

PGECLIMA_R: GERADOR ESTOCÁSTICO PARA SIMULAÇÃO DE CENÁRIOS CLIMÁTICOS BRASILEIROS. II – AUTOMAÇÃO DA ANÁLISE ESTATÍSTICA E VALIDAÇÃO DOS DADOS SIMULADOS

JORIM S. VIRGENS FILHO¹, PIETRO M. OLIVEIRA², MAYSIA L. LEITE³, RODRIGO Y. TSUKAHARA⁴

¹ Matemático, Prof. Associado, Laboratório de Estatística Computacional e Aplicada - LECA/Departamento de Matemática e Estatística, Universidade Estadual de Ponta Grossa - UEPG, Ponta Grossa - PR, Brasil. Fone: (0 xx 42) 3220-3050, jvirgens@globo.com

² Acadêmico de Engenharia de Computação, Desenvolvedor de sistemas, LECA, UEPG, Ponta Grossa - PR, Brasil.

³ Engenheiro Agrônomo, Prof. Associado, Departamento de Biologia Geral, UEPG, Ponta Grossa - PR, Brasil.

⁴ Engenheiro Agrônomo, Pesquisador, Fundação ABC, Castro-PR, Brasil.

Apresentado no XVII Congresso Brasileiro de Agrometeorologia – 18 a 21 de julho de 2011 – SESC Centro de Turismo de Guarapari, Guarapari - ES

RESUMO: Este trabalho objetivou apresentar uma solução computacional para automação do processo de validação estatística, das séries de dados climáticos simuladas pelo PGECLIMA_R. Foram implementadas rotinas na linguagem de programação C++ Builder, que executam cálculos de médias, desvios-padrões, distribuições de frequências com gráficos e índices para analisar o viés estatístico proveniente de representações da realidade por simulações. Verificou-se que, com a validação automática, a análise dos dados ficou simplificada, descartando assim o uso de planilhas eletrônicas como opção para avaliação de séries climáticas geradas sinteticamente.

PALAVRAS-CHAVE: PGECLIMA_R, simulador de clima, estatística climatológica.

PGECLIMA_R: GENERATOR FOR STOCHASTIC SIMULATION OF BRAZILIAN CLIMATE SCENARIOS. II - AUTOMATION OF STATISTICAL ANALYSIS AND VALIDATION OF SIMULATED DATA

ABSTRACT: This study aimed to present a computational solution to automate the process of statistical validation for the series of climatic data simulated by PGECLIMA_R. Routines were implemented in the programming language C++ Builder, running averages, standard deviations, frequency distributions with graphs and ratios for analyzing statistical bias from representations of reality by simulations. It was found that with automatic validation, data analysis was simplified, thus ruling out the use of spreadsheets as an option for assessment of climate series generated synthetically.

KEY-WORDS: PGECLIMA_R, weather generator, climatological statistics.

INTRODUÇÃO: Os modelos para simulação de séries de dados climáticos, que são de grande utilidade na avaliação e planejamento de sistemas agropecuários e hidrológicos, só devem ser considerados adequados para tal fim, após uma extensiva bateria de testes e análises estatísticas, com o intuito de verificar a exatidão entre dados observados (históricos) e simulados. Esta característica, que é prerrogativa de um bom gerador de dados climáticos, geralmente é constatada realizando-se análises estatísticas criteriosas, comparando as distribuições de frequências, valores médios e variabilidade em torno dos valores médios. Poucos softwares de simulação de dados climáticos apresentam esta alternativa, para validar automaticamente os dados gerados. Como exemplo, pode-se citar o gerador CLIGEN (Nicks et al., 1995), que não apresenta uma interface que tenha como saída, um relatório de análise

dos dados simulados. Neste caso, em que o software não faz os cálculos automaticamente, as análises, na maioria das vezes, são realizadas manualmente em planilhas eletrônicas, o que causa um ônus ao analista, por ser um processo dispendioso devido à grande quantidade de procedimentos a serem efetuados. No intuito de resolver este problema de validação, é muito importante incluir no software, um módulo que realize a validação automática dos dados simulados, como é o caso do gerador LARS-WG (Semenov e Barrow, 1997) que oferece uma solução chamada QTEST, que realiza os testes Qui-Quadrado, *t*-Student e F-Snedecor para comparar as séries simuladas e históricas. Desta forma, tendo em vista o desenvolvimento do Gerador Estocástico de Cenários Climáticos – PGECLIMA_R, o objetivo deste trabalho é apresentar uma ferramenta que possibilite automatizar o processo de validação estatística, confrontando os dados observados e simulados, descartando assim o uso de planilhas eletrônicas, como opção para a verificação e a validação das séries sintéticas geradas pelo PGECLIMA_R.

