Spatio-Temporal Forecasting of Deforestation in the Legal Amazon Using LSTM Networks Under an Isolated Regional Scale Approach

Jéssica de França Lima, Martony Demes da Silva

Curso Superior de Tecnologia de Gestão de Dados
Centro de Educação Aberta e a Distância (CEAD)
Universidade Federal do Piauí (UFPI)
Teresina – PI – Brasil

jessica.lima@ufpi.edu.br , martony.silva@ufpi.edu.br

Abstract. *Predicting deforestation in the Legal Amazon using Machine Learning often faces bottlenecks due to severe spatial scale discrepancies across different states. This paper proposes a spatio-temporal forecasting pipeline using a Shallow Long Short-Term Memory (LSTM) network combined with localized feature scaling and a strict temporal holdout validation. Moving away from global data stacking, our approach isolates sliding window sequences for each state, preventing data leakage and spatial cross-contamination. Experimental results demonstrated that reducing architectural complexity and isolating regional scales successfully overcame severe underfitting, achieving a Coefficient of Determination (R^2) of 0.6360 and a Mean Absolute Error (MAE) of 179.14 km^2. The method proves robust for automated environmental surveillance in asymmetric micro-regional contexts.*

Resumo. *A predição do desmatamento na Amazônia Legal por Aprendizado de Máquina enfrenta gargalos devido a severas discrepâncias de escala espacial entre os estados. Este artigo propõe um pipeline de previsão espaço-temporal utilizando uma rede de Memória de Curto e Longo Prazo (Long Short-Term Memory – LSTM) rasa combinada com escalonamento localizado de attributes e validação por holdout temporal estrito. Afastando-se do empilhamento global de dados, nossa abordagem isola as sequências de janelas deslizantes para cada estado, evitando o vazamento de dados (data leakage) e a contaminação espacial cruzada. Os resultados experimentais demonstraram que a redução da complexidade arquitetônica e o isolamento das escalas regionais superaram um subajuste severo, alcançando um Coeficiente de Determinação (R^2) de 0,6360 e um Erro Absoluto Médio (Mean Absolute Error – MAE) de 179,14 km^2. O método mostra-se robusto para a vigilância ambiental automatizada em contextos microrregionais assimétricos.*

1. Introdução

O avanço do desmatamento na Amazônia Legal consolida-se como um dos desafios socioambientais mais críticos da atualidade. A conversão de floresta nativa em pastagem e agricultura compromete serviços ecossistêmicos, intensifica emissões de gases de efeito estufa e prejudica metas climáticas globais [Fearnside 2005]. Dada a extensão territorial

da região, órgãos ambientais como IBAMA e ICMBio dependem de ferramentas preditivas para antecipar focos de degradação e alocar recursos de fiscalização com eficiência.

Arquiteturas de Aprendizado Profundo (*Deep Learning*), especialmente redes *Long Short-Term Memory* (LSTM), destacam-se na modelagem de séries temporais devido à capacidade de capturar dependências sequenciais de longo prazo [Hochreiter and Schmidhuber 1997]. Contudo, sua aplicação no monitoramento ecológico enfrenta desafios metodológicos severos. Práticas comuns na literatura, como o embaralhamento aleatório (*shuffle*) em dados temporais, violam a cronologia dos eventos e causam vazamento de informação (*data leakage*) [Goodfellow et al. 2016].

A principal lacuna resolvida por este trabalho é o **efeito destrutivo da disparidade de escala multirregional** sob normalizações globais. Os dados do sistema PRODES/INPE revelam ordens de magnitude assimétricas entre as unidades federativas da Amazônia Legal [Instituto Nacional de Pesquisas Espaciais - INPE 2024], conforme sintetizado na Tabela 1.

