

MÓDULO COMPUTACIONAL EM PLANILHA ELETRÔNICA PARA O CÁLCULO DE INDICADORES ESTATÍSTICOS PARA AVALIAÇÃO DE ANÁLISES DE REGRESSÃO

Jorge Luiz Moretti de Souza¹;

Universidade Federal do Paraná (UFPR), Curitiba, Paraná.

Matheus Salgado Moretti².

Universidade Federal Tecnológica do Paraná (UTFPR), Curitiba, Paraná.

RESUMO: Teve-se por objetivo no presente estudo levantar os indicadores estatísticos mais utilizados na bibliografia para verificar o desempenho de modelos em análises de regressão entre valores observados e estimados, bem como constituir um módulo computacional em planilha eletrônica para facilitar e agilizar os cálculos desses indicadores estatísticos. O módulo computacional foi desenvolvido no Laboratório de Modelagem de Sistemas Ambientais (LAMOSA) / DSEA / UFPR, em planilha eletrônica, com rotinas implementadas na linguagem de programação *Visual Basic for Applications (VBA-Macro)*. Foram criadas funções para os principais indicadores estatísticos utilizados na literatura: Coeficientes de correlação e determinação; Eficiência de Nash-Sutcliffe; Índices de concordância “*d*” de Willmott e desempenho “*c*” de Camargo-Sentelhas; Erro quadrado médio; Raiz quadrada do erro quadrado médio; Erro absoluto relativo; Erro absoluto médio; Erro relativo percentual médio; Erro relativo percentual absoluto médio; Erro máximo absoluto; e, Razão da média. O módulo desenvolvido foi denominado “Cálculo de indicadores estatísticos para avaliação de análises de regressão”. O módulo implementado permitiu a automatização dos cálculos dos indicadores e proporcionou eficiência, precisão e maior rapidez nas análises. Os resultados obtidos foram analisados e comparados com os padrões estabelecidos na literatura, e confirmou eficácia nos cálculos dos indicadores escolhidos.

PALAVRAS-CHAVE: Métricas. Coeficientes. Índices e Erros.

SPREADSHEET-BASED COMPUTATIONAL MODULE FOR CALCULATING STATISTICAL INDICATORS IN REGRESSION ANALYSIS EVALUATION

ABSTRACT: This study aimed to identify the most widely used statistical indicators in the literature for assessing model performance in regression analyses between observed and estimated values, as well as to develop a spreadsheet-based computational module to facilitate and expedite the calculation of these statistical indicators. The computational module was developed at the Environmental Systems Modeling Laboratory (LAMOSA

/ DSEA / UFPR) using a spreadsheet platform, with routines implemented in the *Visual Basic for Applications (VBA-Macro)* programming language. Functions were created for the main statistical indicators reported in the literature, including: correlation and determination coefficients; Nash–Sutcliffe efficiency; Willmott’s “*d*” concordance index; Camargo–Sentelhas “*c*” performance index; mean squared error; root mean squared error; relative absolute error; mean absolute error; mean relative percentage error; mean absolute relative percentage error; maximum absolute error; and mean ratio. The developed tool, entitled “Calculation of Statistical Indicators for Regression Analysis Evaluation”, enabled automated computations of these indicators, providing efficiency, precision, and faster analysis. The obtained results were analyzed and compared with literature standards, confirming the effectiveness of the implemented calculations.

KEY-WORDS: Metrics. Coefficients. Indices. Errors.

INTRODUÇÃO

Os indicadores estatísticos foram idealizados para a avaliação da precisão e confiabilidade de modelos de regressão, estabelecendo métricas para comparação entre valores observados e estimados (Souza, 2018). Os indicadores ou métricas permitem identificar desvios, melhorar a calibração de modelos e aumentar a robustez das análises preditivas em diferentes áreas de pesquisa. A correta seleção e aplicação dos indicadores (coeficientes, índices e erros) é essencial para garantir a validade das conclusões em estudos que envolvem análise de regressão (Legates e McCabe Junior, 1999; Souza et al., 2023a).

