

Decomposição do Desempenho de Redes Móveis via Redes Neurais Monotônicas Q-Exponenciais

João V. R. Negri, Mike I. C. Domingues, Paulo R. B. da Silva, Gabriel L. E. Pedro, Eliezer S. Flores, Dimas I. Alves, Renato Machado, Debora L. Penha, Paulo R. S. Marcos

Resumo—Avaliar o custo-benefício de investimentos em redes móveis exige separar o efeito da capacidade instalada do efeito da demanda sobre o desempenho, o que abordagens puramente preditivas não tratam adequadamente. Este trabalho propõe um modelo estrutural composto por três redes neurais monotônicas — as redes de limite inferior e superior e de carga — combinadas pelo decaimento Q-exponencial de Tsallis, o qual captura saturação não linear com garantias formais de monotonicidade sobre indicadores de infraestrutura e uso. Aplicado a dados da ANATEL de 531 microrregiões (2022–2025), o modelo é avaliado em três configurações: global com e sem o fator temporal, isto é, para tendências agregadas, e regional, para especificidades locais. O paradigma regional obtém melhor ajuste (RMSE 0,03; R^2 0,71) em relação aos demais, enquanto a comparação entre as configurações sugere contribuição limitada do fator temporal frente às particularidades regionais.

Palavras-Chave—Q-Exponencial, Infraestrutura, Redes Neurais, Monotonicidade

Abstract—Assessing the cost-benefit of investments in mobile networks requires disentangling the effect of installed capacity from that of demand on performance, an issue that purely predictive approaches do not adequately address. This work proposes a structural model composed of three monotonic neural networks—lower-bound, upper-bound, and load networks—combined through the Tsallis Q-exponential decay, which captures nonlinear saturation while providing formal monotonicity guarantees with respect to infrastructure and usage indicators. Applied to ANATEL data from 531 microregions (2022–2025), the model is evaluated under three configurations: global, with and without temporal effects, to capture aggregate trends, and regional, to represent local specificities. The regional paradigm achieves the best fit (RMSE 0.03; R^2 0.71) among the configurations, while comparisons across models suggest a limited contribution of temporal effects relative to regional heterogeneity.

Keywords—Q-Exponential, Infrastructure, Neural Networks, Monotonicity

João Vitor Rosa Negri, Paulo Ricardo Branco da Silva, Gabriel Luiz Espindola Pedro, Eliezer Soares Flores, Dimas Irion Alves e Renato Machado são do Instituto Tecnológico de Aeronáutica (ITA), São José dos Campos, SP, Brasil, e-mails: negri, pbranco, gabrieluizpe, eliezer, dimasirion, rmachado@ita.br. Mike Igor Camilo Domingues é do Instituto Federal de Santa Catarina (IFSC), São José, SC, Brasil, e-mail: mike.icd@aluno.ifsc.edu.br. Debora Luzia Penha e Paulo Roberto Silva Marcos são da Agência Nacional de Telecomunicações (Anatel), Brasília, DF, Brasil, e-mails: debora.penha, paulorsm@anatel.gov.br. Este estudo foi financiado em parte pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Código de Financiamento 001; Pelo Instituto Nacional de Ciência e Tecnologia (INCT Signals); Pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq); pela Anatel, por meio do Termo de Execução Descentralizada (TED) nº 03/2023, Programa de Trabalho: 24.722.2205.20ZD.0001; e pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), processo 2020/09838-0 (BIOS—Brazilian Institute of Data Science)

I. INTRODUÇÃO

A análise da qualidade dos sistemas de telecomunicações é fundamental para o desenvolvimento de políticas públicas eficazes e para o planejamento estratégico de infraestrutura, especialmente em países de dimensões continentais com acentuadas desigualdades regionais [1]. Para isso, a Agência Nacional de Telecomunicações (ANATEL) estabelece marcos regulatórios e campanhas de medição que fornecem um panorama abrangente do setor, com destaque para métricas como o Índice Brasileiro de Conectividade (IBC), o Índice de Qualidade dos Serviços (IQS) e o Índice de Satisfação Geral (ISG)¹. O IBC avalia o estado da conectividade e da infraestrutura física dos municípios, ponderando variáveis como densidade de acessos, percentual de cobertura móvel, adensamento de estações (ERBs) e presença de fibra óptica como *backhaul*. O IQS, por sua vez, reflete o desempenho técnico efetivamente entregue pela rede, mensurando indicadores objetivos como disponibilidade, latência e cumprimento de velocidade de dados. Por fim, o ISG atua como um balizador da percepção dos consumidores, sendo apurado por meio de questionários em pesquisas anuais para medir a qualidade de experiência.

