Comissão Permanente para Análise de Metodologias e Programas Computacionais do Setor Elétrico

GT Metodologia

APRIMORAMENTO NA REPRESENTAÇÃO DA PRODUTIBILIDADE HIDROELÉTRICA E PERDAS HIDRÁULICAS NO PLANEJAMENTO DA OPERAÇÃO ENERGÉTICA DE CURTO PRAZO

Relatório Técnico do GT Metodologia da CPAMP – nº 03-2020

Assessoria técnica

Sumário

1. Resumo executivo

O Grupo de Trabalho para Avaliação dos Dados Cadastrais – GTDP finalizou em 2019 seu primeiro ciclo de trabalhos, culminando na atualização dos valores constantes de produtibilidade específica e perda de carga de diversas usinas contemplados pelos modelos da cadeia de planejamento energético NEWAVE, DECOMP e DESSEM.

O presente trabalho consiste no desenvolvimento de uma metodologia para determinação de representações variáveis destes dois parâmetros, por usina, a partir de resultados intermediários do GTDP. Esta nova representação será incorporada ao modelo DECOMP, especificamente no cálculo da Função de Produção Hidrelétrica Exata e na modelagem da função de produção realizada pelo modelo, substituindo os parâmetros até então constantes por funções da condição operativa de cada usina. O processo que será descrito ao longo deste relatório pode ser resumidamente representado pelos blocos verdes no diagrama apresentado na Figura 1.1.

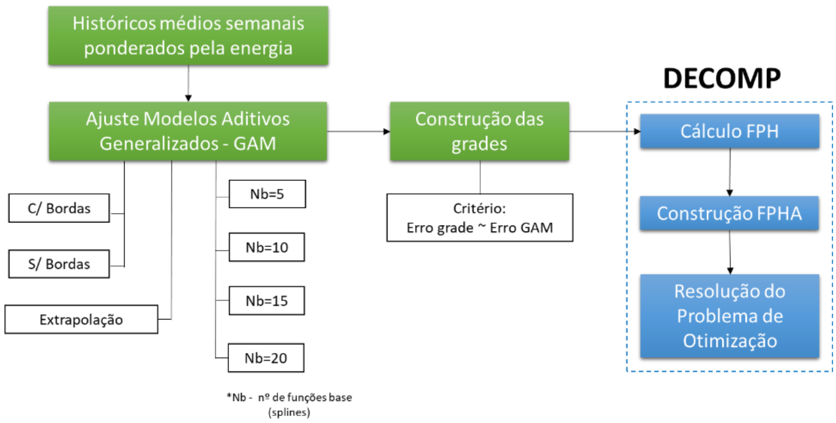

Figura 1.1 – Processo para representação variável da produtibilidade específica e perda hidráulica

Como resultado da aplicação desta nova representação via grade de valores, observou-se valores de produtibilidade específica e perda de carga (perda hidráulica) mais próximos aos valores observados na operação real.

2. Apresentação

Este relatório está inserido no contexto do Grupo de Trabalho de Metodologia da CPAMP – Comissão Permanente para Análise de Metodologias e Programas Computacionais do Setor Elétrico, criada pela Resolução CNPE nº 01/2007 e regulamentada pela Portaria MME n° 47/2008, com a finalidade de garantir coerência e integração das metodologias e programas computacionais utilizados pelas instituições e agentes.

O Grupo de Metodologia da CPAMP é coordenado pela CCEE (representada pela Gerência Executiva de Preços, Modelos e Estudos Energéticos - GEPME) e conta com a participação do MME (representado pelas Secretarias de Energia Elétrica – SEE, Secretaria de Planejamento e Desenvolvimento Energético – SPE e Assessoria Econômica - Assec), da ANEEL (representada pela Superintendência de Regulação da Geração – SRG), do ONS (representado pelas Gerências de Planejamento Energético - PE e de Programação - PR) e da EPE (representada pela Assessoria da Presidência e Superintendência de Planejamento da Geração – SGE). O grupo possui, ainda, a assessoria técnica do CEPEL (representado pelo Departamento de Otimização Energética e Meio Ambiente).

3. Introdução

Em outubro de 2019 foi concluído o primeiro ciclo de estudos do Grupo de Trabalho para Revisão dos Dados Cadastrais Utilizados para Cálculo da Produtibilidade Hidrelétrica – GTDP (NT 0103-2019 - Consolidação das Atividades Realizadas pelo GTDP – Grupo de Trabalho de Avaliação dos Dados Cadastrais), cujo foco era a atualização dos parâmetros de modelagem das usinas hidrelétricas nos modelos de otimização empregados no planejamento energético de curto, médio e longo prazo.

Dentre os resultados contemplados neste primeiro ciclo de estudos estão as perdas de carga hidráulica do circuito e a produtibilidade específica médias das usinas. Embora essas grandezas tenham sido agora recalculadas através de metodologias detalhadas, pautadas em dados operativos históricos e informações criteriosas acerca das curvas colina das turbinas, dos rendimentos de geradores e dos arranjos dos circuitos hidráulicos específicos de cada usina, a consideração destas grandezas como sendo um único valor médio representativo constante, indiferentes às condições operativas de altura de queda e turbinamento, constitui uma simplificação.

Em vista deste aspecto, após a conclusão do primeiro ciclo de estudos do GTDP foi proposto um novo aprimoramento com o objetivo de não somente aumentar a acurácia dos parâmetros de perdas hidráulicas e produtibilidade específica, mas a própria forma como estas grandezas são informadas aos modelos utilizados no planejamento da operação energética. Dadas as características naturais de ambos os parâmetros, foi proposta a representação das perdas hidráulicas em função da vazão turbinada e a produtibilidade específica em função tanto da vazão turbinada, quanto da altura de queda líquida.

Esta nova forma de modelagem se concentra apenas no modelo DECOMP, desenvolvido para aplicação no horizonte de curto prazo, especificamente no cálculo da função de produção energética de cada usina. O modelo NEWAVE de médio/longo prazo tem como premissa que seus reservatórios equivalentes não permitem o monitoramento das condições operativas por usina, impedindo o cálculo das perdas de carga e produtibilidades específicas variáveis. Desta forma, para este modelo, em sua forma de utilização oficial no momento, será mantida a representação através de valores constantes. Posteriormente, esta modelagem poderá ser estendida para a versão já oficialmente utilizada no modelo DESSEM, assim como para o modelo NEWAVE em sua forma “híbrida”. para os períodos em que as usinas forem representadas de forma individualizada

Devido à natureza das atividades propostas, visando o aprimoramento de metodologias atualmente em uso nos modelos de otimização, o presente estudo foi alocado no âmbito da Comissão Permanente para Análise de Metodologias e Programas Computacionais do Setor Elétrico – CPAMP, no Grupo de Trabalho de Metodologias onde foi então criado o Subgrupo de Representação da Produtibilidade Hidroelétrica.

O objetivo do presente relatório é apresentar a metodologia desenvolvida para a representação das produtibilidades específicas e perdas variáveis através de grades, a definição dos parâmetros envolvidos no processo, os principais resultados obtidos durante o procedimento de ajuste das grades e uma análise preliminar da utilização das grades no mdelo DECOMP. Uma avaliação mais aprofundada dos resultados no modelo DECOMP será objeto do segundo relatório do Subgrupo de Produtibilidade do GT Metodologia.

4. Modelagem vigente da função de produção hidrelétrica no modelo DECOMP

Esta seção apresenta um resumo de como o modelo DECOMP considera a função de produção das usinas hidrelétricas na versão vigente oficialmente (30.1), a partir de alguns documentos e publicações [2-5] produzidos pelo CEPEL

A consideração da variação da produtividade das usinas hidrelétricas com a altura de queda é feita por meio da seguinte expressão:

$$FPH_i(V,Q,S) = \rho_i\, Q\left[h_{mon\,i}(V) - h_{jus_i}(Q,S)\right] k_{perdas_i}, \tag{4.1a}$$

$$FPH_i(V,Q,S) = \rho_i\, Q\left[h_{mon\,i}(V) - h_{jus_i}(Q,S) - k_{perdas_i}\right], \tag{4.1b}$$

onde as variáveis de decisão do modelo são o volume armazenado V, a vazão turbinada Q, e a vazão vertida S. Os parâmetros (dados de entrada) são:

ρ_i: produtibilidade específica, que é composta pelo produto entre o rendimento médio do conjunto turbina-gerador (considerado constante), a densidade da água, a aceleração da gravidade e um fator de conversão de unidades;

k_{perdas_i}: perdas nos condutos, que pode ser dada em p.u. (como na expressão 4.1a) ou, de forma alternativa, em metros (como na expressão 4.b).

Com base nessa expressão não linear, o modelo DECOMP constrói um modelo dessa função, denominado "Função de produção hidrelétrica aproximada (FPHA)", que é composto das seguintes etapas:

- determinação de uma grade de discretização para os valores de V e Q;
- cálculo da função de produção exata (FPH) em todos os pontos, dada por (4.1);
- cálculo de uma envoltória convexa para os pontos abaixo desta função exata, e construção de um modelo inicial linear por partes ($FPHA_0$) em V e Q;
- Regressão para minimização das diferenças entre a FPH e $FPHA_0$;
- Aproximação secante para a função no eixo do vertimento.

Desta forma, obtem-se um modelo linear por partes em 4 dimensões para essa função, denominado de FPHA, que é dada pelo conjunto de inequações (4.2):

$$(FPHA_i) \qquad \begin{cases} GH_i^t \leq {\gamma_0}_i^k + {\gamma_V}_i^k V_i^t + {\gamma_Q}_i^k Q_i^t + {\gamma_S}_i^k S_i^t \\ \qquad t = 1,\dots,T, k = 1,\dots,NPF_i \end{cases}, \tag{4.2}$$

onde ${\gamma_0}_i^k$, ${\gamma_V}_i^k$, ${\gamma_Q}_i^k$ e ${\gamma_S}_i^k$ são os coeficientes dessa função para cada corte k da usina i.

5. Dados Utilizados

Conforme mencionado anteriormente, parte das usinas atualmente em operação que participaram do primeiro ciclo de estudos do GTDP tiveram seus parâmetros de perdas de carga e produtibilidades específicas recalculados e atualizados. Para obtenção destes novos valores de perdas de carga e produtibilidades específicas foram utilizados dados históricos operativos em base horária de queda bruta e geração; equação de perdas de carga hidráulica do circuito; curva colina das turbinas; rendimento dos geradores; massa específica da água e aceleração da gravidade locais. Os valores finais são então obtidos através da média dos resultados horários a cada unidade geradora, agregados para a usina e ponderados pela energia gerada. A Figura 5.1 exemplifica estes resultados.

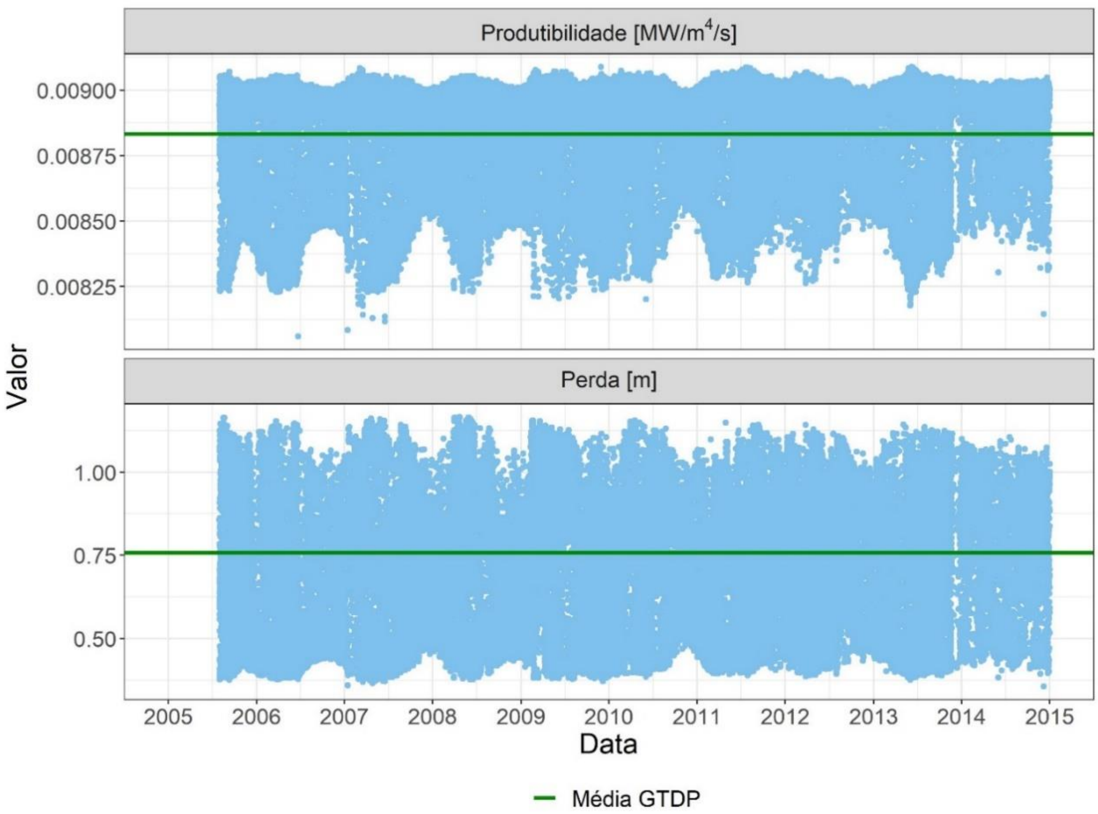

Figura 5.1 - Ilustração do cálculo de produtibilidade específica e perdas de carga médias.

Enquanto se restringe a representação destas duas grandezas através de constantes, a média ponderada obtida pelo GTDP é suficientemente precisa. A observação destes resultados, no entanto, deixa evidente o quão limitante é a redução destes valores para constantes, sendo clara a grande faixa de variação de ambos ao longo do tempo.

Uma proposta natural para cálculo variável das perdas de carga e produtibilidades específicas de acordo com a condição operativa seria o uso das mesmas equações de perdas de carga e curvas colina contempladas nos estudos do GTDP. Mas para que isso fosse possível, seria necessária modelagem de cada unidade geradora individualmente numa usina, dado que a equação de perdas e a curva colina são específicas de cada circuito hidráulico e de cada turbina.

O modelo DECOMP não enxerga unidades geradoras individuais, apenas a usina como um bloco único de produção energética. Essa abstração das unidades impossibilita o uso direto dos parâmetros construtivos em conjunto como a vazão turbinada total para cálculo dos valores médios.

5.1. Comportamento a cada unidade geradora e por usina

As perdas de carga hidráulica por unidade geradora são proporcionais ao quadrado da vazão turbinada, sendo determinada tipicamente em circuitos hidráulicos de adução e de saída individuais para cada turbina por uma equação da forma apresentada em (5.1).

$$H_{perda} = k \times Q_{turb.}^2 \tag{5.1}$$

No caso de usinas com circuitos hidráulicos de adução e de saída compartilhados para um grupo gerador a função de perdas assuma formas mais complexas, por exemplo quando há um trecho de adução compartilhado entre múltiplas máquinas, porém dadas as vazões circulando por cada turbina, é possível determinar todas as perdas com exatidão.

Ao modelar as usinas como blocos singulares de geração a perda de carga deixa de seguir uma forma exata, mesmo que todas suas máquinas possuam a mesma equação de perdas de carga. A Figura 5.2 apresenta um exemplo de perdas de carga médias de uma usina com três unidades geradoras alimentadas por circuitos hidráulicos idênticos (portanto mesma equação).

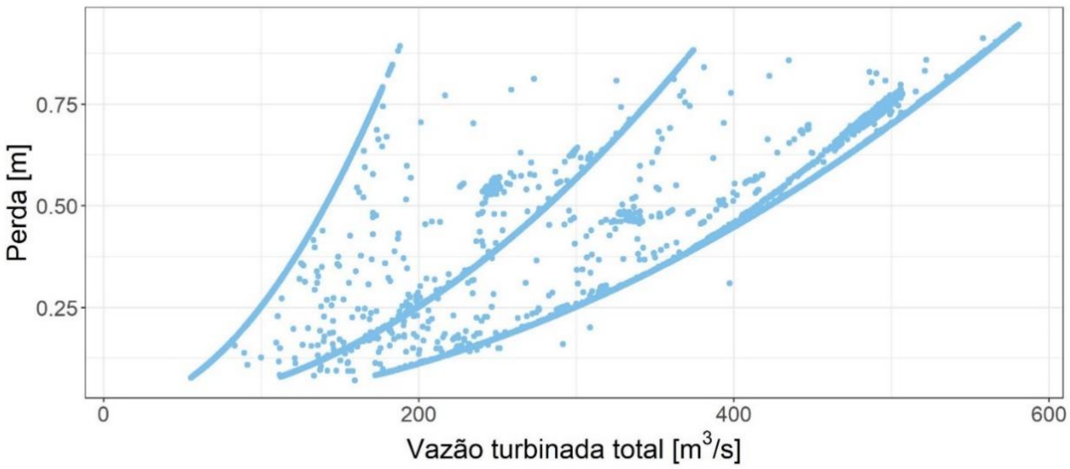

Figura 5.2 - Perdas de carga médias para uma usina de três unidades geradoras.