O objetivo de avaliar o desempenho de modelos preditivos em análises de regressão consiste em identificar os modelos que entregam as previsões mais precisas. Costuma-se comparar dois ou mais modelos e escolher o modelo ou a combinação de modelos com a maior precisão, que proporcione o menor erro (Souza, 2018). Grande parte dos indicadores de desempenho se baseiam na distância entre valores observados e estimados. A distância é conhecida como erro de previsão e o ideal é que esse valor seja o mínimo possível (Chicco et al., 2021). Geralmente, muitos testes e avaliações precisam ser feitos nas análises de regressão envolvendo ajustes de modelos, o que necessita tempo e atenção dos pesquisadores e técnicos nos estudos desenvolvidos. Na engenharia de água e solo, hidrologia, agrometeorologia, entre outras é muito comum a utilização de indicadores para avaliar o desempenho de modelos de estimativa testados em análise de regressão (Sentelhas et al., 1997; Legates e McCabe Junior, 1999; David e Mota, 2022; Souza et al., 2023a).

Na literatura tem-se inúmeros programas ou sistemas estatísticos (GraphPad Prism, Gretl, Minitab, R, SAS, Stata, Statistix, entre outros) que já oferecem diversas ferramentas que podem auxiliar na realização de cálculos estatísticos, como análises de regressão e de indicadores para avaliar o desempenho do modelo testado (Muenchen, 2023). As planilhas

eletrônicas também já trazem ferramentas que podem ser utilizadas para realizar análises estatísticas, incluindo regressão linear entre outras abordagens (Microsoft, 2025).

Para facilitar e agilizar as atividades do usuário, também existem diversas ferramentas e pacotes estatísticos que calculam coeficientes, indicadores e erros estatísticos para avaliar o desempenho de modelos de regressão linear. Algumas das opções mais utilizadas incluem: Pacotes como Metrics, Caret e HydroGOF do R, que oferecem métricas como R^2 , RMSE, MAE e NSE (R Core Team, 2025); Bibliotecas sklearn, metrics, statsmodels e scipy, stats do Python, que permitem calcular erros estatísticos e avaliar modelos (Python Software Foundation, 2025); e, Suplemento Data Analysis ToolPak do Excel, que possui ferramentas de regressão que calcula R^2 , coeficientes e resíduos automaticamente (Microsoft, 2025).

No entanto, se o objetivo for trabalhar com modelos mais específicos ou avançados, a utilização de complementos adicionais ou o desenvolvimento de *macros* na linguagem de programação *Visual Basic for Applications* (VBA-Macro) pode ser uma alternativa extremamente interessante para facilitar, agilizar e expandir as funcionalidades da planilha eletrônica, para a realização de análises repetitivas e que podem levar à erros sistemáticos (Souza e Frizzzone, 2003a; Souza e Frizzzone, 2003b). A avaliação e verificação do desempenho de modelos de estimativa (preditivos) em análises de regressão associando valores observados e estimados é muito utilizada em diversas áreas (Legates e McCabe Junior, 1999; Sentelhas et al., 1997; David e Mota, 2022; Souza et al., 2023b), e acredita-se que ainda tem-se muito por fazer no desenvolvimento de rotinas e módulos para facilitar, agilizar e assegurar a realização de cálculos em planilha eletrônica, concentrando toda atividade ou análise em um mesmo ambiente de trabalho, sem precisar da instalação de pacotes (Souza et al., 2022; Roncovsky e Souza, 2023; Souza et al., 2023a).

OBJETIVO

Diante do contexto apresentado, teve-se por objetivo no presente trabalho levantar os indicadores estatísticos mais utilizados na bibliografia para verificar o desempenho de modelos em análises de regressão entre valores observados e estimados, bem como constituir um módulo computacional em planilha eletrônica para facilitar e agilizar os cálculos desses indicadores estatísticos.

