

Climate Dataset Analyzer: Ferramenta para Extrair, Processar e Comparar Dados de Modelos e Estações Climáticas

Gustavo Machado¹, Warley Junior¹, Maurílio Monteiro² Elton Alves¹

¹Instituto de Geociências e Engenharias
Universidade Federal do Sul e Sudeste do Pará
Marabá – PA – Brasil

²Instituto de Estudos em Desenvolvimento Agrário e Regional
Universidade Federal do Sul e Sudeste do Pará
Marabá – PA – Brasil

{gustavo.paixao, wmvj, maurilio.monteiro, eltonalves}@unifesspa.edu.br

Abstract. *The proliferation and new generation of climate models significantly expand the possibilities for analyzing the impacts of climate change, both for the public and private sectors, at regional and local scales. However, given the large number of available models, each with its own biases, selecting the most suitable model for reference in a research project or initiative becomes a challenging task. In this context, the developed tool Climate Dataset Analyzer (CDA) emerges as an innovative solution, as it will facilitate the selection of the most appropriate model by integrating algorithms capable of comparing locally collected data with data derived from climate modeling.*

Resumo. *A proliferação e a nova geração de modelos climáticos ampliam significativamente as possibilidades de análise dos impactos das mudanças climáticas, tanto para os setores público quanto privado, em escalas regional e local. Contudo, dada a grande quantidade de modelos disponíveis, cada um com seus vieses, a seleção do modelo mais adequado para referência em uma pesquisa ou iniciativa torna-se uma questão desafiadora. Nesse contexto, a ferramenta desenvolvida CDA (Climate Dataset Analyzer) se apresenta como uma solução inovadora, uma vez que facilitará a escolha do modelo mais apropriado ao integrar algoritmos capazes de comparar dados coletados localmente com aqueles derivados de modelagens climáticas.*

1. Contexto

A Sexta Avaliação do Painel Intergovernamental sobre Mudanças Climáticas [IPCC 2021] apresenta evidências científicas e empíricas de que, no primeiro quarto do século XXI, as mudanças climáticas estão ocorrendo em ritmo mais acelerado. São conclusões e previsões baseadas em uma nova geração de modelos climáticos complexos (GCMs) e de modelos do sistema terrestre (ESMs) que aportaram informações acerca da variabilidade da mudança no sistema climático, especialmente em escalas regionais. Um dos esforços que deu suporte a tais avanços inclui o CMIP6 (*Coupled Model Intercomparison Project Phase 6*) [O'Neill et al. 2016], contribuindo para a difusão de modelos com graus diversos de complexidade e o compartilhamento de informações climáticas.

O CMIP6 envolveu uma grande variabilidade de modelos climáticos, com mais de 40 modelos fornecendo uma nova e rica fonte de informações para entender melhor as mudanças climáticas passadas, presentes e futuras [Bock et al. 2020]. Por outro lado, [Davini and d’Andrea 2020] argumentam que a resolução mais acurada por si só não explica ou sana todos os vieses dos modelos. O desempenho do modelo depende da formulação e das parametrizações do modelo tanto quanto da resolução [Eyring et al. 2021, Douville et al. 2021]. Portanto, apesar desses avanços e das melhorias registradas no âmbito do CMIP6, [Bock et al. 2020] indicam que ainda há imensas lacunas em todas as escalas. Assim, há diferenças significativas nas simulações climáticas destes modelos, diferenciações que se expressam também em termos regionais.

A existência dessas lacunas impõe a necessidade de comparações entre os resultados das simulações climáticas dos modelos com dados reais originários de medições locais indicando qual dentre os modelos analisados possuem maior aderência as dinâmicas locais. Assim, o objetivo deste estudo é apresentar o desenvolvimento de uma ferramenta web inovadora denominada CDA (*Climate Dataset Analyzer*) capaz de processar e comparar dados de *datasets* de diferentes modelos climáticos (tal como o CEDA - *Centre for Environmental Data Analysis*) [Harris et al. 2020] com dados reais provenientes do IMNET (Instituto Nacional de Meteorologia) [INMET 2024] através de métricas de regressão tal como o KGE (*Kling-Gupta Efficiency*). Essa ferramenta exige um esforço computacional considerável, devido à complexidade, dinâmica e não linearidade dos processos envolvidos nas variáveis climáticas.

2. Processo Adotado

A Figura 1 contém um conjunto de etapas que são essenciais para o funcionamento do CDA proposto. A etapa *Início da Aplicação* é o ponto de entrada da ferramenta, onde a aplicação é iniciada pelo comando `streamlit run app.py`. Este comando carrega a interface do usuário, como páginas e abas. Na etapa seguinte, *Carregamento das Configurações*, são carregados os arquivos para a configuração da aplicação, tais como, `settings.py`, que armazena configurações globais do projeto e `secrets.toml`, que contém credenciais e informações sensíveis.