É notável a existência de três curvas bem definidas correspondentes à usina operando de uma a três unidades quando a vazão turbinada total pela usina é distribuída igualmente entre as máquinas. Quando a usina está em condições operativas onde cada unidade está turbinando vazões distintas os registros indicam os pontos dispersos entre cada duas curvas.

Outra característica digna de atenção é o compartilhamento de domínio entre curvas. Um mesmo valor de vazão total pode ser repartido entre as três unidades geradoras de infinitas formas, correspondendo a infinitos valores de perda média.

A frequência de pontos correspondentes à repartição assimétrica de vazões e mistura entre domínios das curvas para cada número de unidades são exacerbadas conforme o aumento no número de unidades geradoras da usina, conforme explicitado na Figura 5.3.

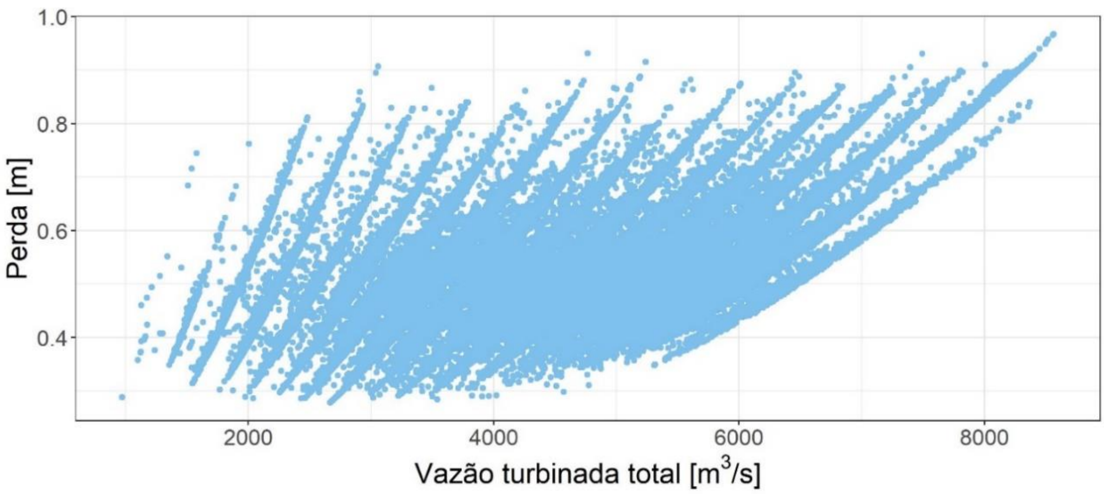

Figura 5.3 - Perdas médias numa usina de grande porte.

A distinção entre o número de unidades em operação, claramente notável no caso anterior, desaparece, mesmo em vazões relativamente baixas para a usina.

Fazendo uso da curva colina, é possível determinar com exatidão o rendimento de uma turbina numa dada vazão de engolimento e altura de queda líquida, e assim, determinar a produtibilidade para uma máquina. Tal como a perda de carga média, a produtibilidade total da usina não é diretamente calculável a partir da vazão turbinada total.

A Figura 5.4 apresenta as produtibilidades específicas médias da mesma usina com três unidades geradoras. Novamente é possível notar três curvas distintas, porém ainda menos bem definidas visualmente se comparadas ao caso das perdas.

Estas análises deixam claro que embora as perdas de carga e produtibilidades específicas sejam calculáveis a cada unidade geradora, não existe uma relação teórica para determinação do valor médio. Para que fosse possível utilizar as equações de perdas de carga e curvas colina, seria necessário conhecer de antemão a estratégia de repartição de carga entre as unidades geradoras e que essa regra fosse aplicada invariavelmente, o que não é viável e não acontece no dia a dia da operação. A forma como cada unidade geradora é operada depende de uma série de fatores, tais como: a queda disponível, a carga a ser tomada nas próximas horas, e a necessidade de armazenamento.

Uma possibilidade seria desenvolver as representações variáveis dessas grandezas com base nos valores ótimos de perdas de carga e produtibilidade específica para uma dada vazão total pois, ao contrário dos valores médios, estes são calculáveis de forma teórica. Esta abordagem, no entanto, constituiria uma superestimação da capacidade de produção energética pelo ponto de vista dos modelos de otimização. A repartição de vazão que leva às perdas de carga e produtibilidade específica ótima é uma condição muito particular, inconsistente com os outros fatores requeridos, mencionados no parágrafo anterior.

A utilização dos resultados operativos obtidos pelo GTDP se torna, assim, a melhor alternativa. Embora estas perdas de carga e produtibilidades específicas sejam consequência de uma programação e despacho determinados pelo ONS, o processo atravessa inúmeras etapas visando a melhor utilização do recurso natural. Desta forma os valores dos resultados operativos obtidos podem ser considerados representativos da capacidade de geração das usinas no horizonte de tempo considerado no primeiro ciclo do estudo.

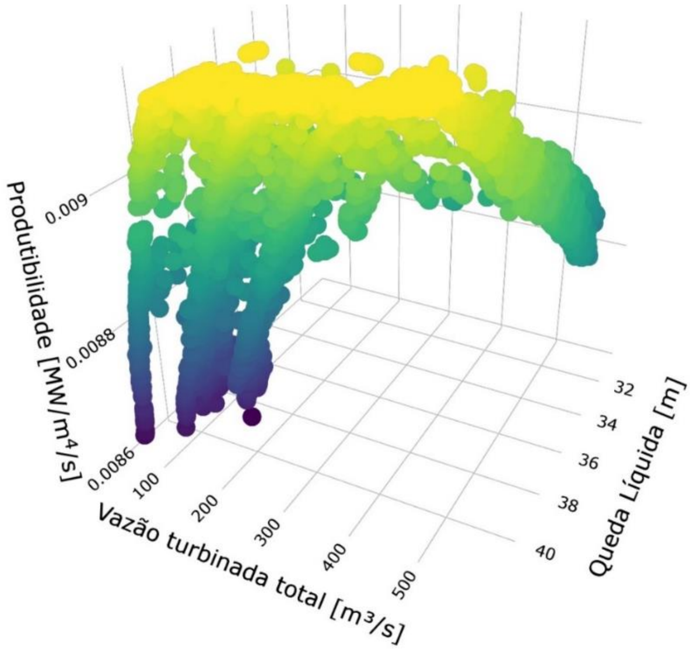

Figura 5.4 - Produtibilidades específicas médias para uma usina de três unidades geradoras.

5.2. Utilização de médias semanais

Os dados obtidos e apresentados na seção anterior foram obtidos em base horária. Entretanto, a discretização temporal do modelo DECOMP é semanal no primeiro mês e mensal no segundo mês, ou seja, o modelo tem por objetivo fornecer como resultado o despacho médio de cada usina ao longo da semana ou mês, em três patamares de carga. Assim, considera-se que, tanto para a curva de perdas nos condutos em função da vazão como para a curva de produtividade específica em função da vazão e da altura de queda, a abordagem de ajuste pelos dados médios semanais é a mais apropriada, para ficar aderente ao período de discretização do DECOMP.

Além disso, uso de dados horários (curva instantânea) ao invés dos dados médios semanais causariam uma subestimativa do valor das perdas nos condutos, pela natureza convexa desta curva. Isto é mostrado na Figura 5.5, que mostra os valores originais horários (pontos azuis) e as médias semanais obtidas (pontos vermelhos) a partir dos dados horários.

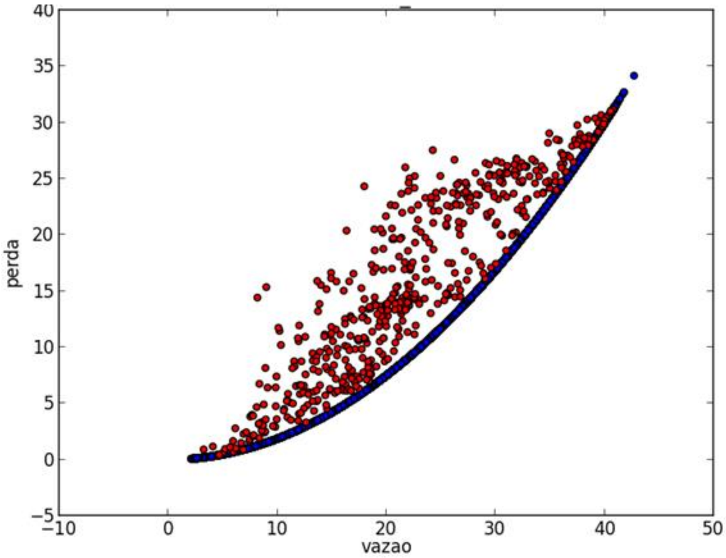

Figura 5.5 – Perdas nos condutos em funçaõ da vazão para determinada usina, considerando os dados originais horários (pontos azuis) e as médias semanais (pontos vermelhos).

Para o caso da produtibilidade específica, o uso de valores horários ocasionaria um efeito inverso, ou seja, poderia superestimar os valores das produtibilidades específicas. Isso é ilustrado na Figura 5.6, que mostra à esquerda os dados originais horários (pontos azuis) e à direita as médias semanais (pontos vermelhos).

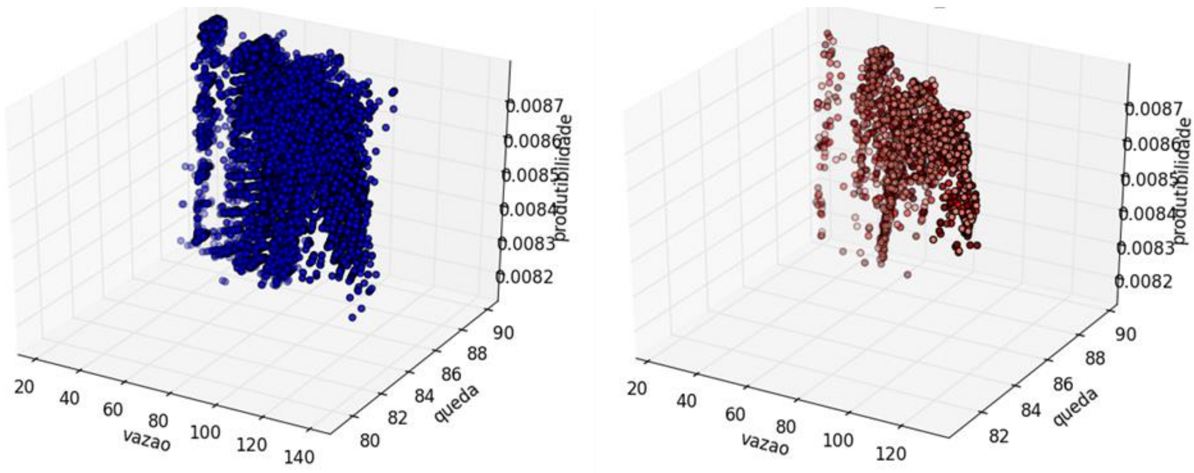

Figura 5.6 – Produtibilidade específica em função da vazão e da altura de queda, considerando os dados originais horários (pontos azuis, à esuqerda) e as médias semanais (pontos vermelhos, à direita).

O relatório técnico [8] analisa com mais detalhes essa questão, e sugere-se as referência [5] sobre o uso de médias semanais para procedimentos deste tipo.

6. Metodologia de Representação Variável

De acordo com os motivos expostos na seção 5.1 tomou-se de base a utilização dos resultados calculados pelo GTDP agregados em valores médios semanais, de modo a ficar aderente às premissas do modelo DECOMP. A análise dos gráficos expostos na seção 5.1 evidencia os comportamentos complexos tanto das perdas de carga quanto da produtibilidade específica, não sendo possível representá-los através de funções simples, como do tipo polinomial. A abordagem adotada consiste no ajuste destes dados com auxílio de Modelos Aditivos Generalizados (*Generalized Additive Models* – GAM), uma forma de modelagem flexível capaz de capturar as nuances dos comportamentos dessas duas grandezas. O ajuste via GAM é então discretizado numa grade de pontos a qual o modelo DECOMP toma como entrada e interpola os pontos necessários.

A decisão pelo uso de grades, apesar das funções ajustadas via GAM possuírem forma paramétrica (como será detalhado na seção a seguir), foi tomada em função de alguns fatores. Primeiramente, estas funções são extremamente complexas, compostas de inúmeros parâmetros, mesmo no caso mais simples univariado. Em segundo lugar, também a demostrado na sequência neste documento, no ajuste das funções de perdas de carga é necessário realizar extrapolações nos tramos inferior e superior das regiões do domínio sem dados observados, onde o ajuste via GAM não é fisicamente viável. Assim, além da complexidade das próprias funções, seria ainda necessário discretizar regiões de aplicabilidade de cada uma.

Visando aumentar a interpretabilidade das formas variáveis de perdas de carga e produtibilidade específica facilitando possíveis alterações para vias de teste ou simulação de cenários, optou-se pela redução dos ajustes a pontos discretos, entre os quais o modelo DECOMP interpola para obtenção dos valores buscados. Pela própria natureza de cada grandeza, as perdas de carga são representadas através de uma grade unidimensional, em função da vazão turbinada, e a produtibilidade específica através de uma grade bidimensional, em função da vazão turbinada e da altura de queda líquida, ilustradas na Figura 6.1.

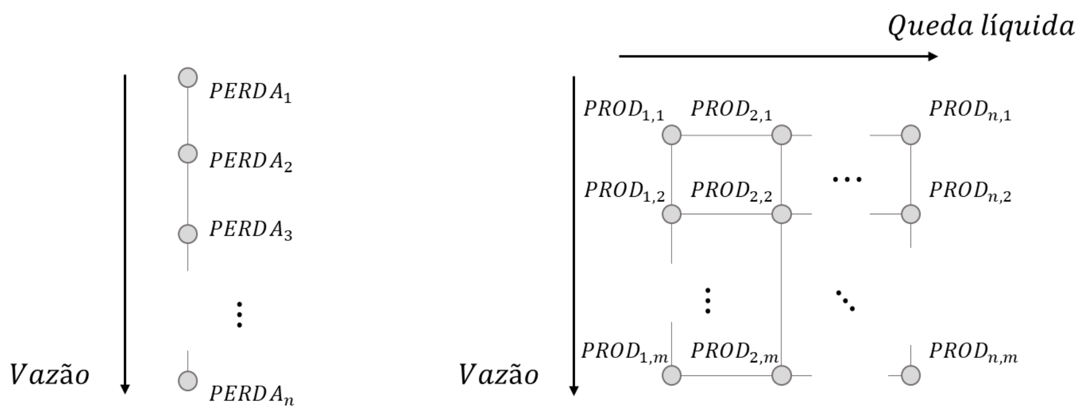

Figura 6.1 - Ilustração das grades de perdas de carga e produtibilidade específica.

O domínio coberto por cada grade é um aspecto que demanda atenção. Ao se utilizar valores constantes, para qualquer vazão turbinada e altura de queda líquida tem-se uma única perda de carga e produtibilidade específica para todo o domínio operativo. Já no caso da representação variável desenvolvida no presente estudo, é necessário ter cautela com as faixas de vazão turbinada e queda líquida cobertas por cada grade, pois como são oriundas de resultados operativos condicionados ao horizonte de tempo do primeiro ciclo do estudo do GTDP (2005-2014), deve-se garantir que toda possível condição operativa indicada pela curva colina esteja contemplada na grade e que possa ser demandada pelo modelo DECOMP.

Foram estipulados limites inferior e superior de vazão turbinada total e altura de queda líquida necessários para as grades de todas as usinas. A vazão turbinada sempre é representada na faixa desde seu engolimento nulo ao máximo engolimento da usina, calculado como o máximo entre a soma das vazões efetivas de cada turbina (cadastradas no arquivo HIDR.dat) e o máximo observado no histórico operativo. O modelo DECOMP calcula a vazão turbinada máxima de cada usina a partir do mesmo dado cadastral, portanto fica garantida a cobertura de todos os valores possíveis para o modelo.

Os valores extremos de queda líquida são determinados através da comparação dos limites da curva colina com correspondentes mínimo e máximo observados no histórico. Como não existem dados cadastrais em respeito à faixa de alturas queda, é possível que o modelo busque valores fora dos limites estipulados. Por outro lado, é importante notar que os limites advindos da curva colina usualmente já ultrapassam a faixa operativa normal de cada turbina. Assim, o planejamento da operação que visite tais pontos será inevitavelmente alterado no momento da operação, de modo que deveriam ser evitados no modelo.

Na seção seguinte será realizada uma exposição teórica detalhada quanto aos Modelos Aditivos Generalizados. Em seguida, será tratada a forma como as grades são produzidas, passando pelo ajuste dos dados, a forma de amostragem para sua extração e procedimentos auxiliares eventualmente necessários.

6.1. Modelos Aditivos Generalizados

Os modelos aditivos generalizados ou GAM (Generalized Additive Models) [6] formam uma família de modelos entre os quais se destaca o modelo aditivo, um modelo de regressão com um preditor definido pela soma de funções suaves das variáveis explicativas $x_1,\ldots,x_p$, conforme indicado em (6.1).

$$y = f_1(x_1) + \cdots + f_p(x_p) + \varepsilon \qquad (6.1)$$

onde y é a variável dependente, $f_j(x_j)$ $\forall j=1,p$ são funções das variáveis explicativas e ε é um ruído branco.