METODOLOGIA

O presente estudo foi realizado seguindo a linha de pesquisa (diagnóstico e análise de dados) da  Plataforma Moretti, tendo como apoio o Laboratório de Modelagem de Sistemas Ambientais / DSEA / SCA / UFPR (Souza et al., 2022; Roncovsky e Souza, 2023; Souza et al., 2023a).

Uma revisão bibliográfica inicial foi realizada selecionando-se artigos científicos que trataram do desenvolvimento ou utilização de indicadores estatísticos com a finalidade de

avaliar a qualidade ou desempenho de modelos em análises de regressão, comparando valores observados versus estimados (Lopes et al., 2023). Os critérios de seleção dos melhores indicadores ou métricas basearam-se na leitura de cada artigo, verificando-se: consistência; qualidade da avaliação realizada; resultados alcançados; e, aceite do indicador (métrica) pela comunicada científica.

O módulo computacional foi desenvolvido em planilha eletrônica. As rotinas computacionais para os procedimentos de execução e cálculo foram implementadas na linguagem de programação *Visual Basic for Applications (VBA-Macro)*. Foram criadas funções para os principais indicadores (métricas) estatísticos levantados na revisão de literatura, considerada no parágrafo anterior.

Os principais indicadores estatísticos (métricas) levantados na literatura e utilizados no módulo computacional foram: Coeficientes de correlação (r); Coeficientes de determinação (R^2); Eficiência de Nash-Sutcliffe (NSE); Índice de concordância “ d ”; Índice de desempenho “ c ”; Erro quadrado médio (EQM); Raiz quadrada do erro quadrado médio ($REQM$); Erro absoluto relativo (EAR); Erro absoluto médio (EAM); Erro relativo percentual médio ($ERPM$); Erro relativo percentual absoluto médio ($ERPAM$); Erro máximo absoluto ($EMaxA$); Razão da média (RM).

Os testes do módulo implementado em planilha eletrônica foram realizados baseando-se em análises estatísticas com dados experimentais obtidos de estudos anteriores, aplicando-se os indicadores mencionados para avaliar a precisão dos modelos de regressão. O módulo foi avaliado quando à sua capacidade de automatização dos cálculos dos indicadores, bem como na sua eficiência, precisão e rapidez nas análises. Os resultados obtidos foram analisados e comparados com os padrões estabelecidos na literatura, para confirmar a eficácia dos indicadores escolhidos para a verificação do desempenho dos modelos de regressão estudados.

RESULTADOS E DISCUSSÃO

O módulo desenvolvido foi denominado “Cálculo de indicadores estatísticos para avaliação de análises de regressão” (Figura 1). Sua utilização pode ser realizada nas seguintes formas: i) Digitando o comando da função, na célula ou “barra de fórmulas” da planilha eletrônica (Figura 2a); ou, ii) Utilizando o respectivo formulário que é criado automaticamente para as funções que são desenvolvidas no editor do *VBA-Macro*, no ambiente da planilha eletrônica (Figura 2b). Clicando-se no botão “inserir função”, que se encontra ao lado da barra de fórmulas, o usuário poderá ter acesso às funções selecionando a opção categoria “Definido pelo usuário”, na caixa de seleção. Posteriormente, escolhe-se a função na lista “Selecione uma função”; ou, iii) Utilizando o formulário desenvolvimento especialmente para essa finalidade, que pode ser chamado com as teclas de atalho “Shift+Ctrl+E” (Figura 2c). No formulário da Figura 2c o usuário faz a entrada dos respectivos dados observados e estimados, indica a célula inicial para dar saída dos resultados. Nas

opções, o usuário pode realizar o cálculo e saída dos resultados de todos os indicadores, ou pode escolher o indicador de sua preferência. Para orientação e identificação dos valores calculados, o usuário também pode escolher se deseja a saída das siglas e unidades dos indicadores.

Figura 1: Tela de abertura do módulo “Cálculo de indicadores estatísticos para avaliação de análises de regressão”.