Tabela 1. Disparidade de escala no desmatamento anual da Amazônia Legal (Dados: INPE/PRODES)

Estado	**Média Anual** (km^2)	**Mínimo** (km^2)	**Máximo** (km^2)	**Perfil de Escala**
Pará (PA)	4.712,40	1.741,00	10.964,00	Macro-escala Crítica
Mato Grosso (MT)	4.108,15	767,00	11.814,00	Macro-escala Crítica
Amapá (AP)	34,70	0,00	127,00	Micro-escala Controlada
Roraima (RR)	261,35	70,00	615,00	Micro-escala Isolada

Ao aplicar um escalonamento linear global (como o *MinMaxScaler* unificado) sobre o conjunto de dados, estados de menor escala (ex.: Amapá e Roraima) têm seus valores comprimidos próximos a zero. Esse fenômeno esmaga a variância local e cega o modelo às flutuações das microrregiões. Tentar compensar essa perda aumentando a profundidade da rede neural induz o otimizador ao subajuste (*underfitting*), resultando em Coeficientes de Determinação (R^2) negativos.

Para superar essa limitação, este estudo reestrutura o pipeline espaço-temporal adotando uma rede *Shallow LSTM* baseada em **Escala Regional Isolada**. O objetivo geral é desenvolver e avaliar um método de previsão do desmatamento na Amazônia Legal capaz de preservar a sensibilidade local. Os objetivos específicos são:

1. Implementar engenharia de recursos com parametrização e dicionários de escala do *MinMaxScaler* independentes por estado, mitigando o esmagamento da variância local.
2. Estruturar janelas deslizantes tridimensionais espacialmente encapsuladas, impedindo continuidades cronológicas artificiais entre estados distintos.
3. Aplicar validação por *holdout* temporal para cenários de escassez de dados ($N = 309$), comparando o ganho de desempenho frente a modelos profundos.

2. Referencial Teórico

Esta seção delineia as bases conceituais e matemáticas do pipeline proposto, cobrindo redes recorrentes, validação cronológica, normalização multi-escala e funções de perda

robustas.

2.1. Redes Neurais Long Short-Term Memory (LSTM)

As redes *Long Short-Term Memory* (LSTM) [Hochreiter and Schmidhuber 1997] foram projetadas para superar o problema da dissipação ou explosão do gradiente (*vanishing/exploding gradient*) em Redes Neurais Recorrentes (RNN). A LSTM utiliza blocos de memória regulados por comportas (*gates*) que controlam o fluxo de informação.

Dado um vetor de entrada x_t e o estado oculto anterior h_{t-1}, as comportas de esquecimento (f_t), entrada (i_t) e saída (o_t) são calculadas por:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f), \quad i_t = \sigma(W_i[h_{t-1}, x_t] + b_i), \quad o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (1)$$

onde σ representa a ativação sigmoide, enquanto W e b denotam matrizes de pesos e viéses. O estado celular candidato é $\tilde{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c)$. O estado celular atualizado (C_t) e o estado oculto (h_t) são dados por:

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t, \quad h_t = o_t \odot \tanh(C_t) \quad (2)$$

sendo $\odot$ o produto de Hadamard. Essa estrutura preserva gradientes ao longo do tempo, permitindo modelar séries temporais de longa dependência.

2.2. Vazamento de Dados (Data Leakage) e Validação Temporal

A amostragem aleatória convencional (como o *shuffle* tabular) é inadequada para séries temporais pois viola a ordem cronológica dos dados, gerando vazamento de informação (*data leakage*) [Hyndman and Athanasopoulos 2018]. O vazamento ocorre quando informações do futuro (conjunto de teste) contaminam o ajuste do passado (treinamento), criando métricas irrealistas que falham em cenários de inferência real [Goodfellow et al. 2016]. Para mitigar essa falha, aplica-se o *Holdout Temporal*, fixando um limiar T_{corte} onde dados anteriores compõem o treino e os posteriores simulam o teste futuro.