Apesar da riqueza desses indicadores regulatórios, a avaliação de desempenho desses sistemas tende a ser complexa, especialmente quando se considera a interação entre a infraestrutura disponível e a demanda sobre ela. Para modelar essa relação de forma precisa, é necessário quantificar o impacto de cada uma dessas componentes no desempenho da rede. Para isso, propomos um modelo estrutural monotônico baseado na função de retenção q-Exponencial, que é capaz de capturar a degradação do desempenho da rede à medida que a ocupação aumenta. Este modelo é treinado conjuntamente, utilizando técnicas de regressão quantílica para garantir que as estimativas do limite inferior e superior sejam consistentes com os dados observados [2]. Com isso, obtém-se uma compreensão mais profunda de como a infraestrutura e a demanda interagem para influenciar o desempenho da rede.

A fundamentação teórica para essa abordagem inspira-se nos Modelos de Fronteira Estocástica (SFA), utilizados para separar a capacidade instalada das ineficiências operacionais [3]. Contudo, como a rigidez paramétrica dessas formulações clássicas limita sua expressividade em dados de alta dimensionalidade, o modelo proposto funde os princípios de SFA com redes neurais, preservando as restrições de forma econômica

¹Embora o IQS e o ISG sejam indicadores amplos que englobam diferentes serviços de telecomunicações, as análises realizadas neste trabalho restringem esses indicadores, usando apenas a parcela referente a telefonia móvel.

e física dentro de arquiteturas flexíveis [4]. Nesse paradigma, a imposição explícita de monotonicidade é essencial para garantir que o modelo respeite relações estruturais conhecidas [5], [6], [7] — como a melhora de desempenho com o aumento de infraestrutura — sem depender de penalizações na função de perda. Adicionalmente, para capturar a saturação não linear característica de sistemas com gargalos correlacionados, a estatística de Tsallis [8] e sua função q-Exponencial oferecem o modelo de degradação de máxima entropia para tais regimes [9], superando o decaimento exponencial clássico. Por fim, a identificação das fronteiras latentes de desempenho é ancorada por regressão quantílica [2], técnica consolidada e extensível a arquiteturas profundas [10], [11], que mapeia os extremos da distribuição sem suposições distribucionais rígidas.

Este trabalho avalia três perspectivas de modelagem: sem estrutura temporal, com dependência temporal explícita e, por fim, um modelo com *embeddings* para capturar efeitos regionais específicos. Os resultados indicam que a inclusão do tempo produz ganhos marginais na adequação do modelo aos dados, porém o uso de *embeddings* apresenta melhoras significativas na captura do comportamento do sistema.

O artigo segue com a descrição do modelo proposto na Seção II. A Seção III apresenta os dados e o pré-processamento. A Seção IV discute os resultados obtidos, e a Seção V apresenta conclusões e perspectivas futuras.

II. O MODELO ESTRUTURAL Q-EXPONENCIAL

O modelo proposto decompõe o desempenho da rede, cujo alvo observável (Y) é medido pelo Índice de Entrega de Infraestrutura (IEI), em quatro componentes físicos interpretáveis. Seja $\mathbf{X} \in \mathbb{R}^{n_x}$ o vetor de variáveis de infraestrutura e $\mathbf{D} \in \mathbb{R}^{n_d}$ o vetor de variáveis de demanda. Cada componente é representado por uma função: $\mathcal{F} : \mathbb{R}^{n_x} \rightarrow [0, 1]$, que fornece o limite inferior garantido; $\mathcal{C} : \mathbb{R}^{n_x} \rightarrow [0, 1]$, que fornece o limite superior alcançável; $\Omega : \mathbb{R}^{n_x} \times \mathbb{R}^{n_d} \rightarrow \mathbb{R}_{\geq 0}$, que fornece a ocupação do sistema; e $\mathcal{R} : \mathbb{R}_{\geq 0} \times \mathbb{R}_{> 0} \times \mathbb{R}_{> 0} \rightarrow [0, 1]$, que fornece a retenção.