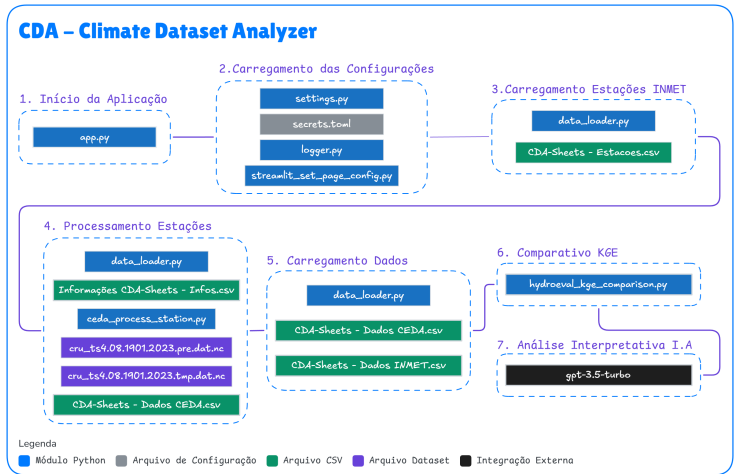


Figure 1. Etapas de funcionamento do CDA.

Na terceira etapa, *Carregamento Estações INMET*, é carregado os dados de estações cadastrados no arquivo excel, dados como longitude e latitude de cada estação. Na quarta etapa, *Processamento Estações*, o módulo `data_loader.py`,carrega os arquivos CSV com informações das estações e dos datasets utilizados, enquanto que o módulo `ceda_process_station.py` realiza o processamento dos dados das estações utilizando datasets do CEDA. Na quinta etapa, *Carregamento dos Dados*, carrega os conjuntos de dados processados do CEDA e dados do INMET para análise, para que na sexta etapa, *Comparativo KGE*, realize a análise comparativa utilizando o indicador KGE (*Kling-Gupta Efficiency*) através do módulo `hydroeval_kge_comparison.py`, que realiza o cálculo e comparação das métricas KGE e NSE entre os dados do CEDA e INMET utilizando a biblioteca `hydroeval`. Por fim, na última etapa, denominada *Análise Interpretativa IA*, o módulo `gpt-3.5-turbo` representa a integração com um modelo de IA (OpenAI GPT-3.5 Turbo) para gerar *insights* e interpretações automáticas dos dados comparados.

3. Solução

O CDA foi desenvolvido em *Streamlit*¹ para extrair e processar dados de estações climáticas e compará-los com *datasets* de modelos climáticos disponíveis. Atualmente, ele é projetado para trabalhar com dados do CEDA e INMET, utilizando ferramentas como *Google Sheets* para armazenamento de *datasets* processados e bibliotecas como *pandas*, *xarray*, *netCDF4* e *hydroeval* para gerenciamento e processamento dos *datasets* do *CEDA Archive* carregados localmente.

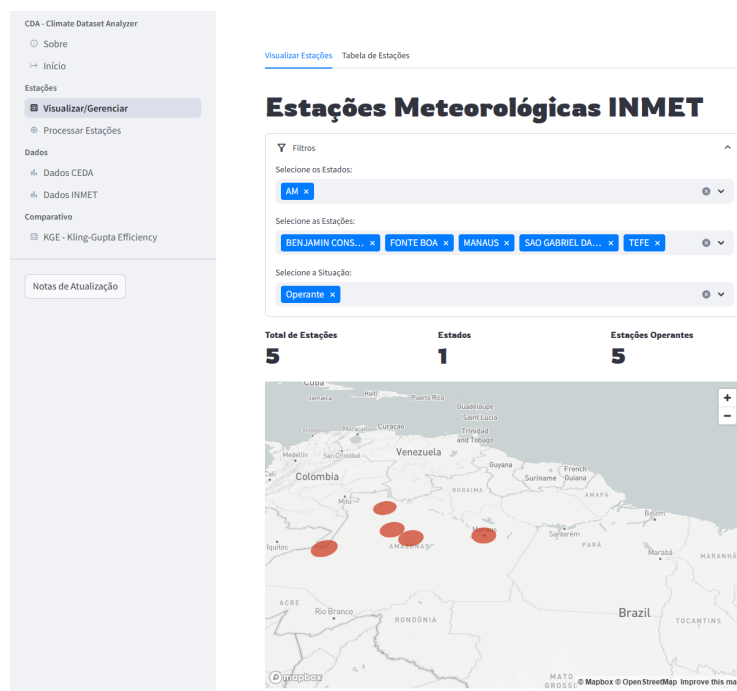


Figure 2. Tela de visualização das estações meteorológicas.