A tela de abertura do módulo apresenta o logotipo da UFPR (Universidade Federal do Paraná) e o nome do departamento: DEPARTAMENTO DE SOLOS E ENGENHARIA AGRÍCOLA. Abaixo, há o logotipo da Plataforma Moretti e o nome do laboratório: LAMOSA - LABORATÓRIO DE MODELAGEM DE SISTEMAS AMBIENTAIS. O título principal do módulo é "PLATAFORMA MORETTI: MÓDULO - CÁLCULO DE INDICADORES ESTATÍSTICOS PARA AVALIAÇÃO DE ANÁLISES DE REGRESSÃO". O autor é identificado como Jorge Luiz Moretti de Souza (LAMOSA / DSEA / SCA / UFPR). Na seção de endereço, são fornecidas as informações de contato: Rua dos Funcionários, 1540 - Curitiba, PR; CEP: 80035-050; Telefone: (041) 3350-5689; e-mail: jmoretti@ufpr.br; e Site: http://www.moretti.agrarias.ufpr.br/. Na parte inferior direita, a versão do software é indicada como Versão 1.0 / 2025, com o código pMoretti_JLM S_2025.

UNIVERSIDADE FEDERAL DO PARANÁ / SETOR DE CIÊNCIAS AGRÁRIAS
DEPARTAMENTO DE SOLOS E ENGENHARIA AGRÍCOLA
UFPR [®] **M** Plataforma Moretti
LAMOSA - LABORATÓRIO DE MODELAGEM DE SISTEMAS AMBIENTAIS

**PLATAFORMA MORETTI: MÓDULO - CÁLCULO DE INDICADORES ESTATÍSTICOS
PARA AVALIAÇÃO DE ANÁLISES DE REGRESSÃO**

Autor: Jorge Luiz Moretti de Souza (LAMOSA / DSEA / SCA / UFPR)

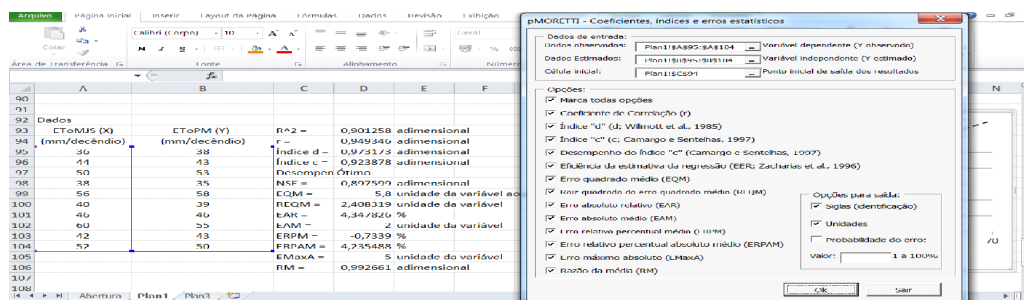
Endereço:
Rua dos Funcionários, 1540 - Curitiba, PR
CEP: 80035-050 Telefone: (041) 3350-5689
e-mail: jmoretti@ufpr.br
Site: http://www.moretti.agrarias.ufpr.br/

Versão 1.0 / 2025
pMoretti_JLM S_2025

Os testes do Módulo implementado em planilha eletrônica permitiram verificar como ocorreu a automatização dos cálculos dos indicadores. O Módulo proporcionou eficiência, precisão e muita rapidez nas análises. Nas Figuras 3 e 4 têm-se exemplos de análises de regressão linear. Na Figura 3 tem-se duas análises de regressão linear, realizando a associação entre a temperatura (°C) de determinada região e o consumo de energia (kWh). Na Figura 4 tem-se também duas análises de regressão linear, realizando a associação entre as evapotranspirações de referências (ET_o) estimadas com os métodos Penman-Monteith (PM ; padrão) e Moretti-Jerszurki-Silva (MJS ; alternativo) ($ET_{o_{MJS}}$ versus $ET_{o_{PM}}$), em que a exigência da proximidade da reta 1:1 ou de 45° pelo modelo de estimativa é importante e necessária.