Denotam-se, por $F \equiv \mathcal{F}(\mathbf{X})$, $C \equiv \mathcal{C}(\mathbf{X})$ e $R \equiv \mathcal{R}(\Omega; \lambda, \gamma)$ os valores escalares resultantes da avaliação das funções nos vetores observados. A decomposição entre um limite superior de capacidade e fatores de ineficiência se inspira em modelos de fronteira estocástica [3], adaptada a uma formulação neural monotônica. O desempenho esperado $\hat{Y}$ é formulado através da seguinte equação:

$$\hat{Y} = \mathcal{F}(\mathbf{X}) + [\mathcal{C}(\mathbf{X}) - \mathcal{F}(\mathbf{X})] \cdot \mathcal{R}(\Omega(\mathbf{X}, \mathbf{D}); \lambda, \gamma). \quad (1)$$

Para garantir analiticamente a coerência da restrição $0 \leq \mathcal{F}(\mathbf{X}) \leq \hat{Y} \leq \mathcal{C}(\mathbf{X}) \leq 1$, parametriza-se o limite superior como uma fração do espaço remanescente acima do limite inferior, tal que $\mathcal{C}(\mathbf{X}) = \mathcal{F}(\mathbf{X}) + (1 - \mathcal{F}(\mathbf{X})) \cdot g(\mathbf{X})$, em que $g(\mathbf{X}) \in [0, 1]$ é a saída sigmoide da sub-rede associada ao limite superior [7]. A ocupação do sistema Ω , por sua vez, é um escalar que sintetiza a exigência imposta à infraestrutura disponível pela demanda exercida sobre ela.

A. Função de Retenção q-Exponencial

A função $\mathcal{R}$ mede a fração da margem máxima de desempenho mantida à medida que $\Omega(\mathbf{X}, \mathbf{D})$ aumenta. Essa função

é parametrizada pela q-Exponencial, uma família de distribuições derivada do princípio da maximização da entropia não-extensiva de Tsallis [8], definida por

$$\mathcal{R}(\Omega; \lambda, \gamma) = (1 + \lambda \Omega^\gamma)^{-1/\lambda}, \quad \lambda > 0, \gamma > 0, \quad (2)$$

em que λ é o parâmetro de curvatura e γ o expoente de alongamento.

A adoção da q-Exponencial fornece o modelo de degradação de máxima entropia para sistemas complexos [8], [9]. Diferentemente do regime exponencial padrão ($\lambda \rightarrow 0$), um parâmetro $\lambda > 0$ indica correlação positiva entre as fontes de degradação de uma rede (gargalos de *backhaul*, congestionamento, interferência), resultando em decaimento racional, no qual o impacto combinado ocorre com menor intensidade do que o produto de efeitos isolados [9].

Por fim, o amortecimento do sistema é estruturalmente garantido pela elasticidade de ocupação. Em regimes racionais ($\lambda > 0$), a sensibilidade máxima da retenção à carga é assintoticamente limitada pela razão γ/λ , assegurando que o desempenho da rede não degrade abruptamente para zero em regimes de demanda extrema.

B. Treinamento Conjunto e Identificabilidade Estrutural

Como (1) decompõe $\hat{Y}$ em múltiplas variáveis latentes não observáveis simultaneamente (F , C , Ω e R), a estimação tradicional resultaria em um problema de não-identificabilidade, uma vez que infinitas combinações de parâmetros poderiam gerar a mesma predição. Para combater essa ambiguidade, o modelo emprega treinamento conjunto baseado em regressão quantílica para fixar as fronteiras da distribuição [10], [11].

A otimização minimiza uma soma ponderada de funções de perda assimétricas do tipo *pinball* [2], definidas por

$$\rho_\tau(u) = u(\tau - \mathcal{I}_{\{u < 0\}}),$$

em que τ é o quantil alvo e $\mathcal{I}_{\{u < 0\}}$ é a função indicadora da condição booleana $\{u < 0\}$, assumindo valor 1 quando o resíduo u é negativo (subestimação) e 0 caso contrário. Através dessa métrica, a rede do limite inferior é ancorada em um quantil inferior (p.ex., $\tau_F = 0, 05$), identificando a fronteira de pior cenário. A rede do limite superior usa um quantil superior (p.ex., $\tau_C = 0, 95$), estabelecendo a capacidade máxima. A expectativa central $\hat{Y}$ converge para a mediana ($\tau = 0, 5$) [2].