Por exemplo, para o caso da modelagem das perdas (y), em função da vazão turbinada (x), tem-se que p=1, logo, (6.1) pode ser escrita como (6.2).

$$y = f_1(x_1) + \varepsilon \qquad (6.2)$$

Já para o caso da modelagem da produtibilidade de uma usina (y), em função da vazão turbinada (x_1) e da queda líquida (x_2), tem-se p=2 e a especificação dada em (6.3).

$$y = f_1(x_1) + f_2(x_2) + \varepsilon \qquad (6.3)$$

Cada função $f_j(x_j)$ pode ser representada por meio de bases de funções suaves. Por exemplo, conforme indicado a seguir, as funções $f_1(x_1)$ e $f_2(x_2)$ são definidas por uma base de *splines* cúbicas $h(x,x^*)$, de acordo com (6.4) e (6.5).

$$f_1(x_1) = \delta_1 + \delta_2 x_1 + \sum_{j=1}^{q_1-2} \delta_{j+2} h_j\left(x_1, x_{1,j}^*\right) \qquad (6.4)$$

$$f_2(x_2) = \mathrm{q}_1 + \mathrm{q}_2 x_2 + \sum_{j=1}^{q_2-2} \mathrm{q}_{j+2} h_j(x_2, x_{2,j}^*) \quad (6.5)$$

em que δ_j $\forall j$=1,q_1 e q_j $\forall j$=1,q_2 denotam os q_1+q_2 parâmetros a serem estimados, enquanto x_1^* e x_2^* denotam as localizações dos nós das funções suaves $h(x,x^*)$.

Uma função polinomial $f(x)$ por partes é obtida dividindo o domínio de x em intervalos contíguos e representando a função $f(x)$, separadamente em cada parte [9], por uma função suave, cujo conjunto forma uma base de funções $h(x,x^*)$. Os pontos x^* que delimitam os intervalos são denominados nós. Os nós podem ser distribuídos uniformemente ou serem posicionados nos quantis das variáveis explicativas. Independentemente da posição e quantidade de nós, em uma base de funções $h(x,x^*)$, formada por *splines* cúbicas, a curva resultante é contínua até a segunda derivada. Por exemplo, para uma base de funções $h(x,x^*)$ formada por q+1 *splines* cúblicas (B-*Splines* de ordem m=3 ou *cubic B-Splines*), o ajuste de um modelo aditivo requer a definição de q+m+2 nós e suas respectivas posições , logo:

$$f(x) = \sum_{j=0}^{q} \mathrm{b}_j B_j^m(x, x_j^*) \quad (6.6)$$

em que β representa o vetor de coeficientes de regressão e as funções $B_j^m(x)$ denotam as bases B-Splines de ordem m, definidas recursivamente, conforme indicado em (6.7).

$$B_j^m(x) = \frac{(x - x_j^*)}{(x_{j+m}^* - x_j^*)} B_j^{m-1}(x) + \frac{(x_{j+m+1}^* - x)}{(x_{j+m+1}^* - x_{j+1}^*)} B_{j+1}^{m-1}(x) \ \forall \mathrm{j}=1,\ldots,q \quad (6.7)$$

Para m =0, $B_j^0(x) = 1$ se $x_j^* \le x \le x_{j+1}^*$, caso contrário $B_j^0(x) = 0$ [11].

O pacote "mgcv" do R disponibiliza as seguintes funções suaves para ajuste de modelos GAM [6]:

- *thin plate regression splines* (TPRS)
- TPRS *shrinkage*
- *cubic regression splines* (CRS)
- CRS *shrinkage*
- *Cyclic* CRS
- P-*splines*

Adicionalmente, note que os parâmetros δ_1 e q_1 em (6.4) e (6.5) são confundidos, logo para evitar o problema de identificação dos parâmetros deve-se fazer q_1=0. Assim, o modelo aditivo em (6.3) pode ser escrito segundo uma especificação linear, i.e., y=X^Tβ+ε, na qual β^T é o vetor de parâmetros $(\delta_1, \ldots, \delta_{\mathrm{q}_1}, \theta_2, \ldots, \theta_{q_2})$ e X é o vetor com as funções suaves das variáveis explicativas, de acordo com (6.8).

$$X = [1, x_1, h_1(x_1, x_{1,1}^*), \ldots, h_{q_1}(x_1, x_{1,q_1-2}^*), x_2, h_2(x_2, x_{2,1}^*), \ldots, h_{q_2}(x_2, x_{2,q_2-2}^*)] \quad (6.8)$$

A partir de n amostras das variáveis dependente e independentes, (y_i, $x_{i,1}$, $x_{i,2}$) $\forall i$=1,n, o ajuste do modelo aditivo, i.e., a estimação do vetor β, é realizada por meio do estimador de mínimos quadrados penalizados (PRSS - *Penalized Residual Sum of Squares*) e envolve a minimização da seguinte função objetivo para um dado vetor de parâmetro de suavização λ, conforme (6.9)

$$\sum_{i=1}^{n}\left[y_i - \sum_{j=1}^{p=2} f_j\left(x_{i.j}\right)\right]^2 + \sum_{j=1}^{p=2} \lambda_j \int f_j''\left(x_j\right)^2 dx_j \tag{6.9}$$

O primeiro termo da função objetivo corresponde a soma dos quadrados dos resíduos, uma medida da aderência do modelo aos dados, enquanto o segundo termo controla o grau de suavização das funções, no qual os parâmetros de suavização $\lambda_j>0$ $\forall j=1,p$ penalizam as curvaturas das funções suaves $f_j(x_j)$ $\forall j=1,p$ e controlam o *trade-off* entre a qualidade do ajuste (soma dos quadrados dos resíduos) e o grau de suavização das funções suaves. A minimização da função objetivo acima permite estimar o vetor de parâmetros β para valores dados de $\lambda_j \forall j=1,p$. Assim, o resultado da minimização da função objetivo em (6.9) é o estimador em (6.10).

$$\hat{\beta} = \left(X^T X + \sum_{j=1}^{p} \lambda_j S_j\right)^{-1} X^T Y \tag{6.10}$$

em que cada linha da matriz X guarda o vetor em (6.8) para a i-ésima unidade da amostra (a matriz X tem n linhas), o vetor Y guarda os valores da variável dependente nas n unidades amostradas e S_j $\forall j=1,p$ são matrizes com valores conhecidos e que dependem das funções suaves $h(x,x^*)$, para maiores detalhes veja [6].

O estimador em (6.10) fornece o vetor $\hat{\beta}$ para parâmetros de suavização λ dados. Assim, o processo de ajuste do modelo aditivo envolve uma busca em grade para encontrar os parâmetros de suavização λ, e a respectiva estimativa do vetor de parâmetros β por meio do estimador em (6.10), que minimizem o escore de validação cruzada (GCV *score* - *Generalized Cross Validation score*) [6,10], definido em (6.11).

$$GCV(\mathrm{l}) = \frac{n\sum_{i=1}^{n}\left[y_i - \sum_{j=1}^{p} f_j\left(x_{i.j}\right)\right]^2}{[traço\,(I-A)]^2} \tag{6.11}$$

em que I é uma matriz identidade e $A = X\left(X^T X + \sum_{j=1}^{p} \lambda_j S_j\right)^{-1} X^T$ a matriz de influência.

Adicionalmente, a busca pela melhor especificação de um modelo GAM pode contar com o critério de Akaike (AIC) [11].

Os parâmetros de suavização podem ser tratados como efeitos aleatórios e neste caso a estimação pode ser efetuada pela maximização da verossimilhança restrita (REML – *Restricted Maximum Likelihood*). Maiores detalhes sobre o GCV score e REML podem ser encontrados em [6] e [12].

6.2. Determinação das grades

O procedimento para obtenção das grades de perdas de carga e produtibilidade específica passa por dois estágios principais: o ajuste dos dados com auxílio do modelo GAM e a amostragem deste ajuste para um conjunto de pontos discretos. Além destas etapas centrais, existem três procedimentos auxiliares que completam o processo: a extrapolação dos tramos inferior e superior dos ajustes de perdas de carga, o acréscimo de pontos da curva colina aos dados históricos de produtibilidades e a amostragem assimétrica de pontos para a grade. Cada um destes estágios e procedimentos auxiliares serão detalhados na sequência.

6.2.1. Parametrização do ajuste do modelo GAM

Na seção anterior ficou evidente a grande flexibilidade do modelo GAM para capturar comportamentos irregulares. O número de splines definidos para composição de cada função dita quanta liberdade ela terá para se ajustar aos dados, sendo cada vez mais maleável conforme mais splines são adicionadas. A Figura 6.2 ilustra este efeito, comparando o ajuste de modelo GAM com uma e dez splines aos dados da mesma usina com três unidades geradoras, agregados semanalmente.

O número de splines é uma parametrização do ajuste, decidido à priori, tal qual a decisão do grau de um ajuste polinomial. Em ajustes do tipo GAM, assim como polinomiais, a otimização do número de splines observando puramente o erro tende a situações de *overfitting*, isto é, o uso de excessivos graus de liberdade. Nestas situações o ajuste final apresenta oscilações excessivas, um comportamento incoerente com aquele fisicamente esperado das perdas de carga e produtibilidade específica de uma usina hidrelétrica. É necessário compor uma métrica que reflita a qualidade do ajuste, mas também seja penalizada pelo comportamento inconsistente no caso de *overfitting*.

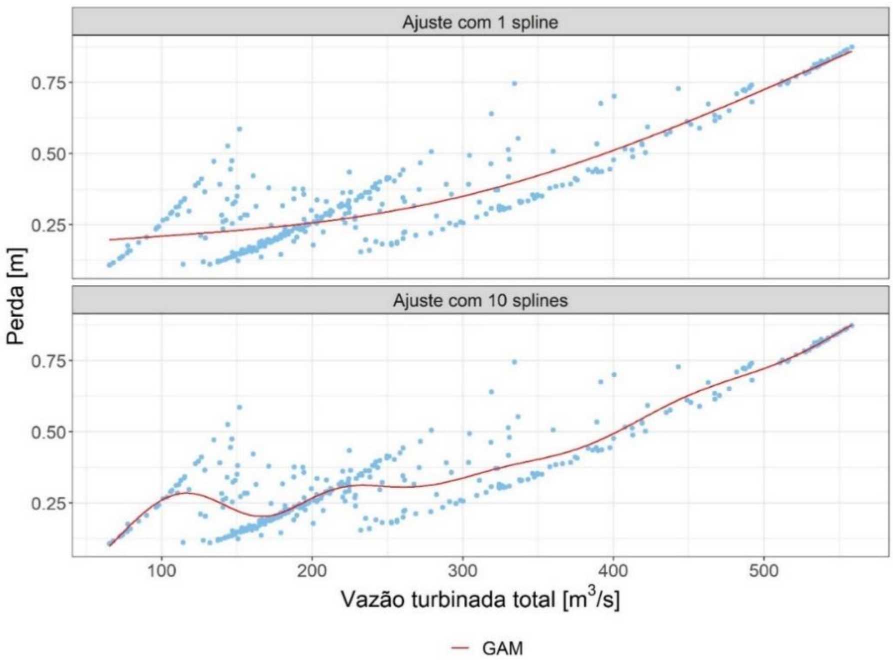

Figura 6.2 - Comparação de ajustes do modelo GAM com uma e com dez splines.

A forma como os parâmetros do modelo GAM são otimizados permite o cálculo de coeficientes de informação, índices de compromisso entre o quão bem modelado foi o dado e quantos parâmetros foram utilizados naquele modelo. Um dos mais populares coeficientes deste tipo é o *Akaike Information Criterion* – AIC, definido conforme (6.12).

$$AIC = 2n - 2\hat{l} \tag{6.12}$$

Onde:

n - Representa o número de parâmetros do modelo; e

$\hat{l}$ - Representa a verossimilhança calculada com os parâmetros ótimos.

De forma geral, a verossimilhança $\hat{l}$ pode ser entendida como uma métrica de qualidade do ajuste, tal qual seria o erro. No caso do modelo GAM, o número de parâmetros n é diretamente relacionado à quantidade de splines consideradas. O coeficiente AIC é comumente utilizado como ferramenta para comparação de modelos. Quanto menor o valor do coeficiente, melhor a proporção de qualidade do ajuste para o número de parâmetros julgado necessário.

Coeficientes de informação são métricas gerais, aplicáveis a diversas formas de modelagem, capturando acima de tudo a parcimônia do modelo. De fato, a penalização do coeficiente AIC (e qualquer outro coeficiente de informação) é função do número de parâmetros, não de aspectos indesejáveis do ajuste especificamente deste tipo de dado. Deve ser lembrado, porém, que estas características inconsistentes com o comportamento físico são, usualmente, consequência de overfitting e, assim, do número elevado de parâmetros. Em função destes aspectos o coeficiente AIC foi adotado como referência para seleção do número de splines em cada dimensão.

Através de análises preliminares foi observado que 20 é um número máximo de splines razoável. Dado este limite a escolha do melhor número de splines é realizada através de otimização combinatória, ajustando o GAM com todas as possíveis combinações de números de splines (de 1 a 20), e selecionando aquela com melhor coeficiente AIC.

Novamente, como o coeficiente AIC não penaliza diretamente os comportamentos inconsistentes não é possível tomá-lo como a palavra final na escolha do número de splines. É possível que o melhor coeficiente AIC esteja associado a um ajuste que ainda apresente aspectos indesejáveis, de modo que é necessária a inspeção visual por parte do analista para definição da parametrização definitiva. Este processo é realizado através da reotimização dos parâmetros num domínio limitado, usualmente reduzindo a faixa de número de splines considerada.

Análises adicionais sobre o uso do modelo GAM para a modelagem dessas curvas, incluindo uma comparação com um modelo de regressão quantílica que também foi avaliado como alternativa para esse trabalho, são apresentadas em [8].

6.2.2. Complementação dos ajustes de perdas

Conforme comentado anteriormente as grades de perdas de carga cobrem, necessariamente, da vazão nula até um máximo, ou teórico a partir da informação do arquivo HIDR.dat ou observado no histórico médio semanal.

A vazão turbinada nula é algo que não existe em nenhum histórico operativo, naturalmente, pois corresponde a usina desativada. Vazões turbinadas baixas, inferiores aos engolimentos mínimos admissíveis em cada unidade geradora, também não constam na maioria dos históricos por corresponderem a pontos muito distantes da zona operativa, prejudiciais à turbina e evitados no dia a dia da operação da usina.

Em função destes fatores os históricos médios semanais raramente apresentam valores de perdas de carga próximos a vazão nula, deixando um trecho de domínio bastante esparso. Além disso, nos casos em que o limite superior da grade vem de informações teóricas, a região final do domínio também será pouco populada.

O uso das funções ajustadas via modelo GAM nestes trechos sem dados tipicamente não resulta em comportamentos razoáveis, de forma que é preciso complementá-las através de outros métodos. Para este fim, foram aplicadas duas formas distintas para composição dos tramos inferior e superior do ajuste de perdas de carga.

A primeira abordagem segue os seguintes estágios, ilustrados na Figura 6.3.

- Para região inferior:
    1. É identificada a menor vazão média semanal observada e amostrada a função ajustada nesta vazão; e
    2. Ajusta-se uma curva do tipo $\mathrm{Perda} = k \times Q_{turb.}^2$ entre a origem e este ponto.

- Para região superior:
    3. É identificado o ponto histórico de maior vazão média semanal; e
    4. Ajusta-se uma curva do tipo $\mathrm{Perda} = k \times Q_{turb.}^2$ entre a origem e este ponto.