MATERIAL E MÉTODOS: Este trabalho foi desenvolvido no Laboratório de Estatística Computacional e Aplicada - LECA, Departamento de Matemática e Estatística, Universidade Estadual de Ponta Grossa - UEPG. Para tanto foi utilizada a linguagem de programação Borland C++ Builder para plataforma Windows. A modelagem estocástica do processo de simulação implementado no PGECLIMA_R, para a geração sintética dos dados diários de precipitação pluviométrica, umidade relativa do ar, radiação solar, temperatura do ar (mínima e máxima), encontra-se descrita em Virgens Filho (2001). Para efetuar a análise das séries simuladas foi necessário classificar os dados conforme o estado de precipitação (dias secos ou chuvosos), o mês do ano e variável climática. Foram implementadas rotinas computacionais para cálculos de estatísticas elementares como frequências simples e acumuladas, tanto absolutas como relativas, médias e desvios-padrões. Para a realização de inferências estatísticas, foram codificados os algoritmos para o teste de Kolmogorov-Smirnov (K-S), que avalia as formas das distribuições de frequências; o teste F, que avalia a homocedasticidade das variâncias e o teste *t*, que foi implementado com o objetivo de comparar as médias observadas e simuladas. Com o intuito de analisar o viés estatístico proveniente de representações da realidade por simulações, para avaliar a qualidade do desempenho de modelos (Willmot, 1982), foram utilizados conforme recomendação de Fox (1981), o cálculo de três índices: O MBE que mensura o erro da tendência das médias; o RMSE que calcula a raiz do erro quadrático médio; o MAE que resulta no erro absoluto das médias. Estes índices foram implementados computacionalmente pelas seguintes equações:

$$MBE = \frac{\sum_{i=1}^N (P_i - O_i)}{N} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (P_i - O_i)^2}{N}} \quad (2)$$

$$MAE = \frac{\sum_{i=1}^N |P_i - O_i|}{N} \quad (3)$$

onde,

P_i é a média dos dados simulados pelo modelo;

O_i é a média dos dados históricos;

N é o número de anos simulados.

Para complementar a análise, incluiu-se também no programa, o índice de concordância "d" proposto por Willmott (1981), que é calculado pela seguinte equação:

$$d = 1 - \left[\frac{\sum_{i=1}^N (P_i - O_i)^2}{\sum_{i=1}^N (|P_i| + |O_i|)^2} \right], 0 \leq d \leq 1 \quad (4)$$

onde,

$$P_i' = P_i - \bar{O};$$

$$O_i' = O_i - \bar{O};$$

\bar{O} é a média total dos dados históricos;

P_i é a média dos dados simulados pelo modelo;

O_i é a média dos dados históricos;

N é o número de anos simulados.

RESULTADOS E DISCUSSÃO: A Figura 1 mostra a tela principal do processo de validação estatística dos dados gerados pelo PGECLIMA_R. Observa-se que os dados são categorizados em 15 classes, por variável climática, mês e estado de precipitação (dia chuvoso ou seco), onde são resumidas as estatísticas diárias e mensais média, desvio-padrão e o tamanho da amostra ("n"), tanto para os dados observados quanto para os dados simulados, que são utilizadas para realização dos testes F e t. Para comparar as distribuições de frequências mensais observadas e simuladas, também é efetuado o teste de K-S aos níveis de significância de 5% e 1%.

