para ser um bom exemplo.

Ao meu marido, Eduardo, que desde o dia que o conheci me encoraja a realizar os meus sonhos. Nesses quase 20 anos juntos você sempre esteve presente apoiando minhas conquistas profissionais e principalmente me ajudando a chegar até aqui.

Aos meus pais, Jane e Afonso, que trabalharam duro para que eu pudesse estudar e trilhar o caminho que escolhi. Muito obrigada pelos valores que vocês me ensinaram.

Ao meu orientador, Prof. Reinaldo Castro Souza por acreditar na minha capacidade, pelos ensinamentos e oportunidades. Foi muito bom conviver e aprender com você nesse período.

Ao Prof. Fernando Luiz Cyrino Oliveira, por ter sido o primeiro a me apresentar análise de séries temporais, despertando a minha curiosidade e interesse. Você é um exemplo de determinação e comprometimento que devemos seguir.

Aos amigos que fiz na PUC-Rio, especialmente a Brenda, pelas conversas e conhecimento compartilhado, aprendi muito com você.

À Paula Maçaira, por compartilhar o seu conhecimento durante esse último ano mostrando os caminhos para alcançar meus objetivos. Obrigada.

À equipe do projeto de pesquisa e desenvolvimento que motivou essa dissertação, aprendi e continuo aprendendo muito com todos vocês.

Aos professores e funcionários do Departamento de Engenharia Industrial da PUC-Rio, pelos ensinamentos e conhecimento compartilhados.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Resumo

Sousa, Margarete Afonso de; Souza, Reinaldo Castro. **Desenvolvimento de Modelos de Previsão de Geração de Energia Elétrica aplicados às Pequenas Centrais Hidrelétricas**. Rio de Janeiro, 2019. 73p. Dissertação de Mestrado – Departamento de Engenharia Industrial, Pontifícia Universidade Católica do Rio de Janeiro.

Uma das principais preocupações mundiais atualmente está relacionada às questões ambientais. Essa preocupação é considerada na seleção de projetos de energia e, como resultado, a geração de energia elétrica a partir de fontes renováveis tem experimentado um forte crescimento em todo o mundo, incluindo o Brasil. Em relação às fontes de energia hidrelétrica, as Pequenas Centrais Hidrelétricas (PCHs) são uma alternativa para reduzir o impacto ambiental. Esses projetos produzem entre 5 e 30 megawatts (MW) e sua instalação tem um baixo custo e respeito ao meio ambiente, principalmente por não existir necessidade de reservatórios de regulação, o que não é o caso de grandes usinas hidrelétricas. Nos últimos anos, o número de PCHs tem aumentado bastante, como consequência dos incentivos para geração de eletricidade a partir de fontes renováveis. Como a geração de energia hidrelétrica é fortemente influenciada por regimes hidrológicos, especialmente no caso de usinas a fio d'água como as PCHs, melhorar a assertividade das previsões de geração de energia elétrica de maneira estocástica torna-se altamente importante para as distribuidoras. Esta dissertação tem como principal objetivo apresentar o desempenho de um grupo de modelos de previsão aplicados para PCHs de uma distribuidora real de energia elétrica. Para isso foram utilizadas diferentes abordagens, incluindo dados de vazão de usinas hidrelétricas vizinhas como variável explicativa em modelos causais, assim como também modelos univariados.

Palavras-chave

Pequenas Centrais Hidrelétricas; Previsão de Geração de Energia Elétrica; Modelos Causais; Séries Temporais; Autocorrelação espacial.

Abstract

Sousa, Margarete Afonso de; Souza, Reinaldo Castro (Advisor). **Development of electric power generation forecasting models applied to Small Hydropower Plants.** Rio de Janeiro, 2019. 73p. Dissertação de Mestrado – Departamento de Engenharia Industrial, Pontifícia Universidade Católica do Rio de Janeiro.

One of the main world concerns nowadays is related to the environment issues. Such concern is considered in the selection of energy projects and, as a result of that, the generation of electricity from renewable sources has experienced a sharp growth all over the world, Brazil included. Concerning hydropower sources, Small Hydropower Plants (SHPs) are an alternative to reduce environmental impact. These projects produce between 5 and 30 megawatts (MW) and its installation has a low cost and respect to the environment, mainly because there is no need of regulation reservoirs, which is not the case in bigger hydroelectric plants. In recent years the number of SHPs is increasing in a great deal, as a consequence of the incentives to generate electricity from renewable sources. Since hydro power generation is heavily influenced by hydrological regimes, especially in the case of run-of-river plants, as SHPs, improving the assertiveness of electric power generation forecasts in a stochastic way becomes highly important for distributing utilities. This master dissertation has as main objective to present the performance of an arrange of forecasting models applied to SHPs of a real distributing utility. It was used different approaches, including inflow data from neighboring hydro plants as exogenous variable, in causal models and also univariate models.

Keywords

Small Hydropower Plants; Electric Power Generation Forecast; Causal Models; Time Series, Spatial Autocorrelation.

Sumário

1	Introdução	11
1.1.	Justificativa e Motivação	13
1.2.	Objetivos	14
1.3.	Estrutura da Dissertação	14
2	Referencial Teórico	16
2.1.	Pequenas Centrais Hidrelétricas (PCHs)	16
2.2.	Variáveis Exógenas e Dependência Espacial	17
3	Metodologia	19
3.1.	Modelos Univaridados	19
3.1.1.	Amortecimento Exponencial Sazonal de Holt-Winters	20
3.1.2.	Autoregressivo Integrado de Médias Móveis (ARIMA)	23
3.2.	Modelos Causais	25
3.2.1.	Função de Transferência de Box & Jenkins	25
3.2.2.	Regressão Dinâmica	28
3.3.	Autocorrelação Espacial	29
4	Análise Descritiva dos Dados	36
4.1.	Dados de Geração de Energia Elétrica das PCHs	36
4.2.	Dados de Vazão Usinas Hidrelétricas Vizinhas	40
4.3.	Dados de Vazão PCHs	47
4.4.	Análise Espacial	49
5	Resultados Obtidos	51
5.1.	Etapa 1 - Previsão dos Modelos Univariados	51
5.2.	Etapa 2 – Obtenção e Tratamento da Variável Explicativa	54
5.3.	Etapa 3 – Aplicação de Modelos Causais	59
5.4.	Etapa 4 – Previsão com Cenário de Vazão	65
6	Conclusão e Trabalhos Futuros	68
	Referências Bibliográficas	70

Lista de Figuras

Figura 1.1 – Matriz Elétrica Brasileira.	11
Figura 3.1 – Comportamento das Séries e o Modelo mais Adequado.	21
Figura 3.2 – Caso “Bispo”, “Torre” e “Rainha”.	31
Figura 3.3 – Metodologia Aplicada.	34
Figura 4.1 – Séries Temporais de Geração de Energia Elétrica das PCHs.	37
Figura 4.3 – Vazões das Usinas Hidrelétricas Vizinhas.	41
Figura 4.4 – Bacias Hidrográficas.	45
Figura 4.5 – Distribuição Geográfica das Usinas Hidrelétricas Vizinhas.	46
Figura 4.6 – Séries de Vazão Próprias das PCHs.	49
Figura 5.1 – Decomposição das Séries Temporais de Geração em Nível, Tendência, Sazonalidade e Resíduo.	52
Figura 5.2 – Gráficos de Dispersão da Vazão Própria e Geração de cada Usina.	55
Figura 5.3 – Gráficos de Linha da Vazão Própria e Geração de cada Usina.	56
Figura 5.4 – Gráficos de Dispersão da Vazão mais Correlacionada e Geração de cada Usina.	57
Figura 5.5 – Comparativo das Melhores Previsões 12 passos a frente para cada PCH.	64
Figura 5.6 – Cenário Vazões Vizinhas – Médias Mensais para 2018.	65
Figura 5.7 – Cenário Vazões Próprias – Médias Mensais para 2018.	66
Figura 5.8 – Previsões da Geração das PCHs para 2018.	67

Lista de Tabelas

Tabela 4.1 – PCHs Consideradas na Pesquisa.	36
Tabela 4.2 – Coordenadas Geográficas das PCHs.	38
Tabela 4.3 – Estatísticas Descritivas das Séries de Geração das PCHs.	40
Tabela 4.4 – Coordenadas das Usinas Hidrelétricas Vizinhas.	45
Tabela 4.5 – Estatística Descritiva das Séries de Vazão das UHEs.	47
Tabela 4.6 – Dados de Vazões das PCHs.	48
Tabela 4.7 – Estatística Descritiva das Séries de Vazão das PCHs.	49
Tabela 5.1 – Divisão das Séries Temporais em Janela de Treinamento e Validação.	51
Tabela 5.2 – Resultado do P-valor para os Testes de Raízes Unitárias.	52
Tabela 5.3 – Resultado MAPE (%) e RMSE (kWmed) – Amortecimento Exponencial e ARIMA.	54
Tabela 5.4 – Novas Janelas de Treinamento e Validação das Séries Temporais para Ajuste do Caso 1.	55
Tabela 5.5 – Correlações entre Geração e Vazão das UHEs Vizinhas.	58
Tabela 5.6 – Usinas Hidrelétricas mais Correlacionadas com Geração.	58
Tabela 5.7 – Novas Janelas de Treinamento e Validação das Séries Temporais para Ajuste do Caso 2.	58
Tabela 5.8 – Novas Janelas de Treinamento e Validação das Séries Temporais para Ajuste do Caso 3.	59
Tabela 5.9 – Modelos – Função de Transferência.	60
Tabela 5.10 – Resultado MAPE (%) e RMSE(kWmed) – Função de Transferência.	61
Tabela 5.11 – Modelos – Regressão Dinâmica.	62
Tabela 5.12 – Resultado MAPE(%) e RMSE(kWmed) – Regressão Dinâmica.	63
Tabela 5.13 – Resultados do MAPE(%) <i>Out-of-Sample</i> Consolidado.	64
Tabela 5.14 – Resultados do MAPE(%) e RMSE (kWmed) de Teste Consolidado.	66

1 Introdução

O Brasil é um país que possui um grande potencial hídrico e, em função desta característica, a matriz elétrica brasileira possui 65,2% da energia gerada através de usinas hidrelétricas (EPE, 2018). Além disso, a energia hidrelétrica é considerada uma fonte de energia renovável que ao lado da energia eólica, biomassa e solar correspondem a 80,4% da matriz elétrica brasileira. Porém, essa fonte de energia elétrica possui uma grande dependência dos regimes hidrológicos causando uma variação na geração durante períodos secos e úmidos. Em função dessa variação ao longo do ano, outras fontes de geração de energia elétrica são necessárias, como as usinas termelétricas, para garantir a demanda por eletricidade (ONS, 2018). A Figura 1.1 apresenta a matriz elétrica por fonte de energia.

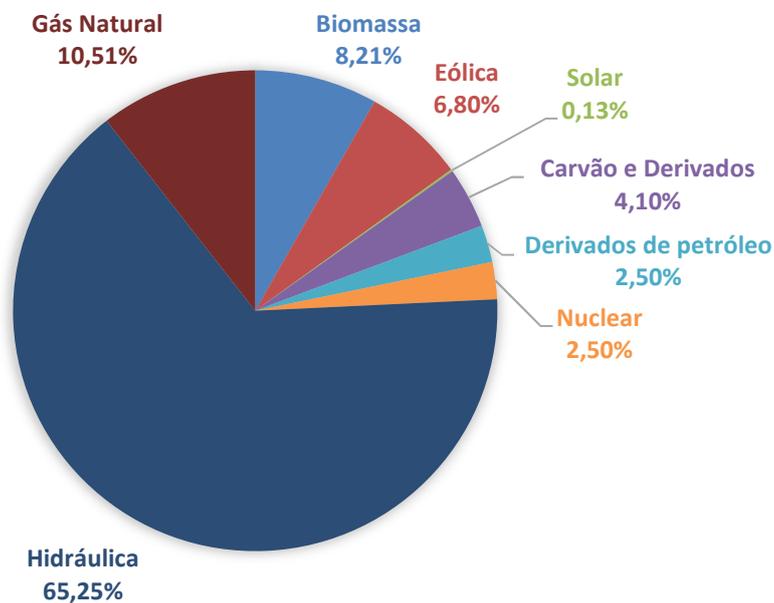


Figura 1.1 – Matriz Elétrica Brasileira.
Fonte: EPE, 2018.

Desde a crise de 2001, o Ministério das Minas e Energia estimula o crescimento de fontes alternativas de energia elétrica e criou programas de

incentivo, como o PROINFA¹, que estimulou empresas a investirem em geração de energia eólica, biomassa e, também, em energia hidrelétrica, porém com baixo impacto ambiental, como as Pequenas Centrais Hidrelétricas (PCHs).

As PCHs são usinas que produzem entre 5 e 30 megawatts (MW) e sua instalação tem um baixo custo e como citado anteriormente, quase nenhum impacto ambiental pois não necessitam de reservatórios para a geração da energia e quando possuem reservatórios sua área não ultrapassa 13km². Além das PCHs, existem outros dois tipos de usinas, as Usinas Hidrelétricas (UHEs) que produzem mais de 30MW e possuem reservatórios e as Centrais Geradoras Hidrelétricas (CGHs) que possuem capacidade inferior a 5 MW e não possuem reservatórios (ABRAPCH, 2018).

Atualmente no Brasil existem 428 PCHs e 683 CGHs em operação, representando 3,5 % da capacidade instalada do Sistema Interligado Nacional – SIN (ANEEL, 2018; ABRAPCH, 2018).

Como as PCHs na maior parte das vezes não possuem reservatórios para garantir a geração, estarão sujeitas a vazão dos rios que por sua vez dependem da quantidade de chuva durante o ano em cada região. Essa dependência em relação aos regimes hidrológicos afeta diretamente a geração deste tipo de usina (Maçaira et al, 2017; Lima et al., 2014).

A estocasticidade da geração hídrica das PCHs e CGHs influencia diretamente a definição da energia que será injetada na rede e utilizada no cálculo das perdas técnicas na rede de distribuição. Atualmente, a Agência Nacional de Energia Elétrica – ANEEL, órgão regulador do setor, estipula o valor (percentual) de perdas técnicas eficiente na rede de distribuição, a partir de informações de equipamentos existentes, assim como a energia injetada na rede para cada concessionária. Esse valor regulatório é calculado durante a Revisão Tarifária

¹ PROINFA – Programa de Incentivo às fontes alternativas de energia elétrica.

Periódica² e acontece a cada 4 ou 5 anos (ANEEL, 2018). Além disso, esse cálculo é feito de forma determinística, a partir de informações passadas.

Devido a esta estocasticidade, relacionada a influência dos regimes hidrológicos, é importante levar em conta na previsão da geração futura das PCHs, além da geração passada, dados de vazão como variável explicativa (Ferreira et al., 2015; Lohmann et al., 2016).

Lohmann et al. (2016) utilizam informações espaciais de bacias vizinhas, em um modelo espaço-temporal chamado SPAR(p), com o objetivo de melhorar as previsões das vazões de usinas hidrelétricas de três bacias brasileiras. Segundo Lohmann et al (2016), como as vazões que estão mais próximas do local que se deseja realizar as previsões tendem a ser mais parecidas do que aquelas que estão muito distantes, essas informações podem ser incorporadas aos modelos de previsão de vazões, assim como as vazões medidas das próprias usinas, a fim de diminuir o erro nas previsões.

Partindo da ideia apresentada por Lohmann et al. (2016) em seu trabalho, e considerando que em muitos casos as PCHs e CGHs não possuem dados de vazão suficientes para realizar as previsões de geração, uma alternativa apresentada nesta dissertação será a utilização de séries de vazão de bacias vizinhas disponíveis como variáveis exógenas na previsão da geração.

1.1. Justificativa e Motivação

Este trabalho possui como motivação a resolução do problema apresentado pela empresa Energisa de Minas Gerais, relacionado ao impacto da previsão da geração hídrica de PCHs e CGHs na definição da meta das perdas técnicas na distribuição e transmissão em função da estocasticidade desta fonte de energia elétrica.

² Esse processo permite a revisão econômico-financeira das concessionárias de energia, permitindo estimar a receita adequada levando em consideração os custos operacionais e os investimentos realizados (ANEEL, 2018).

Por isso, aumentar a assertividade de modelos de previsão de geração de energia elétrica é fundamental para resolução deste problema. Desta forma, os resultados obtidos poderão ser utilizados na criação de cenários que considerem a estocasticidade relacionada aos regimes hidrológicos.

O desenvolvimento desses modelos é o primeiro passo para melhorar a qualidade das estimativas possibilitando assim minimizar o impacto das perdas técnicas na transmissão e distribuição.

1.2. Objetivos

O principal objetivo dessa dissertação é apresentar o desempenho de modelos de previsão de gerações de PCHs utilizando dados de vazão de bacias vizinhas como variável explicativa. Para alcançar esse objetivo serão utilizadas técnicas de modelos estatísticos causais (função de transferência e regressão dinâmica).

O objetivo secundário dessa dissertação é a comparação da capacidade preditiva desses modelos com outros muito utilizados na literatura e que não utilizam variáveis exógenas, como amortecimento exponencial e os modelos ARIMA.

Um terceiro objetivo esperado, visto que em muitos casos a indisponibilidade de dados de vazão pode prejudicar as previsões das gerações futuras, é a utilização da análise da autocorrelação espacial a fim de indicar usinas localizadas na vizinhança que possuam dados suficientes a fim de aumentar a assertividade de modelos causais utilizando a vazão como variável explicativa.

1.3. Estrutura da Dissertação

Esta dissertação está organizada em seis capítulos. Após este Capítulo introdutório, a revisão da literatura é realizada no Capítulo 2, apresentando estudos que indicam a utilização de variáveis explicativas, além da análise espacial aplicada na previsão.

Em seguida no Capítulo 3 é apresentado o arcabouço metodológico que possibilita a realização deste estudo. No Capítulo 4, os dados utilizados nesta pesquisa são detalhados.

Posteriormente no Capítulo 5 os resultados obtidos a partir da aplicação da metodologia descrita são apresentados. As conclusões e sugestões para continuação da pesquisa são apresentadas no Capítulo 6.

2 Referencial Teórico

As usinas hidrelétricas aproveitam a vazão dos rios para gerar energia elétrica transformando a energia potencial da água em mecânica e por sua vez em energia elétrica. A energia gerada é proporcional à vazão que movimenta as turbinas (Lopes et al., 2015). Logo a geração está diretamente associada aos regimes hidrológicos, ou seja, a vazão depende da quantidade de chuva na cabeceira dos rios. Como o principal objetivo desta pesquisa é a apresentação de resultados relacionados a modelos de previsão de geração de energia elétrica de PCHs utilizando dados de vazão de bacias vizinhas como variável explicativa, esta revisão da literatura buscou apresentar estudos que utilizam esta abordagem.

2.1. Pequenas Centrais Hidrelétricas (PCHs)

A energia hidráulica é responsável por 75% da energia elétrica gerada a partir de fontes de energia renováveis no mundo. De acordo com os dados de 2016, as fontes de energia renováveis representam 22,2% da matriz elétrica mundial. Nos últimos anos, outras fontes de energia renováveis aumentaram sua participação na matriz elétrica, por exemplo, em 2017 a energia eólica instalada cresceu no Brasil 21,3% em relação ao ano anterior (EPE, 2018).

Usinas hidrelétricas de grande porte utilizam reservatórios para garantir a geração de energia elétrica, porém a maioria das PCHs opera a fio d'água, ou seja, a água passa continuamente pelas turbinas (Queiroz et al., 2019). Outro ponto comum entre as PCHs é o posicionamento em desníveis dos rios para aproveitamento da força da gravidade proporcionando um maior aproveitamento da vazão dos rios.

O Brasil tem um enorme potencial hidráulico para a implantação de usinas hidrelétricas mas em função dos impactos ambientais decorrentes da construção de

grandes reservatórios, a implantação de PCHs tem sido estimulada para minimizar os impactos dessa fonte de energia.