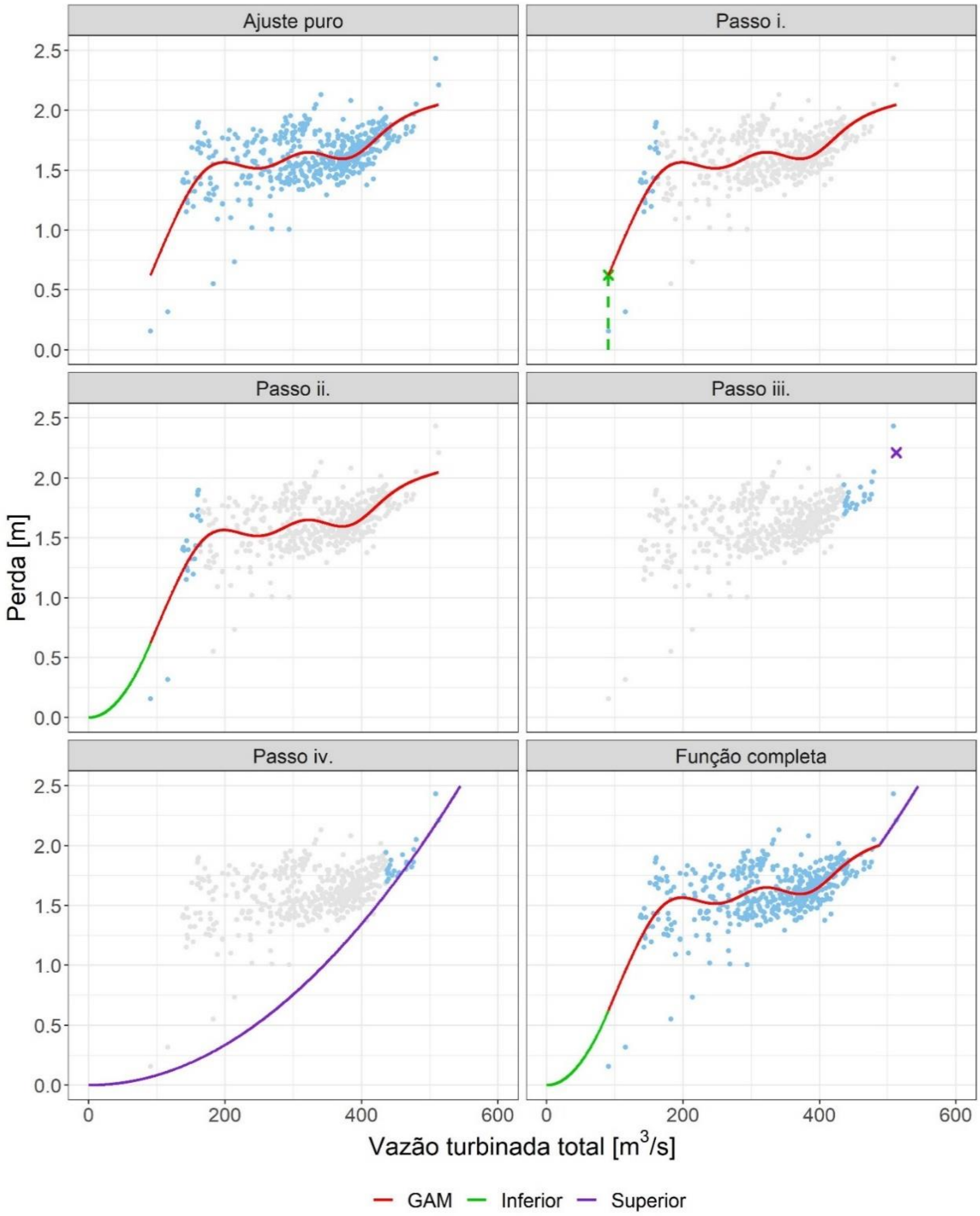

Figura 6.3 - Ilustração do primeiro processo de complementação do ajuste de perdas de carga.

Alternativamente, um segundo procedimento foi desenvolvido:

i. É ajustada uma curva do tipo $\text{Perda} = k_1 + k_2 \times Q_{turb.}^2$ aos pontos com vazões inferiores ao quantil 5%

ii. É ajustada uma curva do tipo $\text{Perda} = k_1 + k_2 \times Q_{turb.}^2$ aos pontos com vazões superiores ao quantil 95%

A Figura 6.4 ilustra o segundo procedimento.

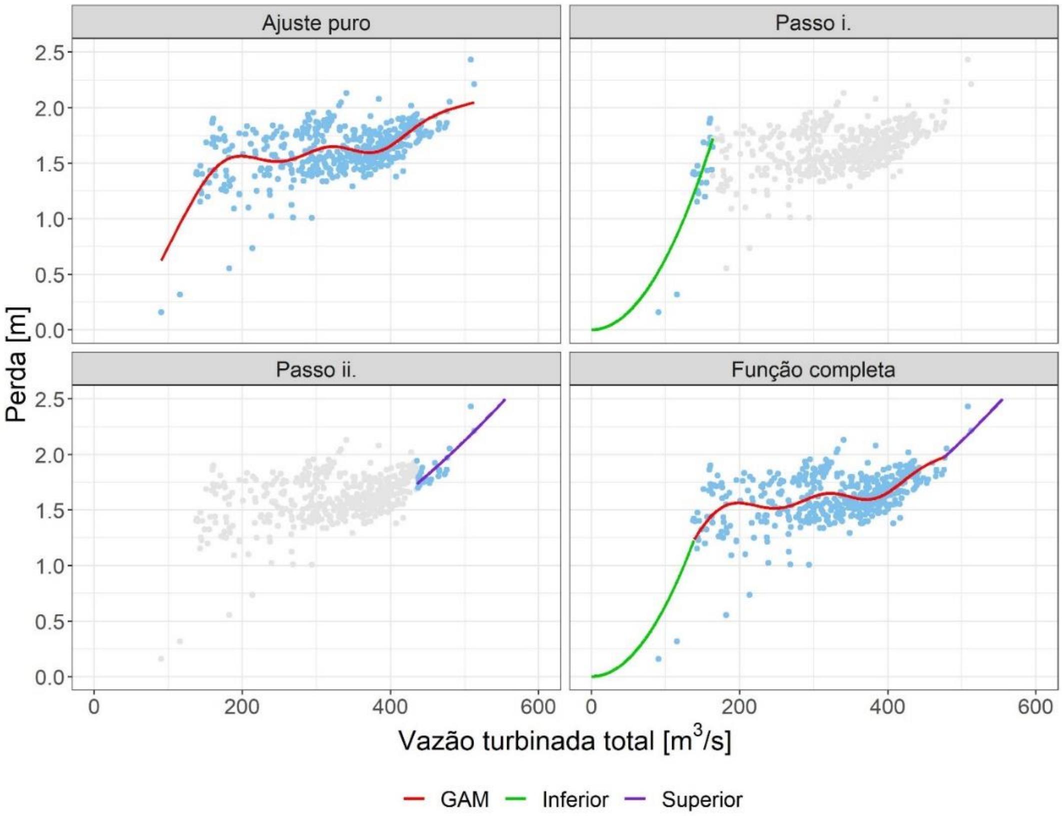

Figura 6.4 - Ilustração do segundo processo de complementação do ajuste de perdas de carga.

Ambos os métodos são pautados em ajustes de funções similares às equações de perdas de carga, o que a princípio é inconsistente com a motivação do ajuste via modelo GAM. No caso das funções ajustadas à região inferior do domínio, é possível assumir que uma eventual operação nesta faixa seria realizada com uma única máquina e, portanto, a equação de perdas seria aceitável. Mesmo no caso de usinas de grande porte, quando esta premissa não é necessariamente verdadeira, a função de perdas ainda precisa necessariamente ser nula na origem e apresentar crescimento monotônico, características manifestadas pelos ajustes quadráticos.

As funções ajustadas nas regiões de vazão elevada se justificam sob argumentos similares. Nesta parte do domínio todas as unidades geradoras da usina já operam próximo ao seu máximo, de modo que eventuais desigualdades na vazão turbinada por cada uma, não geram uma dispersão significativa e é possível aproximá-las pela hipótese de despacho igualmente distribuído.

A Figura 6.5 apresenta uma comparação dos dois métodos em quatro usinas tomadas como exemplo.

Em grande parte dos casos a diferença entre métodos é muito pequena, especialmente na região superior. Em algumas situações, no entanto, é possível que o ajuste final obtido com um dos métodos seja significativamente superior, tanto do ponto de vista de coerência com a função do modelo GAM quanto com o comportamento físico esperado. As regiões inferior e superior não precisam necessariamente advir do mesmo procedimento, ficando a critério do analista responsável a decisão de qual método aplicar em cada caso.

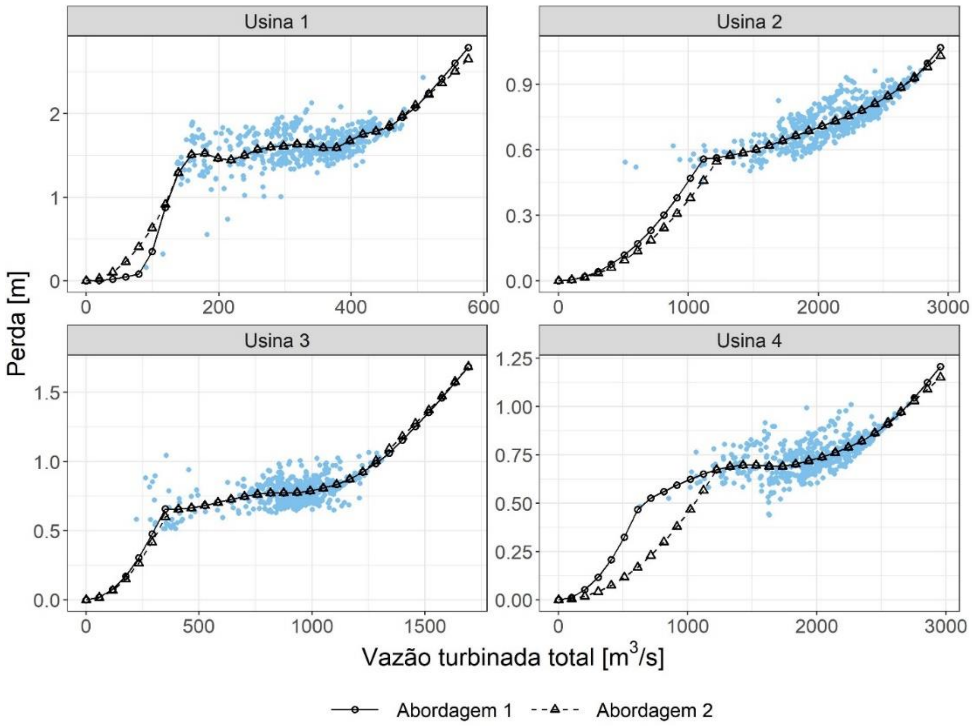

Figura 6.5 - Comparação dos métodos de complementação do ajuste de perdas de carga.

6.2.3. Uso de pontos da curva colina

Os ajustes de produtibilidade específica sofrem das mesmas deficiências que aqueles de perda, descritos anteriormente. De fato, em função da dimensão mais relativa à queda líquida, o problema de esparsidade do domínio é ainda mais significativo quando se trata do ajuste dos dados de produtibilidade específica. A Figura 6.6 apresenta uma projeção dos valores de produtibilidade específica sobre o eixo queda líquida x vazão turbinada para uma usina, junto do domínio a ser coberto pela grade.

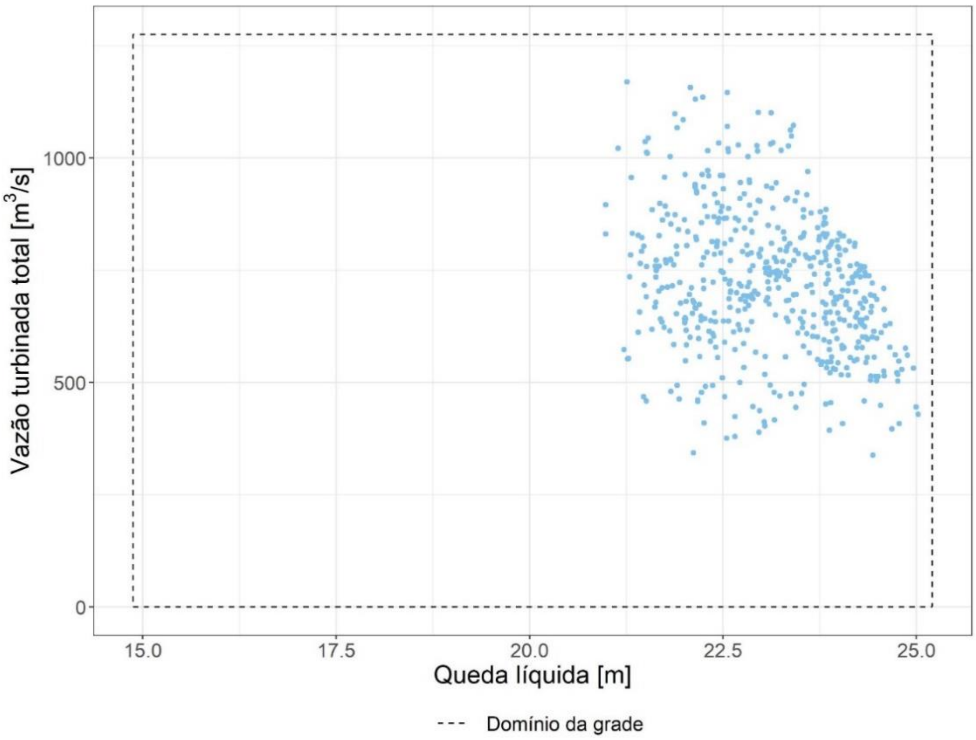

Figura 6.6 - Visualização bidimensional da produtibilidade como função da vazão turbinada e queda líquida.

Fica evidente a diferença de magnitude no problema de esparsidade dos dados. O resultado da extrapolação da função ajustada via modelo GAM para todo o domínio necessário é apresentado na Figura 6.7. É imediatamente observável como a função não decresce conforme a vazão turbinada cai à zero, mas se mantém basicamente constante. Este comportamento é algo estritamente necessário de qualquer ajuste, caso contrário violaria o comportamento físico da produtibilidade específica.

Novamente surge a necessidade de intervir nos ajustes de modo a incorporar os comportamentos fisicamente esperados do dado modelado. No caso da produtibilidade específica, não existe uma função teórica que a relacione à queda líquida e vazão turbinada como acontece com a perda. Ainda, a extensão de domínio sem dados torna ainda mais difícil o ajuste de outras funções que sejam coesas entre si e não levem a descontinuidades bruscas nas mudanças entre regiões de validade.

A abordagem adotada para produtibilidade específica consiste na adição de pontos oriundos da curva colina ao dado médio semanal observado. Na seção 5.1 foi explicitado a não viabilidade em utilizar equações de perdas de carga ou curvas colina diretamente, a não ser que se soubesse de antemão como será dividida a vazão total entre as unidades geradoras. Embora as usinas não sigam regras rígidas para divisão da geração, existem pontos operativos nos quais é possível garantir que todas as unidades operam igualmente.

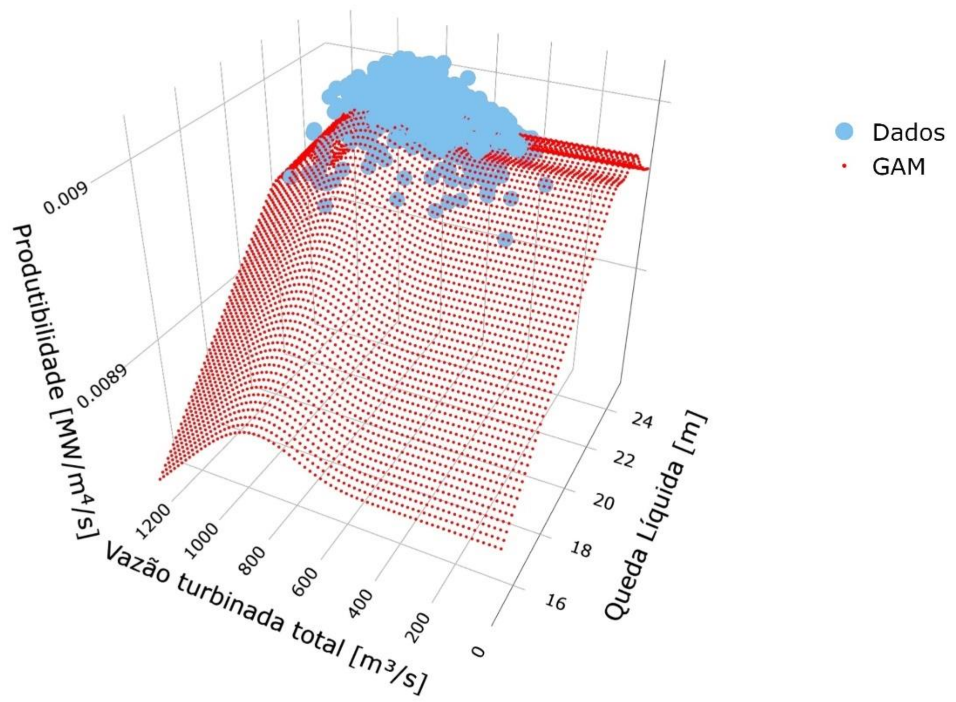

Figura 6.7 - Ajuste do modelo GAM extrapolado para todo o domínio necessário.

A partir da curva colina é possível identificar dois pontos cruciais:

1. Produtibilidade específica na menor queda líquida com maior vazão turbinada; e
2. Produtibilidade específica na maior queda líquida com maior vazão turbinada.

Estes dois pontos correspondem à produtibilidade específica média da usina quando esta turbina o máximo possível através de todas as máquinas. Ainda é possível extrair mais dois pontos da curva colina, associados ainda às quedas limite, porém com a menor vazão turbinável. Ao somar as vazões turbinadas mínimas obtidas na curva colina associada a cada unidade geradora é obtido um valor total que, ao contrário da vazão máxima, não corresponde a um único ponto de operação. Ainda assim estes pontos oferecem alguma informação acreca do comportamento da produtibilidade em vazões baixas. A Figura 6.8 apresenta o ajuste com uso destes quatro pontos, extrapolado para todo o domínio.

É imediatamente notável como o ajuste passa a apresentar todas as qualidades esperadas, como queda brusca conforme a vazão vai à zero e decrescimento para vazões muito elevadas ou quedas muito baixas. Outro ponto importante é a escala dos valores de produtibilidade na função ajustada. No caso sem auxílio da curva colina, a mínima produtibilidade específica está em torno de 0,0089 MW/m4/s, enquanto ao se adicionar os pontos este mínimo vai para 0,0080 MW/m4/s.