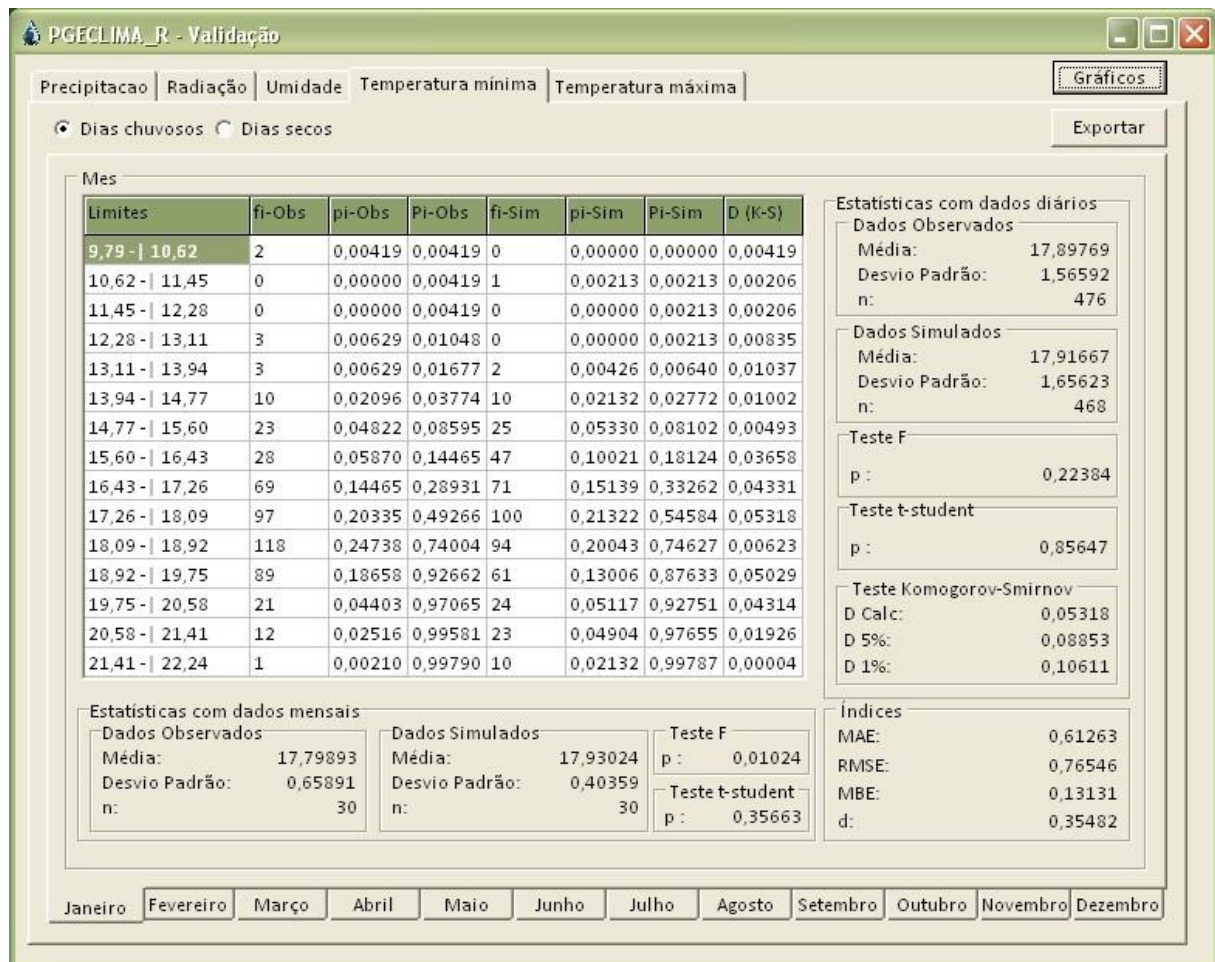


Figura 1 – Tela principal da validação estatística.

No intuito de subsidiar uma análise no viés estatístico das médias mensais, estão disponibilizados os índices MAE, RMSE, MBE e “d” de Willmott. Na Figura 2 são apresentados o “Polígono de frequência” e a “Ogiva relativa”, onde são mostradas graficamente por mês as distribuições de frequências relativas simples e acumuladas, que por sua vez, podem permitir a visualização de possíveis enviesamentos dos dados simulados em relação aos observados (Figura 2-A), bem como das discrepâncias verificadas pelo teste K-S (Figura 2-B). A Figura 3 mostra os gráficos mensais das médias diárias (Figura 3-A) e de seus respectivos desvios-padrões (Figura 3-B), onde se pode observar graficamente o grau de eficiência da simulação dos valores diários, em termos da reprodução da média e da variabilidade em seu entorno.