Além disso, os projetos de implantação de uma PCH são menos complexos quando comparados com a implantação de uma grande usina hidrelétrica. Normalmente esses projetos demandam um prazo de construção menor e uma maior aceitação em relação as questões ambientais (EPE, 2006).

Li et al. (2015) apresentam alguns problemas relacionados a previsão da geração de PCHs, entre eles a falta de dados disponíveis. No estudo apresentam uma análise da correlação entre usinas hidrelétricas e PCHs possibilitando a utilização da previsão das usinas hidrelétricas na previsão da geração das PCHs.

2.2. Variáveis Exógenas e Dependência Espacial

Maçaira et al. (2018) apresentaram uma revisão da literatura mostrando os métodos mais utilizados nos últimos anos que incorporam variáveis explicativas ao modelo. Este artigo apresenta a regressão como o método mais utilizado, seguida de modelos que utilizam rede neurais e máquinas de vetores de suporte.

Alguns estudos recentes apresentam a aplicação de variáveis exógenas para mostrar o impacto das mudanças climáticas no mercado de energia elétrica principalmente em relação a geração das usinas hidrelétricas (Queiroz et al., 2019). Lima et al. (2014) utilizam variáveis climáticas para melhorar a previsão da vazão de reservatórios brasileiros, comparando os resultados de um modelo Bayseano Linear Dinâmico com o modelo Autoregressivo Periódico PAR(p) sem a utilização de variáveis climáticas obtendo melhores resultados.

Maçaira et al. (2017), por sua vez, apresentam um modelo “Causal PAR(p)”, onde variáveis como El Niño e manchas solares são incorporadas ao modelo PAR(p) como variáveis explicativas. Com o estudo, verificou-se que a utilização de variáveis exógenas pode melhorar os resultados de modelos de previsão.

Em função da incerteza dos regimes hidrológicos que afetam a vazão, principalmente no caso das PCHs cuja geração está diretamente associada a vazão,

a estocasticidade na geração precisa ser considerada nos modelos de previsão de geração hídrica. De acordo com Queiroz et al. (2019), quando a usina hidrelétrica possui um reservatório que regula a capacidade de geração esse impacto pode ser minimizado. Porém não é esse o caso quando estudamos usinas a fio d'água. Borges e Pinto (2008) apresentam um modelo que avalia a disponibilidade de geração das PCHs considerando a incerteza da vazão dos rios visto que esse tipo de usina não conta com a vazão garantida pelo reservatório.

Silva et al. (2016) apresentam um estudo sobre a influência do regime de chuvas na geração de energia elétrica para PCHs e Foley (2013) estuda a correlação da precipitação e da geração de energia elétrica em cinco hidrelétricas comprovando essa correlação. Silva et al. (2016) constataram que em períodos de seca a correlação é comprovada porém o inverso não é verdadeiro.

Nesta pesquisa buscou-se além de modelos que incorporam variáveis exógenas melhorando os resultados das previsões, modelos que considerem a dependência espacial entre as variáveis de interesse. Na literatura existem muitas aplicações recentes de modelos espaço-temporais na previsão da demanda turística (Yang e Zhang, 2019; Liu et al., 2019). Yang e Zhang (2019) apresentam a importância das matrizes de pesos espaciais que serão usadas no modelo espaço-temporal e na análise da autocorrelação espacial.

Modelos espaço-temporais são bastante estudados em econometria. Aplicado ao setor elétrico, Cabral et al. (2017) analisam a dependência espacial presente no consumo regional de eletricidade no Brasil e aplicam um modelo espaço-temporal (ARIMASp). Gomez et al. (2013) consideraram os efeitos espaciais para prever a elasticidade preço e renda da demanda por eletricidade apresentando um estudo de caso espanhol de demanda de eletricidade residencial. Neste estudo de caso, um modelo autoregressivo espacial (SAR) foi usado.

Partindo para uma aplicação envolvendo análise espacial e vazões, Lohmann et al. (2016) propõem um modelo espaço-temporal para previsão de vazões. Os autores consideram além da vazão da localidade que se pretende prever, a vazão dos vizinhos e aplicam em um estudo de caso com três bacias brasileiras. O modelo utilizado foi um modelo autoregressivo periódico espacial – SPAR(p).

3 Metodologia

Nesta seção são apresentados os modelos que serão utilizados na metodologia aplicada nesta pesquisa. A fim de verificar a eficácia da utilização da vazão como variável explicativa, dois modelos causais serão aplicados, a função de transferência de Box & Jenkins e a regressão dinâmica. Além desses modelos, o modelo SARIMA e o modelo de Amortecimento Exponencial Holt-Winters também serão utilizados servindo como *benchmarking* a fim de comparar os modelos causais com modelos muito utilizados na literatura. Para a escolha da melhor série de vazão vizinha, que quando incorporada ao modelo possibilita reduzir o erro de previsão, o estudo da autocorrelação espacial é apresentado com o intuito de aprofundar o entendimento sobre os vizinhos. Essa abordagem possibilitará uma aproximação com a realidade minimizando assim o erro de previsão.

3.1. Modelos Univariados

Na análise de séries temporais, os modelos univariados consideram as observações passadas (Y_t) da própria série histórica para explicar os valores futuros (\hat{Y}_{t+h}) (Souza e Camargo, 2004). No caso desta dissertação, para explicar a geração de energia elétrica futura de uma PCH, para os modelos univariados, somente os dados de geração passada serão utilizados na modelagem e previsão.

Os métodos univariados que serão apresentados neste estudo são: método de amortecimento exponencial sazonal de Holt-Winters e o modelo da família de Box & Jenkins, Autoregressivo Integrado de Médias Móveis Sazonal, visto que as séries estudadas apresentam sazonalidade.

3.1.1. Amortecimento Exponencial Sazonal de Holt-Winters

De acordo com Morettin e Toloí (2018), na análise de séries temporais, uma série temporal poderia ser representada como a soma das seguintes componentes:

- a) Nível – o valor médio atribuído a série temporal;
- b) Tendência – é o comportamento da série temporal. Uma série pode ter uma tendência de crescimento ou decrescimento;
- c) Sazonalidade – padrão que se repete durante um determinado período. Por exemplo como se comporta a venda de um determinado produto ao longo do ano; e
- d) Ruído – componente aleatória associada a uma distribuição de probabilidade.

No procedimento de Amortecimento Exponencial pesos são atribuídos as observações passadas e os pesos decrescem conforme a idade da observação. O método de amortecimento exponencial de Holt-Winters é uma evolução do método de Holt que apresenta uma equação de previsão quando os dados possuem uma tendência. Quando uma série temporal apresenta tendência, sazonalidade e nível, Holt-Winters é o método de amortecimento exponencial adequado, pois Winters (1960) acrescentou uma equação de previsão que considera o comportamento sazonal da série (Holt, 2004; Winters, 1960).

Se a série temporal possuir uma variação sazonal constante pode-se utilizar o modelo aditivo e se for observada uma variação sazonal, que aumenta ou diminui ao longo do tempo, o modelo adequado será o multiplicativo (Hyndman e Athanasopoulos, 2018). A Figura 3.1 ilustra o comportamento destes tipos de séries temporais e mostra o modelo mais adequado para cada caso.

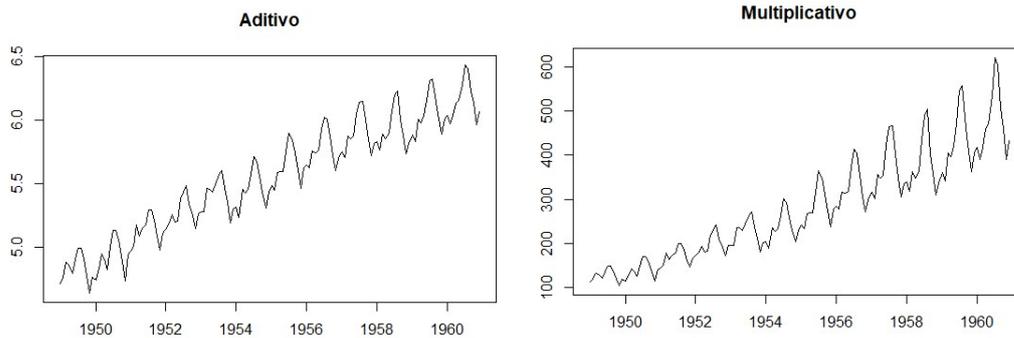


Figura 3.1 – Comportamento das Séries e o Modelo mais Adequado.

O método Holt-Winters possui três equações de suavização, uma para cada componente da série (nível, tendência e sazonalidade) (Hyndman e Athanasopoulos, 2018).

Modelo multiplicativo com tendência:

$$Y_t = \mu(t)\rho_t + \varepsilon_t \quad (3.1)$$

Onde

$$\mu(t) = (a_1 + a_2t), \quad (3.2)$$

a_1 - nível,

a_2 - tendência,

ρ_t - sazonalidade e

$t = 1, 2, \dots, S,$

S – fator sazonal.

Os estimadores são:

$\hat{a}_1(T - 1)$ para o nível no instante $T - 1$.

$\hat{a}_2(T - 1)$ para a tendência no instante $T - 1$.

$\hat{\rho}_{j+kS}(T - 1)$ no j -ésimo fator sazonal no instante $T - 1$.

Segundo Winters (1960), para calcular os estimadores as seguintes equações de suavização são utilizadas:

$$\begin{aligned}\hat{a}_1(T) &= \alpha \left[\frac{Y_t}{\hat{\rho}_{m(T)+kS}(T-1)} \right] + (1 - \alpha)(\hat{a}_1(T-1) + \hat{a}_2(T-1)) \\ \hat{a}_2(T) &= \beta(\hat{a}_1(T) - \hat{a}_1(T-1)) + (1 - \beta)\hat{a}_2(T-1) \\ \hat{\rho}_{m(T)+kS}^*(T) &= \gamma \left[\frac{Y_T}{\hat{a}_1(T)} \right] + (1 - \gamma)\hat{\rho}_{m(T)+kS}(T-1)\end{aligned}\quad (3.3)$$

Modelo aditivo com tendência:

$$Y_t = \mu(t) + \rho_t + \varepsilon_t \quad (3.4)$$

Onde:

$$\mu(t) = (a_1 + a_2t) \quad (3.5)$$

a_1 - nível,

a_2 - tendência,

ρ_t - sazonalidade e

$t = 1, 2, \dots, S$,

S – fator sazonal.

Os estimadores são:

$\hat{a}_1(T-1)$ para o nível no instante $T-1$.

$\hat{a}_2(T-1)$ para a tendência no instante $T-1$.

$\hat{\rho}_{j+kS}(T-1)$ no j -ésimo fator sazonal no instante $T-1$.

As equações de suavização para calcular os estimadores para o modelo aditivo.

$$\begin{aligned}
\hat{a}_1(T) &= \alpha[Y_t - \hat{\rho}_{m(T)+kS}(T-1)] + (1 \\
&\quad - \alpha)(\hat{a}_1(T-1) + \hat{a}_2(T-1)) \\
\hat{a}_2(T) &= \beta(\hat{a}_1(T) - \hat{a}_1(T-1)) + (1 - \beta)\hat{a}_2(T-1) \\
\hat{\rho}_{m(T)+kS}^*(T) &= \gamma[Y_T - \hat{a}_1(T)] + (1 - \gamma)\hat{\rho}_{m(T)+kS}(T-1)
\end{aligned} \tag{3.6}$$

3.1.2. Autoregressivo Integrado de Médias Móveis (ARIMA)

Uma condição para aplicação de modelos da família Box & Jenkins é a estacionariedade da série temporal, ou seja, média e variância devem ser constantes (Souza e Camargo, 2004).

O objetivo da análise de séries temporais é a partir de uma série temporal que possua uma dependência serial encontrar um modelo que descreva essa dependência e transforme o resíduo em um ruído branco, ou seja, uma série temporal estacionária e sem dependência serial (Morettin e Toloí, 2018).

Os modelos da família Box & Jenkins, chamados de modelos autoregressivos integrados de médias móveis (ARIMA), são modelos que consideram as componentes auto-regressiva (AR) da série temporal, médias móveis (MA) e de Integração (I) para quando a série não for estacionária. Neste caso, a aplicação do modelo ARIMA (p, d, q) será possível através da diferenciação da série temporal. Logo, apesar de não ser necessariamente estacionária, a série deve ser homogênea (Box et al, 2016).

O modelo ARIMA multiplicativo ou SARIMA $(p, d, q) \times (P, D, Q)_S$, para séries que possuem componentes sazonais, pode ser definido como (Box et al, 2016; Souza e Camargo, 2004):

$$\Phi(B^S) * \phi(B) * \nabla_S^D * \nabla^d Y_t = \theta(B^S) * \theta(B) * a_t \tag{3.7}$$

Onde:

Operador sazonal auto-regressivo $\Phi(B^S) = (1 - \phi_1 B^S - \dots - \phi_P B^{PS})$ e parâmetros $\phi_i, i = 1, 2, \dots, P$.

Operador não sazonal auto-regressivo (AR) é $\phi(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$ e parâmetros $\phi_i, i = 1, 2, \dots, p$.

Operador de diferença sazonal é $\nabla_S^D = (1 - B^S)^D$ de ordem D , sendo $D = 0, 1$ ou 2 na maior parte das vezes.

Operador de diferença não sazonal é $\nabla^d = (1 - B)^d$ de ordem d , é comum o $d = 0, 1$ ou 2 .

Operador sazonal de médias móveis (MA) é $\theta(B^S) = (1 - \theta_1 B^S - \dots - \theta_Q B^{QS})$ e os parâmetros $\theta_i, i = 1, 2, \dots, Q$.

Operador não sazonal de médias móveis (MA) é $\theta(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$ e parâmetros $\theta_i, i = 1, 2, \dots, q$.

Segundo Box et al. (2016), um modelo será considerado adequado se possuir poucos parâmetros, ou seja, um modelo que atenda ao princípio da parcimônia. Além disso, para realizar a escolha do melhor modelo, deve-se realizar ciclos iterativos. Para atender a esse critério, a metodologia propõe a realização de 4 passos: identificação do modelo, estimação de parâmetros, verificação e por último a previsão.

A etapa de identificação é o momento de encontrar qual a ordem dos polinômios AR e MA, ou seja, a ordem $p + Ps$ e $q + Qs$ que descrevem a série temporal estudada (Z_t) e para isto, as funções de autocorrelação e autocorrelação parcial serão utilizadas. Neste estágio também será definido o grau de diferenciação não sazonal, ordem " d " e o grau de diferenciação sazonal, ordem " D " (Souza e Camargo, 2004).

Na etapa de estimação dos parâmetros, os valores de Φ, ϕ que correspondem ao componente autoregressivo e Θ, θ que correspondem às médias móveis são estimados e, para isto, o método dos mínimos quadrados e da máxima verossimilhança poderão ser utilizados (Souza e Camargo, 2004).

Depois disso, na etapa de verificação o modelo será checado e alguns testes serão realizados para comprovar se está adequado. Neste momento, por exemplo, o teste de sobrefixação pode ser aplicado, testando vários modelos com parâmetros superiores ao modelo escolhido a fim de corrigir possíveis erros de identificação.

Também são realizados testes de análise de resíduos para verificar se o modelo está gerando um ruído branco como resíduo. Depois desta etapa se o modelo não estiver adequado é necessário voltar a etapa de identificação até encontrar o melhor modelo (Souza e Camargo, 2004).

A previsão é a etapa final e o objetivo da metodologia. Trata-se de calcular a previsão h passos à frente - $\hat{Y}_t(h)$, a partir das observações da série temporal estudada $(Y_t, Y_{t-1}, Y_{t-2}, \dots)$.

3.2. Modelos Causais

Como apresentado no item 3.1, os modelos univariados precisam apenas dos valores passados da série estudada em sua modelagem, porém quando outras séries temporais podem explicar o comportamento de uma variável de resposta é possível utilizar modelos causais, dentre eles, a função de transferência proposta por Box & Jenkins e a regressão dinâmica.

3.2.1. Função de Transferência de Box & Jenkins

De acordo com Box et al. (2016), quando o modelo possui uma relação dinâmica entre duas séries temporais, e a forma genérica pode ser representado pela Equação 3.8.

$$\begin{aligned} Y_t &= v_0 + v_1 X_{t-1} + v_2 X_{t-2} + \dots + \varepsilon_t \\ &= (v_0 + v_1 B + v_2 B^2 + \dots) X_t + \varepsilon_t \\ &= v(B) X_t + \varepsilon_t \end{aligned} \quad (3.8)$$

Onde Y_t é a variável de resposta, os pesos v_0, v_1, v_2, \dots representam as respostas à função impulso e ε_t um erro.

A função de transferência $v(B)$ pode apresentar valores de uma ou mais séries temporais além de observações passadas. A forma parcimoniosa, ou seja, que apresenta um menor número de parâmetros, está representada na equação 3.9 (Box et al., 2016).

$$Y_t = \frac{(\omega_0 + \omega_1 B + \omega_2 B^2 + \dots + \omega_s B^s)}{(1 - \delta_1 B - \delta_2 B^2 - \dots - \delta_r B^r)} X_{t-b} + \varepsilon_t \quad (3.9)$$

Importante notar que será necessário descobrir os valores de r, s e b , cujo passo-a-passo será visto mais a frente e também estimar os hiperparâmetros ω e δ (Box et al.,2016).

A formulação geral, Equação 3.10, segundo Box et al. (2016) também substitui o erro ε_t por η_t que pode ser um ruído branco ou mesmo um modelo ARMA (p, q).

$$Y_t = \frac{(\omega_0 + \omega_1 B + \omega_2 B^2 + \dots + \omega_s B^s)}{(1 - \delta_1 B - \delta_2 B^2 - \dots - \delta_r B^r)} X_{t-b} + \eta_t \quad (3.10)$$

Onde:

$$\eta_t = \frac{\theta(B)}{\phi(B)} a_t ;$$

a_t : ruído branco.

O modelo de função de transferência permite como entrada mais de uma variável exógena e a Equação 3.11 representa este caso:

$$Y_t = \sum \frac{\omega_i(B)}{\delta_i(B)} X_{t-b_i} + \eta_t \quad (3.11)$$

Onde:

$$\eta_t = \frac{\theta(B)}{\phi(B)} a_t ;$$

$i = 1, 2, \dots, k$.

a_t : ruído branco.

Para que o modelo seja definido é necessário seguir os passos da metodologia de Box e Jenkins, ou seja, a mesma sequência proposta para o modelo univariado composta por identificação, estimação, verificação e previsão (Box et al.,2016).

Assim como a função de autocorrelação é usada para identificação dos modelos univariados, a função de correlação cruzada (Equação 3.12) entre a

variável de entrada X_t e a variável de saída Y_t é usada na identificação da função de transferência (Box et al.,2016).

$$\rho_{xy}(k) = \frac{\gamma_{YX}(k)}{\sigma_X \sigma_Y} \quad (3.12)$$

Onde:

$\rho_{xy}(k)$ é a função de correlação cruzada entre Y_t e X_t no lag k .

$\gamma_{YX}(k)$ é a covariância cruzada Y_t e X_t no lag k , em geral a covariância cruzada Y_t e X_t no lag k é diferente da covariância cruzada X_t e Y_t no lag k $\gamma_{YX}(k) \neq \gamma_{XY}(k)$.

σ_X é o desvio padrão de X_t .

σ_Y é o desvio padrão de Y_t .

Porém Box e Jenkins verificaram que esse método não se mostrou eficiente pois pode existir uma estrutura de dependência dentro das séries. Por isso, para evitar a autocorrelação de X_t e Y_t um pré-branqueamento nas séries temporais se faz necessário a fim de eliminar essa estrutura de dependência presente nas mesmas (Box et al.,2016).

Antes de identificar os valores de r, s e b , deve-se garantir que as séries sejam estacionárias, para isso é necessário diferenciar as séries eliminando a não estacionariedade sazonal e não sazonal das séries temporais.