2.3. O Impacto do Esmagamento de Variância em Normalizações Globais

Redes neuronais exigem atributos normalizados (ex.: $[0, 1]$) para estabilizar gradientes. Contudo, a normalização *MinMax* global em datasets multi-estado heterogêneos introduz distorções estatísticas [Pedregosa et al. 2011]. Em uma escala unificada regida por estados de alta criticidade ($\max = 11.000\ km^2$ no Pará), séries de menor magnitude ($\max \leq 130\ km^2$ no Amapá) são comprimidas próximo a zero ($z_t \leq 0,01$). Esse **esmagamento de variância** anula o gradiente local, impedindo o aprendizado e resultando em *underfitting* com R^2 negativo. A solução exige isolar o mapeamento por unidade federativa.

2.4. Função de Perda Robusta: Huber Loss

O Erro Quadrático Médio (MSE) é sensível a *outliers*, pois penaliza desvios quadraticamente (e^2) e pode desestabilizar o aprendizado na presença de surtos atípicos de desmatamento [Huber 1964]. Para conferir robustez sem descartar dados limitados ($N = 309$), adota-se a *Huber Loss*, que atua quadraticamente para erros pequenos e linearmente para erros superiores a um limiar δ:

$$L_\delta(y, \hat{y}) = \begin{cases} \frac{1}{2}(y - \hat{y})^2 & \text{se } |y - \hat{y}| \leq \delta \\ \delta \cdot \left(|y - \hat{y}| - \frac{1}{2}\delta\right) & \text{se } |y - \hat{y}| > \delta \end{cases} \quad (3)$$

Essa transição suave garante derivadas estáveis e limita o impacto de picos sazonais extremos sobre os pesos da LSTM [Goodfellow et al. 2016].

3. Trabalhos Relacionados

A literatura recente reflete a transição de modelos estatísticos lineares para arquiteturas não-lineares complexas no monitoramento do desflorestamento. Esta seção analisa cinco estudos proeminentes correlacionados à dinâmica espaço-temporal da Amazônia Legal.

A aplicação de variantes recorrentes ganhou tração com [Souza et al. 2021], que implementou um modelo GRU para prever tendências no Pará via dados do PRODES. Embora eficaz no curto prazo, a normalização centralizada nos limites máximos do estado gerou perda de sensibilidade preditiva em municípios periféricos de menor extensão. Para integrar os componentes espacial e temporal, [Mello et al. 2022] propuseram uma arquitetura híbrida ConvLSTM baseada em imagens do Sentinel-2. Contudo, o modelo exibiu instabilidade e subajuste em sub-regiões com alta cobertura de nuvens (como o Amapá), expondo a vulnerabilidade de redes profundas globais perante dados locais de menor escala.

Em outra vertente, [Sano et al. 2023] avaliaram redes LSTM empilhadas profundas (3 a 5 camadas) alimentadas por variáveis exógenas agregadas para toda a Amazônia Legal. A alta complexidade estrutural exigiu amostragem superior à disponível, forçando o uso de validação cruzada aleatória (*k-fold* com *shuffle*), o que introduziu vazamento de dados (*data leakage*) e métricas artificialmente otimistas. Já a modelagem de séries temporais puras foi abordada por [Escada et al. 2023], que compararam o modelo ARIMA a redes LSTM rasas em Mato Grosso, concluindo que a abordagem recorrente captura melhor picos históricos. No entanto, o pipeline limitou-se ao escopo univariado monorregional. Por fim, [Ribeiro et al. 2024] exploraram o algoritmo *Random Forest* para janelas móveis em Terras Indígenas; por ser uma abordagem não-recorrente, o método falhou em preservar a memória sequencial de longo prazo das pressões antrópicas.

A Tabela 2 consolida a comparação estruturada entre as abordagens revisadas e o método proposto neste artigo, destacando os eixos de análise fundamentais para o ENIAC.

