LABORATORY OF APPLIED MATHEMATICAL PROGRAMMING AND STATISTICS

LAMPS | PUC RIO

Contribuição para CONSULTA PÚBLICA Nº 121 DE 10/02/2022

# Consulta Pública sobre aprimoramentos propostos pela CPAMP

# (ciclo 2021-2022)

Fevereiro 2022

Autor: Alexandre Street, Professor Associado
Departamento de Engenharia Elétrica, PUC-Rio
Diretor e Fundador do *Laboratory of Applied Mathematical Programming and Statistics (LAMPS PUC-Rio)*

# 1 Sumário executivo

Este documento tem como objetivo apresentar algumas contribuições para a CONSULTA PÚBLICA Nº 121 DE 10/02/2022 sobre os aprimoramentos propostos pela CPAMP (ciclo 2021-2022). Esse documento será escrito de maneira breve e objetiva, eventualmente não explicando ou introduzindo jargões largamente utilizados no setor.

Vale ressaltar e parabenizar o grande esforço que todos os membros da CPAMP realizaram para implementar, testar e avaliar todos os casos utilizados nos diversos estudos que permitiram a confecção do relatório técnico dessa chamada pública.

As sugestões de melhoria que serão apresentadas a seguir têm como objetivo incentivar a CPAMP a avançar no seu papel de monitoramento e controle da acurácia dos dados e modelos que impactam em todos os segmentos do nosso setor. Esse controle se faz mais ainda mais crítico e urgente no atual desenho regulatório, onde a visão de futuro e os modelos são centralizadamente geridos. Neste contexto, qualquer melhoria no processo da CPAMP tem o potencial de produzir melhorias em todas as partes do setor elétrico brasileiro.

A seguir, apresento um sumário dos tópicos de contribuição deste documento.

## 1.1 Contribuições

Os seguintes pontos de contribuição foram identificados sobre o documento apresentado na chamada pública:

- Melhoria na precisão da expressão (1) que apresenta a fórmula do PAR(p)-A. Duas fontes de imprecisão foram encontradas e correções sugeridas. Apresento um conjunto de referências e sugestões de aprimoramento para o processo de monitoramento dos modelos de previsão. Trago também algumas reflexões sobre a capacidade de obtermos maior agilidade no processo de gestão dos modelos de previsão e dos dados, nos aproximando minimamente do estado da arte, dentro de um desenho onde modelos e visões são centralizadas.

- Necessidade de métricas de monitoramento do viés das entradas e saídas dos modelos de planejamento. Em especial, a necessidade de acompanhamento contínuo sobre a aderência das previsões de ENAs e das saídas do Newave e Decomp. A aderência das saídas influencia inclusive nos próximos itens.

- Sugestões de melhorias para o processo de avaliação de mudanças na política de planejamento e nas previsões de variáveis operativas e de mercado. Apresento uma discussão sobre os possíveis impactos da falta de métricas de

aderência entre o Newave e a realidade no longo prazo e na expansão e segurança.

- Sugestões de melhorias nas métricas e processos de avaliação e revisão dos parâmetros de aversão a risco. Também apresento reflexões sobre a coerência da métrica atual, os perigos que corremos com ajustes de parâmetros indiretos e com o uso de métricas de risco na ausência de comprovada aderência do Newave à realidade.

Os pontos de contribuição apresentados acima são detalhados em diferentes seções. Nessas seções, todas as vezes em que eu me referir a alguma parte ou seção do relatório técnico produzido pela CPAMP (em consulta pública), especificarei explicitamente "do relatório técnico" logo após a referência. Isso permitirá o leitor diferenciar eventuais referências a partes ou seções anteriores do presente documento, para as quais nada será dito na sequência.

# 2 Imprecisão na expressão (1)

## 2.1 Primeiro ponto de imprecisão

O PAR(p)-A apresentado na expressão (1) do relatório técnico visa separar a média dos lags temporais da ENA dos últimos 12 meses em dois termos. Isso deve ter relação com o uso da expressão no cálculo dos cortes de Benders no modelo Newave, que busca isolar variáveis com mais de um mesmo coeficiente. Entretanto, talvez seja mais interessante, pelo menos na primeira apresentação do PAR(p)-A, que os dois últimos termos estejam agrupados. Assim, fica explícita a novidade com relação ao PAR(p), a saber, a consideração da média móvel anual.

Não obstante, nessa separação, foi encontrada uma imprecisão no terceiro somatório da expressão (1). Ao invés de iniciar em $p_{m_t,i}+1$, outro valor foi utilizado, como mostra a Figura 1. Imprecisões similares foram encontradas em relatórios anteriores.

$$ENA_{t,i} = \sum_{j=1}^{p_{m_t,i}} \phi_{t,j,i}\left(ENA_{t-j,i}\right) + \psi_{t,i}\left(\frac{\sum_{j=1}^{p_{m_t,i}} ENA_{t-j,i}}{12}\right) + \psi_{t,i}\left(\frac{\sum_{j=p_{m_{t+1},i+1}}^{12} ENA_{t-j,i}}{12}\right) + \varepsilon_{t,i} + C_{m.i} \qquad (1)$$

Figure 1: Primeiro ponto de imprecisão na expressão (1).

**Correção:** Nessa expressão, o primeiro termo do terceiro somatório deveria ser $p_{m_t,i}+1$. Com essa correção, os termos do segundo e terceiro somatórios passam

a reconstruir a média móvel dos últimos 12 lags.

**Ponto de atenção na apresentação da expressão (1):** Como essa imprecisão foi encontrada em relatórios metodológicos anteriores, vale verificar se essa imprecisão não está presente no código fonte. Sugiro que o código fonte que sofreu alteração seja publicado em um repositório público para ampla conferência. Imprecisões desse tipo, que aparecem em documentos técnicos, podem também aparecer nos códigos fontes. O problema de uma imprecisão desse tipo é o potencial que ela tem de não ser detectada na compilação do código. Ou seja, o código não apresenta erros ao compilar e passa a ser distribuído e utilizado sem aparentemente apresentar erros, mas os resultados serão inconsistentes. Um exemplo como esse foi detectado (se não me engano) em 2008 no modelo de vazões. A maneira com que a comunidade científica se protege desse tipo de falha humana é através de códigos abertos.