Para assegurar que o modelo respeite a formulação matemática do problema, as sub-redes utilizam uma arquitetura de redes neurais monotônicas [5], [7], [6]. Envolvendo pesos estritamente não-negativos (Softplus) e ativações não-decrescentes (ReLU) [5], [6], o modelo garante de forma estrutural que as funções $\mathcal{C}$, $\mathcal{F}$ e Ω possuam relações monotônicas com infraestrutura e demanda. Tais restrições são expressas por

$$\begin{aligned} \frac{\partial \mathcal{F}(\mathbf{X})}{\partial X_j} &\geq 0, & \frac{\partial \mathcal{C}(\mathbf{X})}{\partial X_j} &\geq 0, \\ \frac{\partial \Omega(\mathbf{X}, \mathbf{D})}{\partial D_k} &\geq 0, & \frac{\partial \Omega(\mathbf{X}, \mathbf{D})}{\partial X_j} &\leq 0. \end{aligned} \quad (3)$$

As duas primeiras restrições garantem que os desempenhos máximo e mínimo da rede aumentem com a infraestrutura. A terceira restrição assegura que a ocupação aumente com o

aumento da demanda, considerando uma infraestrutura fixa. A quarta restrição garante que a ocupação diminua com o aumento da infraestrutura, dada uma demanda fixa.

III. ANÁLISE DOS DADOS

Este trabalho utiliza dados públicos do IBGE (para normalização e agregação) e da Anatel, incluindo o ISG, o IBC e o IQS, bem como indicadores de demanda (assinantes da rede móvel) e de infraestrutura instalada (cobertura móvel). Para definir o IEI, utiliza-se uma combinação linear convexa entre IBC, IQS e ISG, garantindo $IEI \in [0, 1]$. Embora a composição desses índices possa ser explorada mais a fundo para produzir um IEI mais condizente com a realidade, neste trabalho optou-se por uma combinação linear simples e explicável. O trabalho atual concentra-se no estudo de comportamentos espaciais e temporais que correlacionam demanda, infraestrutura e desempenho de rede. Contempla-se a elaboração de um alvo de desempenho otimizado como um objeto de estudo para um trabalho futuro.

Os dados de cobertura ($\mathbf{X}$) e demanda ($\mathbf{D}$) são obtidos de relatórios da Anatel que trazem, respectivamente, o percentual de cobertura populacional e o número de assinantes por período, tecnologia e operadora em cada município. Os períodos de disponibilidade e a granularidade original de cada fonte diferem entre si. Os dados de cobertura são publicados trimestralmente; assinantes mensalmente; o IQS é apurado semestralmente; enquanto IBC e ISG são apurados anualmente. Para viabilizar o treinamento conjunto, todas as séries são harmonizadas ao nível mensal por meio de interpolação do tipo *zero-order hold*, preservando a monotonicidade temporal dos indicadores e evitando suposições de continuidade e de derivadas suaves não justificadas para indicadores apurados por campanhas discretas.

A utilização dos dados ao nível municipal apresenta limitações estatísticas. Dos 5.570 municípios brasileiros, apenas uma fração é consistentemente coberta pelos indicadores da Anatel: o ISG possui validade estatística por unidade da Federação, o IQS possui granularidade municipal, mas depende de um número mínimo de medições para ter validade estatística. Essa esparsidade compromete a representatividade das séries, amplifica a variância em regiões pouco penetradas e inviabiliza a estimação confiável dos componentes estruturais do modelo em escala municipal. Para melhorar a representação estatística dos dados e obter maior consistência amostral e temporal, os municípios são agregados em microrregiões do IBGE. Após a harmonização por microrregião, o conjunto final cobre 42 meses, com 22.302 observações em 531 microrregiões.

IV. RESULTADOS

Para a avaliação do modelo, os dados das 531 microrregiões foram divididos temporalmente a fim de evitar vazamento de informação, sendo os primeiros 36 meses utilizados para treino e os últimos 6 para validação. Três configurações do modelo Q-exponencial são avaliadas de forma progressiva. O modelo *global atemporal* serve como ponto de partida. Nessa configuração, os dados das 531 microrregiões são concatenados em um único conjunto de treinamento ($531 \times 36 = 19.116$

amostras), e a dimensão temporal não é considerada. Em seguida, o modelo *global temporal* introduz a dimensão do tempo, permitindo que F e C variem ao longo dele, mesmo com $\mathbf{X}$ fixo. Por fim, um modelo com *embedding* incorpora um vetor específico por microrregião de dimensão 8. Essa dimensão foi escolhida experimentalmente por representar um bom compromisso entre a complexidade de treinamento e o desempenho do modelo obtido. O vetor aprendido durante o treinamento e somado à saída pré-sigmoide de cada sub-rede possibilita capturar heterogeneidade regional não explicada por $\mathbf{X}$, $\mathbf{D}$ ou t , sem comprometer as garantias de monotonicidade.