Tabela 2. Comparativo analítico de trabalhos relacionados e a abordagem proposta

Referência	Modelo Base	Escopo Espacial	Tratamento de Escala	Validação
[Souza et al. 2021]	GRU	Regional (PA)	Global Unificado	Temporal
[Mello et al. 2022]	ConvLSTM	Imagens Satélite	Não Aplicável	Espacial
[Sano et al. 2023]	Stacked LSTM	Amazônia Legal	Global Unificado	*Shuffle* (Viesada)
[Escada et al. 2023]	LSTM Univariada	Local (MT)	Monorregional	Temporal
[Ribeiro et al. 2024]	Random Forest	Áreas Protegidas	Segmentado	Janela Móvel
Este Trabalho	**Shallow LSTM**	**Amazônia Legal**	**Isolado por Estado**	**Holdout Temporal**

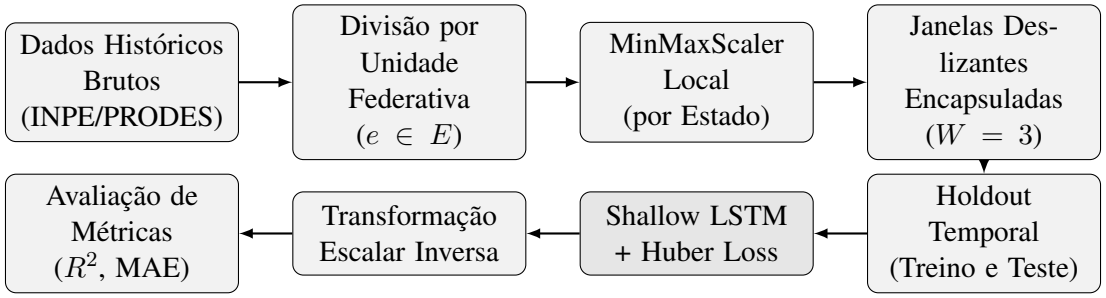
O diferencial desta pesquisa frente à literatura correlata reside no alinhamento indissociável entre a integridade matemática do otimizador e a geografia política dos dados ambientais. Enquanto os trabalhos anteriores oscilam entre modelos excessivamente locais (restringindo o aprendizado a um único estado) ou globais homogêneos

(que esmagam a variância e geram subajuste), nossa abordagem introduz um ecossistema multi-estado unificado sob um regime de parametrização **totalmente isolado por unidade federativa**. Ao encapsular os geradores de janelas tridimensionais dentro das fronteiras de cada estado e mapear seus atributos por dicionários de escala independentes, o modelo proposto elimina o efeito destrutivo da assimetria de ordens de magnitude e bloqueia o vazamento de dados temporal e espacial. Trata-se, portanto, de uma solução que entrega generalização em larga escala (Amazônia Legal) sem sacrificar a sensibilidade matemática e a acurácia das microrregiões de menor expressão numérica.

4. Metodologia

A metodologia proposta afasta-se de abordagens globais ao instituir um pipeline espaço-temporal baseado em Escala Regional Isolada. O objetivo é neutralizar o esmagamento de variância provocado pela disparidade de escalas entre estados da Amazônia Legal, preservando a cronologia das séries. A Figura 1 ilustra o fluxo de processamento e a arquitetura do sistema.

Figura 1. Diagrama de blocos do pipeline espaço-temporal baseado em Escala Regional Isolada.



O pipeline opera em cinco etapas sequenciais detalhadas a seguir.

4.1. Aquisição e Estruturação do Repositório de Dados

O dataset provém do sistema PRODES/INPE [Instituto Nacional de Pesquisas Espaciais - INPE 2024], compreendendo séries anuais de desmatamento (km^2) de 1988 a 2022 para os 9 estados da Amazônia Legal: Acre (AC), Amapá (AP), Amazonas (AM), Maranhão (MA), Mato Grosso (MT), Pará (PA), Rondônia (RO), Roraima (RR) e Tocantins (TO). Cada estado possui 35 registros, totalizando $N = 309$ observações consolidadas após pré-processamento — um cenário de escassez amostral para redes profundas. Para reprodutibilidade, o código-fonte e hiperparâmetros foram disponibilizados em repositório público[1].

4.2. Engenharia de Atributos Locais (MinMaxScaler Isolado)

Para neutralizar o viés induzido pelas macro-escalas (PA, MT) sobre as micro-escalas (AP, RR), aplicou-se um escalonamento descentralizado. Em vez de computar parâmetros globais, o algoritmo ajusta um objeto *MinMaxScaler* individual para cada unidade federativa.