### 2.1.1 Sugestão de aprimoramento

O PAR(p)-A é um caso específico de um modelo mais geral, PAR($p = 12$), onde todos os 12 lags são considerados em todos os meses com coeficientes distintos. Assim, gostaria de destacar os seguintes pontos que podem ser de interesse em investigações futuras:

1. Estimar PAR($p = 12$) utilizando técnicas de regularização. Algumas delas são: LASSO, AdaLASSO (penalização na norma 1)[1], Elastic Net (combinação do LASSO com penalidade na norma 2, *Ridge Regression* – bom para casos com explicativas bastante correlacionadas), pseudo-norma zero (por programação inteira – que regulariza limitando o número de parâmetros diferentes de zero). A utilização de métodos de validação cruzada pode ajudar bastante na seleção de modelos mais robustos, dado que os modelos penalizados anteriormente mencionados são equivalentes a modelos robustos a incerteza nos dados, ou seja, que apresentam, em geral, boa performance fora da amostra.

2. Sugiro a seguinte literatura: Para métodos modernos de regularização e robustificação dos modelos, ver [1] e [2]. Para uma literatura mais geral sobre modelos estatísticos e métodos de machine learning aplicados a séries temporais, ver [3]. Para modelos não Gaussianos com parâmetros variantes no tempo aplicados a séries de geração eólica e ENA, ver [4] e [5]. Para o uso de

[1] LASSO é o acrônimo para *Least Absolute Shrinkage and Selection Operator* e AdaLASSO para sua versão adaptativa, onde diferentes pesos são dados aos diferentes coeficientes.

modelos não paramétricos aplicados a séries de geração eólica, ver [6]. Para um método capaz de realizar previsões probabilísticas de longo prazo em escala mensal e horária, ver os dois primeiros modelos da cadeia de modelos apresentada em [7].[2] Para uma revisão da literatura sobre previsão probabilística, ver [8]. Por fim, para uma ampla revisão da literatura em previsão de renováveis, ver [9].

3. Estudo de variáveis exógenas associadas a previsões climáticas de longo prazo envolvendo sinais de El Ñino, La Ñina, Oscilações Decadais, dentre outras que possam explicar a aparente quebra estrutural nas séries de ENA. Além disso, estudo de explicativas geradas a partir de previsões de métodos de machine learning e modelos não paramétricos podem ser bastante interessantes. Em [8], pode se ver que os principais grupos vencedores de competições de previsão utilizam modelos de regressão quantílica, como o proposto em [6], considerando como variáveis explicativas previsões de modelos de machine learning. É de grande importância que a CPAMP monitore e incentive esforços efetivos nessa área (previsão), que deve se dar primordialmente através de uma maior colaboração, com projetos, consultorias e workshops, com *experts* nacionais e internacionais de reconhecida competência acadêmica e prática. Me coloco à disposição para sugerir uma lista de nomes caso seja do interesse da CPAMP.

Por fim, sugiro a seguinte reflexão. Será que todas essas possibilidades listadas acima, que constituem hoje o estado da arte dos métodos de previsão de recursos renováveis, não nos trazem evidências de que o modelo centralizado é incapaz de gerenciar todos esses dados e todas essas técnicas. Refletir se justamente o fato de continuarmos investigando o modelo PAR(p)-A (estruturalmente ultrapassado e que não utiliza nenhum tratamento nos dados ou explicativas para melhorar as inconsistências nas previsões devido à mudança de regime) em 2022, tendo iniciado essa investigação em 2017, não nos demonstra que o modelo que estamos utilizando não é capaz de colocar o que há de melhor a serviço do setor. Lembrar que o problema real que queremos resolver não envolve um processo PAR(p) ou PAR(p)-A. O modelo real é inacessível e altamente complexo. E a maneira mais eficiente que a humanidade já inventou para endereçar problemas com esse grau de complexidade foi descentralizar as visões entre os agentes e criar desenhos que alinhem a busca por melhores modelos e processos de gestão de dados com o benefício econômico dos agentes. Sugiro ampla reflexão sobre esse tema, pois o

[2] Este modelo foi desenvolvido em um projeto de P&D com a Energisa está sendo aperfeiçoado em um segundo projeto atualmente em curso. O aperfeiçoamento expande o modelo anterior par considerar diferentes regimes (safra ou fora safra no caso de biomassas, dia ou noite no caso de fotovoltaicas, etc.). Isso pode ser particularmente interessante para permitir incorporar efeitos de incertezas e restrições de curto prazo no Newave.

problema real que queremos resolver está bem longe do problema que resolvemos no Newave. E talvez seja melhor resolver um problema mais próximo do real através de heurísticas individuais em um ambiente onde os agentes competem por eficiência na previsão do que resolver perfeitamente um problema irreal de forma centralizada e com viés de uma única visão.