¹É um framework escrito em Python de código aberto para cientistas de dados. Pode ser acessado em <https://docs.streamlit.io/>

Conforme ilustrado nas Figuras 2 e 3, a ferramenta possui as seguintes funcionalidades: (1) Estações: exibe os dados das estações em um mapa e em formato tabular, permitindo visualização e gerenciamento das estações, e utiliza *datasets* do *CEDA Archive* para processar as estações e capturar valores de precipitação e temperatura correspondentes às estações selecionadas; (2) Dados: visualização dos dados de captação das coordenadas de temperatura e precipitação, subdividida em dados do CEDA, que contém informações processadas a partir do *dataset* `cru_ts4.08.1901.2023.pre.dat.nc` (Precipitação) e do *dataset* `cru_ts4.08.1901.2023.tmp.dat.nc` (Temperatura) disponíveis no CEDA e dados das estações do IMNET, ou seja, apresenta dados reais de estações INMET para visualização; (3) Comparativo: utiliza métricas de regressão tais como KGE e NSE (*Nash-Sutcliffe Efficiency*) para análise de similaridade entre os *datasets* processados do CEDA e os dados reais das estações do IMNET. A ferramenta pode ser acessada pela url: <http://5.161.202.226:8501/>

A ferramenta desenvolvida se apresenta como uma solução eficiente e inovadora para grupos de pesquisa e órgãos estaduais que necessitam realizar estudos climáticos em escalas regionais ou locais. Ela facilitará a escolha do modelo mais apropriado ao integrar algoritmos capazes de comparar dados coletados localmente com aqueles derivados de modelagens, oferecendo diagnósticos estatísticos paramétricos e não paramétricos sobre o desempenho dos modelos climáticos nos pontos geográficos de interesse dos usuários.

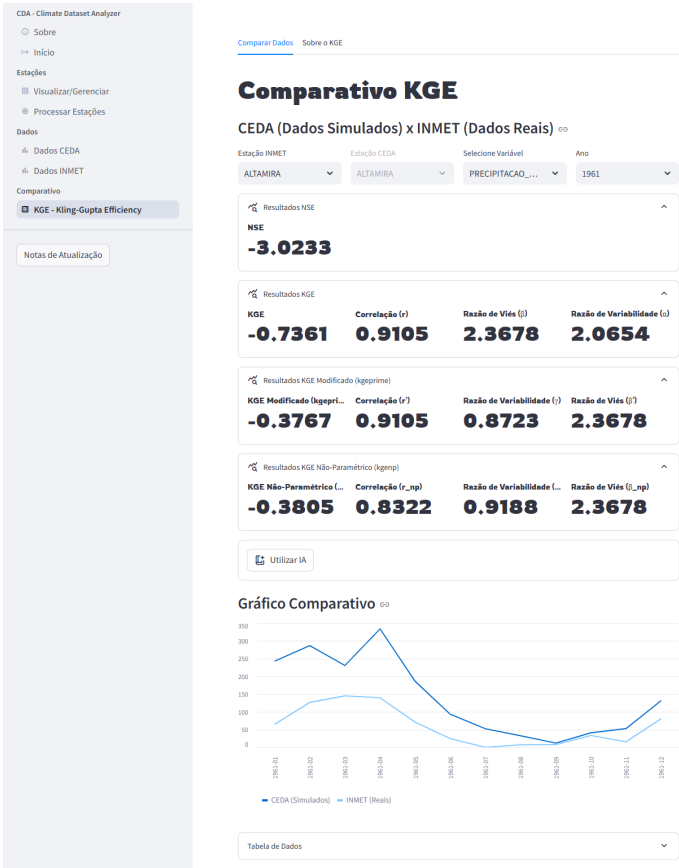


Figure 3. Comparação de dados simulados (CEDA) com dados reais (IMNET).

References

- Bock, L., Lauer, A., Schlund, M., Barreiro, M., Bellouin, N., Jones, C., Meehl, G., Pre-
doi, V., Roberts, M., and Eyring, V. (2020). Quantifying progress across different
cmip phases with the esmvaltool. *Journal of Geophysical Research: Atmospheres*,
125(21):e2019JD032321.
- Davini, P. and d'Andrea, F. (2020). From cmip3 to cmip6: Northern hemisphere at-
mospheric blocking simulation in present and future climate. *Journal of Climate*,
33(23):10021–10038.
- Douville, H., Raghavan, K., Renwick, J., Allan, R. P., Arias, P. A., Barlow, M., Cerezo-
Mota, R., Cherchi, A., Gan, T., Gergis, J., et al. (2021). Water cycle changes.
- Eyring, V., Gillett, N. P., Achuta Rao, K. M., Barimalala, R., Barreiro Parrillo, M., Bel-
louin, N., Cassou, C., Durack, P. J., Kosaka, Y., McGregor, S., et al. (2021). Human
influence on the climate system (chapter 3).
- Harris, I., Osborn, T. J., Jones, P., and Lister, D. (2020). Version 4 of the cru ts monthly
high-resolution gridded multivariate climate dataset. *Scientific data*, 7(1):109.
- INMET (2024). Banco de dados meteorológicos.
- IPCC (2021). *Climate Change 2021 – The Physical Science Basis: Working Group I
Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate
Change*. Cambridge University Press.
- O'Neill, B. C., Tebaldi, C., Van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G.,
Knutti, R., Kriegler, E., Lamarque, J.-F., Lowe, J., et al. (2016). The scenario model
intercomparison project (scenariomip) for cmip6. *Geoscientific Model Development*,
9(9):3461–3482.