(A)

(B)

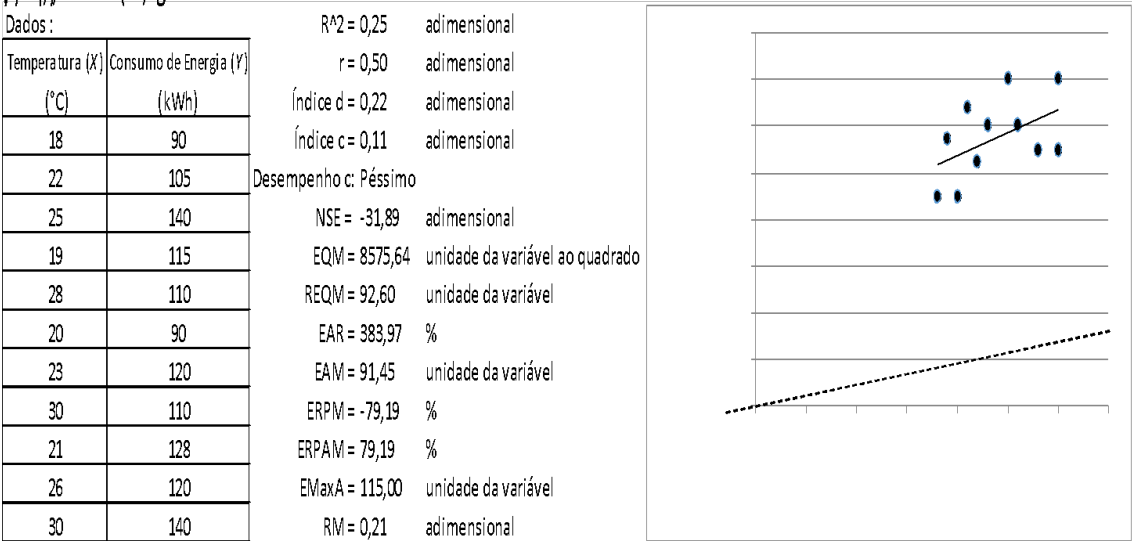


Nas análises de regressão linear da Figura 3 os melhores indicadores ou métricas precisam auxiliar o usuário na verificação de como os respectivos pares de pontos (estimados versus medidos) se encontram em relação à reta ou curva do modelo de regressão. Neste caso, é importante que os pontos se aproximem o máximo possível da reta ou curva de regressão. Um modelo preditivo realizará estimativas perfeitas se todos os pontos caírem sobre a reta ou curva de regressão analisada, algo próximo ao que está apresentado na Figura 3b. Nos exemplos da Figura 3, a utilização dos índices “*d*” e “*c*”, eficiência *NSE* e a maioria dos erros disponíveis não faz sentido de serem utilizados ou analisados, pois quantificam a proximidade entre os respectivos valores das variáveis dos eixos *X* e *Y*. Assim, não tem importância ou não é natural que valores de temperatura do ar sejam iguais ao valores de consumo de energia. O que se deseja é apenas verificar o grau de associação entre as variáveis estimadas (*X*; temperatura) e observadas (*Y*; consumo de energia), e para isso o mais indicado é utilizar os coeficientes de correção (*r*) ou determinação (*R*²). Assim, na Figura 3a tem-se uma associação fraca e dispersa, enquanto na Figura 3b tem-se uma associação estreita quase perfeita (*r* e *R*² estão muito próximos de 1).