A utilização dos pontos da curva colina leva o ajuste a manifestar todas as características esperadas, mas isto não é suficiente. A queda brusca de produtibilidade conforme a vazão se aproxima de zero deveria, teoricamente, terminar em valores nulos, o que não acontece. Como não é possível introduzir restrições no ajuste via modelo GAM, a nulidade das produtibilidades específicas em vazão turbinada zero é forçada, independente dos valores originalmente ajustados, ao final de todos os procedimentos para desenvolvimento das grades.

Uma última observação deve ser feita quanto ao uso destes pontos. Naqueles de vazão baixa, como já foi explicitado, não há um único ponto operativo associado, de modo que seu uso constitui uma aproximação. Em segundo lugar, deve ser lembrado que a curva colina detalha o comportamento instantâneo da produtibilidade, não médio semanal como os dados modelados. Desta forma, mesmo que uma vazão média semanal igual ao máximo da usina corresponda a uma única distribuição de vazões turbinadas entre máquinas ao longo da semana, a direta associação desta condição operativa àquela da curva colina demandaria que a queda líquida, também, se mantivesse constante ao longo de toda a semana, o que é pouco provável na operação real.

Em função destes aspectos, é possível que a informação da curva colina seja incoerente com os dados médios semanais em determinados casos. O uso destes pontos, parcialmente ou em sua totalidade, fica a critério do analista durante a realização dos ajustes do GAM.

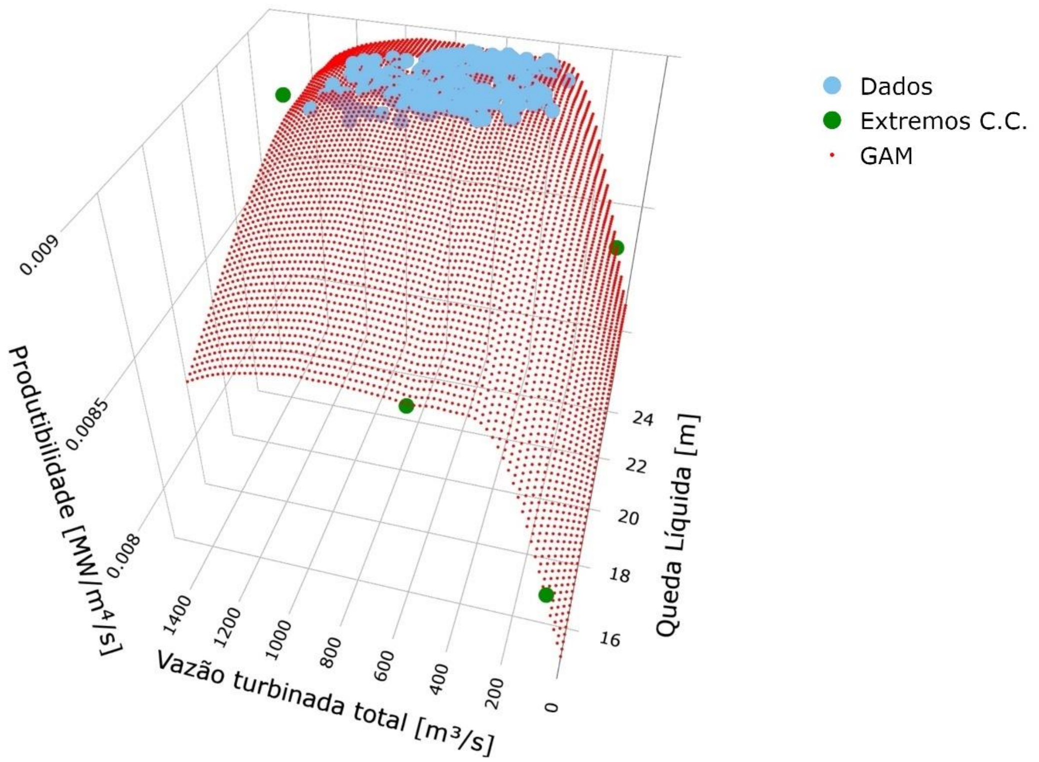

Figura 6.8 - Ajuste GAM com adição de pontos da curva colina extrapolado.

6.2.4. Parametrização das grades

A partir do ajuste do modelo GAM definitivo é extraída uma grade de pontos através da amostragem em determinados pontos do domínio. Assim como a definição do número de splines descrito na seção anterior, a definição das grades demanda a escolha de quantas segmentações, ou seja, quantos pontos, a compõem.

Para um determinado número de divisões, o domínio é particionado igualmente e a extração das grades é realizada através da amostragem do ajuste do modelo GAM aos nós, isto é, junções entre as partes. O modelo DECOMP realiza consulta ao resto do domínio através de interpolações.

Este processo constitui, essencialmente, uma aproximação linear por partes do ajuste original, de modo que com infinitas partes, isto é, infinitos pontos componentes, a grade retorna à representação contínua original. A Figura 6.9 demonstra esse argumento.

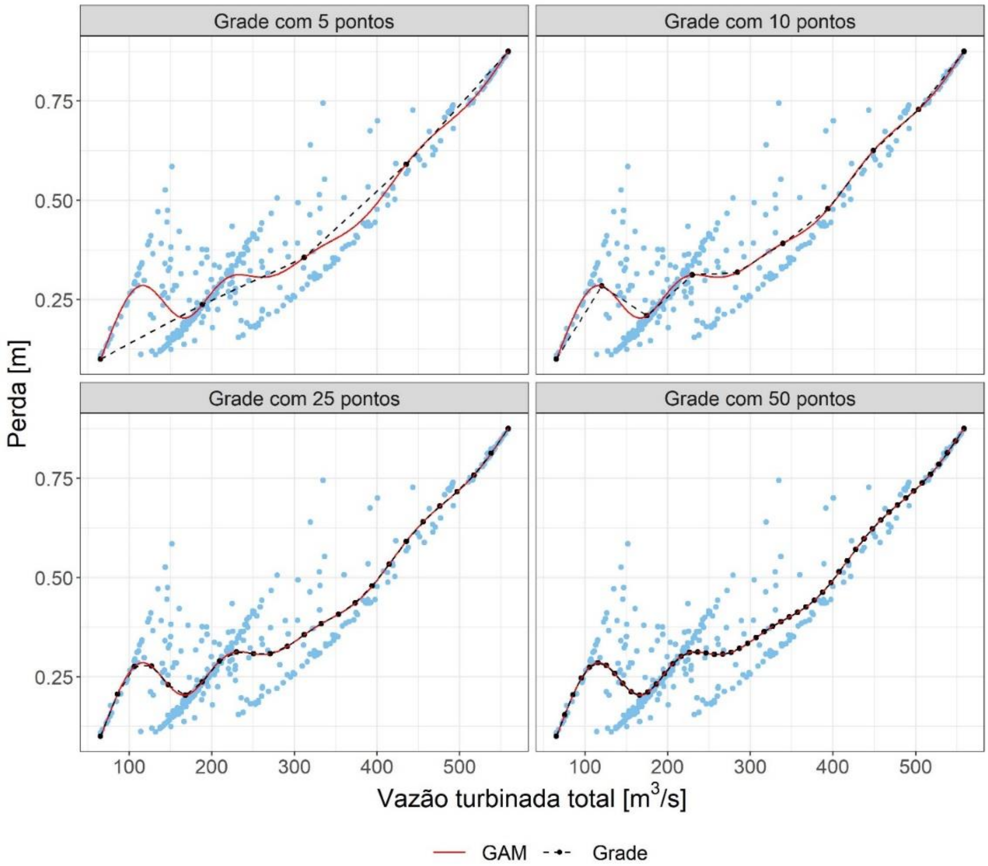

Figura 6.9 - Aproximação por partes conforme o detalhamento da grade aumenta.

Sob esta luz, podemos definir dois erros conforme (6.13).

$$
\begin{aligned}
ERRO_{grid} &= \sum_{i=1}^{N} |y_i - f(\tilde{x}_i;n)| \\
ERRO_{GAM} &= \sum_{i=1}^{N} |y_i - f(\tilde{x}_i)|
\end{aligned}
\qquad (6.13)
$$

Onde:

- y é a variável de resposta – perdas de carga ou produtibilidade específica
- $\tilde{x}$ representa vazão, para perdas de carga, ou vazão turbinada e queda líquida, para produtibilidade específica
- f_{GAM} é a função ajustada via modelo GAM
- n é o número de segmentações da grade de pontos

- N é o número de observações no dado médio semanal
- f_{grid} é a interpolação do vetor x na grade de n pontos

Conforme n se eleva e a representação por grade converge à representação contínua original, a razão entre estes dois erros tende a 1, de acordo com (6.14).

$$\lim_{n \to \infty} \frac{ERRO_{grid}}{ERRO_{GAM}} = 1 \tag{6.14}$$

A Figura 6.10 apresenta a análise da razão de erro para ajustes de grade de perdas de carga em duas usinas tomadas como exemplo com diferentes graus de variabilidade no dado sendo ajustado.

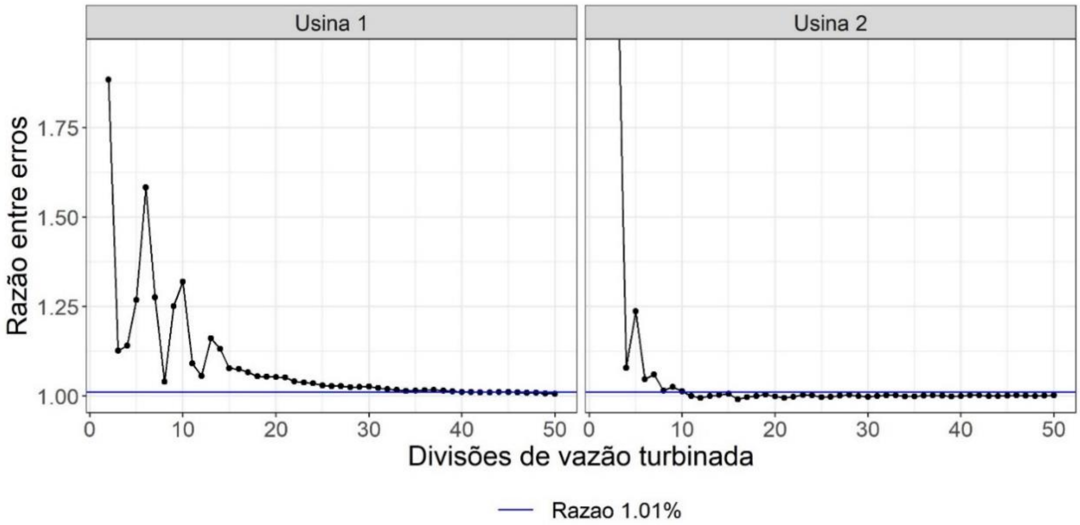

Figura 6.10 - Análise da razão de erros no ajuste de perdas de carga para duas usinas.

A análise destes dois comportamentos deixa claro que, apesar da representação por grades ser de fato uma simplificação, o número de divisões necessárias para alcançar uma razão de erros baixa não é elevado. A oscilação em baixos números de divisões do caso à esquerda se deve ao fato do ajuste do modelo GAM apresentar, também, muitas oscilações.

Como o domínio de vazão turbinada é particionado em segmentos de mesmo tamanho, ao aumentar o número de partes, todas as existentes são deslocadas. Isto faz com que no caso de uma função com muitos picos e vales, a qualidade da aproximação varie bruscamente conforme o número de partes varia. À medida que este número cresce, no entanto, o deslocamento das partes existentes passa a ser muito pequeno, causando pouco efeito na razão.

Com base nestas análises, é possível determinar uma regra de decisão do número de segmentações que garante boa representatividade e parcimônia:

- Identifica-se um número y_i de segmentações de vazão turbinada cuja razão de erro de estimação é inferior a 1,01%; e
- Caso todas as razões para número de segmentações y, $y_i \leq y \leq y_{max}$, permaneçam inferiores, y_i é escolhido.

O número máximo y_{max} foi considerado por padrão igual a 30, sendo flexível em casos como o apresentado na Figura 6.10. A Figura 6.11 apresenta a aplicação deste procedimento nos mesmos dois casos expostos anteriormente, definindo y_{max} igual a 60 para melhor ilustração do resultado.

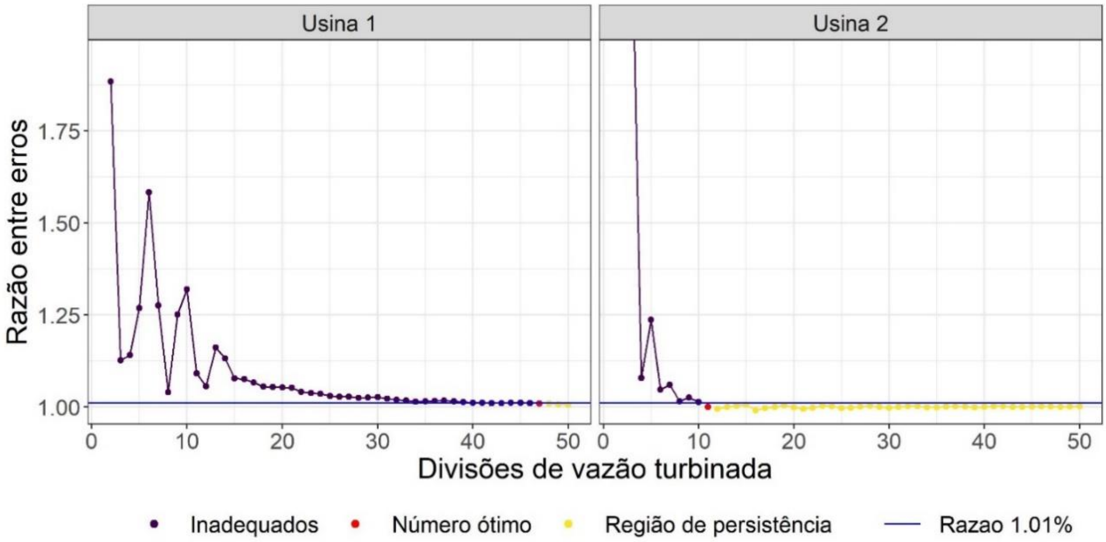

Figura 6.11 - Avaliação do melhor número de divisões de vazão em grades de perdas de carga.

Para definição do melhor número de divisões na grade de produtibilidade específica é necessário adaptar o procedimento aplicado às grades de perda. Com a adição de uma segunda dimensão, da queda líquida, há inúmeros pontos a partir dos quais qualquer incremento de número de divisões, queda ou vazão, mantém a razão de erros abaixo de 1,01%. O procedimento adaptado segue os passos

i. Identificam-se números x_i e y_i de segmentações de queda e vazão cuja razão de erro de estimação é inferior a 1,01%
ii. Caso todas as razões para número de segmentações x, $x_i \leq x \leq x_{max}$ e y, $y_i \leq y \leq y_{max}$, permaneçam inferiores, o par (x_i, y_i) é salvo
iii. Uma vez avaliado todo o domínio, é escolhido o par dentre os salvos com menor produto

Tipicamente se consideram os valores máximos em cada dimensão iguais a 30, assim como no caso das perdas de carga, o que foi suficiente para a maioria das usinas.

O terceiro passo garante a grade mais parcimoniosa possível dentre todas as possibilidades que atendem a razão mínima entre erros. A Figura 6.12 apresenta uma análise da razão entre erros para múltiplas combinações de divisões de queda líquida e vazão turbinada. Pode ser observado que o comportamento é similar ao das perdas de carga, com a razão decrescente conforme o número de divisões em alguma dimensão cresce.

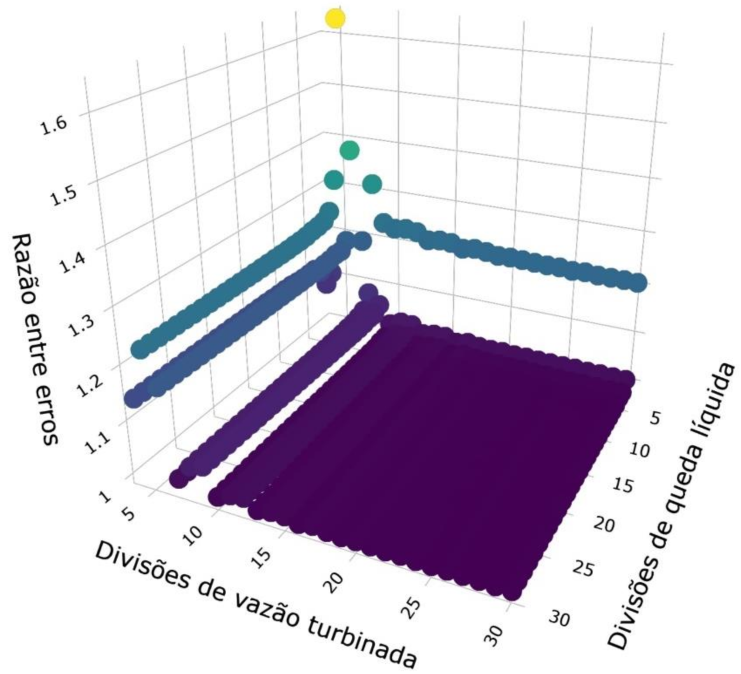

Figura 6.12 - Avaliação do melhor número de divisões de vazão turbinada e queda líquida.