Todos os modelos são treinados com o otimizador AdamW [12] ($\eta = 5 \times 10^{-4}$ e decaimento 10^{-5}), tamanho de *batch* 128, e escalonador de cosseno anelado [13] por 300 épocas. As sub-redes possuem três camadas com arquitetura [32, 64, 32]. A Tabela I apresenta métricas de desempenho dos três modelos avaliadas no conjunto de validação. Além de métricas clássicas (MAE, RMSE e R^2), são reportadas métricas específicas do modelo estrutural: a cobertura do limite inferior mede a proporção de observações acima do limite inferior estimado, isto é, acima de $\tau_F = 0,05$. A cobertura do limite superior mede a proporção abaixo do limite superior, isto é, abaixo de $\tau_C = 0,95$. A contenção indica a fração de observações contidas no intervalo $[F, C]$.

TABELA I
MÉTRICAS DE DESEMPENHO POR MODELO (CONJUNTO DE VALIDAÇÃO)

Métrica	Global Atemporal	Global Temporal	<i>Embedding</i>
MAE	0,0369	0,0362	0,0234
RMSE	0,0499	0,0483	0,0325
R^2	0,3256	0,3691	0,7134
cobertura do limite inferior	0,9454	0,9422	0,9438
cobertura do limite superior	0,9231	0,9639	0,9601
contenção	0,8685	0,9062	0,9040

O modelo obtido com *embedding* supera os demais em todas as métricas de ajuste pontual, com $R^2 = 0,71$ contra 0,37 do modelo *global temporal*. Como parte da variância do IEI entre microrregiões decorre de características locais não capturadas por $\mathbf{X}$ e $\mathbf{D}$, as métricas poderiam ser melhoradas pela incorporação de indicadores de infraestrutura e demanda mais granulares, o que poderia se tornar um critério para futuras coletas de dados da Anatel.

A adição da dimensão temporal no modelo global traz ganho modesto em MAE e R^2 em relação ao caso atemporal, sugerindo que a variação temporal dentro de uma mesma microrregião é pequena frente à variação entre microrregiões. Esse comportamento é coerente com a natureza dos dados: indicadores de infraestrutura evoluem lentamente, e o período de 42 meses pode ser insuficiente para que tendências temporais de expansão ou saturação de cobertura se manifestem de forma estatisticamente relevante. Essa leitura é reforçada pela Figura 1, que apresenta a evolução mediana nacional das três componentes do sistema: ao longo do período, os limites inferior e superior e o desempenho permanecem em patamares estáveis, sem tendências de ascensão ou declínio pronunciadas, indicando que o sistema opera em regime estável no intervalo.

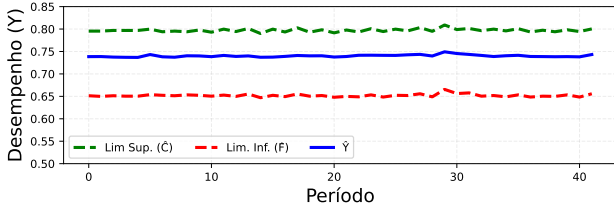


Fig. 1. Evolução temporal das previsões medianas nacionais do modelo *global temporal*.

A. Efeitos isolados do sistema

Além das métricas numéricas, a adequação do modelo ao comportamento esperado do sistema pode ser avaliada com gráficos do tipo *ceteris paribus*: todos os parâmetros de entrada são fixados em suas medianas e apenas uma característica é variada por vez, evidenciando sua influência isolada sobre as redes estimadas. A Figura 2 apresenta esses perfis para o modelo *global atemporal*, em que os painéis (a–d) correspondem às variáveis de demanda e (e–h) às variáveis de infraestrutura.