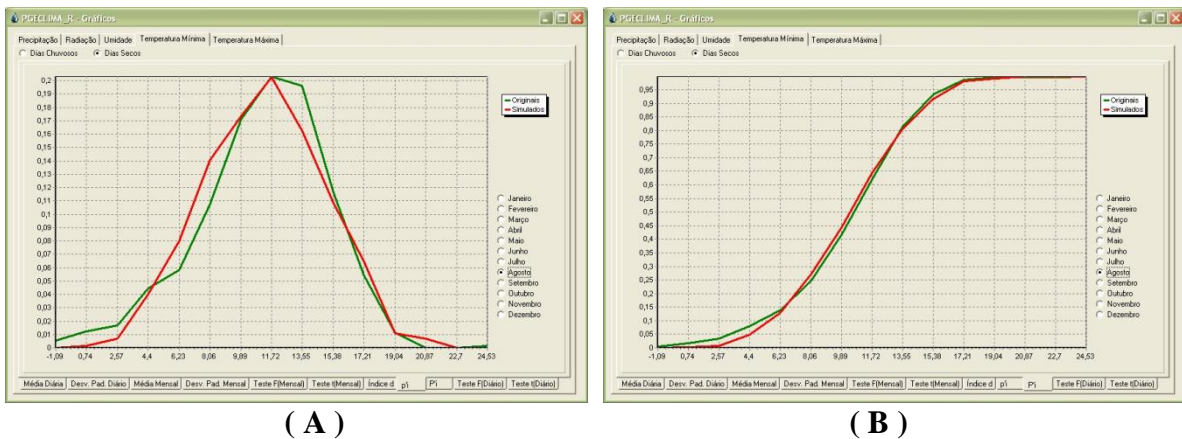


Figura 2 – Polígono de frequência (A) e Ogiva relativa (B) apresentando as distribuições de frequências observadas (linha verde) e simuladas (linha vermelha).



Figura 3 – Gráficos mensais das médias diárias (A) e seus respectivos desvios-padrões (B).

Na Figura 4 são disponibilizados graficamente os valores mensais de p (significância estatística) verificados pelos teste F (Figura 4-A) e t (Figura 4-B). É importante frisar que os tipos de gráficos apresentados nas Figuras 3 e 4, também estão disponíveis no PGECLIMA_R, para todas as variáveis e estado de precipitação, tendo por base as distribuições de médias mensais, exceto no caso da precipitação pluviométrica em que são avaliadas as distribuições dos totais mensais. Comparando-se os cálculos realizados pelo PGECLIMA_R com os realizados manualmente em planilhas eletrônicas, pode-se observar que a eficiência com que o programa efetua os mesmos é bastante significativa, pois o tempo gasto programando os cálculos nas planilhas para a uma localidade com as mesmas variáveis climáticas, foi aproximadamente de 8 horas, sendo que o programa realizou os mesmos cálculos em aproximadamente 30 segundos.



Figura 4 – Gráficos apresentando os valores de p mensais dos testes F (A) e t (B) para dados diários.

Em relação ao simulador LARS-WG (Semenov e Barrow, 1997) que por meio da solução QTEST, realiza a validação de dados simulados, pode-se perceber que a interface de validação automática se mostrou muito mais ágil, prática e intuitiva, no sentido que as análises são apresentadas diretamente na tela do computador, o que não acontece no LARS-WG que gera um relatório com as análises em um arquivo do tipo texto.

CONCLUSÕES: A ferramenta de validação estatística automática dos dados climáticos simulados pelo PGECLIMA_R se mostrou muito eficiente e intuitiva, disponibilizando gráficos e testes estatísticos, os quais dispensaram o uso de planilhas eletrônicas como opção para avaliação de séries climáticas geradas sinteticamente.

AGRADECIMENTOS: Os autores agradecem o CNPq, a Fundação Araucária de Apoio ao Desenvolvimento Científico e Tecnológico do Paraná e a FINEP pelo auxílio financeiro concedido para o desenvolvimento desta pesquisa.

REFERÊNCIAS BIBLIOGRÁFICAS:

FOX, D.G. Judging air quality model performance: A summary of the AMS Workshop on Dispersion Model Performance. **Bulletin American Meteorological Society**, v.62, n.5, p.599-609, 1981.

NICKS, A.D., LANE, L.J., GANDER, G.A. **Weather Generator**. In: USDA-Water erosion prediction project (WEPP). West Lafayette: USDA/ARS, 1995. p.2.1-2.22.

SEMENOV, M.A., BARROW, E.M. Use of a stochastic weather generator in the development of climate change scenarios. **Climatic Changes**, v.35, p.397-415, 1997.

VIRGENS FILHO, J.S. **Ferramenta computacional para simulação de séries climáticas diárias, baseada na parametrização dinâmica das distribuições de probabilidade**. Botucatu, 92p., 2001. Tese (Doutorado em Agronomia/ Energia na Agricultura) - Faculdade de Ciências Agrônomicas, Universidade Estadual Paulista.

WILLMOTT, C.J. On the validation of models. **Physical Geography**, v.2, 184-194, 1981.

WILLMOTT, C.J. Some comments on the evaluation of model performance. **Bulletin American Meteorological Society**, v.63, n.11, p.1309-1313, 1982.