É interessante notar como o aumento de divisões de queda líquida parece ter pouco efeito em comparação ao aumento na vazão turbinada. Este comportamento é esperado e comum entre todas as usinas, pois a vazão turbinada é um elemento mais determinante na produtibilidade específica do que a queda líquida.

A Figura 6.13 apresenta o procedimento de decisão do melhor número de divisões aplicado à esta mesma usina. O ponto de menor produto nesta análise.

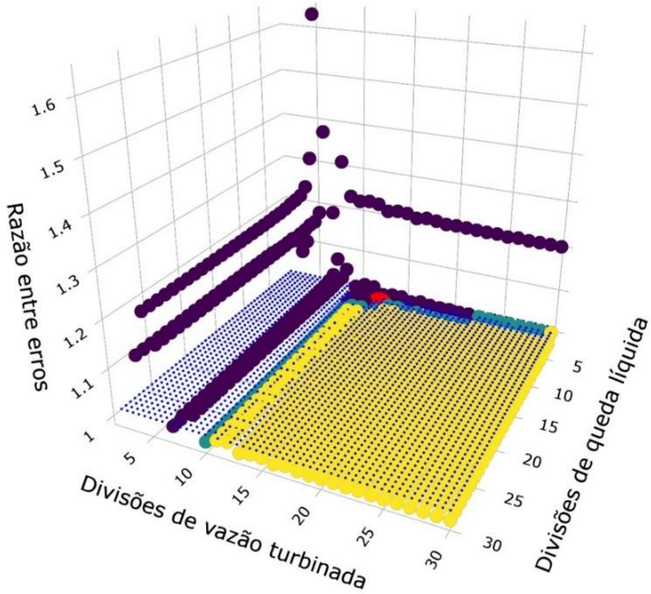

Figura 6.13 - Avaliação do melhor número de divisões de queda líquida e vazão turbinada.

6.2.5. Ajuste de grades assimétricas

Os limites superiores de divisões estipulados na seção anterior se provaram suficientes para a maioria das usinas consideradas neste estudo. Nos ajustes de perdas, ainda nos casos onde foi necessário determinar uma grade com mais de 30 pontos, estas não chegam a alcançar tamanhos excessivos que possam prejudicar o desempenho computacional do modelo DECOMP. O mesmo não ocorre para grades de produtibilidade específica.

Os métodos para determinação de números de divisões iniciam com a discretização do domínio em pontos igualmente distribuídos. Este primeiro passo considera o domínio completo, estipulado como descrito no início do item 3 deste documento, e não somente a região populada de dados. Por outro lado, os erros definidos no item 3, só podem ser calculados naturalmente na região populada por dados. Isto cria uma condição na qual o processo eleva o número de divisões no domínio completo para alcançar maior detalhamento numa região central.

Para grades de perdas de carga, onde a parte populada do domínio é consideravelmente maior que a não populada, esta condição não leva a tamanhos críticos (e, novamente, mesmo as maiores grades ainda são relativamente pequenas). No caso das grades de produtibilidade específica a taxa de domínio não populado pode ser significativa, obrigando o método a elevar o número de divisões desnecessariamente para que se alcance o detalhamento necessário na região de dados. A Figura 6.14 ilustra este efeito.

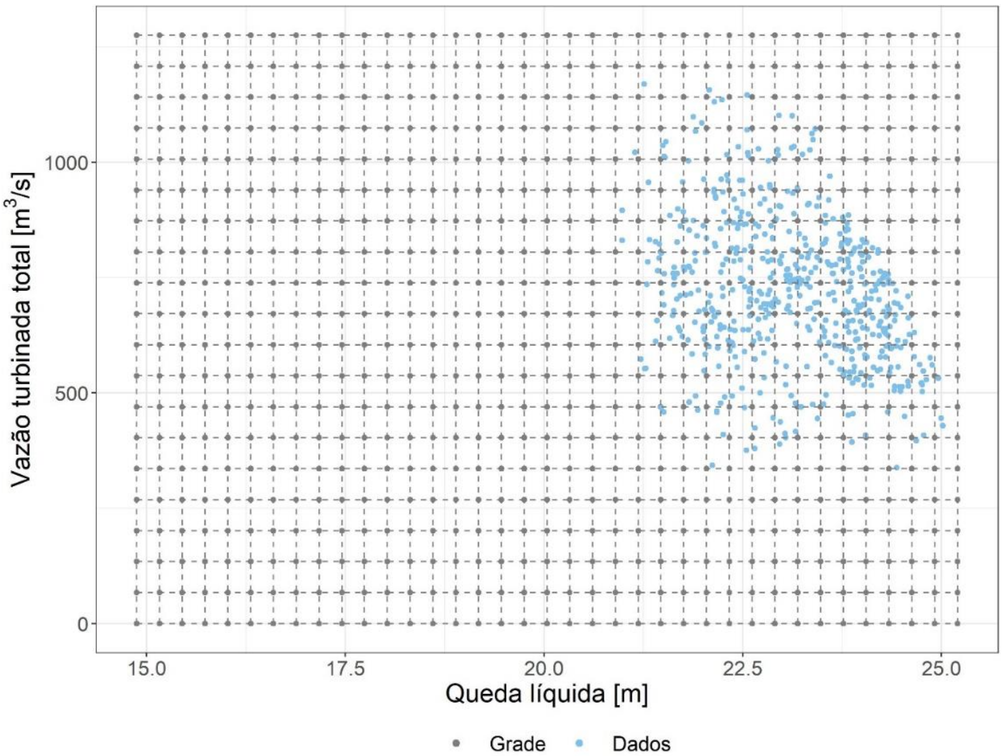

Figura 6.14 - Visualização bidimensional dos dados e grade de produtibilidade específica como função da vazão turbinada e queda líquida.

Em alguns casos esta condição levou o tamanho das grades a números proibitivos, o que tornou necessário o desenvolvimento de uma metodologia para contornar o problema. Isto levou à criação de um método heurístico no qual a dimensão da grade é calculada com base no domínio onde há dados, possivelmente estendido em alguma direção ao limite do domínio completo caso esteja muito próximo. Uma vez determinada esta grade interna, a densidade em cada eixo é propagada para o resto do domínio, reduzida de um fator proporcional à taxa de ocupação do domínio completo. A Figura 6.15 uma grade desenvolvida com esta assimetria.

É importante observar que ainda existe regularidade na dispersão da grade. O domínio ainda é segmentado em cada eixo e os pontos da grade jazem sobre as interseções entre as fronteiras destas divisões. Seria possível reduzir as grades ainda mais ao se abandonar esta regularidade, porém o procedimento de interpolação utilizado pelo DECOMP depende desta estrutura para funcionar adequadamente.

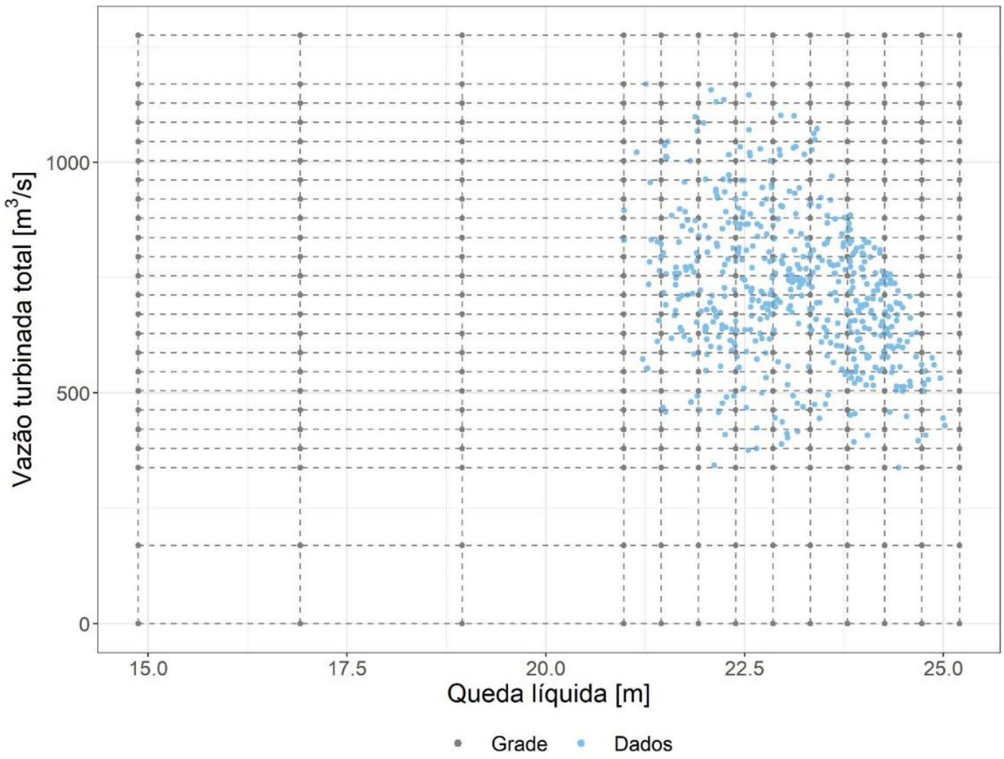

Figura 6.15 - Visualização bidimensional dos dados e grade assimétrica de produtibilidade como função da vazão turbinada e queda líquida.

6.2.6. Avaliação das grades ajustadas

Nesta seção é apresentada uma comparação entre os erros de ajuste considerando os valores de perda e produtibilidade constantes e aqueles obtidos pelas grades. O erro para cada usina é calculado como a média dos desvios absolutos entre o valor observado (registro histórico) e o valor calculado (constante ou grade), conforme (6.15).

$$\begin{aligned} ERRO_{grid} &= \sum_{i=1}^{N} |y_i - f(\tilde{x}_i; n)| \\ ERRO_{constante} &= \sum_{i=1}^{N} |y_i - k_{GTDP})| \end{aligned} \tag{6.15}$$

Onde:

- N é o número de observações do registro histórico utilizado na calibração dos modelos
- y representa a variável a ser avaliada (perda ou produtibilidade)
- k_{GTDP} representa a perda ou produtibilidade obtida no primeiro ciclo de cálculos do GTDP
- $f_{grid}(\tilde{x}; n)$ representa a interpolação do ponto $\tilde{x}$ no grid de n células

Na Figura 6.16 são apresentados os erros para todas as usinas considerando a representação por um valor constante e pela grade. Pode-se observar uma redução do erro ao se utilizar a última opção.

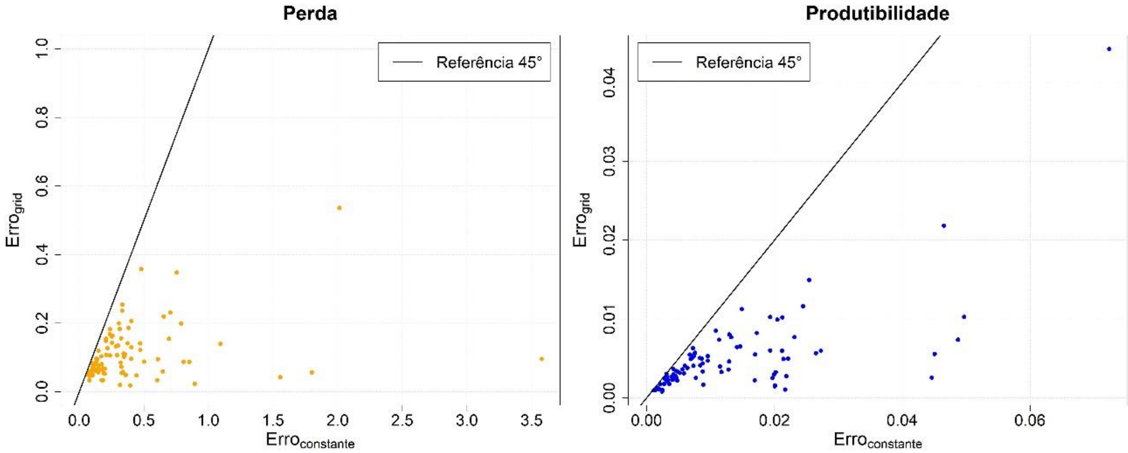

Figura 6.16 - Comparação dos erros de ajuste.

7. Incorporação das produtibilidades e perdas variáveis no cálculo da função de produção exata

Nas seções anteriores foi discutido em detalhe como as grades são obtidas a partir dos dados operativos e resultados do GTDP. Ao longo destas exposições, foi brevemente mencionado como o modelo DECOMP faria uso destas informações, mas de forma vaga. Nesta seção será discutido como foi adaptado o cálculo da função de produção exata.

A Figura 7.1 apresenta um diagrama ilustrativo do cálculo da função de produção exata com utilização das grades, a partir de uma vazão turbinada total Q_i e volume inicial V_i quaisquer. Este processo difere do original, dado pela equação (4.1b), apenas no valor de perdas de carga ($k_{perdas\,i}$) e produtibilidade específica (ρ_i) utilizado, antes uma constante, agora interpolado na grade adequada.

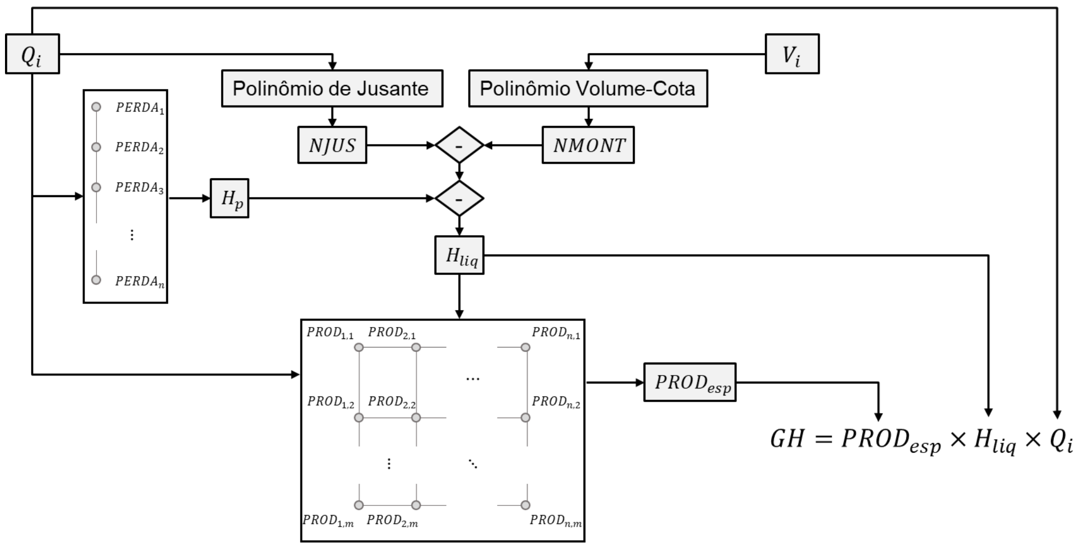

Figura 7.1 - Cálculo da função de produção incorporando representações variáveis.

Como função exclusivamente da vazão turbinada, a perda de carga média associada a Q_i pode ser obtida de imediato através de interpolação linear simples, definida em (7.1) e ilustrada na Figura 7.2.

$$PERDA_i = PERDA_a + \frac{PERDA_b - PERDA_a}{Q_b - Q_a} \times (Q_i - Q_a) \tag{7.1}$$

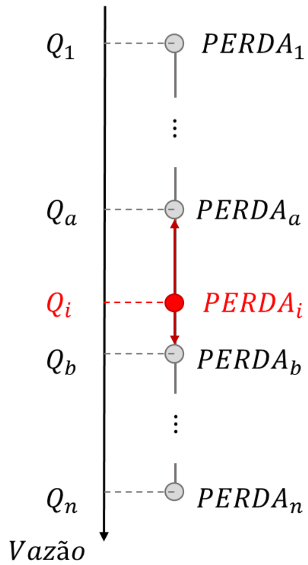

Figura 7.2 - Ilustração da interpolação na grade de perdas de carga.