Na análise das variáveis de demanda, dois comportamentos se destacam, mostrando-se consistentes com as restrições de monotonicidade impostas pela arquitetura: o decaimento monotônico de $\hat{Y}$ com o aumento da demanda, refletindo a degradação do desempenho induzida pela maior ocupação do sistema, e a independência das funções F e C em relação à demanda, pois ambas dependem exclusivamente de $\mathbf{X}$. Na análise das variáveis de infraestrutura, observa-se melhora monotônica tanto do desempenho quanto dos limites estruturais com o aumento da cobertura, com exceção da cobertura 5G, cujo impacto é modesto, o que pode ser explicado pelo estágio inicial de penetração dessa tecnologia no Brasil.

A Figura 3 apresenta os mesmos perfis para o modelo com *embeddings*. Nesse caso, as variações de $\mathbf{X}$ e $\mathbf{D}$ produzem efeitos de menor magnitude sobre $\hat{Y}$, F e C , preservando a monotonicidade. Esse comportamento indica que o vetor de *embedding* parece capturar parte relevante da heterogeneidade explicável do sistema, atribuindo aos indicadores de infraestrutura e demanda disponíveis um papel complementar. Isso não invalida o modelo, mas sugere uma limitação das variáveis de entrada na representação da heterogeneidade regional, levantando a hipótese de que fatores não observados, possivelmente associados a aspectos estruturais, socioeconômicos ou contextuais das microrregiões, podem influenciar o desempenho do sistema.

B. Análise das microrregiões com menor desempenho

As Figuras 4 e 5 apresentam as 30 microrregiões com menor desempenho estimado no último período disponível, mês 12 de 2025, com seus respectivos valores de $\hat{Y}$, F e C . A comparação entre os modelos *global temporal* e *embedding* revela diferenças estruturais relevantes na forma como cada configuração representa a heterogeneidade regional.

No modelo *global temporal* da Figura 4, os limites F e C variam pouco entre as microrregiões ordenadas, convergindo para patamares próximos à mediana nacional. Esse comportamento indica que o modelo aprende predominantemente um regime médio de operação, com capacidade limitada de

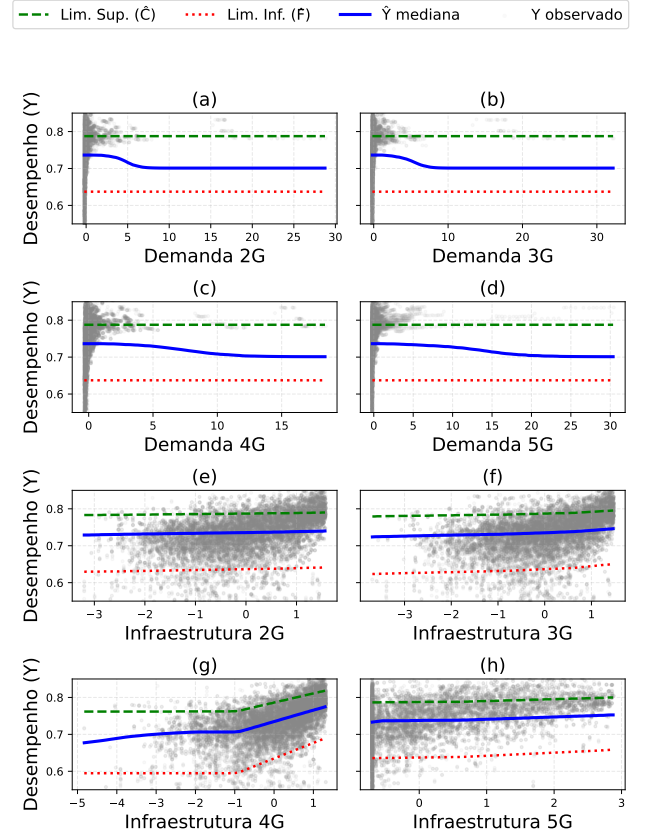
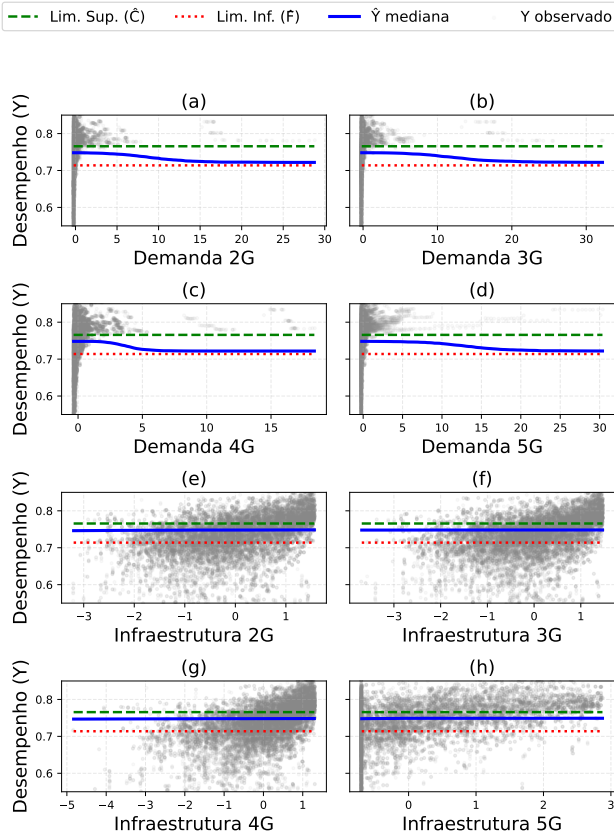