## 2.2 Segundo ponto de imprecisão

Os coeficientes $\phi$, $\psi$, e $C$ sublinhados na Figura 2 possuem índices inconsistentes e imprecisos que precisam ser ajustados.

$$ENA_{t,i} = \sum_{j=1}^{p_{m_t,i}} \underline{\phi_{t,j,i}} \left(ENA_{t-j,i}\right) + \underline{\psi_{t,i}} \left( \frac{\sum_{j=1}^{p_{m_t,i}} ENA_{t-j,i}}{12} \right) + \underline{\psi_{t,i}} \left( \frac{\sum_{j=p_{m_{t+1},i+1}}^{12} ENA_{t-j,i}}{12} \right) + \varepsilon_{t,i} + \underline{C_{m.i}} \qquad (1)$$

Figure 2: Segundo ponto de imprecisão na expressão (1)

Os dois primeiros coeficientes, $\phi$, $\psi$, não deveriam ser indexados por $t$, mas sim por $m_t$. Isso porque eles não podem assumir qualquer valor ao longo de toda a série temporal como em modelos de coeficientes variantes no tempo. No caso da família PAR(p), os coeficientes estão restritos a uma variação periódica (com coeficientes diferentes para cada mês do ano). Obviamente que isso resulta em coeficientes que variam no tempo, mas tal variabilidade deveria se dar de maneira consistente com a notação, ou seja, com $m_t$ ao invés de $t$. Já o terceiro coeficiente, $C$, não deveria ser indexado por $m$, mas também por $m_t$. Vale ressaltar que o fato de existirem múltiplas configurações pode ser endereçado por um índice adicional que define a configuração vigente do período $t$, por exemplo, $c[t] \in \mathcal{C}$, onde $\mathcal{C}$ é o conjunto de configurações. Assim, estima-se o PAR(p) ou PAR(p)-A para cada configuração $c \in \mathcal{C}$ e posteriormente podemos nos referenciar ao coeficiente correto através desse novo índice.

**Correção:** Dessa forma, os coeficientes da expressão (1) deveriam ser: $\phi_{i,j,m_t}$, $\psi_{i,m_t}$, e $C_{i,m_t}$. Aqui, ignorando as diferentes configurações.

### 2.2.1 Sugestão de aprimoramento

Sugiro que os índices quando utilizados sigam uma ordem alfabética, $i, j, m$, e que o índice do mês do período $t$, $m_t$, seja substituído por uma notação vetorial, $m[t]$. Isso tornaria a expressão muito mais fácil de se compreender, evitando subíndice de subíndice.

# 3 Sobre o objetivo das análises e as métricas de controle

Do sumário executivo do relatório técnico destaco e comento as seguintes passagens:

1. **Página 8, segundo parágrafo:** *"As análises e estudos presentes no documento foram feitos tendo em vista que em 02/09/2020 o Comitê de Monitoramento do Setor Elétrico (CMSE) deliberou para que a CPAMP avaliasse mecanismos visando elevação estrutural dos níveis de armazenamento dos reservatórios das usinas hidrelétricas, sobretudo aos finais dos períodos secos, bem como que propusesse transição capaz de minimizar os impactos no GSF (Fator de Ajuste do MRE – Mecanismo de Realocação de Energia) e na tarifa do consumidor de energia elétrica."*

2. **Página 8, terceiro parágrafo:** *"Neste contexto, a metodologia de aperfeiçoamento do modelo GEVAZP1 (modelo de geração de séries sintéticas de energias e vazões), denominada PAR(p)-A, traz melhorias importantes para a representação da hidrologia recente na geração de cenários de vazões e de Energias Naturais Afluentes (ENAs). Constatou-se, ainda, que o modelo proposto tende a gerar cenários continuamente mais extremos (positivos e negativos) ao longo do horizonte estudado, em comparação ao modelo vigente (PAR(p)). Destaca-se também que testes estatísticos formais indicaram que o PAR(p)-A reproduz as caraterísticas do histórico de ENAs e que tem ganhos quando comparado com o PAR(p)."*

A conexão entre a o objetivo das instruções dadas pelo CMSE (para a busca de mecanismos que elevem estruturalmente os reservatórios e mitiguem os impactos no GSF), destacada no item 1, e a metodologia sendo avaliada nesta consulta pública, introduzida no parágrafo destacado no item 2, pode nos trazer bons *insights* se explorada mais profundamente. Esse ponto é de grande importância, pois o reconhecimento da causalidade entre a proposta de melhorias nos modelos de afluência e a elevação estrutural dos reservatórios é crucial por diversos fatores que explico ao longo dessa seção. Além disso, essa conexão nos permite ver que existem outras demandas prioritárias, que se não endereçadas, prejudicam a avaliação do PAR(p)-A, da segurança sistêmica, dos os impactos no PLD, nas tarifas, e da própria mudança nos parâmetros de risco.

Se reconhecemos que a instrução de elevação dos níveis de reservatório pode ser parcialmente endereçada através de um aperfeiçoamento no modelo de previsão

de vazões utilizado no Newave (cuja função é calcular o valor da água no médio e longo prazo), estamos dizendo que existe, em alguma extensão, uma causalidade entre os níveis baixos dos reservatórios observados nos últimos anos e um eventual viés (erro persistente) de previsão no PAR(p). O reconhecimento explícito dessa lógica esclarece, dentre outas coisas, como devemos medir e controlar a qualidade das previsões para endereçar a parcela da questão levantada pelo CMSE (item 1) que concerne aos modelos de vazão.

Nesse contexto, é importante lembrar que o esvaziamento dos reservatórios é fruto de um acumulado de ordens de despacho decorrentes de toda uma cadeia de modelos, que por sua vez utilizam as previsões de ENAs para um horizonte de 1 a 60 meses à frente. Assim, precisamos entender um pouco mais sobre esses modelos. Essa cadeia é constituída pelo Newave, que é utilizado para calcular a função de custo futuro (FCF) ao final do mês subsequente ao mês corrente da operação até o final de um horizonte de 60 meses (com etapas mensais e patamares de carga), pelo Decomp, que é utilizado para calcular a FCF para o final da semana subsequente à semana de operação (com etapas semanais) e pelo Dessem, que é utilizado para definir a programação do despacho de curto prazo, indo do dia seguinte (com etapas semi-horárias) até o resto da semana operativa (com etapas diárias, não cronológicas, por patamar de carga). Dessem recebe, portanto, ao final da semana operativa, a FCF do Decomp que contempla toda a informação de sensibilidade e custo de 1 a 60 meses à frente. Vale ressaltar aqui que o valor da água, que tem a interpretação de "custo do combustível água", ou custo marginal de oportunidade do uso da água, pode ser obtido como o negativo da derivada da FCF. Dentro dessa cadeia, o GEVAZP é o software utilizado para gerar previsões das ENAs no Decomp e Newave através do modelo de séries temporais PAR(p), ou PAR(p)-A. Ou seja, no final do dia, estamos discutindo o efeito no volume armazenado nos reservatórios (e outras variáveis importantes como o PLD, geração térmica, etc.) decorrentes de um modelo de previsão de ENAs que afeta a visão de futuro, caracterizada pela FCF, sobre a operação do sistema do mês $m+1$ até $m+60$ (onde $m$ é o mês corrente).