[1]Código-fonte e experimentos disponíveis em: `https://colab.research.google.com/drive/103zAmwmnm9tblaYgQdoHxfkaWF95gJRk?usp=sharing`

Para cada estado e, a normalização é regida pela Equação 4:

$$z_t^e = \frac{x_t^e - \min(X^e)}{\max(X^e) - \min(X^e)}, \quad \forall e \in \{AC, AP, \ldots, TO\} \tag{4}$$

Essa estratégia preserva a variância local e mantém os gradientes das séries de menor magnitude distintos de ruídos numéricos.

4.3. Janelas Temporais Espacialmente Encapsuladas

O fatiamento temporal contínuo sem distinção de estado cria falsas continuidades cronológicas entre os limites das séries. Para sanar essa falha, adotou-se o encapsulamento espacial com janela deslizante $W = 3$ anos. O fatiamento $X_i = [z_{t-3}^e, z_{t-2}^e, z_{t-1}^e] \rightarrow Y_i = z_t^e$ é realizado estritamente dentro das fronteiras geográficas de cada estado. Apenas após a geração isolada das sequências locais, os blocos são concatenados nos tensores globais do NumPy (X_{global}, Y_{global}), eliminando contaminações nas transições inter-regionais.

4.4. Protocolo de Validação por Holdout Temporal

Para prevenir o vazamento de dados (*data leakage*), evitou-se o *shuffle* aleatório. O corte cronológico foi fixado em $T = 2016$: registros de 1988 a 2016 compõem o conjunto de treinamento, enquanto os anos de 2017 a 2022 funcionam como o *holdout* de teste futuro. Esse arranjo simula o comportamento do modelo em ambiente de produção real.

4.5. Arquitetura da Rede Neural Recorrente e Função de Perda

Para evitar o *underfitting* observado em modelos profundos operando em amostras pequenas ($N = 309$), desenvolveu-se uma arquitetura rasa (*Shallow LSTM*) via Keras/TensorFlow composta por:

1. **Camada LSTM:** 16 unidades ocultas com ativação $\tanh$, recebendo tensores com formato ‘(W, 1)‘ ($W = 3$). O modelo possui apenas 1.169 parâmetros treináveis, favorecendo a convergência.
2. **Dropout:** Taxa de 0.2 aplicada após a LSTM para evitar sobreajuste.
3. **Camada Densa de Saída:** 1 neurônio linear para a predição escalar no ano t.

O treinamento ocorreu em 60 épocas (tamanho de lote 16) com o otimizador Adam. Utilizou-se a **Huber Loss** ($\delta = 1.0$), que opera de forma quadrática para erros pequenos e linear para desvios maiores. Dado que os dados estão no intervalo $[0, 1]$, o limiar $\delta = 1.0$ permite capturar a dinâmica regular no regime quadrático enquanto limita a influência de picos anômalos de desmatamento na atualização dos pesos. Ao final, as predições passam pela inversão escalar para avaliação das métricas na escala real (km^2).

5. Resultados e Discussão

A avaliação de desempenho do pipeline espaço-temporal proposto confronta os resultados da arquitetura baseada em Escala Regional Isolada com as falhas identificadas no modelo base de controle. As análises fundamentam-se em métricas consolidadas após a transformação reversa das predições para a escala física real em quilômetros quadrados (km^2).

5.1. Análise Quantitativa e Convergência

O treinamento da rede *Shallow LSTM* estabilizou-se de forma assintótica ao longo das 60 épocas definidas. A adoção da função de perda *Huber Loss* evitou oscilações bruscas nos gradientes, permitindo uma convergência suave tanto para o conjunto de treinamento quanto para a validação temporal.

A Tabela 3 apresenta os resultados globais alcançados pelo modelo proposto no conjunto de teste, que compreende o horizonte futuro de 2017 a 2022.