Fig. 2. Perfis *ceteris paribus* do modelo *global atemporal*. Em cada painel, o eixo x contém a variável de entrada (z-score), mantendo as demais na mediana, e o eixo y apresenta $\hat{Y}$, F e C .

individualizar os limites estruturais de cada região. Como consequência, a variação no desempenho estimado entre as microrregiões é ditada principalmente pela função de carga $\Omega(\mathbf{X}, \mathbf{D})$, e não por diferenças nos limites F e C , o que é consistente com o baixo valor de R^2 apresentado pelo modelo.

O modelo com *embedding* da Figura 5 apresenta variância maior tanto em $\hat{Y}$ quanto em F e C . Microrregiões como o Nordeste de Roraima, Paranatinga e Alto Solimões, que figuram entre as piores colocações de desempenho $\hat{Y}$ em ambos os modelos, exibem no modelo com *embedding* limites superiores mais baixos e limites inferiores mais elevados do que a média nacional, sugerindo que *embeddings* capturam corretamente a restrição estrutural dessas regiões, ou seja, que o potencial máximo alcançável pela infraestrutura local é inferior ao de regiões mais desenvolvidas, independentemente da demanda.

Essa distinção tem implicações para políticas públicas. Em regiões onde o desempenho é baixo, mas a distância $C - F$ é ampla, a infraestrutura instalada possui potencial técnico não realizado, possivelmente por sobrecarga de demanda ou por ineficiências operacionais. Nesses casos, ações das prestadoras de telefonia móvel voltadas à gestão ou à melhoria da qualidade de serviço podem ser mais eficazes do que novos investimentos em cobertura. Por outro lado, em regiões onde tanto $\hat{Y}$ quanto C são baixos, o diagnóstico aponta para insuficiência estrutural de infraestrutura, demandando


 Fig. 3. Perfis do modelo com *embedding*, mesma convenção da Fig. 2.

expansão da cobertura e capacidade instalada.

V. CONCLUSÃO

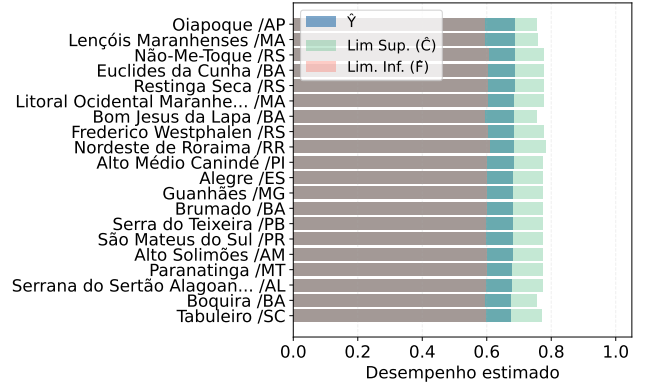
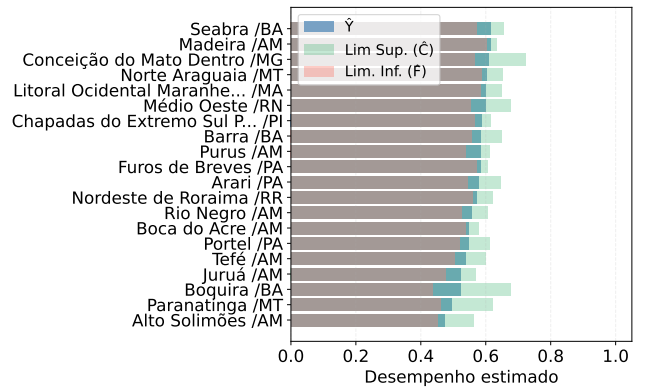
O trabalho propõe um modelo estrutural interpretável para análise do desempenho de redes móveis baseado na retenção Q-exponencial de Tsallis, decompondo o desempenho em três componentes fisicamente interpretáveis — limite inferior, limite superior e ocupação — com garantias formais de monotonicidade e restrições estruturais que dispensam penalizações adicionais na função de perda para impor comportamento monotônico. Aplicado a dados da ANATEL sobre 531 microrregiões (2022–2025), a configuração com *embedding* por microrregião supera as globais, evidenciando que a heterogeneidade regional é um fator importante de variação do IEI.