A identificação dos valores de r, s e b , ocorrerá nas séries pré-branqueadas, ou seja, serão identificadas as respostas impulsiais estimadas. O pré-branqueamento, na verdade, é a determinação de um modelo ARIMA que transformará a série X_t em uma série de ruídos α_t . Posteriormente o modelo encontrado será aplicado a série de saída Y_t que fornecerá a série de resíduos β_t . É nesse momento que serão estimadas as respostas impulsiais a partir da correlação cruzada entre os resíduos α_t e β_t (Ferreira, 2018).

O valor de b é tal que $v_0 = v_1 = v_{b-1} = 0$, ou seja, o primeiro lag significativo da função de correlação cruzada. Para o valor de s será a quantidade de lags crescentes depois do b , ou seja, as demais defasagens X que serão consideradas no modelo. Em relação ao r , se ocorrer um corte brusco a partir de $s + b + 1$, o r será igual a zero ($r = 0$). Se ocorrer um decaimento exponencial a

partir de $s + b + 1$, o r será igual a um ($r = 1$). Se a partir de $s + b + 1$ ocorrer uma senóide amortecida, o r será igual a dois ($r = 2$) (Box et al.,2016).

Para identificar o η_t a partir da equação $\hat{\eta}_t = Y_t - \hat{v}_0 X_t - \hat{v}_1 X_{t-1} - \hat{v}_2 X_{t-2} - \dots$, a metodologia de Box & Jenkins é aplicada para encontrar o modelo $ARMA(p, q)$ (Box et al.,2016).

A estimação dos hiperparâmetros $\omega, \delta, \theta, \phi$ será calculada a partir do método de estimação de mínimos quadrados ordinários. Para verificar se o modelo está adequado, assim como no modelo univariado, executa-se os testes de análise de resíduos, além de verificar se o modelo gerou um ruído branco como resíduo, também é possível verificar a correlação cruzada entre o resíduo e a série α_t (Ferreira, 2018).

3.2.2. Regressão Dinâmica

Para que um modelo de regressão dinâmica seja utilizado verifica-se a existência de uma defasagem (dinâmica) temporal da variável dependente nas variáveis explicativas do modelo (Gujarati, 2000). A Equação 3.13 apresenta um exemplo de modelo dinâmico:

$$Y_t = \alpha + \theta_0 X_t + \theta_1 X_{t-1} + \dots + \theta_q X_{t-q} + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + a_t \quad (3.13)$$

Nesta equação além da variável explicativa X_t e suas defasagens X_{t-q} teremos a defasagem de Y_{t-p} (valor passado da variável de resposta) que poderá explicar a variável dependente, α será o intercepto, θ e ϕ os parâmetros e a_t um ruído branco. Esse modelo também é chamado de modelo auto-regressivo com defasagem distribuída (Gujarati, 2000; Ferreira, 2018).

Não existe na literatura uma forma definida para determinar as defasagens de X_t e Y_t (p e q), por isso a melhor estratégia é determiná-las empiricamente. Uma primeira abordagem é a definição de um número alto de defasagens e determinar através do teste de significância se o coeficiente da maior defasagem é significativo

e deve continuar no modelo, caso contrário estima-se o modelo com uma defasagem a menos até que o coeficiente da maior defasagem seja significativo. É possível também realizar a verificação da significância começando por uma defasagem e acrescentar uma nova defasagem até que o maior coeficiente de defasagem deixe de ser significativo (Ferreira, 2018).

Uma outra abordagem utiliza critérios de informação para escolha do modelo, neste caso, vários modelos com valores altos e baixos de defasagens são testados e seleciona-se o melhor modelo comparando os resultados. Por exemplo, o modelo que possuir o menor Critério Akaike (AIC) ou menor critério de informação bayesiano (BIC) será escolhido (Ferreira, 2018). Para mais informações sobre os critérios de informação consultar Akaike, 1974 e Vrieze, 2012.

3.3. Autocorrelação Espacial

De uma forma geral, a autocorrelação espacial na verdade indica que uma variável em uma região i pode estar correlacionada com a mesma variável em regiões vizinhas j (Almeida, 2012).

Para realizar a análise da autocorrelação espacial levando em consideração a vazão das usinas é necessário definir um arranjo espacial através da construção de uma matriz de pesos relacionada à localização de cada usina analisada (Almeida, 2012).

Segundo Cliff et al. (1975) existem algumas formas de construção da matriz de pesos, baseada na distância geográfica (d_{ij}) ou baseada na contiguidade. A matriz de pesos é uma matriz $n \times n$.

Matriz de Pesos Baseada na Distância Geográfica

Este método é mais utilizado quando a distribuição espacial é feita através de pontos, porém também pode ser usada para calcular a distância entre os centróides de regiões. Calcula-se a distância entre o ponto de interesse i e o ponto vizinho j e posteriormente define-se a função da distância que será utilizada na preparação da matriz.

Duas funções são comuns neste caso: (i) onde os elementos da matriz serão representados pela distância, $w_{ij} = d_{ij}$, (ii) ou como o inverso da distância, $w_{ij} = 1/d_{ij}^2$ (Downey, 2006).

Uma matriz muito utilizada baseia-se nos k vizinhos mais próximos, ou seja, define-se através de uma matriz binária quais são os vizinhos que atendem a um critério relacionado a distância em quilômetros entre o vizinho j e o ponto de interesse i (Almeida, 2012).

$$w_{ij}(k) = \begin{cases} 1 & \text{se } d_{ij} \leq d_i(k) \\ 0 & \text{se } d_{ij} > d_i(k) \end{cases} \quad (3.14)$$

Onde:

$d_i(k)$ é a menor distância até a localidade i para que ela possua k vizinhos

d_{ij} é a distância entre i e j .

$w_{ij}(k)$ é o elemento da matriz.

Dessa forma todas as regiões possuirão o mesmo número de vizinhos e as “ilhas”, quando um ponto não possui vizinhos, não existirão (Almeida, 2012).

Outra forma de definir os pesos da matriz binária utiliza uma distância fixa \bar{d} . Neste caso a quantidade de vizinhos vai variar, então é importante definir um valor de \bar{d} que evite o surgimento de “ilhas”.

$$w_{ij}(d) = \begin{cases} 1 & \text{se } d_{ij} \leq \bar{d} \\ 0 & \text{se } d_{ij} > \bar{d} \end{cases} \quad (3.15)$$

Nos dois casos apresentados acima é possível utilizar uma matriz baseada na distância inversa no lugar da matriz binária. Dessa forma dependendo da distância um determinado vizinho j pode influenciar mais o ponto de interesse i se estiver mais próximo. A Equação 3.16 apresenta os pesos para a distância $d_i(k)$ e a Equação 3.17 para a distância fixa \bar{d} .

$$w_{ij}(k) = \begin{cases} 1/d_{ij} & \text{se } d_{ij} \leq d_i(k) \\ 0 & \text{se } d_{ij} > d_i(k) \end{cases} \quad (3.16)$$

$$w_{ij}(d) = \begin{cases} 1/d_{ij} & \text{se } d_{ij} \leq \bar{d} \\ 0 & \text{se } d_{ij} > \bar{d} \end{cases} \quad (3.17)$$

A diagonal da matriz que apresenta os pesos w_{ii} possui peso zero visto que um ponto não pode ser vizinho dele mesmo (Anselin, 1995).

Matriz de Pesos Baseada na Contiguidade

Quando se está analisando cidades ou regiões que estão definidas como áreas, a escolha dos pesos levará em consideração se a região vizinha j faz fronteira com a região de interesse i pelos vértices, pelas bordas ou pelos vértices e bordas. Quando o vizinho for escolhido quando faz fronteira pelo vértice, diz-se que a escolha do peso seguirá o caso “bispo”, quando a fronteira se dá pela borda, trata-se do caso “torre” e se a escolha for pelas bordas e vértices, esse caso é chamado de “rainha”. Essa nomenclatura foi inspirada nos movimentos realizados pelas peças do xadrez e pode ser melhor visualizada na Figura 3.2 (Almeida, 2012).

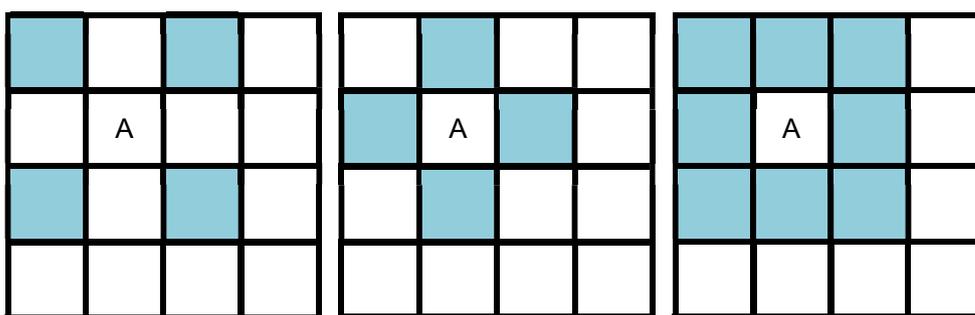


Figura 3.2 – Caso “Bispo”, “Torre” e “Rainha”.

A matriz nesse caso, será composta de zeros (0) e uns (1), ou seja, com valores zero quando os vizinhos não forem contíguos ou um quando forem (Almeida, 2012).

$$w_{ij} = \begin{cases} 1 & \text{se } i \text{ e } j \text{ são contíguos} \\ 0 & \text{se } i \text{ e } j \text{ não são contíguos} \end{cases} \quad (3.18)$$

Antes de utilizar os dados de vazão nos modelos causais é necessário fazer uma análise exploratória de dados espaciais, ou seja, analisar como os dados estão distribuídos espacialmente, se existe um padrão entre as localidades ou mesmo se existem localidades que destoam da área analisada. Para realizar este estudo serão utilizadas as estatísticas *I de Moran (Moran I)* e *I de Moran Local (Local Moran I)* (Almeida, 2012; Anselin, 1995). A primeira permite identificar se a variável de interesse possui um comportamento parecido ou não globalmente, ou seja, se os dados se comportam da mesma forma em toda área estudada e a segunda estatística permite identificar grupos com comportamento similares, permitindo a criação de *clusters*. Por exemplo, se para uma variável de interesse os valores altos estiverem localizados próximos de vizinhos que também possuem valores altos e os valores baixos dessa mesma variável estiverem próximos a valores baixos em localidades vizinhas, a estatística *I de Moran* será positiva (Almeida, 2012).

Quando se estuda uma região não se pode contar apenas com a observação visual da mesma, neste caso deve-se utilizar alguma estatística que permita inferir sobre a distribuição espacial de forma global. Na Equação 3.19 pode-se visualizar a estatística *I de Moran*:

$$I = \frac{n \sum_i \sum_j w_{ij} X_i X_j}{S_0 \sum_{i=1}^n X_i^2} \quad (3.19)$$

Onde:

i é a localidade de interesse.

j é a localidade vizinha.

X_i é a variável de interesse da localidade de interesse.

X_j é a variável de interesse da localidade vizinha

S_0 é a soma de todos os elementos da matriz W .

w_{ij} é o peso espacial entre i e j .

Segundo Anselin (1995), a estatística I de Moran Local é indicada para verificar a similaridade de localidades próximas, um valor significativo e positivo indica que existe similaridade entre a variável de interesse e os vizinhos indicando que existem *clusters* espaciais. A estatística I de Moran Local pode ser escrita como mostrada na Equação 3.20.

$$I_i = \frac{(x_i - \bar{x})}{m_0} \sum_{j=1}^n w_{ij}(x_j - \bar{x}) \quad (3.20)$$

$$m_0 = \sum_{i=1}^n (x_i - \bar{x})^2 / (n) \quad (3.21)$$

Onde:

i é a localidade de interesse.

j é a localidade vizinha.

x_i é a variável de interesse.

$\bar{x} = \sum_i x_i / n$ é a média de x .

w_{ij} é o peso espacial entre i e j .

Na Figura 3.3 a metodologia que será aplicada nesta pesquisa é apresentada em 3 etapas. A primeira etapa concentra-se na apresentação de modelos univariados que servirão como *benchmarking* e comparação dos resultados. A segunda etapa diz respeito ao tratamento dos dados de vazão e a escolha das séries temporais vizinhas. Na terceira etapa o ajuste e previsão, utilizando modelos causais, apresentarão os resultados atendendo ao objetivo proposto nesta dissertação. Na quarta etapa será testado um cenário de vazão para aplicação do modelo que apresentou melhor resultado.

O software R foi utilizado para realização dos testes, ajustes e previsões dos modelos utilizados nesta dissertação (R Team, 2015; Ferreira, 2018).

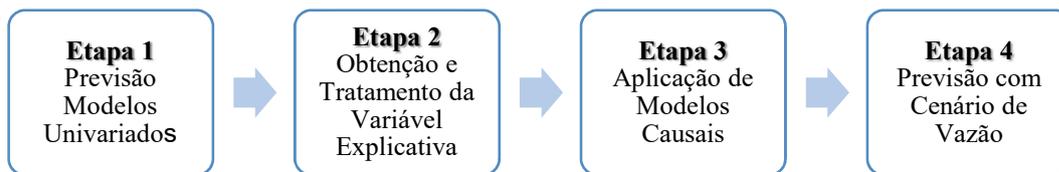


Figura 3.3 – Metodologia Aplicada.

O detalhe de cada etapa está descrito a seguir.

Etapa 1 - Previsão dos Modelos Univariados:

1. Escolher as séries de geração de energia elétrica para estudo.
2. Fazer análise descritiva das séries.
3. Dividir as séries em treinamento, validação e teste.
4. Aplicar modelo de amortecimento exponencial e ARIMA e fazer previsão.

Etapa 2 – Obtenção e Tratamento da Variável Explicativa:

5. Pesquisar dados de vazão para as séries de geração escolhidas no sistema de telemetria do ANA.
6. Pesquisar séries de vazão localizadas em regiões vizinhas disponíveis na ONS.
7. Realizar estudos de correlação das séries de geração com vazão de regiões vizinhas.
8. Realizar estudos de autocorrelação espacial a fim de identificar os melhores vizinhos.

Etapa 3 – Aplicação de Modelos Causais:

9. Aplicar modelos de função de transferência e regressão dinâmica usando a vazão da própria usina, como variável explicativa.
10. Aplicar modelos de função de transferência e regressão dinâmica usando com variável explicativa, a melhor série de vazão correlacionada com a série de geração da PCH.

11. Aplicar modelos de regressão dinâmica e função de transferência usando os melhores vizinhos, identificados na etapa anterior, como variável explicativa.
12. Comparar Resultados.

Etapa 4 – Previsão com Cenário de Vazão:

13. Criar cenário para vazão própria pela média.
14. Criar cenário de vazões para vizinhos espaciais pela média.
15. Realizar previsão para 2018 com os melhores modelos identificados na etapa anterior.
16. Comparar e analisar resultados.

4 Análise Descritiva dos Dados

Neste capítulo será apresentada a análise descritiva dos dados de geração de energia elétrica das PCHs disponibilizado pela Energisa Minas Gerais que possui a concessão das usinas. Além disso foram analisados os dados de Vazão das Usinas Hidrelétricas Vizinhas e Vazão Própria considerados nesta pesquisa.

4.1. Dados de Geração de Energia Elétrica das PCHs

Para aplicar a metodologia proposta 5 PCHs foram escolhidas. O critério utilizado para seleção das usinas que iriam participar do estudo foi a disponibilidade de dados de vazão própria no site da Agência Nacional de Águas, a capacidade da usina para verificação da aplicabilidade em usinas de tamanhos distintos e o número de observações dos dados de geração.

A Tabela 4.1 apresenta as usinas que fazem parte desta pesquisa. As usinas de Anna Maria e Guary estão localizadas no Rio Pinho; Ivan Botelho 1, Zé Tunin e Ivan Botelho 2 estão localizadas no Rio Pomba. Esses dois rios fazem parte da sub-bacia do Paraíba do Sul, na bacia Atlântico Sudeste.

Nesta amostra apenas a usina Anna Maria é classificada como CGH (< 5MW), as demais usinas são PCHs (> 5MW). Para facilitar a leitura, tanto a CGH como as PCHs serão chamadas a partir de agora como PCHs.

Tabela 4.1 – PCHs Consideradas na Pesquisa.

PCH	Rio	Data de entrada em Operação	Capacidade (MW)	Número de observações
Anna Maria	Rio Pinho	out/00	1,5	219
Guary	Rio Pinho	nov/10	5,3	98
Ivan Botelho 1	Rio Pomba	mai/03	24,4	188
Ivan Botelho 2	Rio Pomba	out/03	24,4	183
Zé Tunin	Rio Pomba	dez/12	8,0	73

A Figura 4.1 apresenta as séries de geração de energia elétrica que serão consideradas na pesquisa. Observando os gráficos é possível perceber que as séries possuem um comportamento sazonal e uma aparente estacionariedade.

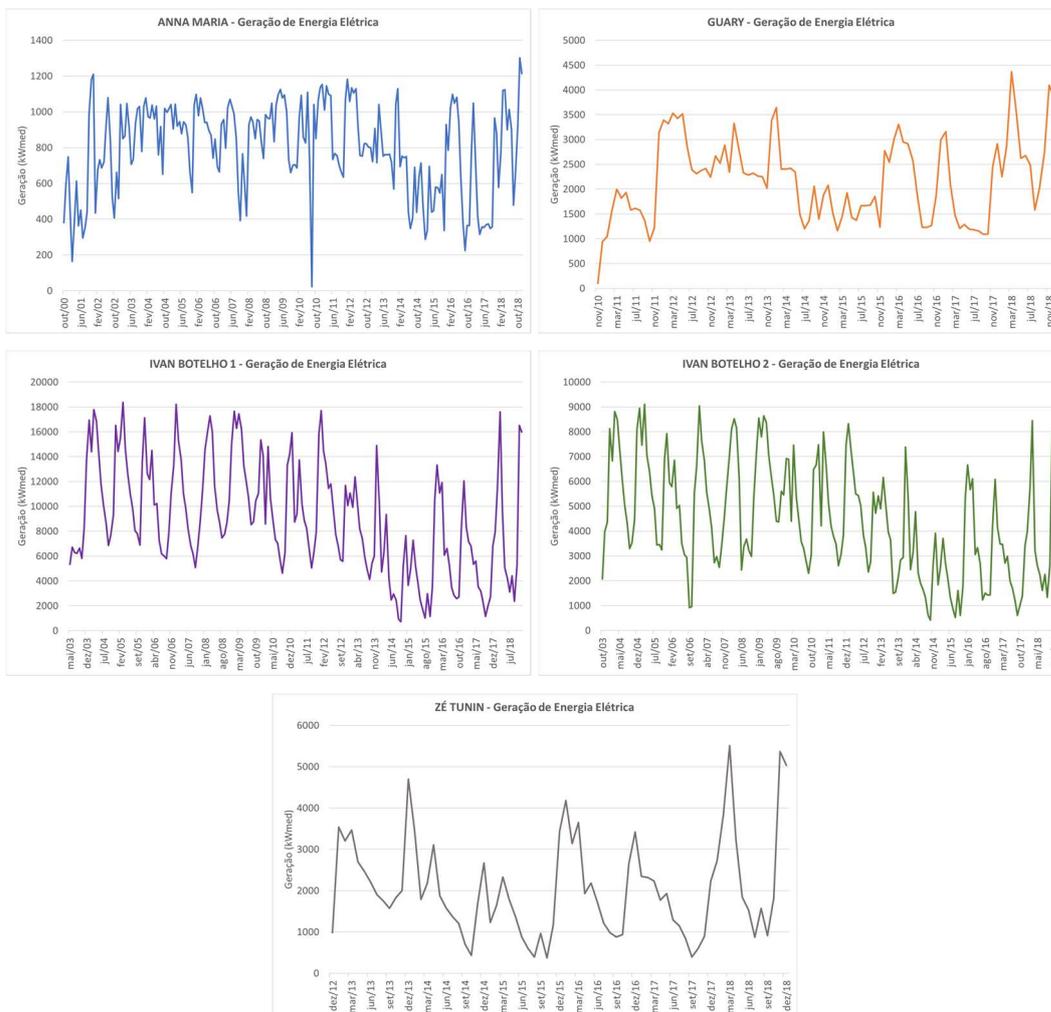


Figura 4.1 – Séries Temporais de Geração de Energia Elétrica das PCHs.