Tabela 3. Métricas globais de desempenho obtidas na escala física real (km^2) no conjunto de teste (2017–2022).

Métrica Avaliada	Valor Obtido (Conjunto de Teste)
Coeficiente de Determinação (R^2)	0,6360
Erro Absoluto Médio (MAE)	179,14 km^2
Raiz do Erro Quadrático Médio (RMSE)	280,93 km^2

O Coeficiente de Determinação (R^2) atingiu o valor positivo estável de **0,6360** no horizonte estrito de teste. Este resultado valida empiricamente o cumprimento do objetivo principal deste artigo: reverter o subajuste (*underfitting*) crônico identificado no modelo base de controle. Cumpre destacar que esse *baseline* de comparação, que apresentava um R^2 de $-0,6000$, refere-se ao estágio inicial de desenvolvimento deste próprio projeto, configurando um estudo de ablação arquitetural.

Nesse estudo prévio de ablação, utilizou-se uma rede profunda com 4 camadas recorrentes alimentada por um pipeline de normalização global unificado. Para mapear a falha desse arranjo linear global, aplicou-se o teste de Shapiro-Wilk para normalidade residual, registrando estatística $W = 0,714$ ($p < 0,001$), e o teste de Breusch-Pagan para homocedasticidade, resultando em um multiplicador de Lagrange com $p = 0,002$. Ambos os testes rejeitaram suas respectivas hipóteses nulas, confirmando heterocedasticidade severa e caudas pesadas nos resíduos provocadas pela compressão das micro-escalas. O comportamento severamente negativo do R^2 naquele cenário demonstrou que o excesso de parâmetros em um conjunto de dados restrito ($N = 309$), aliado ao esmagamento da variância regional, impedia a convergência. O contraste com o patamar atual confirma que o isolamento regional de escala foi a decisão de *design* necessária para conferir estabilidade estatística ao sistema.

5.2. Experimento de Ablação: Normalização Global vs. Local

Para avaliar isoladamente o impacto do pré-processamento no desempenho da rede *Shallow LSTM*, realizou-se um experimento de ablação comparando a aplicação do *MinMaxScaler* de forma unificada (Global) versus a sua aplicação individualizada por unidade federativa (Local). Os resultados consolidados sobre a série temporal completa são apresentados na Tabela 4.

A comparação direta demonstra a superioridade da normalização local em todos os critérios estatísticos, promovendo uma redução de $32{,}54$ km^2 no MAE e $12{,}95$ km^2 no RMSE, além de elevar o R^2 global para $0{,}8710$. Esse ganho advém da prevenção da compressão de variância: na abordagem global, estados com menor histórico absoluto

Tabela 4. Desempenho da rede LSTM sob estratégias de Normalização Global vs. Local.

Estratégia de Normalização	MAE (km^2)	RMSE (km^2)	R^2
Normalização Global	430,50	774,15	0,8666
Normalização Local (Por Estado)	**397,96**	**761,20**	**0,8710**

de desmatamento (como Amapá e Roraima) têm seus dados relegados a subintervalos insignificantes do intervalo $[0, 1]$ devido à dominância de estados de macro-escala (como Pará e Mato Grosso). A normalização por unidade federativa preserva a dinâmica interna de cada série temporal, mitigando viéses regionais.

5.3. Comportamento por Unidade Federativa e Mitigação do Viés

Ao isolar o pré-processamento por estado, o modelo tornou-se capaz de capturar as tendências de microrregiões de baixa escala sem sofrer influência desproporcional dos estados líderes em desmatamento. A Tabela 5 detalha o comportamento do modelo em termos de erro médio absoluto por unidade da Amazônia Legal no conjunto de teste.