De maneira mais objetiva, a lógica de funcionamento dos modelos nos levam a crer que se o PAR(p) prevê sistematicamente acima do realizado, especialmente quando olhamos para horizontes de 3 a 60 meses à frente, o Newave carregará todo o acumulado desse erro de aporte adicional de água para a FCF. Para o leitor interessado, sugiro ver a discussão e os experimentos preliminares em [10]. Além disso, vale notar as fortes evidências desse viés na Figura 7 do relatório técnico, onde desde 2015, a ENA (linha azul) é sistematicamente inferior à MLT (linha pontilhada), valor para o qual as previsões do PAR(p) convergem poucos passos à frente. Dessa forma, a FCF que o Dessem recebe do Newave e Decomp pode estar trazendo "um conto de fadas" sobre o futuro do custo operativo (viés otimista,

com valor da água subestimado). Tal visão faz com que o Dessem considere em suas decisões que a água que hoje por ventura esteja escassa, chegará nos meses futuros próxima à média histórica de longo prazo. Mas esse problema não ocorre somente por causa do viés otimista nas previsões de vazão. As restrições individuais dos diversos reservatórios, as restrições da rede elétrica, as restrições horárias de rampa e *unit commitment* não são consideradas no cálculo da FCF, ainda que todos saibam que tais restrições tornam a operação altamente complexa e cara. É nesse contexto, de alta complexidade hoje e (em parte falsa) perspectiva de flexibilidade, abundância e simplicidade no futuro, que as ordens de despacho das hidrelétricas são mantidas pelo Dessem, mesmo quando todos já sabem que deveríamos armazenar a água e ligar algumas térmicas[3].

Assim, qualquer avaliação de aderência de modelos de previsão de ENA precisa considerar métricas de erro (entre o realizado e o previsto) para um horizonte de 1 a 24 meses à frente, pelo menos. Isso porque, de acordo com nosso objetivo e com a lógica da relação entre níveis de reservatórios e previsão das vazões nos modelos que calculam o valor da água, a integral desse erro desempenha um papel muito relevante. Nesse sentido, os testes estatísticos mencionados no paragrafo destacado no item 2 poderiam contemplar métricas que quantificassem o viés de longo prazo de ambos os modelos, PAR(p) e PAR(p)-A. Isso alinharia a medida de qualidade das previsões de ENA com o papel que elas desempenham dentro da cadeia de modelos do setor até chegar nos níveis dos reservatórios.

**Dito isso, parece preocupante que, até hoje, não tenhamos implementado nenhum estudo ou monitoramento sistemático, através de relatórios ou *dashboards* públicos, sobre a aderência entre o que se planeja**

[3]É importante não cair na tentação de buscar a solução desse problema de "visão otimista" aumentando o nível de aversão a risco do Newave. Esse problema que acabo de descrever não é uma problema de perfil de risco, é um problema de modelagem da realidade. Existem diversos perfis de despacho decorrentes de restrições elétricas, de *unit commitment*, etc, que jamais seriam obtidos com o Newave, qualquer que seja o nível de aversão a risco. Para ilustrar, imagine que você enxergue o mundo de forma totalmente pixelada, com uma resolução tão baixa que tudo que esteja a mais de poucos centímetros de distância seja enxergado por você como blocos com faces de 1 $m^2$, cada uma contendo apenas uma única cor. Nessa situação, jamais você conseguirá identificar, por exemplo, um degrau de escada. Neste exemplo, estaríamos sempre tropeçando por conta da nossa visão sobre os percalços que o mundo nos apresenta. Ou, trazendo um pouco mais da dinâmica dos modelos do setor para esse exemplo ilustrativo, você estaria constantemente levando grandes sustos ao se deparar com calçadas e bueiros abertos entrando no seu campo de visão de poucos centímetros à frente; equivalente ao operador passando a enxergar a diferença entre a realidade do dia seguinte através do Dessem e a previsão que os Newaves dos meses e anos anteriores traziam. Aumentar a aversão a risco neste caso pode ser inclusive ineficiente frente ao que se poderia fazer aumentando a resolução da visão de futuro. Assim, se realmente queremos avançar na gestão eficiente dos recursos eletro-energéticos, não adianta aumentar o nível de aversão a risco no Newave sem antes aumentar minimamente a resolução da sua visão da realidade.

**com o Newave e o que se implementa na operação; mesmo debaixo de todas as evidências como os volumosos despachos fora da ordem de mérito e as enormes distorções reportadas em artigos e observadas na prática (ver [10]). Podemos estar calculando o valor da água de maneira completamente desconectada da realidade e não temos a menor ideia. Medir e controlar essa discrepância é vital, pois o sinal que o Newave passa para o Dessem através da sua FCF acumula a soma de todo o viés otimista embutido nos custos operativos do segundo mês até o final do horizonte de 60 meses. Essa ausência de controle e transparência pública se estende a métricas de performance já bastante discutidas, como probabilidade de racionamento, dentre outras. O setor precisa ter mais transparência nas suas decisões de operação, até mesmo para imunizar decisões de um possível viés político conjuntural.** Ver "A crise energética de 2015", Valor Econômico (Fev. 2015), "Hora de rever a operação e expansão do setor elétrico", Valor Econômico (Nob. 2015) e "Métricas de avaliação e monitoramento", Canal Energia (Jan. 2019) em [11]