Os dados históricos da geração estão apresentados em kWmed e estão disponíveis em base mensal a partir da data de entrada em operação de cada usina apresentada na Tabela 4.1 até dezembro de 2018.

Tabela 4.2 – Coordenadas Geográficas das PCHs.

Usina	Latitude	Longitude
Anna Maria	-21.481551	-43.462153
Guary	-21.488948	-43.455189
Ivan Botelho 1	-21.374056	-42.993778
Zé Tunin	-21.314941	-42.944981
Ivan Botelho 2	-21.342164	-42.969641

Na Tabela 4.2 também são apresentadas as coordenadas geográficas das PCHs que serão utilizadas para realizar a análise espacial. Sua localização no mapa do Brasil pode ser visualizada na Figura 4.2. As PCHs estão localizadas no estado de Minas Gerais próximas a cidade de Ubá e Santos Dumont.



(a)



(b)

Figura 4.2 – Localização das PCHs do Rio Pinho (a) e Rio Pomba (b) - Visão Macro e Detalhe.

Fonte: Google, 2019.

Para finalizar, a Tabela 4.3 apresenta as estatísticas descritivas de geração de energia elétrica das PCHs estudadas. Em relação ao coeficiente de variação Anna Maria possui o menor coeficiente (32%) e Zé Tunin o maior (54%). Essa medida indica a variabilidade dos dados em relação à média. Em relação a assimetria das séries, ou seja, como os dados se comportam, a série de geração de Anna Maria possui um valor negativo para assimetria o que indica que a série possui uma quantidade maior de dados com valores mais altos. Outras medidas também ajudam a perceber esse comportamento como a mediana.

Tabela 4.3 – Estatísticas Descritivas das Séries de Geração das PCHs.

Estatísticas	Anna Maria	Guary	Ivan Botelho 1	Ivan Botelho 2	Zé Tunin
Média	790,31	2075,93	9123,28	4489,34	1892,79
Mediana	823,45	2036,97	8623,53	4215,05	1787,51
1º quartil	656,25	1405,25	5994,64	2772,27	1147,46
3º quartil	1004,95	2568,65	12210,21	6169,59	2345,87
Desvio Padrão	251,56	776,39	4410,86	2248,93	1013,06
Coefficiente de Variação	32%	37%	48%	50%	54%
Assimetria	-0,57	0,17	0,21	0,20	0,61
Curtose	2,47	2,25	2,20	2,08	2,82
Mínimo	22,75	100,99	725,06	414,09	373,55
Máximo	1210,10	3647,76	18365,27	9111,45	4703,89

4.2.

Dados de Vazão Usinas Hidrelétricas Vizinhas

Para esta dissertação optou-se pela obtenção dos dados de vazão através do site do Operador Nacional do Sistema Elétrico - ONS na página relacionada ao histórico da operação (ONS, 2019). Nessa página apesar das séries não possuírem dados desde 1931, como é possível acessar dentro do *deck* de preços do NEWAVE³, a maioria das séries históricas apresenta observações desde janeiro de 1999.

As séries de geração das PCHs, conforme apresentado na Tabela 4.1, possuem datas de entrada em operação diferentes. A usina Anna Maria é a mais antiga e entrou em operação em outubro de 2000. Para aplicar os modelos que possuem variáveis explicativas, os dados de vazão obtidos foram considerados suficientes.

As usinas hidrelétricas vizinhas que possuem histórico de vazão no site da ONS estão disponíveis em base mensal e são medidas em m³/s. A Figura 4.3 mostra as séries de vazões. Todas as séries de vazão apresentam uma componente sazonal e estacionariedade, com exceção da série de Nilo Peçanha que possui uma tendência decrescente e a série de Pereira Passos que varia em nível.

³ O *deck* de preços do NEWAVE é composto pelos arquivos necessários para que os programas NEWAVE e DECOMP possam ser processados. O modelo NEWAVE é o “Modelo de Planejamento da Operação de Sistemas Hidrotérmicos Interligados de Longo e Médio Prazo” enquanto o modelo DECOMP é o “Modelo de Planejamento da Operação de Sistemas Hidrotérmicos Interligados de Curto Prazo” (CEPEL, 2018).

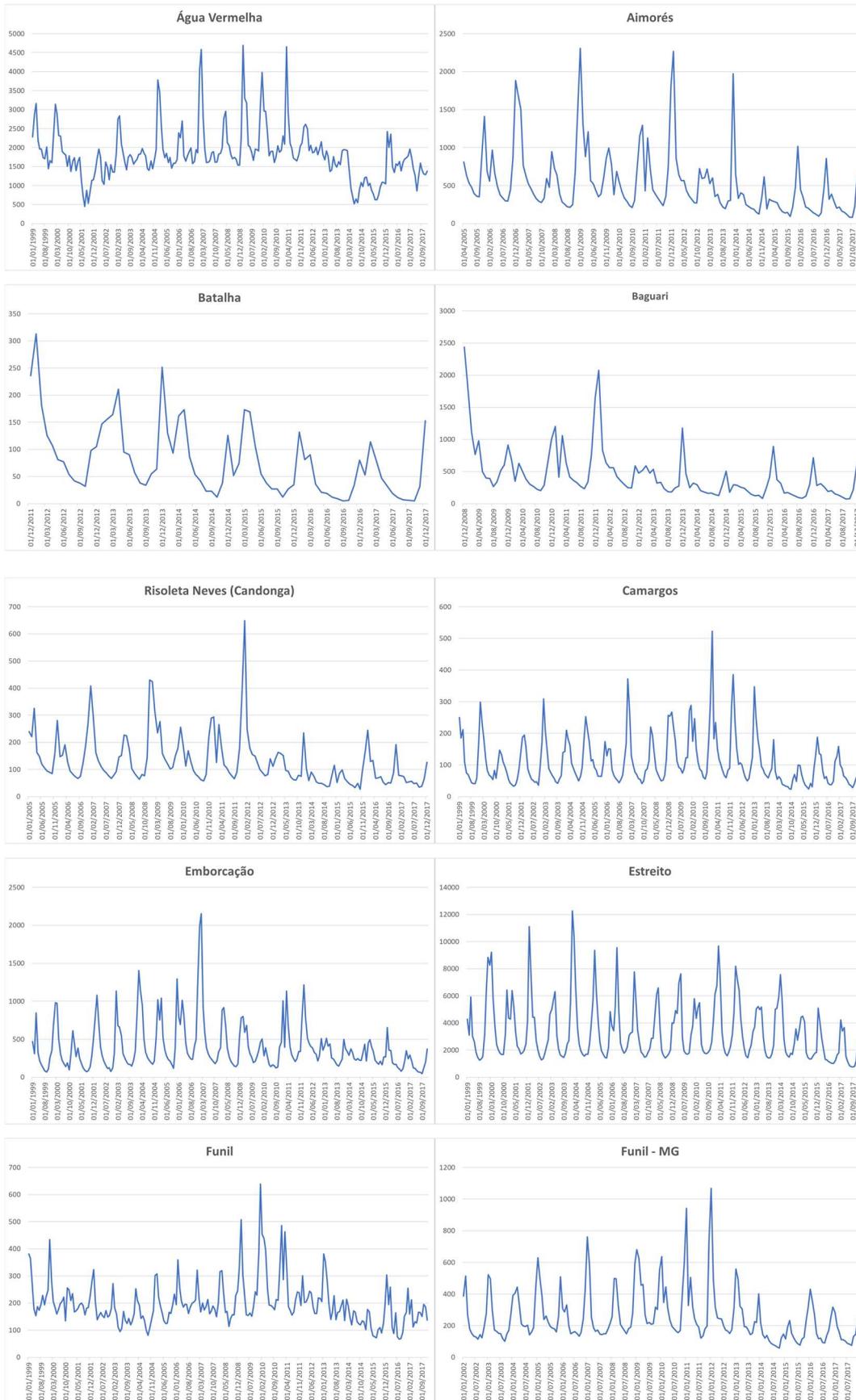


Figura 4.3 – Vazões das Usinas Hidrelétricas Vizinhas.

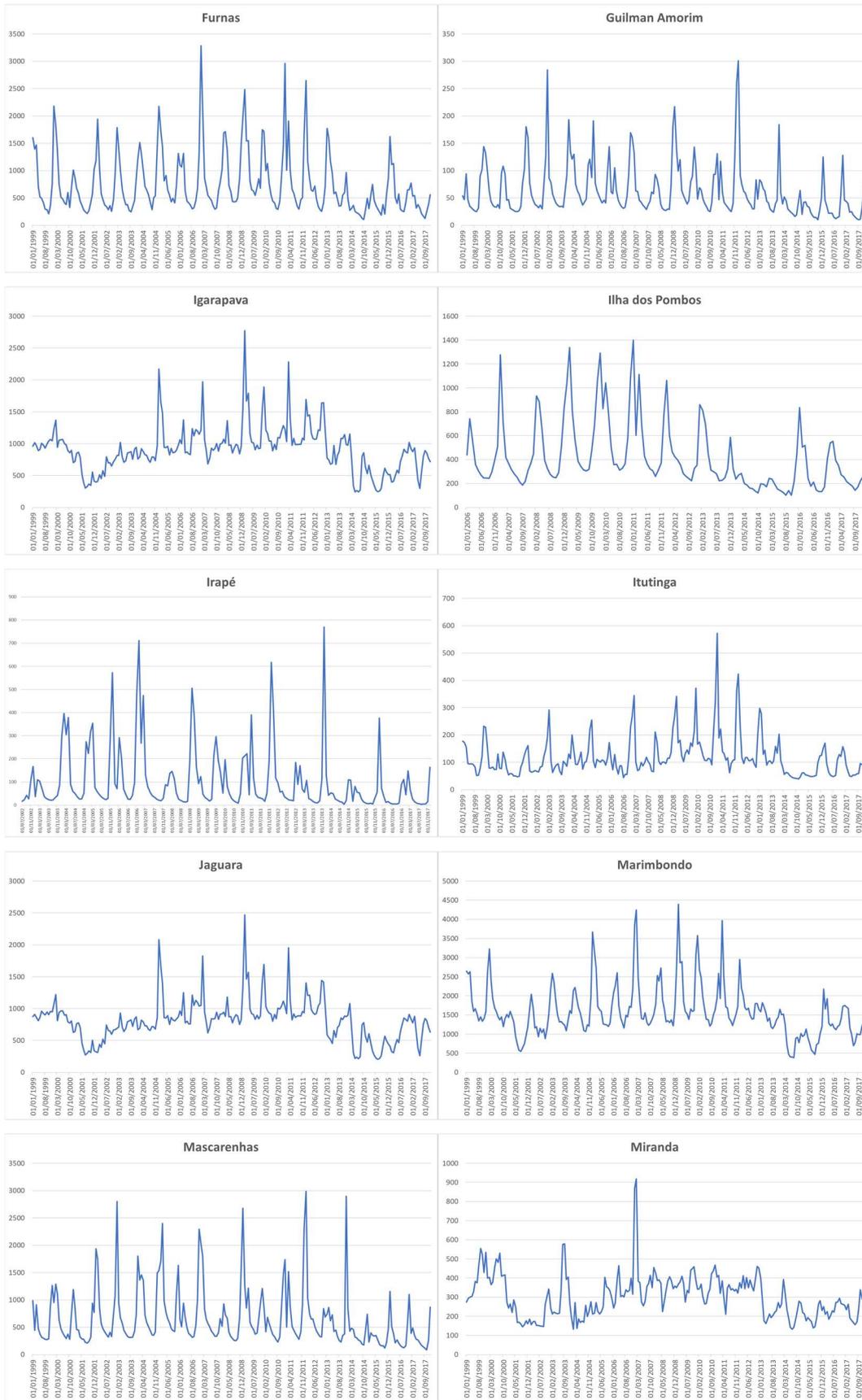


Figura 4.3 – Vazões das Usinas Hidrelétricas Vizinhas (continuação).

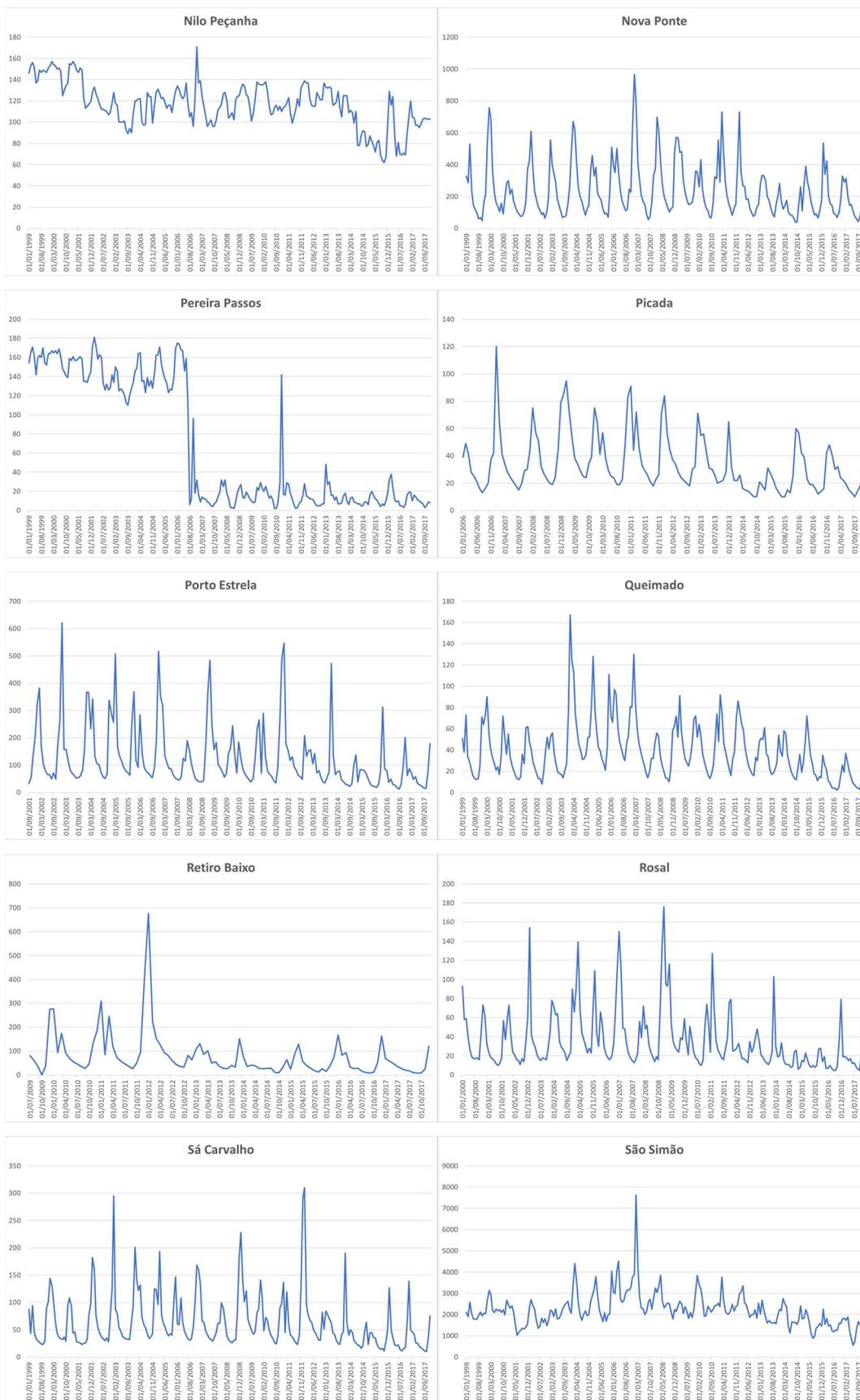


Figura 4.3 – Vazões das Usinas Hidrelétricas Vizinhas (continuação).

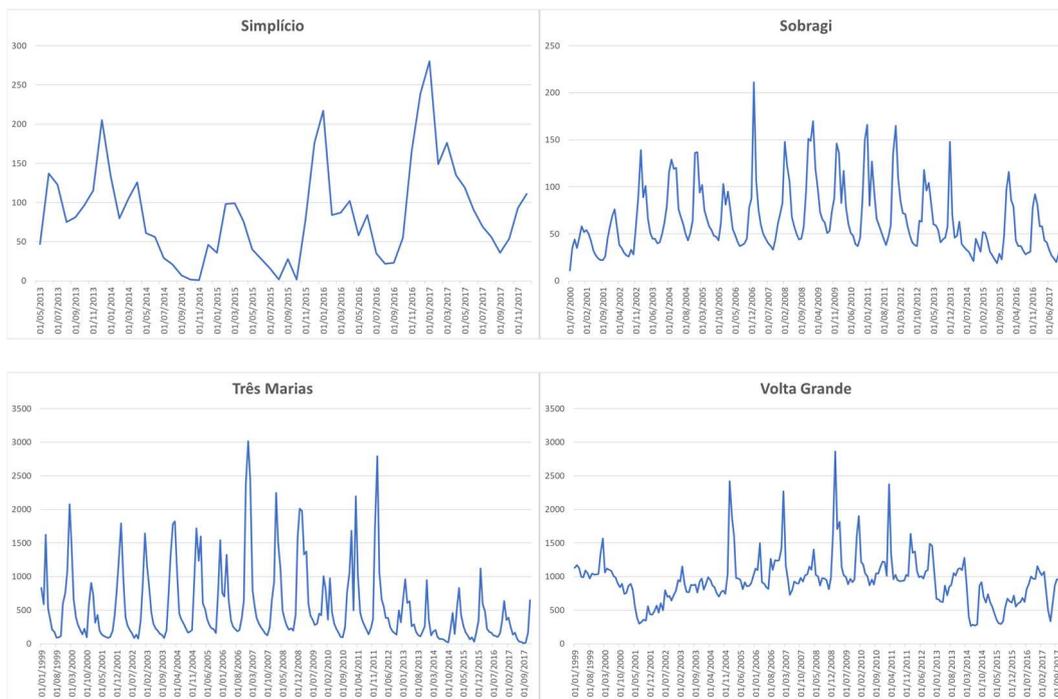


Figura 4.3 – Vazões das Usinas Hidrelétricas Vizinhas (continuação).

De acordo com a Primeira Lei de Tobler, “*everything is related to everything else, but near things are more related than distant things.*” (Tobler, 1970, p. 236), ou seja, os pontos que se localizam mais próximos são mais correlacionadas espacialmente do que os pontos mais distantes, por essa razão apenas as usinas hidrelétricas que estão geograficamente posicionadas nos estados de Minas Gerais, Rio de Janeiro e Espírito Santo, nas bacias hidrográficas Paraná, Atlântico Sudeste e São Francisco foram consideradas (Figura 4.4).

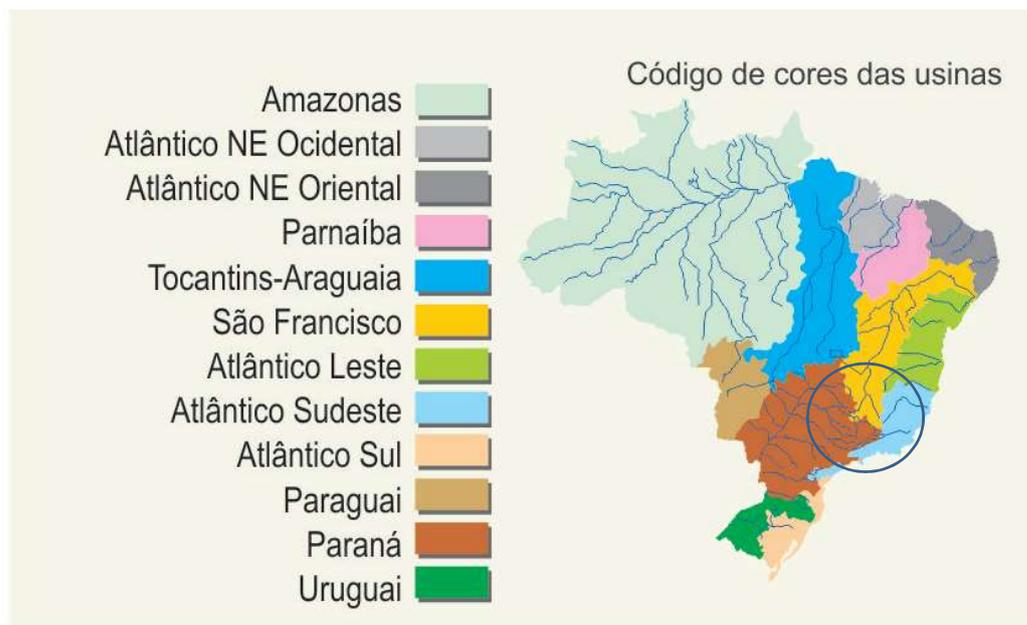


Figura 4.4 – Bacias Hidrográficas.
Fonte: ONS, 2019b.