Nas análises de regressão linear da Figura 4 os melhores indicadores ou métricas precisam auxiliar o usuário na verificação de como os respectivos pares de pontos (estimados versus medidos) se encontram em relação à reta do modelo de regressão. Neste caso, como as unidades dos eixos *X* e *Y* são as mesmas (mm decêndio⁻¹) e o modelo precisa estimar valores muito próximos aos observados, somente verificar se os pontos se aproxima o máximo possível da reta ou curva de regressão não é suficiente (como na Figura 4a). Na Figura 4a a associação foi boa, mas o modelo não está estimando bem os valores de *Y* (*ETo_{PM}*). É importante também que o modelo proposto ou estudado realize estimativas que se proximidade o máximo possível da reta 1:1 ou de 45°, evidenciando proximidade entre os respectivos valores estimados (*X*; *ETo_{JMS}*) e observados (*Y*; *ETo_{PM}*). Assim, um modelo preditivo somente realizará estimativas perfeitas se todos os pontos caírem sobre a reta de regressão analisada, e ao mesmo tempo ficarem sobre a reta 1:1 ou de 45° do gráfico de dispersão *XY*, contendo as variáveis estimadas (*X*; *ETo_{JMS}*) e observadas (*Y*; *ETo_{PM}*). Como se encontra apresentado na Figura 4b. Nas análises de regressão da Figura 4b, a utilização de todos os indicadores ou métricas disponíveis no Módulo computacional fazem sentido, e o usuário poderá escolher os indicadores ou métricas que se mostrarem mais promissores para estudar a área ou atividade de seu interesse.

Figura 3: Indicadores (métricas) das análises de regressão linear, associando “Temperatura (°C) versus Consumo de energia (kWh)”: a) Modelo linear não se ajustou ou explicou bem a relação entre temperatura e consumo de energia; b) Modelo linear se ajustou ou explicou bem a relação entre temperatura e o consumo de energia.

(A)



(B)

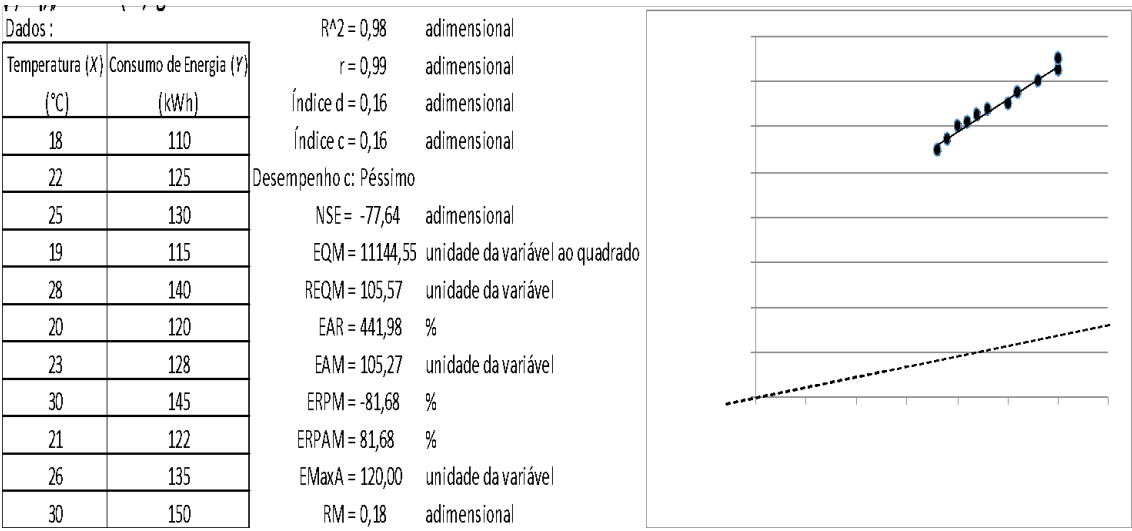
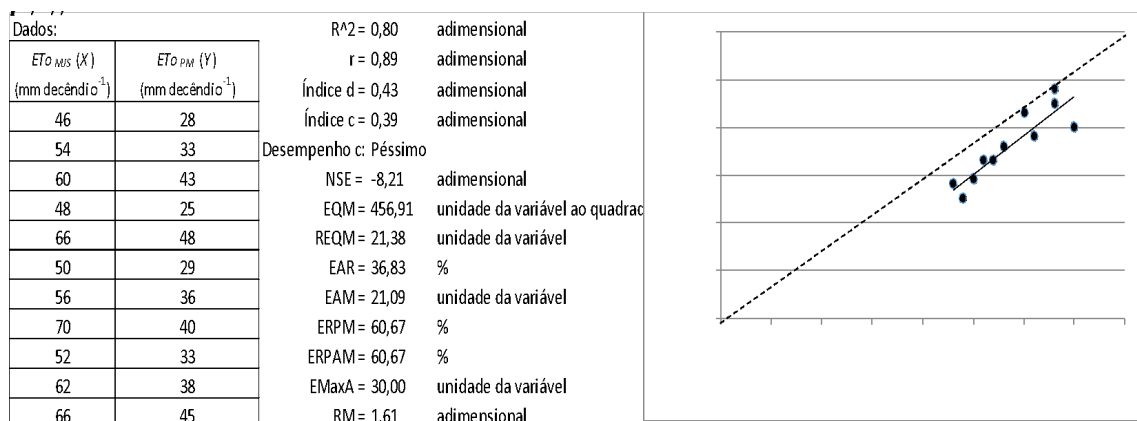
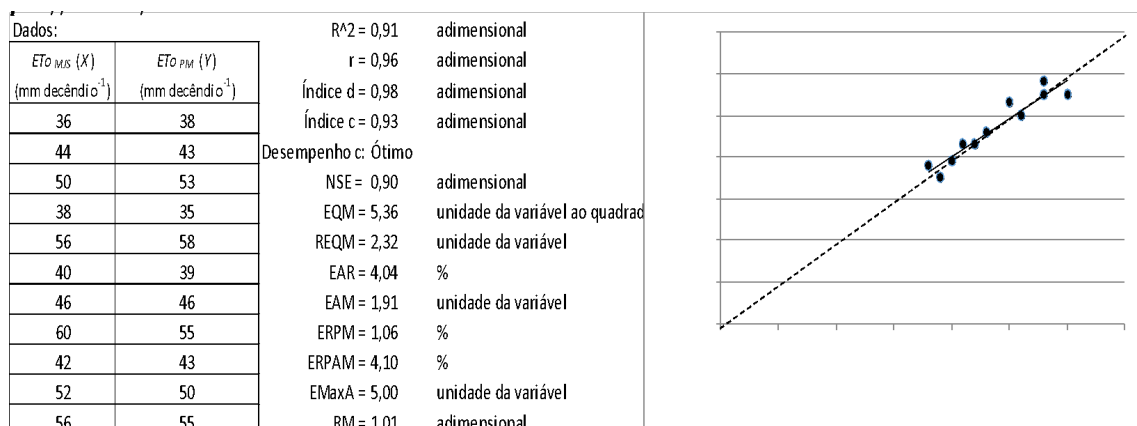