Em [10], destaco um conjunto de discussões e recomendações de construção de *dashboards* de acompanhamento de entradas (ENAs) e saídas (variáveis importantes previstas pelo Newave) enviado à CPAMP em janeiro de 2022. Neste link, o leitor encontrará uma discussão e reflexão mais ampla sobre os temas aqui colocados. Copio a seguir apenas o trecho relativo ao email enviado para a CPAMP que coloca a necessidade do monitoramento do viés de previsões das ENAs e do planejamento do Newave:

> *"Seria interessante termos um monitor tipo dashboard que a CPAMP poderia todo mês acompanhar, deixar público e até utilizar de termômetro para demandar aperfeiçoamentos. Com o tempo ele poderia incluir outras métricas importantes, também já fruto de muita crítica, como a probabilidade de racionamento, etc.*
>
> *Para os Inputs: cálculo do viés (erro entre previsto e realizado) 1 a 12 passos à frente para todas as ENAs. O que entra no Newave para cálculo do valor da água não é uma previsão 1 passo à frente, mas de 1 a 60 passos à frente. E é esse erro, ou o viés nessas previsões de k passos à frente é que importam. O modelo pode estar ótimo 1 passo à frente mas péssimo 6 passos à frente. Monitorar isso é vital. Esse resultado poderia ser acompanhado através de 3 gráficos:*
>
> - *um gráfico de barras contendo a média de janeiro de 2011 até hoje (eixo x - k passos à frente, y o erro médio).*
> - *mesmo gráfico, porém considerando apenas os últimos 3 anos.*

- *um gráfico das séries temporais (x é tempo aqui) dos erros evoluindo no tempo nos últimos 3 anos.*

*Para os outputs: cálculo de erro entre o planejamento e realizado. Selecionar variáveis relevantes do planejamento (armazenamento, geração térmica, hídrica e CMO) realizado pelo newave de 12 meses atrás para os último 6 meses e comparar com o que de fato foi implementado nesses últimos 6 meses. Esses 6 valores de erro (e eventualmente o total acumulado também) poderiam ser monitorados temporalmente e gráficos como os dos pontos 1 a 3 acima poderiam ser utilizados. Esse cálculo poderia ser dividido em dois: a) com os dados de carga, expansão, e tudo mais que é fixo no modelo (premissas) que foram de fato utilizadas no planejamento 12 meses atrás e b) com as premissas ajustadas para o histórico (dado observado). Em ambos os casos, a conta deveria ser feita com o newave sendo rodado para o histórico de ENA realizado. A diferença entre a e b também nos mostraria se esses valores que consideramos fixos contribuem pouco ou muito, com viés ou não, para o erro."*

## 3.1 Sugestões de melhoria

1. Apresentar a motivação da análise a partir de um objetivo (como, por exemplo, a orientação do CMSE destacada no item 1 desta subseção).

2. Apresentar a lógica que motiva a alteração dos modelos de ENA em função do impacto que suas previsões produzem no objetivo selecionado.

3. Desenhar experimentos controlados preliminares e métricas que corroborem essa visão.

4. Fundamentar e validar essa lógica com especialistas isentos, de reconhecido renome e comprovada contribuição para área em questão.

5. Selecionar especialistas nacionais e internacionais, fora do âmbito do CEPEL (que nesse caso desempenha o papel de fornecedor do software), que possam contribuir com ideias externas para o CEPEL e CPAMP no processo de compreensão do problema e validação das soluções.

6. Realizar workshop com especialistas e sociedade/agentes que permitam uma discussão sobre como endereçar o problema de maneira efetiva. Com base nos resultados do workshop e estudos realizados, propor métricas e fundamentos para a avaliação dos modelos de previsão.

7. Realizar esses estudos, apresentações, contratações de consultores externos (quando necessário) e workshops com maior participação dos agentes e garantir o envolvimento da academia (inspirar-se no modelo Chileno de avaliação e validação de mudanças regulatórias).

8. Vale a pena pensar em duas cadeiras para membros da academia na CPAMP. Inicialmente como parte de um conselho consultivo e posteriormente como membro integrante. Estamos falando de modelos, processos e estudos de grande complexidade e a CPAMP ganharia bastante respaldo.

9. Criação de *dashboards* públicos (e outros produtos que criem legados) de acompanhamento que permitam os controles criados serem observados semanalmente. Criar processos oficiais (como procedimentos de rede) que utilizem essa métricas para que ações sejam tomadas. Os próprios *dashboards* podem implementar os processos e alertar sobre a violação de limites estabelecidos pelo ONS, CCEE, EPE e CMSE.

# 4 Sobre mudanças na política de planejamento e previsões de variáveis operativas e de mercado

Enquanto que o impacto da mudança do modelo de previsão de ENAs nas garantias físicas (GF) faz sentido de ser avaliado pelas próprias simulações do Newave, pois as GFs são geradas pelo próprio Newave, a avaliação ou previsão do impacto de qualquer mudança na FCF no PLD e em variáveis operativas, em geral, não devem ser realizadas com os resultados do próprio Newave. O relatório técnico já traz, de certa forma, essa percepção ao realizar estudos com rodadas encadeadas entre o Newave e o Decomp para apresentar resultados de impactos no PLD e na tarifa. Contudo, é importante avançar nesse processo pois mesmo o encadeamento entre o Decomp e o Newave ainda tem o potencial de produzir largas discrepâncias com relação ao que será realizado. Isso porque nem o PLD e nem a operação são determinados pelo Newave ou pelo Decomp. Note que i) o Newave e o Decomp simplificam aspectos importante do sistema (como a transmissão, critério de segurança, rampas, e o *unit commitment*) e ii) o acumulo dos desvios decorrentes dessas simplificações podem levar, poucos meses à frente, a um estado final de armazenamento bastante diferentes daqueles que são obtidos pela sequência de modelos que é empregada na prática hoje – onde a operação considera aspectos detalhados do sistema no curto prazo (via Dessem), mas utiliza uma FCF altamente simplificada calculada pelo Decomp e Newave (ver [10] para uma ampla discussão e larga fonte de material técnico que suportam essa visão).