Na Tabela 4.4 são apresentadas as coordenadas geográficas das usinas que são necessárias para realizar a análise espacial.

Tabela 4.4 – Coordenadas das Usinas Hidrelétricas Vizinhas.

Usina	Latitude	Longitude
Água Vermelha	-19,8653	-50,3467
Aimorés	-19,4564	-41,0937
Baguari	-19,018	-42,1228
Batalha	-17,3542	-47,5014
Camargos	-21,3247	-44,6164
Candongá (Risoleta Neves)	-20,2071	-42,8543
Emborcação	-18,4454	-47,9895
Funil – MG	-21,1434	-45,0362
Funil – RJ	-22,5288	-44,568
Furnas	-20,6698	-46,3178
Guilman Amorim	-19,6779	-42,9175
Igarapava	-19,9918	-47,758
Ilha dos Pombos	-21,8435	-42,5794
Irapé	-16,7375	-42,575
Itutinga	-21,2922	-44,6251
Jaguara	-20,0238	-47,4337
Luiz Carlos Barreto de Carvalho (Estreito)	-20,1531	-47,2794
Marimondo	-20,3039	-49,1978
Mascarenhas de Moraes	-20,2862	-47,0645
Miranda	-18,909	-48,0423
Nilo Peçanha	-22,6833	-43,8739

Nova Ponte	-19,1286	-47,6984
Pereira Passos	-22,6864	-43,8234
Picada	-21,9078	-43,5367
Porto Estrela	-19,1169	-42,6636
Queimado	-16,2104	-47,3229
Retiro Baixo	-18,8779	-44,7786
Rosal	-20,9538	-41,7167
Sá Carvalho	-19,6378	-42,822
São Simão	-19,0193	-50,4998
Simplício	-21,9223	-42,7906
Sobragi	-21,9813	-43,3502
Três Marias	-18,2142	-45,2628
Volta Grande	-20,0323	-48,2215

Na Figura 4.5 é possível ter uma ideia da distribuição geográfica das UHEs vizinhas (pontos azuis) em relação as PCHs estudadas (pontos verdes). O Brasil é uma país de grandes dimensões, mesmo restringindo a área considerada para Minas Gerais, Espírito Santo e Rio de Janeiro, algumas usinas estão muito distantes. Por exemplo São Simão está a mais 800 Km de distância de Zé Tunin.

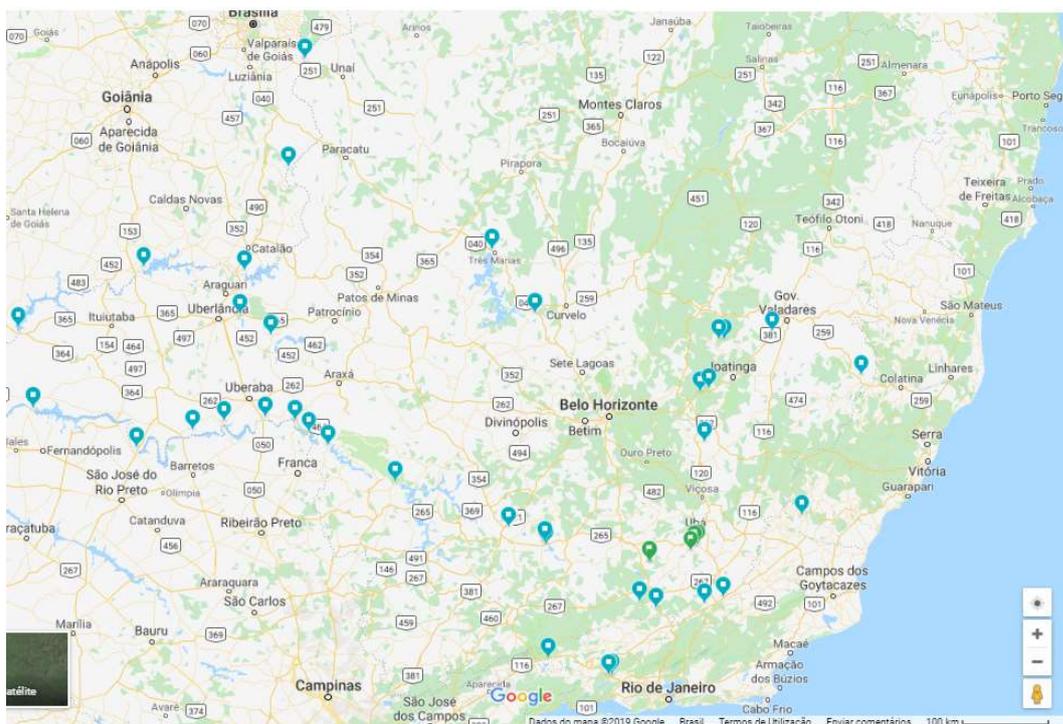


Figura 4.5 – Distribuição Geográfica das Usinas Hidrelétricas Vizinhas.

Em relação aos dados de Vazão das UHEs vizinhas, Nilo Peçanha é a série que possui a menor variabilidade (18%) em relação à média, os dados nesse caso

são mais homogêneos, como pode ser observado na Figura 4.3. A maior variabilidade ocorre com a série de Irapé (132%). Observando a assimetria das séries apenas a Vazão da UHE Nilo Peçanha possui uma maior concentração de valores maiores, as demais possuem assimetria positiva, indicando uma maior concentração de valores menores (Tabela 4.5).

Tabela 4.5 – Estatística Descritiva das Séries de Vazão das UHEs.

UHE	Média	Mediana	1º quartil	3º quartil	Desvio padrão	CV	Assimetria	Curtose	Mín	Máx
Água Vermelha	1817,81	1747,50	1484,75	1982,00	678,00	37%	1,39	6,84	448	4693
Aimorés	529,23	387,00	274,00	651,00	420,81	80%	2,06	7,62	83	2306
Baguari	443,61	314,00	203,00	534	400,22	90%	2,60	10,95	74	2438
Batalha	1817,81	1747,50	32,00	107,00	678,00	37%	1,39	4,38	448	4693
Camargo	110,19	85,50	57,00	143,75	76,50	69%	1,80	7,26	23	523
Candonga	130,08	99,50	73,75	159,00	91,45	70%	2,21	9,99	27	649
Emborcação	391,22	291,00	192,75	463,25	314,95	81%	2,24	9,82	48	2153
Funil	198,13	183,00	152,00	222,00	82,30	42%	1,82	7,99	67	639
Funil-MG	254,66	195,00	148,00	311,75	164,49	65%	1,86	7,21	58	1068
Furnas	741,81	550,50	367,50	950,75	544,08	73%	1,72	6,29	104	3282
Guilman Amorim	62,29	46,00	32,75	78,00	47,43	76%	2,12	8,65	10	301
Igarapava	909,80	908,00	730,50	1044,00	357,45	39%	1,19	7,21	243	2774
Ilha Pombos	418,91	321,00	238,00	507,00	281,16	67%	1,56	4,91	102	1399
Irapé	102,17	47,50	24,00	114,50	135,24	132%	2,45	9,46	4	770
Itutinga	118,18	102,50	72,25	134,50	72,51	61%	2,43	11,55	39	572
Jaguará	811,84	829,50	658,25	924,25	323,08	40%	1,16	7,21	207	2471
Estreito	3358,67	2571,50	1723,75	4444,25	2173,37	65%	1,40	4,80	763	12250
Marimbondo	1587,61	1452,00	1214,00	1754,25	669,28	42%	1,42	6,04	383	4394
Mascarenhas	651,50	454,00	325,25	830,25	527,53	81%	2,13	7,97	86	2986
Miranda	302,17	288,50	216,75	375,25	113,89	38%	1,39	7,95	132	918
Nilo Peçanha	116,56	117,00	103,75	129,25	21,13	18%	-0,24	2,95	62	171
Nova Ponte	237,66	178,00	114,75	325,50	168,71	71%	1,46	5,05	39	965
Pereira Passos	68,00	20,50	10,00	141,25	67,33	99%	0,43	1,32	2	181
Picada	33,64	27,50	19,75	42,00	20,44	61%	1,51	5,20	10	120
Porto Estrela	125,74	81,50	55,00	153,50	113,63	90%	1,98	6,86	12	621
Queimado	39,25	34,00	20,00	53,00	26,09	66%	1,42	5,97	2	167
Retiro Baixo	77,69	47,50	28,00	91,75	92,48	119%	3,66	20,46	0	676
Rosal	35,53	24,00	16,00	44,50	30,22	85%	1,98	7,26	5	176
Sá Carvalho	63,24	45,00	32,00	78,25	49,35	78%	2,19	9,17	10	310
São Simão	2224,19	2141,50	1785,75	2528,50	770,93	35%	1,99	13,10	551	7629
Simplicio	85,09	78,50	36,00	116,00	62,25	73%	1,00	3,82	1	280
Sobragi	64,39	54,00	40,00	79,75	35,25	55%	1,31	4,53	11	211
Três Marias	546,20	345,50	174,75	705,50	561,55	103%	1,87	6,39	8	3017
Volta Grande	941,79	938,50	760,75	1062,25	362,16	38%	1,48	8,40	267	2863

4.3. Dados de Vazão PCHs

Nesta dissertação as PCHs consideradas possuem três anos de dados de vazão disponíveis no site da ANA Telemetria (SNIRH, 2019). A distribuidora, que forneceu os dados estudados, possui 32 PCHs em sua área de concessão, porém apenas 18 possuem dados disponíveis de vazão. Em função disso é possível

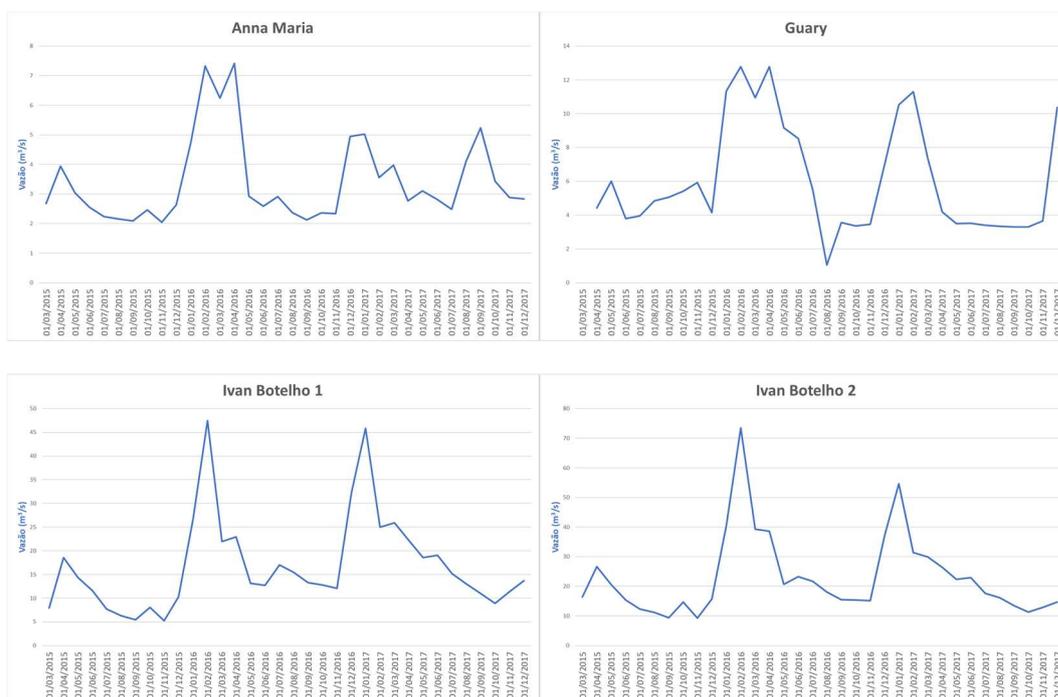
perceber a importância desta pesquisa para verificar se as abordagens para identificação das vazões vizinhas possibilitarão a obtenção de boas previsões, visto que nem sempre as vazões próprias estarão disponíveis.

As estações de telemetria, data da primeira observação disponível e quantidade de observações estão apresentadas na Tabela 4.6.

Tabela 4.6 – Dados de Vazões das PCHs.

Estação	Localização	Início	Número de observações
Anna Maria	Montante	fev/15	35
Guary	Jusante	mar/15	34
Ivan Botelho 1	Montante	fev/15	35
Zé Tunin	Montante	fev/15	35
Ivan Botelho 2	Jusante	fev/15	35

Apesar do número de observações ser pequeno, as séries serão testadas para fins de comparação com os modelos utilizando vazões vizinhas.



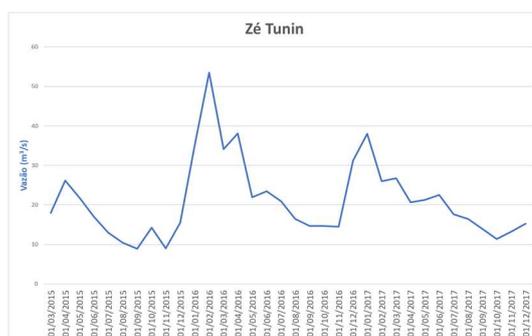


Figura 4.6 – Séries de Vazão Próprias das PCHs.

Finalizando a análise descritiva dos dados, em relação as séries de vazão das PCHs, todas possuem um valor de assimetria positivo, característica que se repete na grande maioria das séries de vazão. Comparando as séries a partir do coeficiente de variação, a PCH Ivan Botelho 2 possui a maior variabilidade em relação à média. As medidas de dispersão são apresentadas na Tabela 4.7.

Tabela 4.7– Estatística Descritiva das Séries de Vazão das PCHs.

	V. Anna Maria	V. Guary	V. Ivan Botelho 1	V. Ivan Botelho 2	V. Zé Tunin
Média	3,44	6,23	17,04	22,74	21,17
Mediana	2,88	4,95	13,70	17,60	17,95
1º quartil	2,48	3,53	11,17	14,69	19,99
3º quartil	4,03	9,00	22,04	26,38	38,50
Desvio padrão	1,42	3,33	9,91	13,71	9,72
CV	41%	53%	58%	60%	46%
Assimetria	1,51	0,71	1,60	2,11	1,38
Curtose	4,35	2,08	5,37	7,03	4,83
Mínimo	2,04	1,05	5,27	9,23	8,89
Máximo	7,42	12,78	47,50	73,50	53,48

4.4. Análise Espacial

Um passo muito importante da análise da autocorrelação espacial é a preparação da matriz de pesos. Para preparar a matriz que define quem é vizinho e quem não é vizinho, algumas informações foram consideradas como: a distância entre as usinas e se a usina está no mesmo rio e bacia.

Para que todas as localidades tenham ao menos 2 vizinhos, a distância máxima definida foi 120 km.

Outra informação necessária diz respeito a variável de interesse para o cálculo da estatística I de *Moran* e I de *Moran* Local. Nesta dissertação foi escolhida a média das vazões.

5 Resultados Obtidos

Neste capítulo serão apresentados os resultados obtidos seguindo a metodologia proposta na seção 3.

5.1. Etapa 1 - Previsão dos Modelos Univariados

Alguns passos relacionados a primeira etapa, ou seja, a escolha das usinas (passo 1) e a análise descritiva das séries (passo 2) foram apresentadas na seção 4.

No passo seguinte as séries foram divididas em treinamento e validação de acordo com a Tabela 5.1. O objetivo da utilização dessa estratégia é possibilitar o ajuste do modelo para a série de treinamento e depois realizar a previsão 12-passos a frente e comparar o resultado obtido com o valor observado no ano de 2017 através do erro fora da amostra (*out-of-sample*)

Tabela 5.1 – Divisão das Séries Temporais em Janela de Treinamento e Validação.

PCH	Treinamento	Validação
Anna Maria	out/00 – dez/16	jan/17 – dez/17
Guary	nov/10 – dez/16	jan/17 – dez/17
Ivan Botelho 1	mai/03 – dez/16	jan/17 – dez/17
Ivan Botelho 2	out/03 – dez/16	jan/17 – dez/17
Zé Tunin	dez/12 – dez/16	jan/17 – dez/17

Para um melhor entendimento do comportamento das séries, a Figura 5.1 apresenta a decomposição das séries em nível, tendência, sazonalidade e resíduo. Todas as séries apresentam uma componente sazonal bem marcada em relação a estacionariedade. Apesar do gráfico aparentemente indicar que as séries são estacionárias, os testes Augmented Dickey-Fuller e Philips-Perron, que verificam a presença de raiz unitária foram realizados e os resultados são apresentados na Tabela 5.2. Nos dois testes a estacionariedade da série é a hipótese alternativa, ou

seja, H_0 (Dickey e Fuller, 1979; Philips e Perron, 1988). De acordo com os testes realizados pode-se dizer que a PCH Guary não é estacionária.

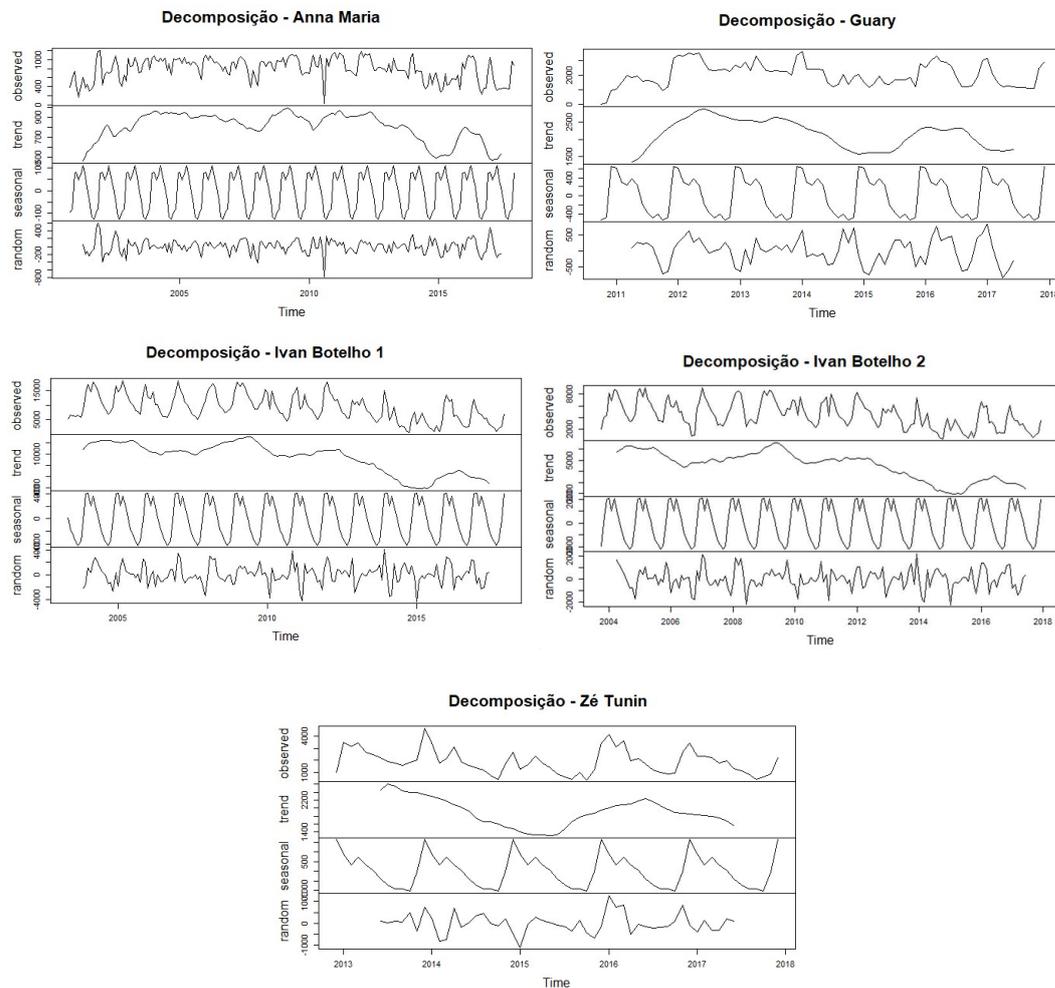


Figura 5.1 – Decomposição das Séries Temporais de Geração em Nível, Tendência, Sazonalidade e Resíduo.