Figura 4: Indicadores (métricas) das análises de regressão linear, associando as evapotranspirações de referência (ET_o) estimadas com os modelos Penman-Monteith (PM ; padrão) e Moretti-Jerszurki-Silva (MJS ; alternativo), sendo: a) Modelo linear não se ajustou ou explicou bem a relação entre a “ $ET_{o_{MJS}}$ versus $ET_{o_{PM}}$ ”; b) Modelo linear se ajustou ou explicou bem a relação entre a “ $ET_{o_{MJS}}$ versus $ET_{o_{PM}}$ ”.

(A)



(B)



CONSIDERAÇÕES FINAIS

O Módulo implementado em planilha eletrônica permitiu a automatização dos cálculos dos indicadores e proporcionou eficiência, precisão e maior rapidez nas análises, sendo todo processo realizado em um mesmo ambiente de trabalho.

As rotinas e Módulo incrementará a biblioteca de ferramentas computacionais da Plataforma MORETTI, e poderão ser disponibilizadas à comunidade em geral.

REFERÊNCIAS

CHICCO, D.; WARRENS, M. J.; JURMAN, G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis

evaluation. PeerJ Computer Science, v. 5, n. 7, e623, 2021. <https://peerj.com/articles/cs-623/>

DAVID, L. M.; MOTA, T. M. Quality assessment of small urban catchments stormwater models: a new approach using old metrics. Hydrology, v. 9, n. 87, 28p., 2022. <https://doi.org/10.3390/hydrology9050087>

LEGATES, D. R.; McCABE JUNIOR, G. J. Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. Water Resources Research, v. 35, n. 1, p. 233-241, 1999.