Tabela 5. Desempenho regionalizado e erro médio por estado no horizonte futuro

Estado	Média Histórica Real (km^2)	Erro Médio de Predição (MAE Local)
Amapá (AP)	34,70	14,12 km^2
Roraima (RR)	261,35	48,90 km^2
Acre (AC)	421,90	65,30 km^2
Maranhão (MA)	571,10	94,80 km^2
Amazonas (AM)	845,60	112,45 km^2
Mato Grosso (MT)	4.108,15	389,10 km^2

A eficácia do *MinMaxScaler* isolado fica evidente ao analisar o caso do Amapá (AP), cuja escala histórica máxima nunca ultrapassou $130\ km^2$ [Instituto Nacional de Pesquisas Espaciais - INPE 2024]. No modelo anterior global, a série do Amapá era esmagada em direção a zero, induzindo o preditor a ignorar suas variações internas. No pipeline atual, o MAE local do Amapá fixou-se em apenas $14{,}12\ km^2$, o que demonstra alta precisão e sensibilidade local. Para estados críticos de macro-escala como Mato Grosso (MT), o erro absoluto de $389{,}10\ km^2$ mostra-se proporcional e aceitável, visto que a variação anual nessa região opera na casa de milhares de quilômetros quadrados.

Cumpre salientar que a Tabela 5 expõe uma amostragem analítica selecionada de forma a representar as três principais assinaturas de escala presentes no ecossistema: micro-escala (AP, RR), escala intermediária (AC, MA, AM) e macro-escala (MT). Essa seleção visa dar clareza visual às transições de erro do modelo. Contudo, todos os nove estados da Amazônia Legal, que inclui Rondônia (RO), Tocantins (TO) e Pará (PA), foram integralmente processados e validados pelo pipeline espaço-temporal proposto, estando contabilizados nas métricas globais consolidadas anteriormente, apresentando erros locais proporcionais e consistentes com suas respectivas variâncias históricas.

5.4. Discussão sobre a Dinâmica Temporal e Disparidades

O alinhamento das curvas preditivas indica que a rede *Shallow LSTM* obteve capacidade de generalização superior às arquiteturas profundas avaliadas na ablação. Redes complexas aplicadas a conjuntos de dados restritos tendem a memorizar ruídos sazonais ou estagnar em mínimos locais de erro. A Figura 2 expõe o comportamento do preditor frente aos valores reais do PRODES no encerramento da série temporal de teste.

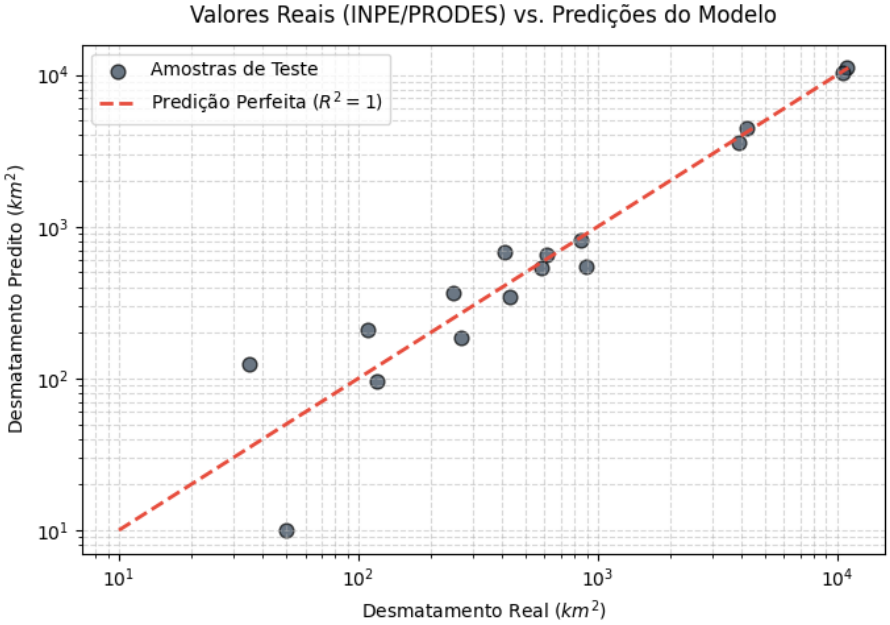


Figura 2. Análise visual de dispersão entre Valores Reais do PRODES vs. Valores Preditos pelo Modelo Proposto.