Tabela 5.2 – Resultado do P-valor para os Testes de Raízes Unitárias.

PCH	Dickey-Fuller	Philips-Perron
Anna Maria	0,01	0,01
Guary	0,0248	0,0238
Ivan Botelho 1	0,01	0,01
Ivan Botelho 2	0,01	0,01
Zé Tunin	0,0197	0,01

O próximo passo (passo 4) desta etapa é ajustar os modelos de amortecimento exponencial e ARIMA e depois realizar a previsão 12 passos a frente para o ano de 2017. Conforme visto na Figura 5.1, as séries de geração possuem um comportamento sazonal, logo o ajuste do modelo será feito pelo método de

amortecimento exponencial de Holt-Winters. As séries foram ajustadas tanto para o modelo aditivo como para o multiplicativo apesar da variação sazonal apresentar um comportamento constante.

Para escolher o modelo SARIMA que apresentará o menor erro de previsão de cada usina, foi utilizada a função *auto.arima* do pacote *forecast* que automatiza os passos da metodologia Box & Jenkins permitindo a escolha do melhor modelo a partir dos critérios AIC, AICc ou BIC (Hyndman e Khandakar, 2008).

Os modelos ARIMA selecionados foram: Anna Maria $(1,1,1) \times (0,0,2)_{12}$, Guary $(1,0,0) \times (2,0,0)_{12}$, Ivan Botelho 1 $(1,0,0) \times (1,1,0)_{12}$, Ivan Botelho 2 $(1,0,2) \times (1,1,1)_{12}$ e Zé Tunin $(0,1,1) \times (1,1,0)_{12}$.

Os resultados *out-of-sample* são apresentados na Tabela 5.3.

Foram utilizadas as métricas *Mean Absolute Percentage Error* (MAPE) e *Root Mean Square Error* (RMSE) para avaliar o erro *out-of-sample*, ou seja, para verificar a precisão das previsões comparando com os dados reais separados para validar o modelo. A escolha dessas duas métricas ocorreu devido a sua ampla utilização reportada na literatura (Hyndman e Athanasopoulos, 2018).

As equações são mostradas abaixo:

$$MAPE = 100 \frac{1}{N} \sum_{t=1}^N \left| \frac{\hat{Y}_t - Y_t}{Y_t} \right| \quad (5.1)$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^N (\hat{Y}_t - Y_t)^2}{N}} \quad (5.2)$$

Onde \hat{Y}_t é previsão da geração no tempo t e Y_t é o valor da geração observada e N é o horizonte de previsão que será de 12 passos a frente.

Tabela 5.3 – Resultado MAPE (%) e RMSE (kWmed) – Amortecimento Exponencial e ARIMA.

		Anna Maria	Guary	Ivan Botelho 1	Ivan Botelho 2	Zé Tunin
HW - A	MAPE	54	84	95	93	57
	RMSE	280	1211	3577	1918	698
HW - M	MAPE	46	61	129	116	62
	RMSE	262	896	4371	2089	815
ARIMA	MAPE	55	58	41	80	74
	RMSE	261	839	2162	1653	915

5.2.

Etapa 2 – Obtenção e Tratamento da Variável Explicativa

Para ajustar e prever as gerações de energia elétrica das PCHs através da aplicação dos modelos causais, as abordagens foram divididas em 3 casos:

Caso 1 – a variável exógena será a vazão medida no ponto de telemetria da própria PCH.

Caso 2 – a variável exógena será a vazão vizinha que possui o maior coeficiente de correlação com a geração da PCH.

Caso 3 – todas as vazões das UHEs escolhidas a partir da análise espacial serão incorporadas aos modelos como variáveis exógenas.

Em todos os casos, durante a etapa 3, serão utilizadas as vazões reais observadas para o ano de 2017.

Para aplicação da primeira abordagem (Caso 1), a série de vazão precisa apresentar o mesmo número de observações da série de geração. As usinas selecionadas possuem dados a partir de 2015, logo as séries de treinamento foram cortadas conforme apresentado na Tabela 5.4. Na Figura 5.2 são apresentados os gráficos de dispersão e na Figura 5.3 os gráficos de linha mostrando o comportamento das séries.

Tabela 5.4 – Novas Janelas de Treinamento e Validação das Séries Temporais para Ajuste do Caso 1.

Geração PCH	Vazão	Treinamento	Validação
Anna Maria	Anna Maria	fev/15 – dez/16	2017
Guary	Guary	mar/15 – dez/16	2017
Ivan Botelho 1	Ivan Botelho 1	fev/15 – dez/16	2017
Ivan Botelho 2	Ivan Botelho 2	fev/15 – dez/16	2017
Zé Tunin	Zé Tunin	fev/15 – dez/16	2017

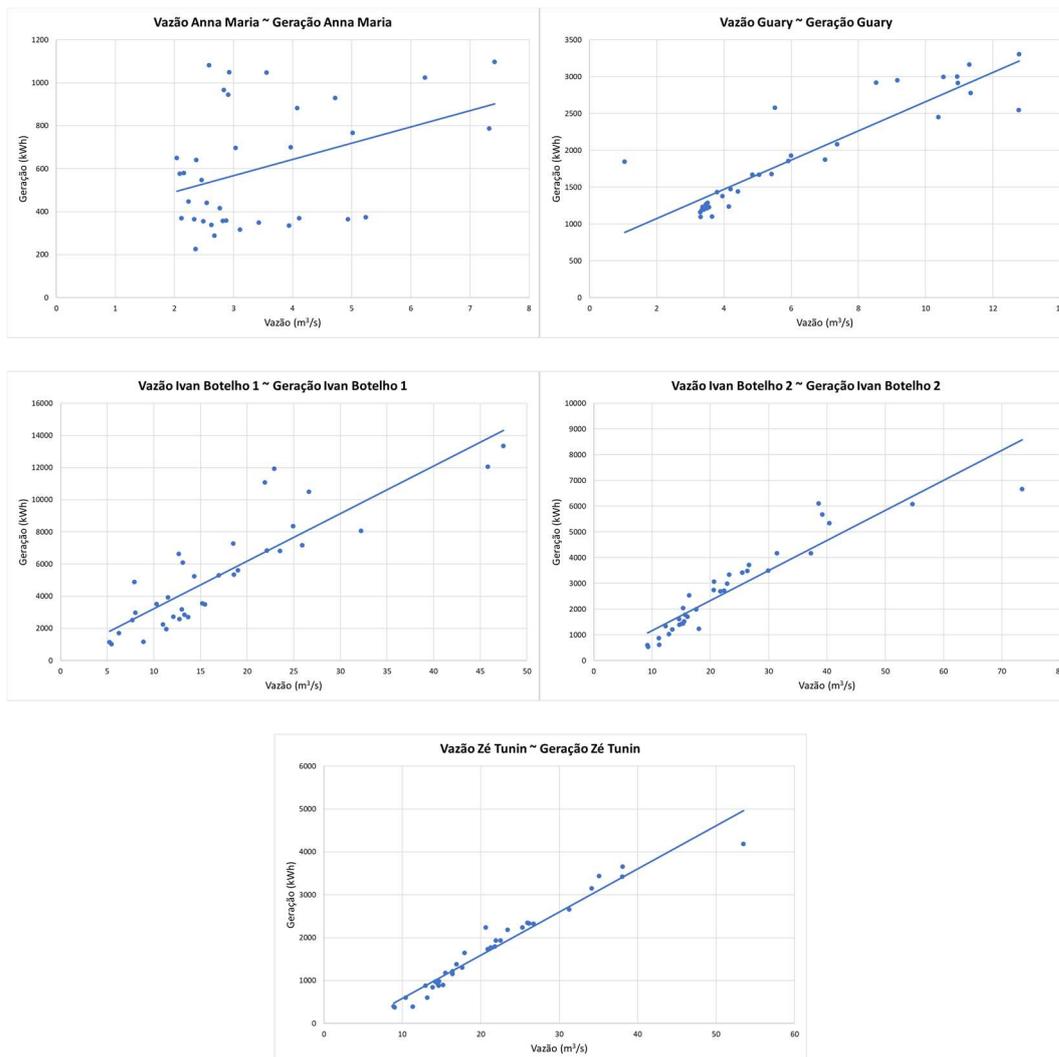


Figura 5.2 – Gráficos de Dispersão da Vazão Própria e Geração de cada Usina.

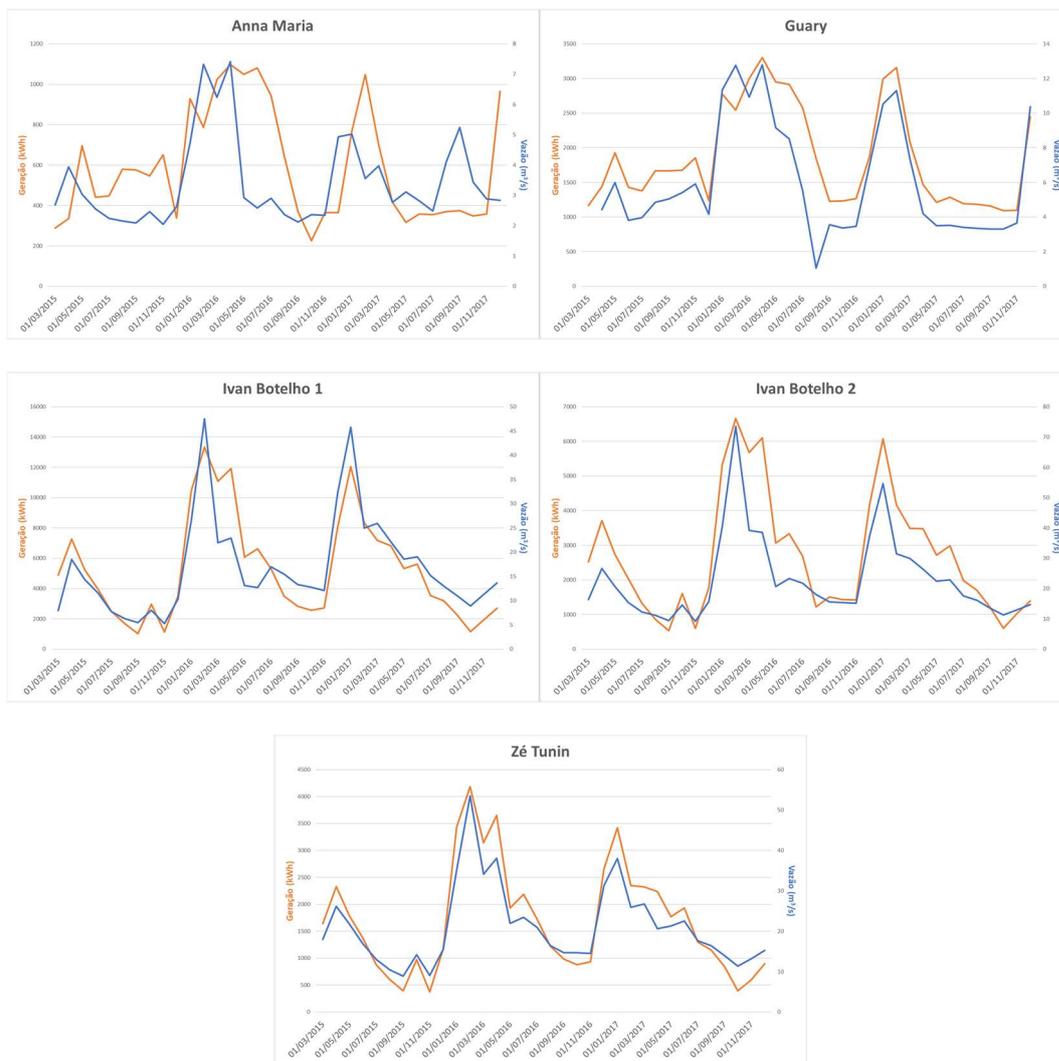


Figura 5.3 – Gráficos de Linha da Vazão Própria e Geração de cada Usina.

Na segunda abordagem (Caso 2), para identificar qual a vazão vizinha que será utilizada no modelo como variável exógena, o coeficiente de correlação de Pearson foi adotado. Neste caso, pretende-se verificar qual a relação que existe entre as séries de geração das PCHs e as 34 séries de vazão localizadas nos estados de Minas Gerais, Espírito Santo e Rio de Janeiro. O valor do coeficiente de correlação de Pearson varia entre -1 e 1. Quando mais próximo de 1 maior a correlação positiva entre as variáveis e quanto mais próximo de -1 maior a correlação negativa entre elas. Na Figura 5.4 são apresentados os gráficos de dispersão da Vazão e Geração.

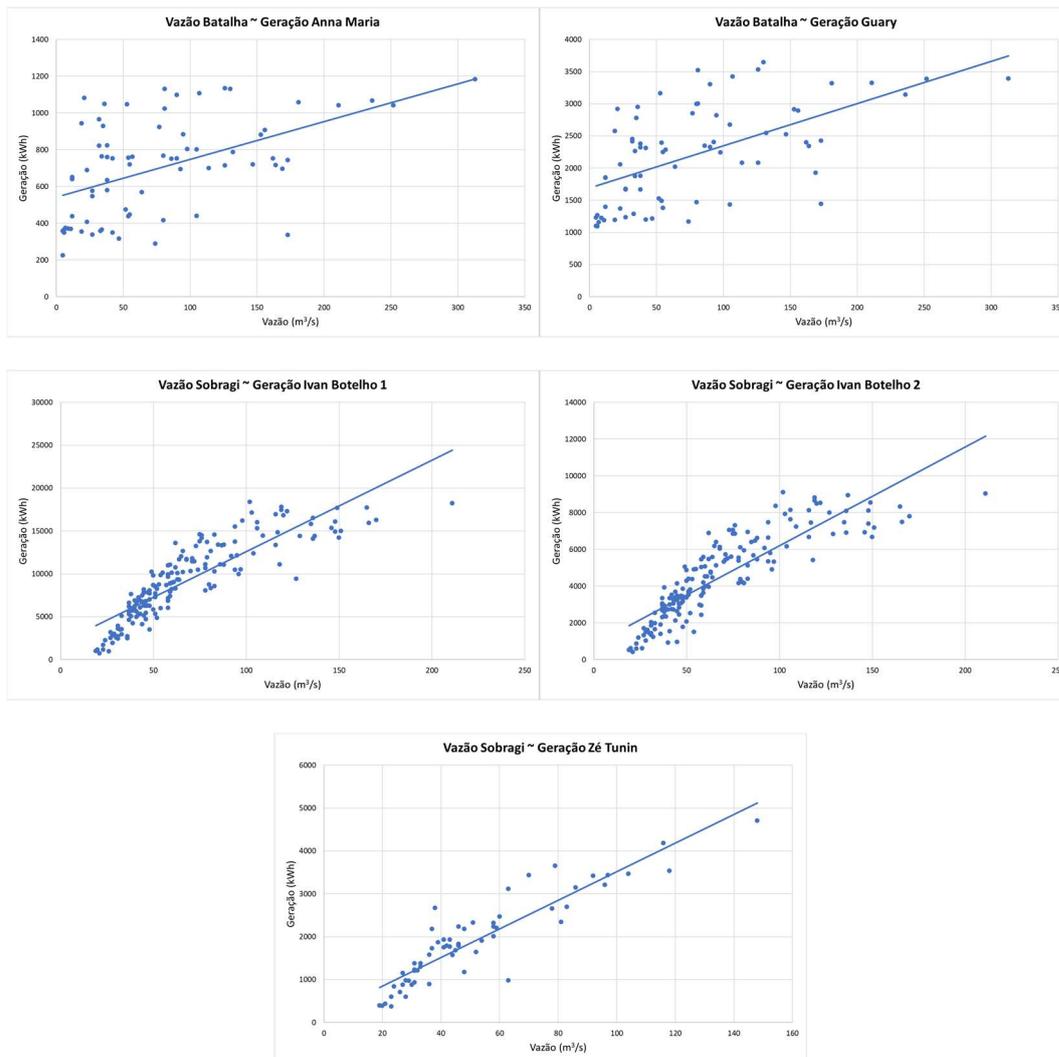


Figura 5.4 – Gráficos de Dispersão da Vazão mais Correlacionada e Geração de cada Usina.

Apesar desse coeficiente não mostrar se as variáveis possuem uma relação de causalidade, é possível perceber se as variáveis estudadas possuem um comportamento similar. Através dos gráficos de dispersão observa-se uma maior correlação entre a Vazão de Sobragi e as Gerações de Ivan Botelho 1, Ivan Botelho 2 e Zé Tunin, que podem ser confirmadas na Tabela 5.5 que apresenta o resultado do cálculo do coeficiente e as usinas hidrelétricas cujas vazões serão usadas nos modelos.

Tabela 5.5 – Correlações entre Geração e Vazão das UHEs Vizinhas.

Geração PCH	Vazão UHE	Correlação
Anna Maria	Batalha	0,48
Guary	Batalha	0,58
Ivan Botelho 1	Sobragi	0,86
Ivan Botelho 2	Sobragi	0,86
Zé Tunin	Sobragi	0,91

A usina hidrelétrica de Sobragi está localizada no Rio Paraibuna e a de Batalha no Rio São Marcos. Na Tabela 5.6 são apresentados o número de observações e a data da primeira observação das usinas hidrelétricas que serão consideradas no modelo.

Tabela 5.6 – Usinas Hidrelétricas mais Correlacionadas com Geração.

Vazão UHE	Rio	Data Primeira Observação	Número de Observações
Batalha	Rio São Marcos	dez/11	73
Sobragi	Rio Paraibuna	jul/00	210

Para realizar o ajuste dos modelos causais as séries de geração e vazão precisam ter o mesmo número de observações. Para isso as séries foram cortadas respeitando o tamanho da menor série temporal, ou seja, mesmo que a série de vazão possua um número grande de observações, as janelas de treinamento ficaram com o mesmo tamanho das séries de geração. As novas janelas de treinamento podem ser visualizadas na Tabela 5.7.

Tabela 5.7 – Novas Janelas de Treinamento e Validação das Séries Temporais para Ajuste do Caso 2.

G. PCH	V. UHE	Treinamento	Validação
Anna Maria	Batalha	dez/11 – dez/16	2017
Guary	Batalha	dez/11 – dez/16	2017
Ivan Botelho 1	Sobragi	mai/03 – dez/16	2017
Ivan Botelho 2	Sobragi	out/03 – dez/16	2017
Zé Tunin	Sobragi	dez/12 – dez/16	2017

Para aplicar a abordagem para o caso 3, o primeiro passo realizado foi a identificação das séries de vazão que seguem o critério escolhido na preparação da matriz de pesos espaciais.

A estatística I de *Moran* encontrada nessa análise foi 0.4739 indicando que os dados possuem uma similaridade global, ou seja, valores altos estão rodeados de valores altos e vice-versa.

A estatística I de *Moran* Local para a matriz definida, ou seja, mostrando como vizinhos Ilha dos Pombos, Picada, Simplício e Sobragi de cada uma das PCHs confirmam que esses vizinhos podem ser considerados adequados visto que o valor de todas as localidades é positivo e significativo. Por exemplo, os vizinhos de Anna Maria possuem valores da variável de interesse similar aos valores dos vizinhos. A estatística I de *Moran* Local de Anna Maria é 0,3719 com p-valor < 0,01.