Para fins de comparação de variáveis operativas ou de preços, precisamos de

um simulador de referência. Um simulador que possua garantias de aderência. Caso contrário, corremos o risco de estamos comparando impactos operativos e de preços em sistemas totalmente desconectados da realidade. Conforme comentado anteriormente, o caso da garantia física, por conta da regulação vigente determinar o uso do Newave para o seu cálculo, faz sentido utilizar esse software para comparações. Mas neste ponto, vale refletir se a mesma discussão sobre a aderência dos resultados do Newave e a realidade operativa não se aplicaria também às garantias físicas. Faz sentido calcular números tão relevantes para a segurança do nosso setor sem que exista um controle e monitoramento sistemático de aderência entre os cenários operativos preditos pelo Newave e a realidade? O que as garantias físicas representam sem esse controle de qualidade? Como seriam as garantias físicas se elas fossem calculadas com cenários mais realistas? Acredito que exista um excelente dever de casa a ser realizado nesse sentido que tem o potencial de impactar em uma série de problemas patológicos do setor, como o GSF, segurança, etc.

## 4.1 Sugestão de melhoria do processo de avaliação de novas políticas operativas

De forma a conferir um nível de confiança às avaliações realizadas entre os modelos, é importante que seja criado um processo de avaliação de referência que seja amplamente testado e agnóstico à FCF sendo avaliada. Nesse sentido, deveríamos criar um simulador e um conjunto de dados históricos para os quais ao "plugarmos" uma FCF, as simulações resultantes teriam certas garantias de aderência à realidade. Isso permitiria avaliar, com um dado nível de confiança, o impacto em variáveis operativas (volumes, risco de racionamento, etc), no PLD e nas tarifas a partir de qualquer proposta de mudança no Newave ou Decomp, em seus modelos auxiliares, ou em seus dados. A maneira de se implementar esse processo pode variar dependendo da cadeia de modelos sendo utilizada e de possíveis estratégias de economia de esforço computacional. Por exemplo, no contexto atual do setor, esse processo de simulação poderia ser implementado encadeando os modelos Newave-Decomp-Dessem para o horizonte de estudo, caso ficasse claro que essa trinca de modelos produz despachos suficientemente próximos ao que se implementa, utilizando para isso métricas não muito agregadas. A cadeia de modelos seria então aplicada a uma série de conjuntos de dados, cada um considerando períodos cronologicamente acoplados e horizonte suficientemente grande para que o efeito da composição de erros possa ser avaliado, e contemplando diferentes situações de estresse do sistema. Neste contexto, seria possível medirmos de maneira $\epsilon$–acurada (com certificado de qualidade conhecido, definido por um parâmetro de erro máximo $\epsilon$) o real benefício de trocarmos partes relevantes que determinam em grande parte a performance

operativa e dos preços em um horizonte de longo prazo.

É claro que esse processo é computacionalmente custoso e trabalhoso de ser implementado. Não obstante, entendo que é de responsabilidade das entidades que desenvolvem os modelos (CEPEL), que monitoram os seus usos (CPAMP) e que impõe a sua utilização (MME) ao ONS, CCEE e EPE investir e endereçar essa importante tarefa, que apesar de ser trabalhosa, não é impossível. As práticas e procedimentos computacionais adotados para se realizar previsões mais acuradas para o PLD no segmento de comercialização podem ser um primeiro passo nesse sentido.

Alternativamente, esse processo pode ser pensado e desenhado através de um P&D em duas fases. Na primeira haveria um estudo amplo realizado por uma entidade executora de pesquisa em conjunto com agentes qualificados e com as entidades do setor (CPAMP, EPE, ONS e CCEE) para a elaboração de uma proposta inicial e testes computacionais preliminares. Na segunda fase, seria selecionado uma entidade para implementar o procedimento proposto na fase um em código aberto e realizar a instalação e manutenção das aplicações que seriam construídas em nuvem para a CPAMP, EPE, ONS e CCEE. *Dashboards* públicos deveriam ser criados e alimentados com os estudos realizados por esses agentes para todo o setor. O código estaria aberto para todos os agentes poderem replicar em suas próprias infraestruturas caso necessário. Esse processo permitiria não só os estudos que interessam para fins dessa consulta pública, mas também a uma gama de outras finalidades, da geração à distribuição. Logo, a verba de PD e um produto em código aberto seriam mais do que justificáveis e poderiam ser obtidos de um pool de agentes, que passariam a ter acesso a um simulador das variáveis operativas e de preços de alta precisão.

# 5 Sobre a reavaliação do perfil de aversão a risco

1. **Página 8, quarto parágrafo:** *"Dada a melhor representatividade hidrológica dos cenários gerados pelo PAR(p)-A, observou-se modificação no comportamento das principais variáveis operativas, tornando importante reavaliar a percepção de risco do setor frente a este comportamento mais aderente ao observado no histórico de afluências."*

2. **Página 8, primeira frase do quarto parágrafo:** *"Dessa forma, este relatório contempla simulações que incorporam aos modelos o aprimoramento da representação hidrológica empregando o PAR(p)-A, avalia a necessidade de alteração do critério de parada e recalibração dos atuais parâmetros* $\alpha = 50\%$ *e* $\lambda = 35\%$ *do CVaR para serem utilizados nos processos de planejamento da operação e formação de preço a partir de janeiro de 2023."*

Conforme a discussão na nota de roda-pé da página 9 deste documento e as discussões anteriores sobre a acurácia do modelo Newave, a avaliação do perfil de risco com base em comparações de simulações do Newave (ou encadeamento do Newave com Decomp) me parecem ter pouca sustentação prática[4]. Isso porque 1) não existem estudos controlados de aderência entre as previsões de longo prazo (muitos passos à frente) do Newave ou Decomp e a realidade implementada e 2) os estudos de aversão a risco são baseados justamente na avaliação de um risco medido por variáveis operativas desses softwares. Nesse contexto, qualquer estudo operativo e de definição de parâmetros de aversão a risco torna-se questionável.