LOPES, R. D.; VEZZANI, F. M.; PARAGUAIO, E. V. Abordagem da Qualidade do Solo nos trabalhos publicados no Brasil. **Revista Brasileira de Meio Ambiente**, v. 11, n. 1. p. 87-105, 2023.

MICROSOFT. Use the analysis ToolPak to perform complex data analysis. Microsoft Support. Disponível em: <https://support.microsoft.com/en-us/office/use-the-analysis-toolpak-to-perform-complex-data-analysis-6c67ccf0-f4a9-487c-8dec-bdb5a2cefab6>. Acesso em: 18 fev. 2025.

MUENCHEN, R. A. The popularity of data science software. 2023. Data Science Software Reviews. Disponível em: <https://r4stats.com/articles/popularity/>. Acesso em: 17 fev. 2025.

PYTHON SOFTWARE FOUNDATION. Python programming language. 2025. Disponível em: <https://www.python.org/downloads/>. Acesso em: 18 fev. 2025.

R CORE TEAM. The R Project for Statistical Computing. 2025. Disponível em: <https://cran.r-project.org/>. Acesso em: 18 fev. 2025.

RONCOVSKY, V.; SOUZA, J. L. M. Levantamento e análise dos parâmetros da equação de Angström-Prescott de localidades brasileiras para a constituição de um módulo computacional voltado à disponibilização de dados e estimativa da radiação solar incidente. Anais. CONVIBRA, 2023. 8p. (Evento Online: www.convibra.org)

SENTELHAS, P. C.; MORAES, S. O.; PIEDADE, S. M. S.; PEREIRA, A. R.; ANGELOCCI, L. R.; MARIN, F. R. Análise comparativa de dados meteorológicos obtidos por estações convencional e automática. Revista Brasileira de Agrometeorologia, v. 5, n. 2, p. 215-221, 1997.

SOUZA, J. L. M. Fundamentos de matemática e estatística para formulação de modelos e análise de dados: aplicado às ciências agrárias. Curitiba, Plataforma Moretti/DSEA/SCA/UFPR, 2018. (Manual didático)

SOUZA, J. L. M.; FRIZZONE, J. A. Modelo aplicado ao planejamento da cafeicultura irrigada. III. Análise de risco econômico da cafeicultura em dois sistemas de irrigação. Acta Scientiarum Agronomy, v. 25, n. 2, p. 399-408, 2003a.

SOUZA, J. L. M.; FRIZZONE, J. A. Modelo aplicado ao planejamento da cafeicultura irrigada. I. Simulação do balanço hídrico e do custo com água. Acta Scientiarum Agronomy, v. 25, n. 1, p. 103-112, 2003b.

SOUZA, J. L. M.; KIMURA, B. H.; UHLIK, C. F.; MAZIERO, R. O.; MORETTI, M. S. Módulo computacional para disponibilização do coeficiente kRs de Hargreaves e Samani de localidades brasileiras e cálculo da radiação solar incidente. Anais. CONVIBRA, 2023a. 8p. (Evento Online: www.convibra.org)

SOUZA, J. L. M.; MICHAOVSKI, R. E.; MOCELIN, L. K.; RONCOVSKY, V.; BILYK, C. G.; MORETTI, M. S. Levantamento de parâmetros e desenvolvimento de módulo para cálculo da intensidade, duração e frequência da precipitação pluviométrica de localidades brasileiras. Anais. CONVIBRA, 2022. 10p. (Evento Online: www.convibra.org)

SOUZA, J. L. M.; OLIVEIRA, C. T.; ROSA, S. L. K.; TSUKAHARA, R. Y. Calibration and validation of the AquaCrop model to estimate soybean production in the Campos Gerais, Parana State, Brazil. CIGR Journal, v. 25, n. 4, p. 1-13, 2023b.