A interpolação da produtibilidade específica média depende de duas variáveis: a vazão turbinada Q_i e queda líquida H_{liq}, sendo esta dependente do cálculo das perdas de carga. Uma vez de posse das duas variáveis, o valor médio é obtido através de interpolação bilinear na grade, definida em (7.2) e ilustrada na Figura 7.3.

$$PROD_i = \frac{1}{(H_b - H_a)(Q_b - Q_a)} \times [H_b - H_{liq} \quad H_{liq} - H_a] \times \begin{bmatrix} PROD_{a,a} & PROD_{a,b} \\ PROD_{b,a} & PROD_{b,b} \end{bmatrix} \times \begin{bmatrix} Q_b - Q_i \\ Q_i - Q_a \end{bmatrix} \quad (7.2)$$

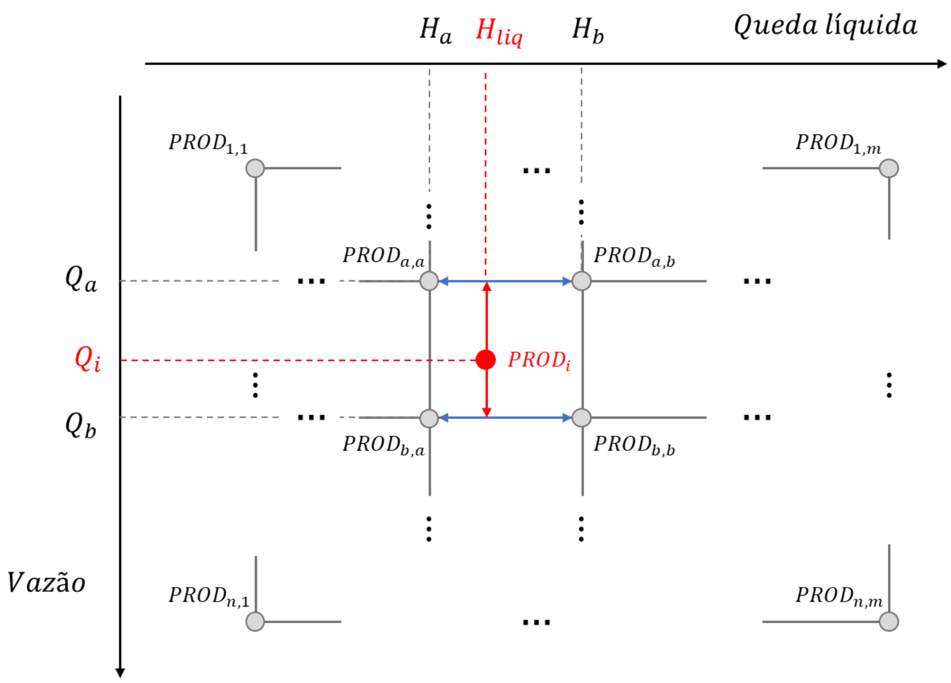

Figura 7.3 - Ilustração da interpolação na grade de produtibilidades específicas.

8. Avaliação preliminar da operação de curto prazo

Nesta seção serão apresentados resultados preliminares obtidos com o modelo DECOMP, as análises completas sobre a utilização das grades de produtilibilidade e perda serão objeto do segundo relatório do subgrupo de Produtibilidade.

A operação de curto prazo ao considerar dados de produtibilidade específica e perdas variáveis está sendo avaliada a partir de execuções independentes (não encadeadas) de todos os decks do ano de 2019 com a versão 30.3 do modelo DECOMP totalizando em 52 casos. Para efeitos de comparação, os decks oficiais do ONS, que consideram produtibilidade específica e perdas fixas, também foram processados na versão 30.3.

As Figura 8.1, Figura 8.2, Figura 8.3 e Figura 8.4 apresentam a comparação da energia armazenada dos submercados SE/CO, S, NE e N, respectivamente, para os dados de produtibilidade específica e perdas fixas e variáveis. Não foram observadas mudanças significativas para os armazenamentos de nenhum subsistema. No subsistema SE/CO a maior diferença encontrada corresponde a uma redução de 0.20 %EARmax na revisão 2 de março de 2019 ao considerar produtibilidade específica e perdas variáveis. No subsistema Sul a maior diferença ocorreu na revisão 0 de maio, com aumento de 1.1 % EARmax. No subsistema NE, a maior diferença não ultrapassou 0.1 % EARmax. Finalmente, no subsistema Norte, a maior redução foi de 0.6 % EARmax.

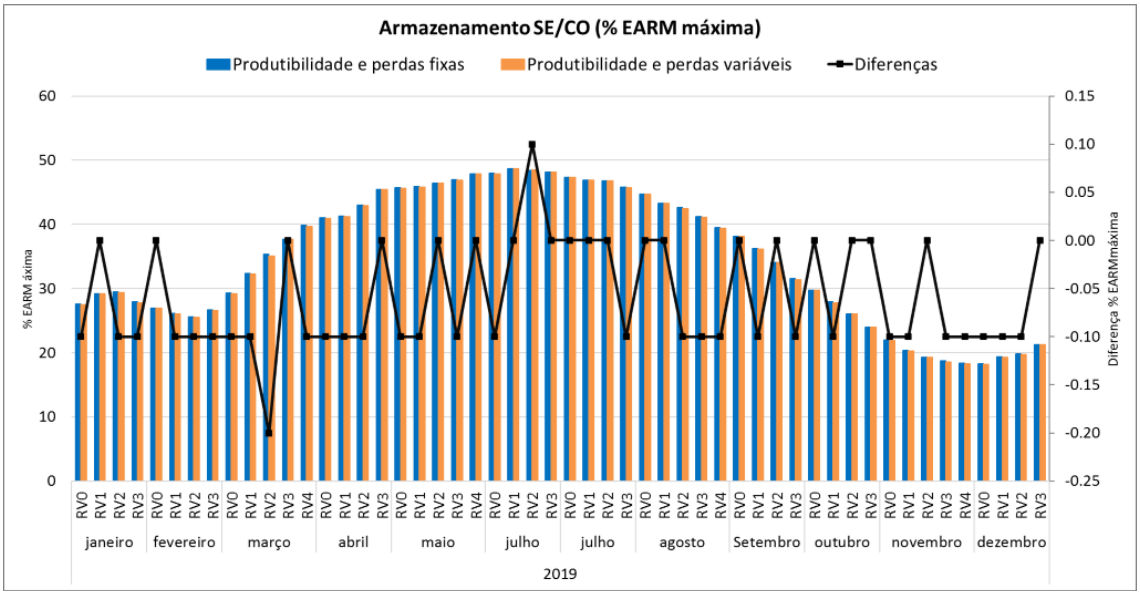

Figura 8.1 – Energia armazenada no subsistema SE/CO.

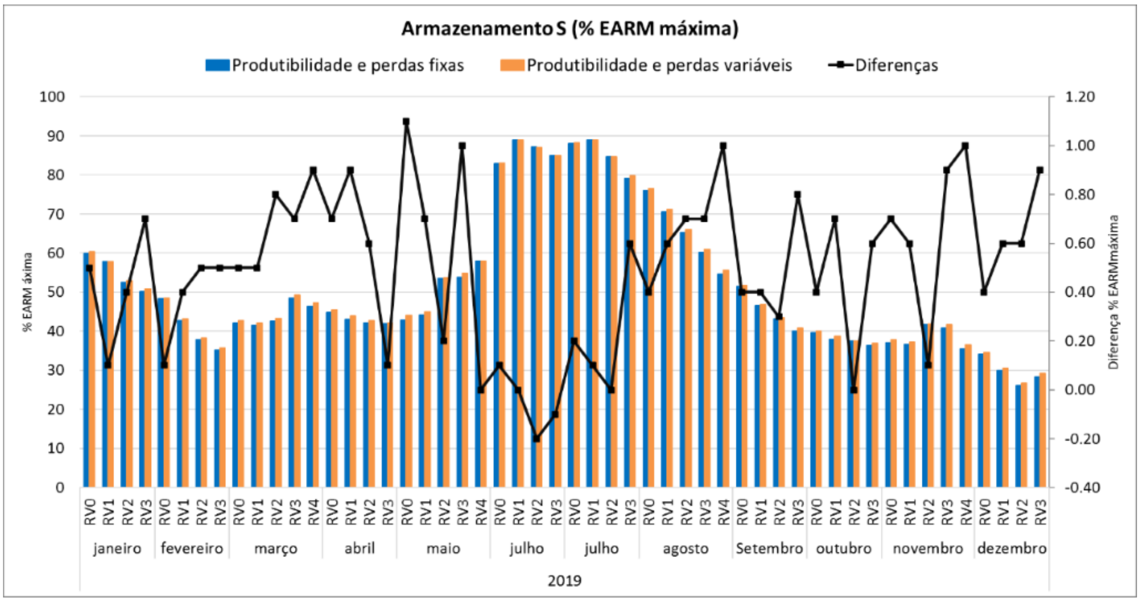

Figura 8.2 – Energia armazenada no subsistema S.

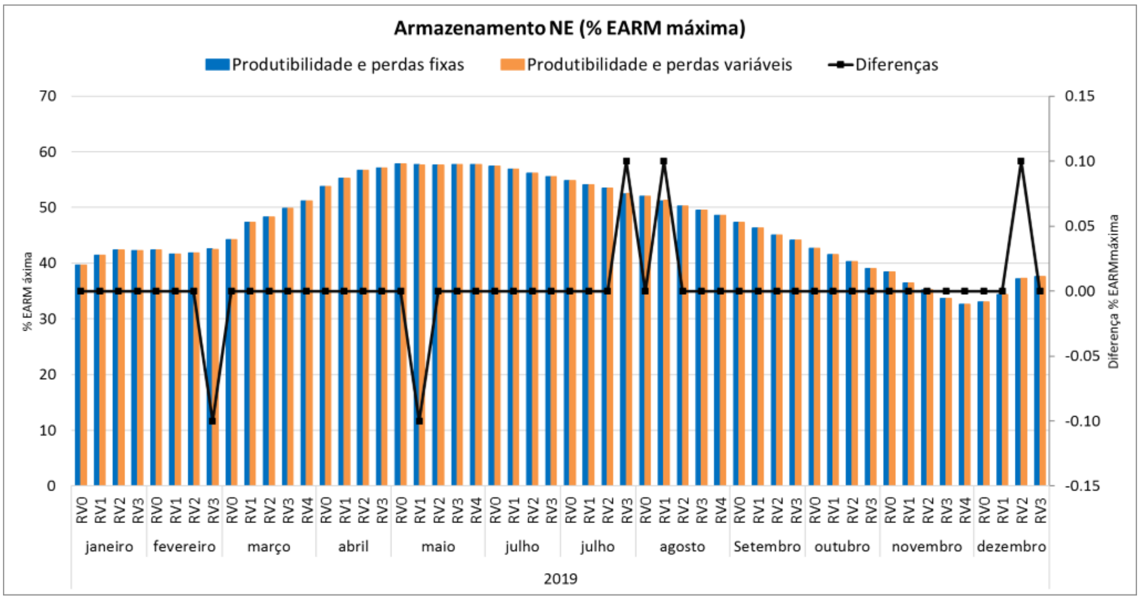

Figura 8.3 – Energia armazenada no subsistema NE.

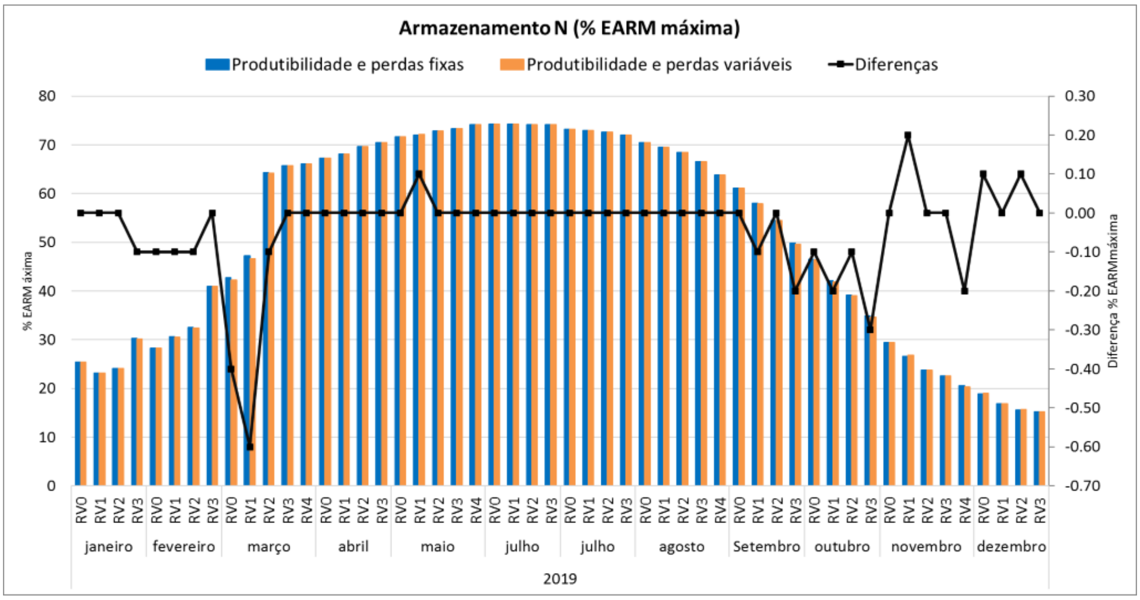

Figura 8.4 – Energia armazenada no subsistema N.

Da Figura 8.5 à Figura 8.8 estão apresentados os custos marginais de operação de todas as revisões do ano de 2019, comparando casos executados com produtibilidade específica e perdas variáveis com os casos executados com estes parâmetros fixos. Não foram observadas mudanças significativas para os custos marginais de operação dos subsistemas. Nos subsistemas SE/CO e S a maior diferença corresponde ao aumento de aproximadamente 9,00 R$/MWh em fevereiro, que equivale a um aumento percentual de 2% com relação a produtibilidade específica e perdas fixas. Nos subsistemas NE e N, a maior diferença corresponde a uma redução de aproximadamente 11,00 R$/MWh na revisão 4 de agosto, que corresponde a uma diferença percentual de aproximadamente 6% .

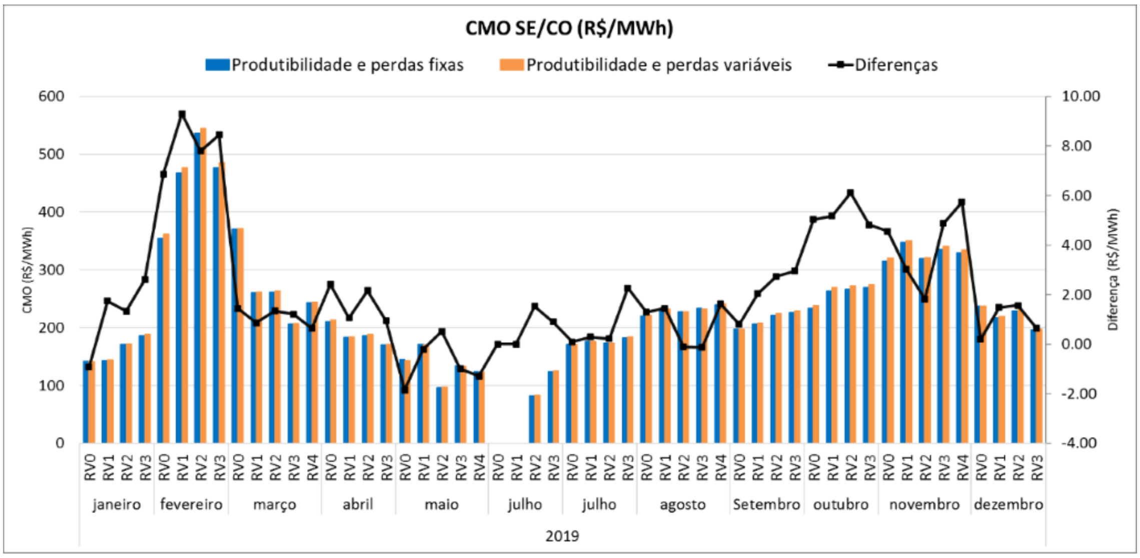

Figura 8.5 – Custo marginal de operação do subsistema SE/CO.

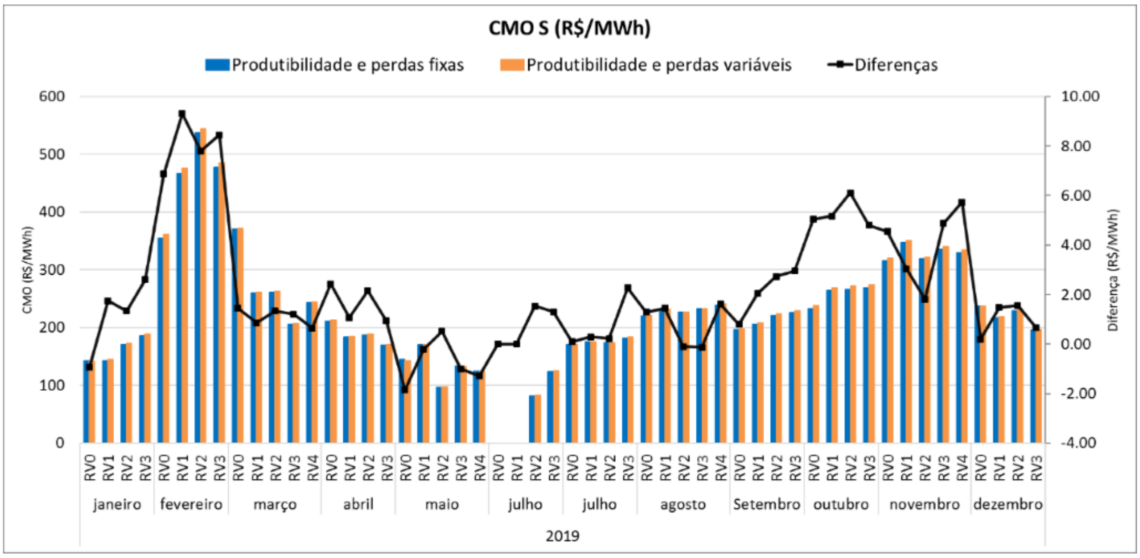

Figura 8.6 – Custo marginal de operação do subsistema S.