Vale ressaltar que a medida de risco utilizada hoje no Newave é baseada no *Conditional Value at Risk* (CVaR) encadeado aplicado ao custo da função objetivo. Porém, como essa métrica não possui uma interpretação direta sobre as variáveis que o operador é capaz de opinar e de definir limites de segurança (por exemplo, o volume armazenado ou déficit), todos os estudos de calibração dos parâmetros de aversão a risco ($\lambda$, da combinação convexa entre o CVaR e o valor esperado e o nível de confiança, $\alpha$, do CVaR) são realizados de maneira indireta. Basicamente os estudos buscam pares $(\lambda, \alpha)$ que produzam custos esperados (médios) e níveis de risco aceitáveis no longo prazo. Contudo, toda a discussão a respeito da ausência de procedimentos de monitoramento da aderência entre os resultados de longo prazo que o Newave produz e a realidade tornam essas análises também questionáveis e sem comprovação de acurácia. Além disso, pelo princípio da eficiência entre "risco e retorno", um menor custo de operação esperado deve vir acompanhado de um maior risco (por exemplo, violação dos volumes mínimos ou cortes de carga). Assim, se escolhemos os parâmetros $(\lambda, \alpha)$ de tal forma que o custo esperado seja minimizado, essa otimização deverá levar a pares que coloquem o risco no limiar da região aceitável. Nesse contexto, se as evidências apontadas em [10] se fizerem presentes e os resultados do Newave estiverem otimistas com relação à realidade, existe a chance de selecionarmos parâmetros que na prática nos colocam fora da região aceitável de risco. Isso porque, sob uma visão otimista, o risco estaria no limiar da região aceitável, mas na realidade mais complexa e adversa, os recursos disponíveis para uso nessa política não seriam suficientes para sustentar uma operação dentro do limiar aceitável de risco.

## 5.1 Sugestão de melhoria

Na minha opinião, a maneira indireta que utilizamos para quantificar o risco cria processos complexos e utiliza uma parametrização que pode não capturar políticas relevantes para o operador. Além disso, como a aversão a risco é endógena ao

[4]Esse ponto é um problema antigo, que vem da gênesis do atual modelo de aversão a risco adotado no Newave.

Newave (está programada nele e utiliza as suas variáveis sendo otimizadas), este precisa, por definição, ter uma garantia de aderência com a realidade para que a métrica de risco nos permita alguma interpretação prática. Caso contrário, estamos com um termômetro com uma escala desconhecida. O primeiro passo é medir a discrepância, como já sugerido nas seções anteriores. O segundo passo é buscar modelagens que capturem, ainda que de forma simplificada (mas de maneira pessimista), aspectos relevantes da operação prática que permitam reduzir a métrica de viés entre planejamento e operação. A consideração de restrições de rede, incertezas e restrições de curto prazo, e reservatórios individualizados é *sine qua non* e não depende de nenhuma ciência desconhecida dos técnicos do setor. Essas *features* estão disponíveis em softwares comerciais de prateleira que podem ser utilizados de benchmark. Por último, é fortemente recomendado que o processo de aversão a risco não seja realizado a partir de uma métrica indireta.

A superfície de aversão a risco (SAR) e o controle do corte de carga com restrições locais de risco, por exemplo, me parecem critérios muito mais práticos, mais coerentes com as variáveis costumeiras da operação, e evitaria uma gama de complicações que temos hoje. Não obstante, a SAR não pode ser simplesmente vista como alternativa ao CVaR na função objetivo. Ela deve vir acompanhada de garantias quantitativas de que, para um dado cenário de ENA, carga e demais premissas, os armazenamentos dos reservatórios previstos para um ano à frente dentro do Newave são críveis caso aquele cenário ocorra. Ou seja, precisa vir acompanhado do incremento do realismo/acurácia na representação do sistema comprovado por estudos controlados. Caso contrário, ainda estaremos achando hoje, de maneira otimista, que atenderemos à uma SAR do ano que vem e quando chegar lá a realidade será diferente da prevista.Note que quando digo "realidade diferente da prevista" nesse parágrafo não me refiro somente a um erro estatístico gerado por viés de previsão das ENAs. Me refiro também a um erro decorrente da modelagem que o Newave utiliza para prever a operação do sistema ao longo de um mês. Se eu sei exatamente a carga, a ENA, a expansão, a geração renovável, e todas as disponibilidades de todos os equipamentos, o Newave deveria prever sem viés a quantidade de geração térmica e hidrelétrica de cada reservatório ou susbsistema. Assim, me refiro à parcela do erro de previsão decorrente da acurácia do Newave.

Por fim, o processo de atualização dos parâmetros modifica, o que será implementado no futuro. Portanto, tem potencial de introduzir mais inconsistência entre o planejamento e a operação e grandes solavancos no PLD e no despacho térmico. Nesse sentido, é importante que no mesmo mês em que uma data futura seja programada para a mudança dos parâmetros, o Newave desse mês já passe a considerar os novos parâmetros a partir da data especificada. Por exemplo, caso dia 31 de abril de 2022 seja deliberado que em janeiro de 2023 novos parâmetros

de risco serão utilizados, o PMO de maio de 2022, que contempla os próximos 5 anos, já deve considerar os novos valores para os parâmetros de risco de janeiro de 2023 em diante.