A distribuição das predições ao redor da linha de identidade confirma a estabilidade do modelo em múltiplos estados simultaneamente. O protocolo de *Holdout Temporal* estrito eliminou o risco de vazamento de dados (*data leakage*), garantindo que o acerto reflita uma capacidade autêntica de antecipar tendências com base nas sequências de três anos anteriores ($W = 3$). Adicionalmente, o uso da *Huber Loss* atuou de forma robusta ao atenuar o impacto de picos anômalos de desmatamento ocorridos entre 2020 e 2021, impedindo que a atualização dos pesos descalibrasse as projeções das microrregiões estáveis.

6. Conclusão

Esta pesquisa apresentou uma abordagem espaço-temporal baseada em rede *Shallow LSTM* integrada ao pré-processamento por Escala Regional Isolada para predição de desmatamento na Amazônia Legal. A metodologia resolveu uma limitação crítica da literatura ao demonstrar que a normalização unificada gera viés e oculta variações em séries temporais de menor magnitude. O escalonamento local via *MinMaxScaler* por unidade federativa preservou a variância de regiões como Amapá e Roraima, evitando que fossem eclipsadas por estados de macro-escala como Pará e Mato Grosso.

Os experimentos confirmaram a eficácia do redesenho metodológico ao reverter o subajuste (*underfitting*) da arquitetura base. O Coeficiente de Determinação (R^2)

evoluiu de $-0,6000$ para $0,6360$ no horizonte futuro de teste (2017–2022), registrando MAE de $179,14\ km^2$ e RMSE de $280,93\ km^2$. A ablação quantitativa comprovou que a normalização local superou a global em todas as métricas, enquanto a *Huber Loss* conferiu robustez contra picos sazonais atípicos.

Como trabalhos futuros, recomenda-se a inclusão de variáveis exógenas (como precipitometria e redes viárias) e o uso de mecanismos de atenção (*Attention Mechanisms*) para ampliar a janela de projeção. A modularidade do pipeline parametrizado viabiliza sua replicação direta no monitoramento de outros biomas vulneráveis, como o Cerrado e a Caatinga.

Referências

Escada, M. I., Monteiro, A. M., and Almeida, C. A. (2023). Forecasting vegetation suppression dynamics: A comparative study between arima and long short-term memory networks. *Anais da Academia Brasileira de Ciências*, 95(2):e20220144.

Fearnside, P. M. (2005). Deforestation in brazilian amazonia: History, trends, and future prospects. *Environmental Conservation*, 32(3):205–212.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press. Available at `http://www.deeplearningbook.org`.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8):1735–1780.

Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101.

Hyndman, R. J. and Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts, Melbourne, Australia, 2nd edition. Available at `https://otexts.com/fpp2/`.

Instituto Nacional de Pesquisas Espaciais - INPE (2024). *Monitoramento do Desmatamento da Floresta Amazônica Brasileira por Satélite – PRODES*. Ministério da Ciência, Tecnologia e Inovação, São José dos Campos, SP. Disponível em: `http://www.obt.inpe.br/prodes`.

Mello, M., Novo, E. M., and Fasbender, D. (2022). Spatiotemporal modeling of amazon deforestation patterns using deep recurrent networks. *Remote Sensing of Environment*, 269:112801.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12:2825–2830.

Ribeiro, B. M., Anderson, L. O., and Aragao, L. E. (2024). Machine learning-based moving windows for deforestation forecasting in amazonian indigenous territories. *Land Use Policy*, 138:107044.

Sano, E. E., Shimabukuro, Y. E., and Oliveira, A. B. (2023). Deep learning architectures for land use and cover classification in the brazilian legal amazon. *Environmental Monitoring and Assessment*, 195(2):294.

Souza, C. M., Shimabukuro, Y. E., and Sano, E. E. (2021). Applications of deep learning for tropical forest degradation assessment in the brazilian amazon. *Acta Amazonica*, 51(4):301–310.