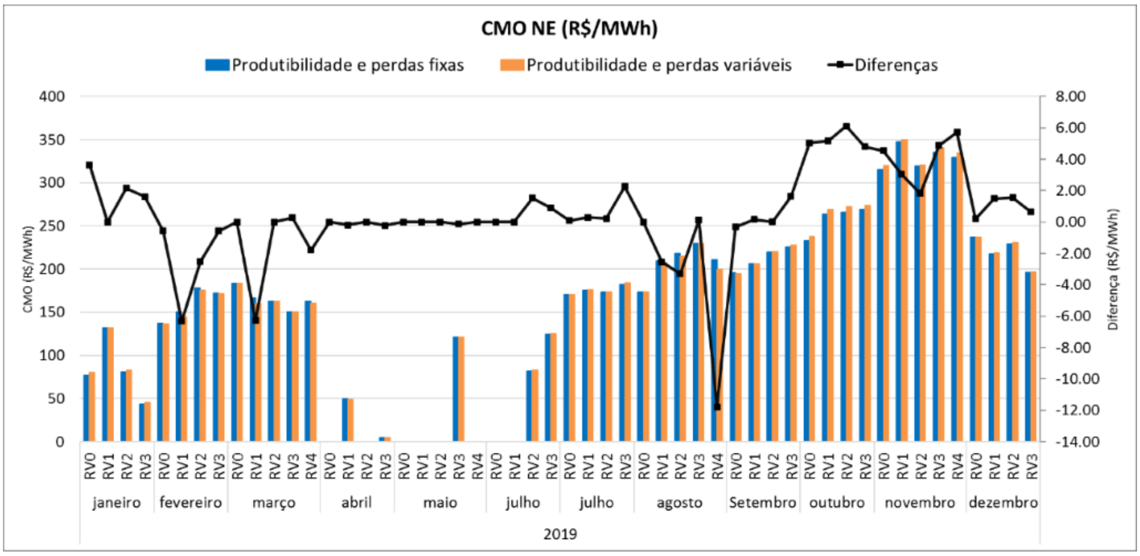

Figura 8.7 – Custo marginal de operação do subsistema NE.

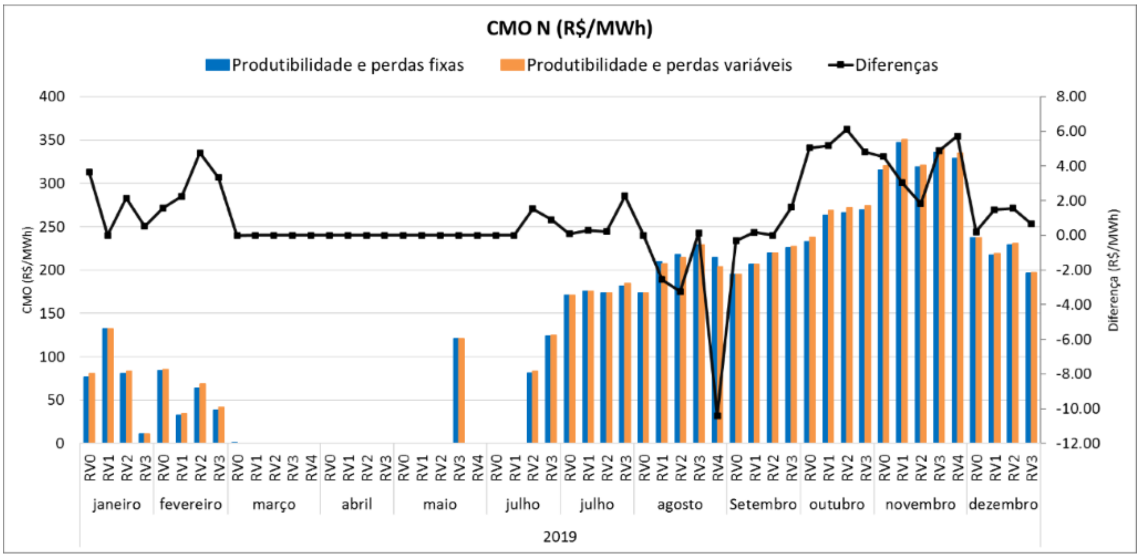

Figura 8.8 – Custo marginal de operação do subsistema N.

Da Figura 8.9 até a Figura 8.12 estão apresentados as gerações térmicas de todas as revisões do ano de 2019, comparando casos executados com produtibilidade específica e perdas variáveis com os casos executados com estes parâmetros fixos. Não foram observadas mudanças significativas para as gerações térmicas dos subsistemas.

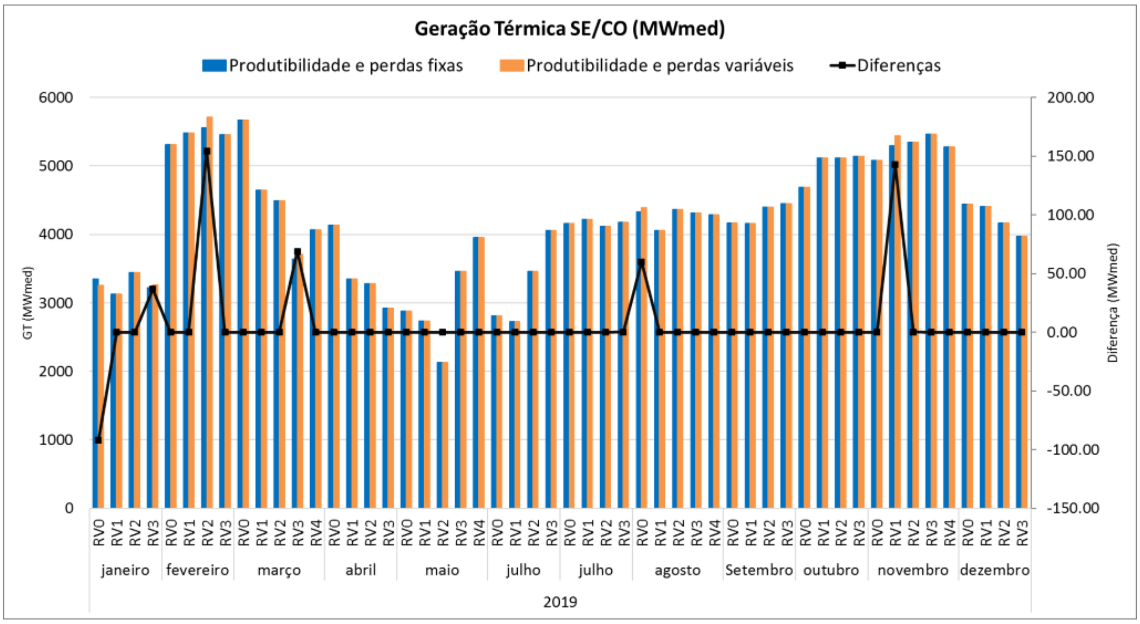

Figura 8.9 – Geração térmica do subsistema SE/CO.

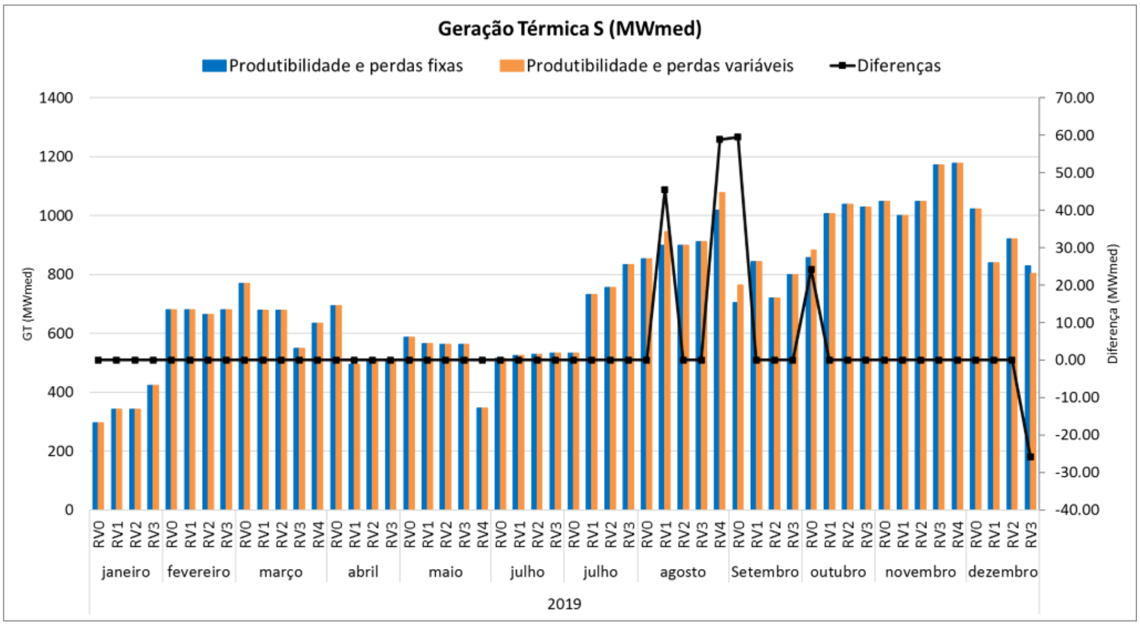

Figura 8.10 – Geração térmica do subsistema S.

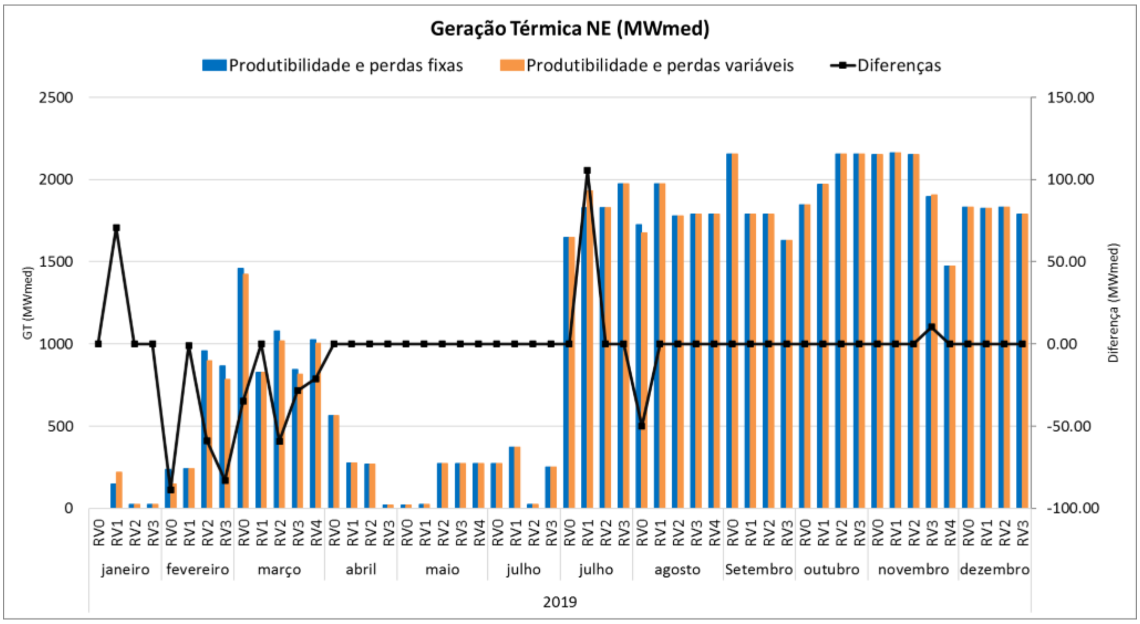

Figura 8.11 – Geração térmica do subsistema NE.

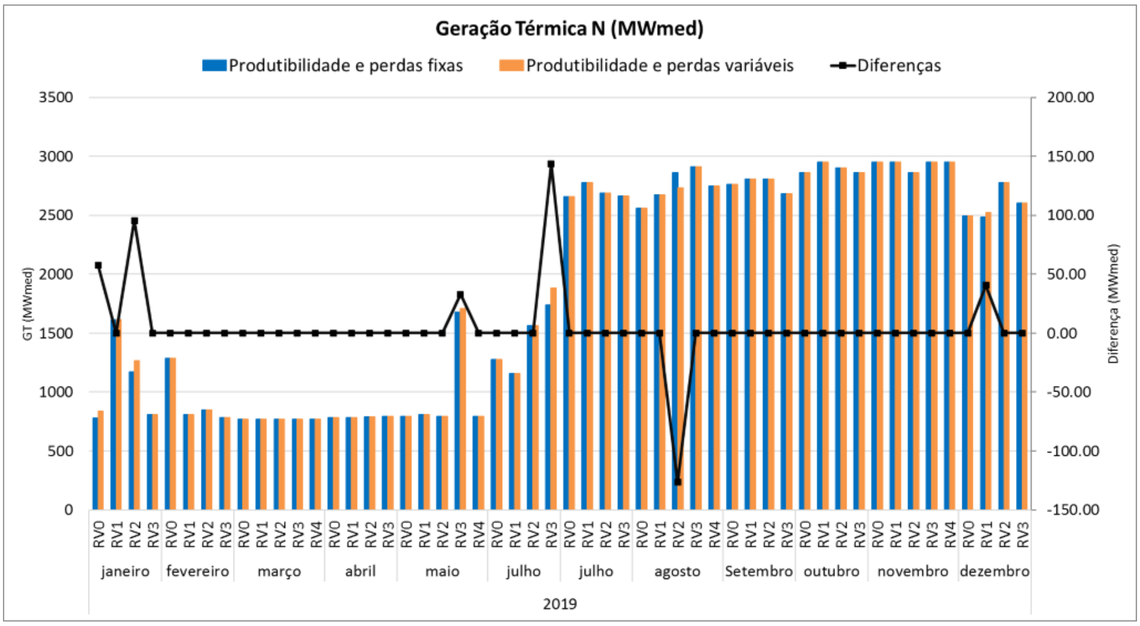

Figura 8.12 – Geração térmica do subsistema N.

9. Conclusões

Frente aos resultados e análises expostos até o momento, o GT-MET concluiu que a metodologia utilizada para ajuste das grades de perdas e produtibilidades específicas representa de maneira adequada a variabilidade destas grandezas frente às diferentes situações operativas observadas na janela analisada neste estudo.

Dando continuidade às avaliações deste subgrupo, as grades serão aplicadas no modelo de planejamento de curto-prazo, DECOMP, e serão analisados os impactos na representação da função de produção das usinas e, consequentemente, na operação do sistema.

10. Referências

[1] ONS, NT 0103-2019 - Consolidação das Atividades Realizadas pelo GTDP – Grupo de Trabalho de Avaliação dos Dados Cadastrais.

[2] CEPEL - Centro de Pesquisas de Energia Elétrica, Manual de metodologia do modelo DECOMP, Abril/2018. (disponível em http://srvlumis02.cepel.br/pt_br/sala-de-imprensa/noticias/documentacao-tecnica-das-metodologias-e-modelos-de-otimizacao-energetica-do-cepel.htm)

[3] CEPEL - Centro de Pesquisas de Energia Elétrica, "Manual de metodologia do modelo DESSEM", Abril/2020 (disponível em http://srvlumis02.cepel.br/pt_br/sala-de-imprensa/noticias/documentacao-tecnica-das-metodologias-e-modelos-de-otimizacao-energetica-do-cepel.htm)

[4] A.L. Diniz, M.E.P. Maceira, "A four-dimensional model of hydro generation for the short-term hydrothermal dispatch problem considering head and spillage effects", *IEEE Trans. Power Syst.*, v. 23, n.3, pp. 1298-1308, Aug. 2008.

[5] Yocogawa L., Brandão L., Diniz A. - Modelagem da Função de Produção das Usinas Hidroelétricas Considerando as Características Operativas das Unidades Geradoras, XXV SNPTEE, Belo Horizonte, nov. 2019.

[6] Wood, S. N. Generalized Additive Models: An introduction with R, Chapman & Hall, 2006.

[7] CEPEL - Centro de Pesquisas de Energia Elétrica, "Metodologia orientada por dados para modelagem da produtibilidade específica e perdas nos condutos das usinas hidrelétricas", Relatório Técnico CEPEL (em elaboração)

[8] CEPEL - Centro de Pesquisas de Energia Elétrica, "Consideração do Engolimento Máximo das Turbinas em Função da Altura de Queda na Representação da Vazão Turbinada Máxima no Modelo DECOMP", Nota Técnica Nº 1 / 2015 , disponível no endereço: http://srvlumis02.cepel.br/pt_br/sala-de-imprensa/noticias/documentacao-tecnica-das-metodologias-e-modelos-de-otimizacao-energetica-do-cepel.htm.

[9] Hastie, T.; Tibshirani, R; Friedman, J. The Elements of Statistical Learning: Data Mining, Inference and Prediction, Springer, 2009.

[10] Ahamada, I; Flachaire, E. Non-Parametric Econometrics, Oxford University Press, 2010.

[11] Wood, S.; Pya, N.; Safken, B. Smoothing Parameter and Model Selection for General Smooth Models, Journal of the American Statistical Association, v. 111, n. 516, pp. 1548-1575, 2016.

[12] Wood, S. N. Package "mgcv", Agosto, 2020.