# 6 Discussão final sobre os aperfeiçoamentos propostos

Gostaria de iniciar destacando a consagrada frase de William Thomson – ou Lord Kelvin, como ficou mais conhecido – ***"Aquilo que não se pode medir, não se pode melhorar"***, que se faz absolutamente presente em todas as discussões que permearam este documento até aqui. **Essa frase deveria ser vista como um princípio para todas as entidades que utilizam modelos no nosso setor, principalmente pela enorme dependência que o desenho regulatório atual possui com uma única cadeia de modelos e uma única visão de futuro decorrente dos dados utilizados nestes modelos. Essa enorme dependência de um único modelo e visão cria para todas essas instituições uma enorme responsabilidade na gestão de dados e modelos.**

O nosso setor tem uma larga tradição na disciplina de otimização. Essa tradição nos permitiu grandes avanços metodológicos. Contudo, precisamos agora aumentar nosso contato com outras disciplinas, a saber, a Estatística e *Machine Learning*. Precisamos introduzir nos processos do setor, que utilizem ou não otimização, a ideia de que todos os modelos precisam ter controle de aderência, ou erro com relação aos dados/medições que tem como objetivo prever. Não podemos criar modelos, enchê-los de dados nominais, e esperar que os seus resultados sejam consistentes com a realidade. Os modelos são meras aproximações de uma realidade muito mais complexa. E os dados precisam ser tratados, monitorados e validados para que os resultados desses modelos nos levem para onde queremos[5]. Assim, todos os modelos do setor devem ser monitorados em termos de acurácia com dados tratados e bem gerenciados. Inclusive, alguns dos parâmetros e premissas dos modelos que utilizamos podem ser ajustados para reduzir o erro entre a previsão do planejamento e a realidade observada. Essa é a ideia por trás de muitos modelos de *Machine Learning*, ajustar adaptativamente os parâmetros para reduzir o erro com a realidade. É importante, contudo, saber desenhar os experimentos de controle e ajuste de parâmetros. Essa tarefa requer uma renovação na maneira de pensar do setor sobre o uso e papel dos modelos.

Acredito que para os casos abordados no relatório técnico, o primeiro passo

[5]Existem vários exemplos de dados cadastrais completamente desatualizados e não aderentes (lembrar do caso das produtibilidades das hidros) e outros que parece que esqueceram no modelo, como o fator de desconto no tempo.

mais imediato, ainda pensando dentro do desenho centralizado atual do setor, seja implementar processos de monitoramento e métricas de aderência como sugerido na Seção 3 deste documento. Contudo, acredito que esse processo deveria ser realizado de maneira realmente profissional em termos de gestão, processos e *compliance*. Isso deveria ser rapidamente implementado e com o auxílio de consultorias especializadas de renome mundial e, preferencialmente, em forma de sistemas e *dashboards* de livre acesso no site do operador. Adicionalmente, ainda dentro do arcabouço regulatório atual, o desenvolvimento do simulador proposto na Seção 4.1 poderia constituir um segundo passo de melhoria que permitiria ganhos em diversas atividades do setor. Contudo, mudanças mais profundas e conceituais podem envolver 1) a avaliação de propostas de quebra de monopólio no fornecimento dos modelos computacionais por parte do CEPEL e transferência de responsabilidade de escolha dos softwares para o ONS (ver discussão sobre essa ideia e seus fundamentos no artigo "A inconsistência do setor e a repetida crise", Valor Econômico (Jun. 2021), disponível em [11]) e 2) avançar na agenda de discussões sobre a descentralização dos modelos e visões através de um mercado por ofertas de preços e quantidade para o curto prazo.

# References


[1] D. Bertsimas and J. Dunn, *Machine learning under a modern optimization lens*. Dynamic Ideas LLC, 2019.

[2] J. Blanchet, Y. Kang, and K. Murthy, "Robust wasserstein profile inference and applications to machine learning," *Journal of Applied Probability*, vol. 56, no. 3, pp. 830–857, 2019.

[3] R. P. Masini, M. C. Medeiros, and E. F. Mendes, "Machine learning advances for time series forecasting," *arXiv preprint arXiv:2012.12802*, 2020.

[4] H. Hoeltgebaum, C. Fernandes, and A. Street, "Generating joint scenarios for renewable generation: The case for non-gaussian models with time-varying parameters," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 7011–7019, 2018.

[5] G. Bodin, R. Saavedra, C. Fernandes, and A. Street, "Scoredrivenmodels. jl: a julia package for generalized autoregressive score models," *arXiv preprint arXiv:2008.05506*, 2020.

[6] M. Ruas, A. Street, and C. Fernandes, "A multi-quantile regression time series model with interquantile lipschitz regularization for wind power probabilistic forecasting," *arXiv preprint arXiv:2003.09983*, 2020.

[7] A. Street, A. L. Silva, C. Fernandes, A. Milhorance, E. Telles, G. Bodin, R. Saavedra, C. Fernandes, and A. Street, "Methods for optimal risk-averse demand contracting strategy in distribution companies: A brazilian case study," *Soon in the Power System Computational Conference, Porto*, 2022.

[8] J. Nowotarski and R. Weron, "Recent advances in electricity price forecasting: A review of probabilistic forecasting," *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 1548–1568, 2018.

[9] T. Hong, P. Pinson, Y. Wang, R. Weron, D. Yang, and H. Zareipour, "Energy forecasting: A review and outlook," *IEEE Open Access Journal of Power and Energy*, vol. 7, pp. 376–388, 2020.

[10] A. Street, "Experimentos, evidências e custos da inconsistência nos modelos do setor elétrico brasileiro," *LAMPS PUC-Rio website*, Janeiro 2022. [Online]. Available: http://www.lamps.ind.puc-rio.br/noticia/modelingrisk/

[11] ——, "A inconsistência do setor e a repetida crise," *Valor Econômico*, Junho 2021. [Online]. Available: http://www.lamps.ind.puc-rio.br/publicacao/compilacao-artigos-de-opiniao/