Assim como nos casos 1 e 2 as séries foram cortadas para que todas as séries tenham o mesmo número de observações. As novas janelas de treinamento, Tabela 5.8, iniciam em maio de 2013.

Tabela 5.8 – Novas Janelas de Treinamento e Validação das Séries Temporais para Ajuste do Caso 3.

G. PCH	V. UHE	Treinamento	Validação
Anna Maria	Ilha dos Pombos, Picada, Simplício e Sobragi	mai/13 – dez/16	2017
Guary	Ilha dos Pombos, Picada, Simplício e Sobragi	mai/13 – dez/16	2017
Ivan Botelho 1	Ilha dos Pombos, Picada, Simplício e Sobragi	mai/13 – dez/16	2017
Ivan Botelho 2	Ilha dos Pombos, Picada, Simplício e Sobragi	mai/13 – dez/16	2017
Zé Tunin	Ilha dos Pombos, Picada, Simplício e Sobragi	mai/13 – dez/16	2017

5.3.

Etapa 3 – Aplicação de Modelos Causais

Nesta etapa para cada caso serão ajustados modelos causais de Função de Transferência e de Regressão Dinâmica.

Para aplicação do modelo de Box & Jenkins de função de transferência alguns passos são necessários para identificação dos parâmetros que serão usados no modelo. A identificação dos parâmetros (r,b,s) ocorre após o pré-branqueamento das séries. A Tabela 5.9 apresenta os modelos que foram identificados e utilizados para realização da previsão 12 passos à frente. Para o Caso 3 X_{1t} – Ilha dos Pombos, X_{2t} – Picada, X_{3t} – Simplício, X_{4t} – Sobragi. Para o Caso 1, X_{1t} será a vazão

própria da PCH e para o Caso 2 será a vazão correlacionada identificada na etapa 2.

No último passo para verificação do modelo de Função de Transferência são executados testes de análise de resíduos para verificar se o modelo gerou um ruído branco. O teste de Box-Ljung foi realizado para confirmar se os modelos geraram ruído branco e não foram encontradas evidências estatísticas comprovando a dependência dos resíduos.

Tabela 5.9 – Modelos – Função de Transferência.

PCH	Caso	Modelo de Função de Transferência
	1	$Y_t = 60,01 X_t + \frac{1}{(1 - 0,49B - 0,41 B^2)} a_t$
	2	$Y_t = 1,38 X_t + \frac{1}{(1 + 0,32B - B^{12})} a_t$
	3	$Y_t = -0,18 X_{1t} + 0,63 X_{2t} - 0,09 X_{3t} - 0,06 X_{4t} + \frac{1}{(1 - 0,93B)} a_t$
Anna Maria		
	1	$Y_t = 175,36 X_{1t} + \frac{1}{(1 - 0,93B)} a_t$
	2	$Y_t = 4,06 X_{1t} - (2,52 B) a_t$
	3	$Y_t = -0,03 X_{1t} + 0,58 X_{2t} - 0,07 X_{3t} - 0,12 X_{4t} + \frac{1}{(1 - 0,96B)} a_t$
Guary		
	1	$Y_t = 255,87 X_{1t} + \frac{1}{(1 - 0,75 B)} a_t$
	2	$Y_t = 73,27 X_{1t} + \frac{(-0,86 B)}{(1 - 0,39 B - 0,19 B^{12} - 0,28 B^{24})} a_t$
	3	$Y_t = 0,27 X_{1t} + 0,014 X_{2t} + 0,12 X_{3t} + 0,79 X_{4t} + \frac{1}{(1 - 0,28B)} a_t$
Ivan Botelho 1		
	1	$Y_t = 115,84 X_{1t} + a_t$
	2	$Y_t = 0,28 X_{1t} + 0,04 X_{2t} - 0,09 X_{3t} - 0,77 X_{4t} + \frac{1}{(1 - 0,27B)} a_t$
	3	$Y_t = 0,28 X_{1t} + 0,04 X_{2t} - 0,09 X_{3t} - 0,77 X_{4t} + \frac{1}{(1 - 0,27B)} a_t$
Ivan Botelho 2		

1	$Y_t = -333,17 + 96,98 X_{1t} + a_t$
2	$Y_t = 35,33 X_{1t} + \frac{1}{(1 - 0,44 B)} a_t$
Zé Tunin	
3	$Y_t = 0,23 X_{1t} - 0,09 X_{2t} + 0,10 X_{3t} + 0,98 X_{4t} + \frac{1}{(1 - 0,27 B)} a_t$

Na Tabela 5.10 são apresentados os resultados *out-of-sample* do modelo de função de transferência para os três casos. Percebe-se que o melhor resultado, obtido quando a vazão foi adicionada como variável exógena, utiliza a vazão da própria da usina apesar dos poucos dados disponíveis. O segundo melhor resultado foi obtido quando foram adicionadas as quatro séries escolhidas (Ilha dos Pombos, Sobragi, Picada e Simplício) a partir da análise espacial como variáveis exógenas. A única exceção encontramos para o caso de Zé Tunin que apresenta o melhor MAPE quando a vazão mais correlacionada com a geração é adicionada ao modelo. Vale ressaltar que para essa usina o coeficiente de correlação de *Pearson* encontrado foi 0,91 com a vazão da usina de Sobragi que está a 84 km de distância da PCH e está entre as usinas escolhidas pela análise espacial.

Tabela 5.10 – Resultado MAPE (%) e RMSE(kWmed) – Função de Transferência.

		Anna Maria	Guary	Ivan Botelho 1	Ivan Botelho 2	Zé Tunin
FT Caso 1	MAPE	28	10	25	23	33
	RMSE	294	205	861	416	321
FT Caso 2	MAPE	54	55	48	33	27
	RMSE	261	816	1403	549	359
FT Caso 3	MAPE	30	49	35	30	31
	RMSE	274	750	1263	593	430

Conforme apresentado no referencial teórico para escolher as defasagens utilizadas na regressão dinâmica, vários modelos foram testados empiricamente e o menor BIC, assim como a significância dos parâmetros, foram utilizados para selecionar o modelo.

Os modelos testados são apresentados a seguir:

- 1) $Y_t = \alpha + \phi_1 Y_{t-2} + \theta_0 X_{1t-2} + \theta_1 X_{2t-2} + \theta_2 X_{3t-2} + \theta_3 X_{4t-2} + a_t$
- 2) $Y_t = \alpha + \phi_1 Y_{t-1} + \theta_0 X_{1t-1} + \theta_1 X_{2t-1} + \theta_2 X_{3t-1} + \theta_3 X_{4t-1} + a_t$

$$3) Y_t = \alpha + \phi_1 Y_{t-1} + \theta_0 X_{1t} + \theta_1 X_{2t} + \theta_2 X_{3t} + \theta_3 X_{4t} + a_t$$

$$4) Y_t = \alpha + \phi_1 Y_{t-1} + \theta_0 X_{1t} + \theta_1 X_{1t-1} + \theta_3 X_{2t} + \theta_4 X_{2t-1} + \theta_5 X_{3t} + \theta_6 X_{3t-1} + \theta_7 X_{4t} + \theta_8 X_{4t-1} + a_t$$

$$5) Y_t = \alpha + \phi_1 Y_{t-1} + \theta_0 X_{1t} + \theta_1 X_{1t-1} + \theta_3 X_{1t-2} + \theta_4 X_{2t} + \theta_5 X_{2t-1} + \theta_6 X_{2t-2} + \theta_7 X_{3t} + \theta_8 X_{3t-1} + \theta_9 X_{3t-2} + \theta_{10} X_{4t} + \theta_{11} X_{4t-1} + \theta_{12} X_{4t-2} + a_t$$

Sendo $X_{1t}, X_{2t}, X_{3t}, X_{4t}$ as quatro séries de vazão escolhidas seguindo os critérios apresentados na etapa 2, ou seja, Ilha dos Pombos, Picada, Simplício e Sobragi, respectivamente.

Nos Caso 1 e 2 apenas uma variável explicativa foi acrescentada ao modelo, de acordo com os resultados encontrados na etapa 2, logo os modelos testados não possuem as variáveis X_{2t}, X_{3t}, X_{4t} .

As variáveis defasadas foram acrescentadas ao modelo permitindo assim uma melhor representação do impacto das variáveis explicativas na variável de resposta Y_t (Geração).

Os modelos que obtiveram o menor BIC estão na Tabela 5.11. Depois de ajustar o modelo todas as séries apresentaram ruído branco nos resíduos e para confirmar, o teste de Box-Ljung foi realizado. Para todos os modelos o p-valor foi maior do que 0,05 não apresentando evidências estatísticas comprovando a dependência dos resíduos.

Tabela 5.11 – Modelos – Regressão Dinâmica.

PCH	Caso	Modelo Regressão Dinâmica
Anna Maria	1	$Y_t = \alpha + \phi_1 Y_{t-1} + \theta_0 X_{1t+1} + a_t$
	2	$Y_t = \alpha + \phi_1 Y_{t-1} + \theta_0 X_{1t} + a_t$
	3	$Y_t = \alpha + \phi_1 Y_{t-1} + \theta_0 X_{1t} + \theta_1 X_{2t} + \theta_2 X_{3t} + \theta_3 X_{4t} + a_t$
Guary	1	$Y_t = \alpha + \theta_0 X_{1t} + \theta_1 X_{1t+1} + a_t$
	2	$Y_t = \alpha + \phi_1 Y_{t-1} + \theta_0 X_{1t} + a_t$
	3	$Y_t = \alpha + \phi_1 Y_{t-1} + \theta_0 X_{1t} + \theta_1 X_{2t} + \theta_2 X_{3t} + \theta_3 X_{4t} + a_t$

Ivan Botelho 1	1	$Y_t = \alpha + \phi_1 Y_{t-1} + \theta_0 X_{1t} + \theta_1 X_{1t-1} + a_t$
	2	$Y_t = \alpha + \theta_0 X_{1t} + \theta_1 X_{1t-1} + \theta_3 X_{1t-2} + a_t$
	3	$Y_t = \alpha + \phi_1 Y_{t-1} + \theta_0 X_{1t} + \theta_1 X_{1t-1} + \theta_3 X_{2t} + \theta_4 X_{2t-1} + \theta_5 X_{3t} + \theta_6 X_{3t-1} + \theta_7 X_{4t} + \theta_8 X_{4t-1} + a_t$
Ivan Botelho 2	1	$Y_t = \alpha + \theta_0 X_{1t} + \theta_1 X_{1t-1} + \theta_3 X_{1t-2} + a_t$
	2	$Y_t = \alpha + \theta_0 X_{1t} + \theta_1 X_{1t-1} + \theta_3 X_{1t-2} + a_t$
	3	$Y_t = \alpha + \phi_1 Y_{t-1} + \theta_0 X_{1t} + \theta_1 X_{2t} + \theta_2 X_{3t} + \theta_3 X_{4t} + a_t$
Zé Tunin	1	$Y_t = \alpha + \theta_0 X_{1t} + \theta_1 X_{1t-1} + \theta_3 X_{1t-2} + a_t$
	2	$Y_t = \alpha + \phi_1 Y_{t-1} + \theta_0 X_{1t} + a_t$
	3	$Y_t = \alpha + \phi_1 Y_{t-1} + \theta_0 X_{1t} + \theta_1 X_{1t-1} + \theta_3 X_{2t} + \theta_4 X_{2t-1} + \theta_5 X_{3t} + \theta_6 X_{3t-1} + \theta_7 X_{4t} + \theta_8 X_{4t-1} + a_t$

De acordo com os resultados apresentados na Tabela 5.12 as previsões das PCHs Anna Maria, Ivan Botelho 1 e Ivan Botelho 2 apresentaram ótimos resultados para o Caso 3. Já as PCHs Guary e Zé Tunin apresentaram melhores resultados quando incluíram a vazão própria como variável explicativa no modelo de regressão dinâmica.

Tabela 5.12 – Resultado MAPE(%) e RMSE(kWmed) – Regressão Dinâmica.

		Anna Maria	Guary	Ivan Botelho 1	Ivan Botelho 2	Zé Tunin
RD Caso 1	MAPE	67	18	54	18	19
	RMSE	300	276	1659	369	229
RD Caso 2	MAPE	53	50	64	48	26
	RMSE	265	741	1960	795	365
RD Caso 3	MAPE	31	34	17	17	24
	RMSE	256	679	1060	526	399

Comparando todos os modelos testados é possível observar na Tabela 5.13 que os modelos que incorporam, como variável exógena a vazão própria, apresentam os melhores MAPE. Para verificar porém qual o melhor modelo quando não existe dado de vazão para uma determinada PCH, observa-se que a Regressão Dinâmica utilizando na escolha dos vizinhos a análise espacial apresentou o melhor resultado. Confirmando assim essa abordagem como satisfatória para resolver o problema desta pesquisa.

Tabela 5.13 – Resultados do MAPE(%) *Out-of-Sample* Consolidado.

	Anna Maria	Guary	Ivan Botelho 1	Ivan Botelho 2	Zé Tunin
FT - Vazão Própria	28	10	25	23	22
RD - Vazão Própria	67	18	54	19	19
RD - Vizinha Espacial	31	34	17	17	24
FT - Vizinha Espacial	30	50	35	30	31
Arima	55	58	41	80	74
FT - Vazão Vizinha	54	77	49	33	27
RD - Vazão Vizinha	53	50	65	48	26
HW - Aditivo	54	92	96	94	57
HW - Multiplicativo	46	92	130	117	63

Na Figura 5.5 são apresentados os gráficos comparativos das melhores previsões com o valor real da geração observado para o ano de 2017.

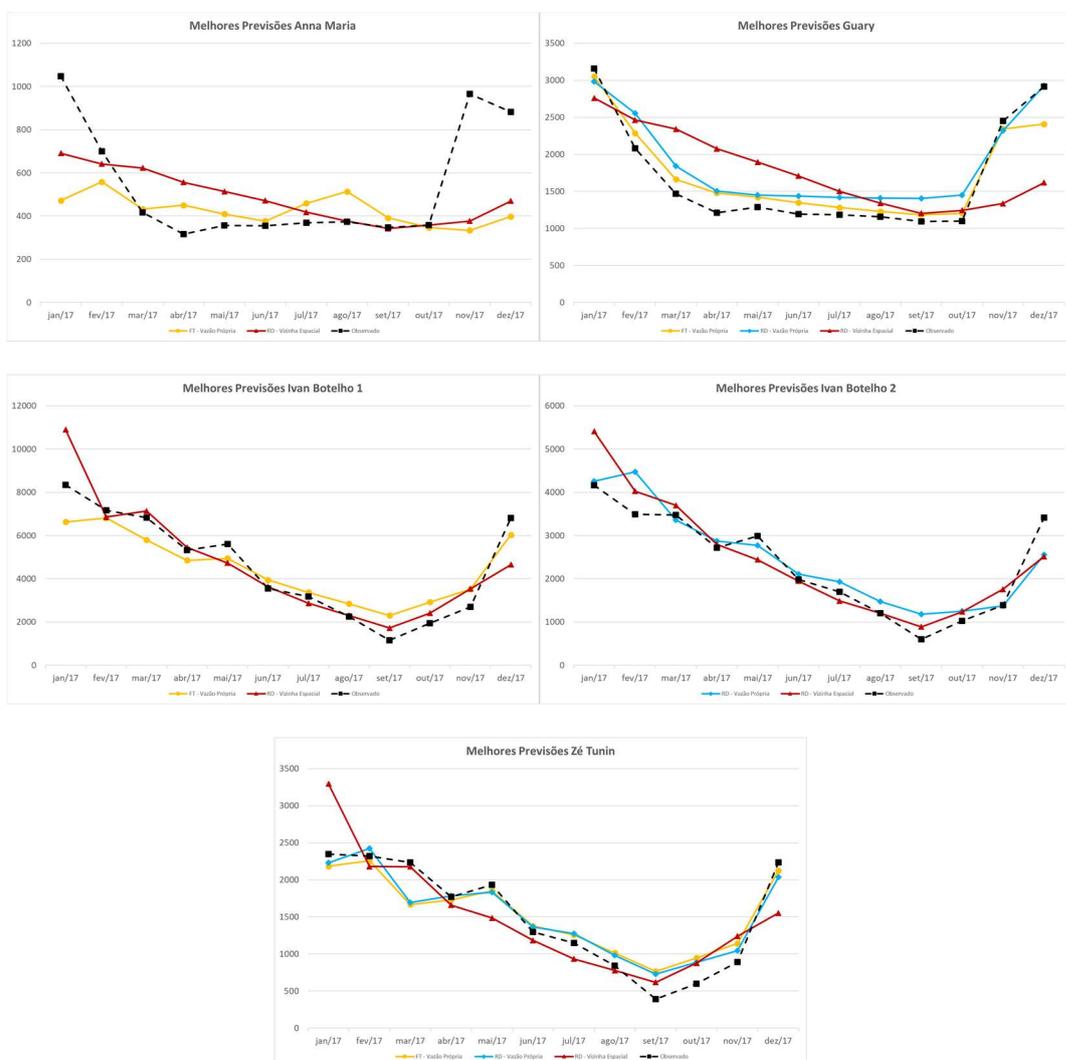


Figura 5.5 – Comparativo das Melhores Previsões 12 passos a frente para cada PCH.

5.4. Etapa 4 – Previsão com Cenário de Vazão

Para simular uma situação real, onde deseja-se prever a geração da PCH 12 passos a frente para o ano de 2018, um cenário sintético das vazões precisa ser gerado. Para atender a essa condição um cenário a partir das médias mensais foi criado.

Serão utilizados os modelos de Regressão Dinâmica com as Vazões Vizinhas depois da análise espacial, assim como a Função de Transferência com as Vazões Próprias, pois esses modelos apresentaram os melhores resultados na etapa anterior.

Nas Figuras 5.6 e 5.7 são apresentados os cenários das vazões próprias e vizinhas para o ano de 2018.

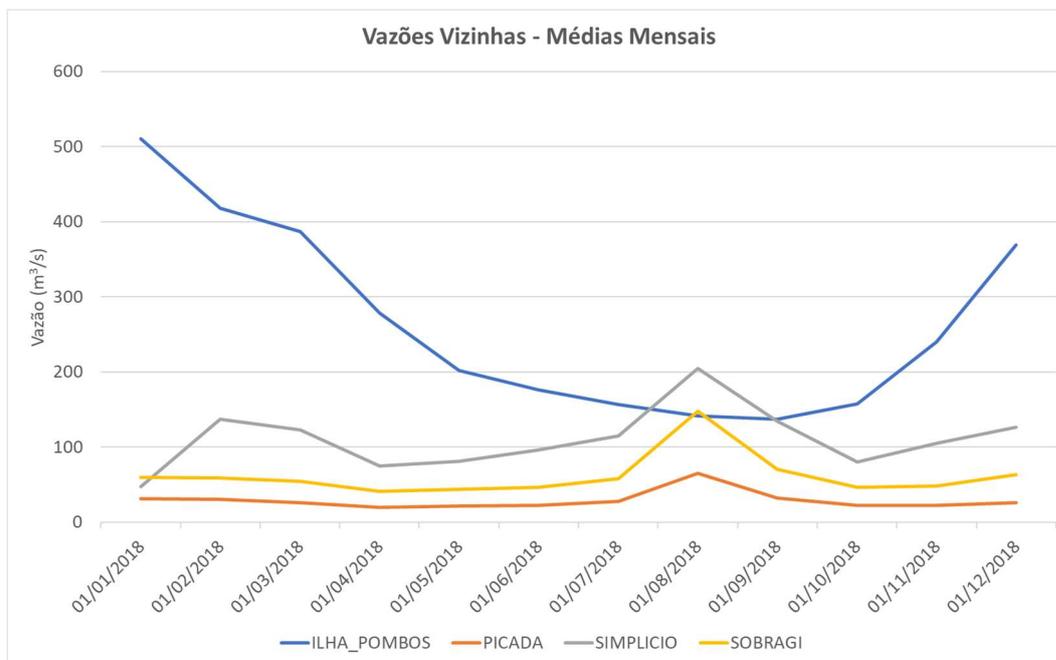


Figura 5.6 – Cenário Vazões Vizinhas – Médias Mensais para 2018.