AGRADECIMENTOS

Os autores agradecem à Anatel, em particular a Kai de Andrade Lima, pelo apoio e pelas discussões técnicas, bem como a Bartolomeu F. Uchôa-Filho pelas contribuições e discussões relacionadas a este trabalho.

REFERÊNCIAS

- [1] N. Gozzi, N. Comini, and N. Perra, “Bridging the digital divide: mapping internet connectivity evolution, inequalities, and resilience in six brazilian cities,” *EPJ Data Science*, vol. 13, no. 1, p. 69, 2024.
- [2] R. Koenker and G. Bassett, “Regression quantiles,” *Econometrica*, vol. 46, no. 1, pp. 33–50, 1978.
- [3] D. Aigner, C. A. K. Lovell, and P. Schmidt, “Formulation and estimation of stochastic frontier production function models,” *Journal of Econometrics*, vol. 6, no. 1, pp. 21–37, 1977.


 Fig. 4. Ordenação das 30 microrregiões com menor desempenho estimado no modelo *global temporal* (último período). Barras indicam $\hat{Y}$, com marcadores do limite inferior e do limite superior.

 Fig. 5. Ordenação das 30 microrregiões com menor desempenho estimado no modelo com *embedding* (último período)

- [4] Z. Wei, H. Sang, A. Prokhorov, and Y. Ma, “Shape-aware deep learning for models of production,” *Journal of Productivity Analysis*, vol. 65, no. 2, p. 13, Feb. 2026. [Online]. Available: <https://doi.org/10.1007/s11123-026-00797-3>
- [5] J. Sill, “Monotonic networks,” in *Advances in Neural Information Processing Systems (NIPS) 10*, 1997, pp. 661–667.
- [6] D. Runje and S. M. Shankaranarayana, “Constrained monotonic neural networks,” in *International Conference on Machine Learning*. PMLR, 2023, pp. 29 338–29 353.
- [7] V. Wargnier-Dauchelle, T. Grenier, F. Durand-Dubief, F. Cotton, and M. Sdika, “Explainable monotonic networks and constrained learning for interpretable classification and weakly supervised anomaly detection,” *Pattern Recognition*, vol. 160, p. 111186, 2025.
- [8] C. Tsallis, “Possible generalization of boltzmann-gibbs statistics,” *Journal of Statistical Physics*, vol. 52, no. 1–2, pp. 479–487, 1988.
- [9] T. Mukherjee, S. Bebertta, and D. Senapati, “Stochastic modeling of q-lognormal fading channels over tsallis’ entropy: Evaluation of channel capacity and higher order moments,” *Digital Signal Processing*, vol. 133, p. 103856, 2023.
- [10] H. Tyralis, G. Papacharalampous, N. Dogulu, and K. P. Chun, “Deep huber quantile regression networks,” *Neural Networks*, vol. 187, p. 107364, 2025.
- [11] C. Wan, J. Lin, J. Wang, Y. Song, and Z. Y. Dong, “Direct quantile regression for nonparametric probabilistic forecasting of wind power generation,” *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 2767–2778, 2017.
- [12] I. Loshchilov and F. Hutter, “Decoupled weight decay regularization,” in *International Conference on Learning Representations (ICLR)*, 2019.
- [13] —, “SGDR: Stochastic gradient descent with warm restarts,” in *International Conference on Learning Representations (ICLR)*, 2017.