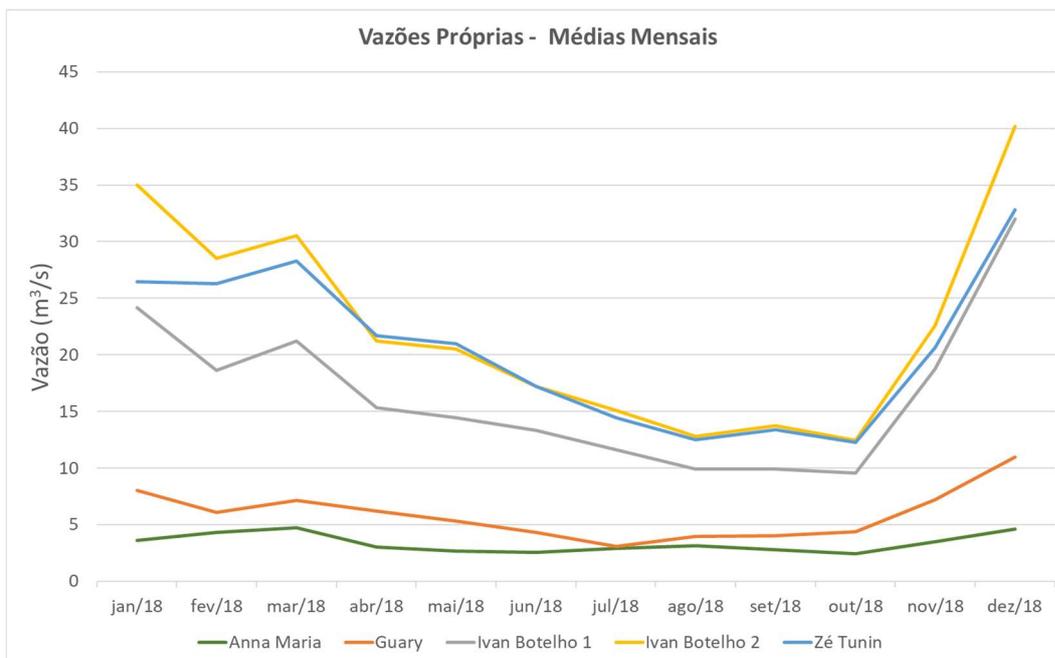


Figura 5.7 – Cenário Vazões Próprias – Médias Mensais para 2018.

Na Tabela 5.14 é possível verificar, comparando esses dois modelos, que os menores erros de previsão foram obtidos com a regressão dinâmica. Essa constatação confirma que a utilização das vazões vizinhas após análise espacial podem ser utilizadas em substituição da vazão própria quando for inexistente.

Para o modelo de Função de Transferência com Vazão Própria todos os resultados pioram com a utilização desse cenário de vazão. O MAPE encontrado para a Regressão Dinâmica da usina Anna Maria na etapa 3 foi 31% e passou para 29% na etapa 4 e para usina Guary passou de 34% para 24%. Para as demais usinas os erros aumentaram.

Tabela 5.14 – Resultados do MAPE(%) e RMSE (kWmed) de Teste Consolidado.

		Anna Maria	Guary	Ivan Botelho 1	Ivan Botelho 2	Zé Tunin
FT Vazão Própria	MAPE	39	42	46	30	35
	RMSE	471	1463	6182	2397	1665
RD Vazão Espacial	MAPE	29	24	29	27	31
	RMSE	293	902	4768	2063	1352

Complementando a análise dos resultados das previsões, a Figura 5.8 apresenta os gráficos comparativos dos modelos para o ano de 2018.

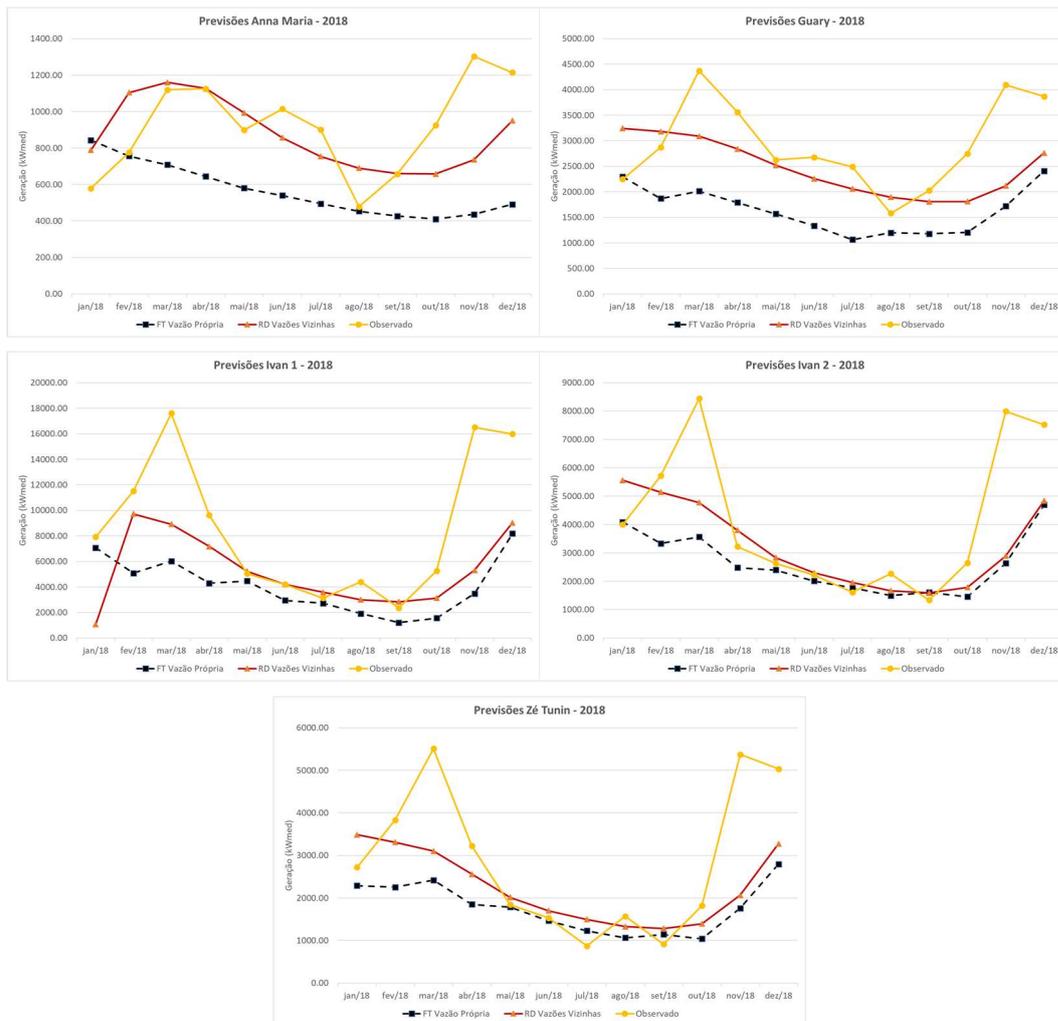


Figura 5.8 – Previsões da Geração das PCHs para 2018.

6 Conclusão e Trabalhos Futuros

Este trabalho possui como principal motivação a apresentação de modelos de previsão das gerações de PCHs que utilizem dados de vazão de bacias vizinhas como variável explicativa. Para isso foram utilizadas duas abordagens para definição da melhor série de vazão vizinha ou das melhores séries de vazão. A primeira abordagem utilizou um estudo de correlação, utilizando o coeficiente de Pearson, onde 34 séries de vazão de usinas hidrelétricas localizadas nos estados de Minas Gerais, Espírito Santo e Rio de Janeiro foram testadas e a melhor correlação entre as Vazões e a Geração da PCH estudada foi selecionada.

Na segunda abordagem, um estudo de autocorrelação espacial foi realizado para definir as séries de vazão vizinhas que apresentam similaridade e que estejam localizadas a uma distância definida que permita que a PCH estudada possua mais de um vizinho próximo. A escolha dessa segunda abordagem foi resultado da revisão da literatura que identificou, em alguns artigos que estudam o comportamento das vazões, a aplicação de modelos espaço-temporais, considerando além da dependência temporal, uma dependência espacial dos dados.

Partindo da premissa que as PCHs, usinas consideradas como fio-d'água, possuem uma dependência em relação aos regimes hidrológicos, que por sua vez afetam diretamente a geração deste tipo de usina, seria bem interessante, antes de partir para utilização de vizinhos com comportamento similar, utilizar a própria vazão medida em estações de telemetria na PCH. Por isso, neste trabalho optou-se por apresentar uma terceira abordagem utilizando a vazão própria e comparar os resultados com as abordagens dos vizinhos. Uma limitação deste trabalho está na quantidade de observações das séries de vazão própria.

Além dos objetivos acima, em um segundo momento este trabalho se propôs a comparar os resultados das abordagens utilizando variáveis exógenas com modelos univariados clássicos.

Para obtenção dos resultados, os modelos clássicos univariados de amortecimento exponencial de Holt-Winters e Arima foram testados e foram comparados aos modelos causais de Função de Transferência de Box & Jenkins e Regressão Dinâmica, incorporando a vazão própria como variável explicativa, a melhor vazão vizinha correlacionada com a geração e as vazões resultantes da análise espacial dos vizinhos.

Pode-se concluir que a utilização de variáveis exógenas de vazão na previsão da geração melhora os resultados dos modelos como a revisão da literatura aponta. A utilização das vazões vizinhas se mostrou uma boa alternativa para o problema de falta de dados de vazão própria muito comum neste tipo de usina. Além deste ponto, a escolha das vazões vizinhas após análise espacial melhora a capacidade preditiva dos modelos quando comparados a abordagem da correlação da vazão com a geração.

Os modelos de regressão dinâmica se mostraram mais eficientes que os modelos de função de transferência pois esse modelo permitiu a inclusão da variável explicativa defasada, além da parte autoregressiva da variável de resposta. Quando comparados com os modelos de função de transferência os resultados melhoraram. Apesar do resultado satisfatório e esperado da utilização da vazão própria nos modelos causais, de acordo com a literatura, o número de observações é considerada pequena para capturar adequadamente o comportamento sazonal das séries.

Pode-se dizer que o objetivo proposto nesta dissertação foi atingido, apresentando alternativas para o problema apresentado pela distribuidora de energia relacionado a apresentação de modelos de previsão de geração que possuam um desempenho adequado para permitir um posterior cálculo de perdas técnicas mais adequado a realidade estocástica desse tipo de usina.

Como proposta para continuação deste trabalho uma evolução natural para o apresentado até o momento seria a utilização de um modelo espaço-temporal multivariado, aplicando os lags espaciais além dos temporais em um modelo de regressão dinâmica. Outra abordagem a ser explorada seria a utilização de outros modelos utilizando aprendizado de máquina, além de combinação de previsões para aumentar a assertividade da previsão.

Referências Bibliográficas

ALMEIDA, E. **Econometria Espacial Aplicada**. Alínea Editora, Campinas, 2012.

ABRAPCH. O que são PCHs e CGHs. **Associação Brasileira de Pequenas Centrais Hidrelétricas e Centrais Geradoras Hidrelétricas**. 2016. Disponível em: [.http://www.abrapch.org.br/pchs/o-que-sao-pchs-e-cghs](http://www.abrapch.org.br/pchs/o-que-sao-pchs-e-cghs). Acesso em: 15 jun. 2018.

ANEEL. 2017. PRODIST. **Agência Nacional de Energia Elétrica**. <http://www.aneel.gov.br/prodist>. Acessado em 25 jan. 2019.

ANEEL. 2018. Metodologia de Cálculo Tarifário da Distribuição. **Agência Nacional de Energia Elétrica**. Disponível em: http://www.aneel.gov.br/metodologia-distribuicao/-/asset_publisher/e2INtBH4EC4e/content/perdas/654800?inheritRedirect=false. Acesso em: 24 jan. 2019.

ANSELIN, L.. Local Indicators of Spatial Association — LISA. **Geographical Analysis**, 27: 93-115, 1995.

AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, 19:716-723, 1974.

CABRAL, J.A., LEGEY L.F. L., CABRAL M. V. F.. Electricity consumption forecasting in Brazil: A spatial econometrics approach, **Energy**, Volume 126, 2017, p. 124-131, 2017.

BORGES, Carmen L.T., PINTO, Roberto J. Small Hydro Power Plants Energy Availability Modeling for Generation Reliability Evaluation. **IEEE Transactions on Power Systems**, Vol. 23, no. 3, August 2008.

BOX, G.E.P., JENKINS, G.M., **Time Series Analysis, Forecasting and Control**. Holden Day, San Francisco. 1970.

BOX, G. E. P., JENKINS, G. M., REINSEL, G. C., LJUNG, G. M. **Time Series Analysis : Forecasting and Control** - Fifth Edition. Wiley Series in Probability and Statistics. John Wiley & Sons, New Jersey. 2016.

CEPEL. NEWAVE - Modelo de Planejamento da Operação de Sistemas Hidrotérmicos Interligados de Longo e Médio Prazo. 2018. Disponível em: http://www.cepel.br/pt_br/produtos/newave-modelo-de-planejamento-da-operacao-de-sistemas-hidrotermicos-interligados-de-longo-e-medio-prazo.htm. Acesso em: 13 jul 2019.

CLIFF, A.D.,HAGGETT, P.,ORD, J.K., BASSETT, K., DAVIES, R. Elements of Spatial Structure – A Quantitative Approach. **Cambridge University Press**. 1975.

DICKEY, D. A. ; FULLER, W. A. Distribution of the estimators for autoregressive time series with the unit root. **Journal of the American Statistical Association**, v.74, n.366a, p.427-431,1979.

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

EPE (BRASIL). Plano Nacional de Energia – 2030 - PNE 2030 - Geração Hidrelétrica. 2006. Disponível em: <http://www.epe.gov.br/pt/publicacoes-dados-abertos/publicacoes/Plano-Nacional-de-Energia-PNE-2030>. Acesso em: jan. 2019.

EPE (BRASIL). Balanço Energético Nacional 2018: Ano base 2017 / **Empresa de Pesquisa Energética**. – Rio de Janeiro: EPE, 2018. Disponível em : http://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-303/topico-419/BEN2018__Int.pdf . Acesso em : 21 jan. 2018.

FERREIRA, P. G. C.; OLIVEIRA, F. L. C.; SOUZA, R. C. The stochastic effects on the Brazilian Electrical Sector. **Energy Economics**, 2015, 49: 328-335.

FERREIRA, P. G. C. **Análise de Séries Temporais em R: curso introdutório**. 1. ed. Rio de Janeiro: Elsevier: FGV IBRE, 2018.

FOLEY, R. **Correlation between the precipitation and energy production at hydropower plants to mitigate flooding in the Missouri River Basin**. Thesis (S.B.)--Massachusetts Institute of Technology, Department of Mechanical Engineering, 2013.

GOOGLE. **Google Maps**. Disponível em: <https://www.google.com/maps>. Acesso em: jan. 2019.

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

GUJARATI, D. N.. **Econometria Básica**. 3ª ed. São Paulo: Pearson Education do Brasil, 2000.

HOLT, C. C.. Forecasting seasonals and trends by exponentially weighted moving averages. **International Journal of Forecasting**, 20, 5 – 10, 2004.

HYNDMAN, R., ATHANASOPOULOS, G.. **Forecasting: principles and practice**. 2018. Disponível em: <http://otexts.com/fpp2/>. Acesso em: 18 dez. 2018.

HYNDMAN, R. J., KHANDAKAR, Y.. Automatic time series forecasting: The forecast package for R. **Journal of Statistical Software**, 27(1), 1–22, 2008.

LI, G., LIU, C. X., LIAO, S. L., CHENG, C. T.. Applying a correlation analysis method to long-term forecasting of power production at small hydropower plants. **Water**, 7(9), 4806-4820, 2015.

JENKINS, G. **Practical Experiences with Modeling and Forecasting Time Series**. St. Helier: Gwilym Jenkins and Partners (Overseas) Ltd. 1979.

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

LIU, P.; ZHANG, H.; ZHANG, J.; SUN, Y.; QIU, M.. Spatial-temporal response patterns of tourist flow under impulse pre-trip information search: From online to arrival. **Tourism Management**, Volume 73, p. 105-114, 2019.

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

LOPES, V. S., BORGES, C. L. T. 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, July 2015.

MAÇAIRA, P. M.; CYRINO OLIVEIRA, F. L.; FERREIRA, P. G. C.; ALMEIDA, F. V. N.; SOUZA, R. C. Introducing a causal PAR(p) model to evaluate the influence of climate variables in reservoir inflows: a Brazilian case. **Pesquisa Operacional** (Impresso), v. 37, p. 107-128. 2017.

MAÇAIRA, P. M., THOMÉ, A.M.T., OLIVEIRA, F. L. C, FERRER, A. L. C. Time series analysis with explanatory variables: A systematic literature review. **Environmental Modelling & Software**, Volume 107, Sep. 2018, Pages 199-209, 2018.

MORETTIN, P. A.; TOLOI, C. M. C.. **Análise de Séries Temporais. Volume 1 – Modelos lineares univariados**. 3ª ed. São Paulo: Blucher, 474p, 2018.

OLIVEIRA, F. L C.; FERREIRA, P. G. C.; SOUZA, R. C. A parsimonious bootstrap method to model natural inflow energy series. **Mathematical Problems in Engineering**, 2014.

ONS. O que é o SIN. **Operador Nacional do Sistema Elétrico**. Disponível em: <http://ons.org.br/paginas/sobre-o-sin/o-que-e-o-sin..> Acesso em 25 abr. 2018.

ONS. Resultados da Operação – Histórico da Operação. **Operador Nacional do Sistema Elétrico**. Disponível em : <http://ons.org.br/paginas/resultados-da-operacao/historico-da-operacao>. Acesso em: 10 mai. 2019a.

ONS. Mapas – Hidroelétricas do SIN - Rede de Operação - Horizonte 2023. **Operador Nacional do Sistema Elétrico**. Disponível em : <http://www.ons.org.br/paginas/sobre-o-sin/mapas>. Acesso em: mar. 2019b.

PHILLIPS, P. C. ; PERRON, P. Testing for a unit root in time series regression. **Biometrika**,v.75, n. 2, p.335-346, 1988.

QUEIROZ, A. R. ; FARIA, V. A. D. ;LIMA, L. M. M. ; LIMA, J. W. M. Hydropower revenues under the threat of climate change in Brazil. **Renewable Energy**, 133, 873-882. 2019.

R Core Team. **The R Project for Statistical Computing**. 2015. Disponível em: <https://www.r-project.org/>. Acesso em: 20 mar. 2018.

SILVA, J. M., VASCONCELLOS B. T. C., BRAGA I. R..Correlação Entre o Índice Padronizado de Precipitação e Geração de Centrais Hidrelétricas de Pequeno Porte. **Revista Brasileira de Energias Renováveis**. v. 5, n. 3 (2016).

SNIRH. Sistema Hidro – Telemetria. Sistema Nacional de Informações sobre Recursos Hídricos. 2019. Disponível em: <http://www.snirh.gov.br/hidrotelemetria/Mapa.aspx>. Acesso em: 25 mar. 2019.

SOUZA, R. C., CAMARGO, M. E., **Análise e previsão de séries temporais: Os modelos ARIMA**. 2ª ed. 2004.

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

VRIEZE, S. Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). **Psychological Methods**, 2012.

WINTERS, P. R. Forecasting sales by exponentially weighted moving averages. **Management Science**, 6, 324– 342, 1960.

YANG, Y; ZHANG, H. Spatial-temporal forecasting of tourism demand. **Annals of Tourism Research**, Vol. 75, p. 106-119, 2019.