

Totalmente Online - 14 a 19 de novembro de 2020

GT: Estudos e Modelagem do Tempo e Clima

AVALIAÇÃO DO DESEMPENHO DO MICE NO PREENCHIMENTO DE FALHAS EM DADOS DE TEMPERATURA MÁXIMA DO AR NA REGIÃO METROPOLITANA DE SÃO PAULO

Tiago Bentes Mandú¹, Mary Toshie Kayano², Giovana Deponte Galetti³, Nicole Cristine Laureanti⁴,
Eduardo Almeida da Silva⁵.

RESUMO

Um dos maiores desafios relacionados ao uso de registros meteorológicos obtidos a partir de estações meteorológicas de superfície é a incompletude destes dados. Diante disso, é necessário utilizar métodos eficazes para preencher essas falhas, na literatura científica existe uma ampla variedade de abordagens que podem ser empregadas para realizar o preenchimento destas lacunas. O objetivo do presente estudo foi analisar a eficiência do algoritmo *Multivariate Imputation by Chained Equation* (MICE) no preenchimento de registros de temperatura máxima do ar na Região Metropolitana de São Paulo (RMSP). Foram utilizados dados diários de temperatura máxima e mínima do ar coletados a partir da estação meteorológica convencional nº 83781 pertencente ao Instituto Nacional de Meteorologia (INMET), localizada na área central da RMSP, para o período de 1961 a 2019. O controle de qualidade dos dados foi realizado a partir da variação interquartil. O MICE é uma metodologia estatística baseada no amostrador de Gibbs, uma técnica de simulação Bayesiana que utiliza as distribuições condicionais dos dados a fim de obter amostras de distribuição conjunta usadas para gerar múltiplas imputações a partir de interações em conjuntos de dados univariados utilizando cadeias de Markov via Monte Carlo, possui 24 diferentes métodos de ligação que podem ser utilizados para realizar as imputações. Neste estudo, 20 interações foram comparadas três diferentes abordagens: Média Preditiva Correspondente Ponderada (MPCP), Imputação via Florestas Aleatórias (IFA) e Regressão Linear Bayesiana (RLB) um total de 15 % dos dados foi utilizada, que é a porcentagem recomendada, realizada mês a mês a fim de reduzir os efeitos de sazonalidade. Além disso, foi avaliada a qualidade das estimativas a partir do Erro Médio Absoluto (EMA), Raiz do Erro Quadrático Médio (RMSE) e Viés Percentual (VP). Com isso, se observou que o maior desempenho nos três índices para a MPCP foi observado em janeiro, junho, julho, novembro e dezembro, em agosto para a IFA e nenhum mês para a RLB. O EMA foi menor (3,31°C) no IFA em março e maior (6°C) em junho para a RLB, o RMSE mais baixo (4,14°C) foi observado para a RLB também no mês de março, enquanto o maior (7,4°C) foi observado para a MPCP em agosto. Para o VP, o menor registro de -0,5% foi visto em dezembro para a MPCP, em março e dezembro para a RLB, enquanto que os maiores percentuais de -24,6% (janeiro), -19,0% (junho) e -18,5% (julho), para a IFA, RLB e MPCP, respectivamente, mostrando que o modelo apresentou subestimação nos três métodos de ligação comparados. Em relação à média, os menores EMA (4,45°C) e RMSE (4,97°C) foram para o IFA enquanto que para o VP o valor mais baixo (-1,18%) foi visto para a MPCP. Os resultados apontam que as três ligações do MICE apresentam alta variabilidade na precisão entre os meses, assim como baixos erros, sendo esses menores para a MPCP e IFA, o algoritmo MICE apresenta desempenho satisfatório considerando a limitação de se utilizar apenas informações do próprio conjunto de dados.

PALAVRAS-CHAVE: Imputação mês a mês, Estatística aplicada a meteorologia, Algoritmo, Modelagem, Dados meteorológicos.

¹ Estudante de Pós-Graduação em Meteorologia, Instituto Nacional de Pesquisas Espaciais.

² Docente do programa de Pós-Graduação em Meteorologia, Instituto Nacional de Pesquisas Espaciais.

³ Estudante de Pós-Graduação em Meteorologia, Instituto Nacional de Pesquisas Espaciais.

⁴ Estudante de Pós-Graduação em Meteorologia, Instituto Nacional de Pesquisas Espaciais.

⁵ Estudante de Graduação em Meteorologia, Universidade Federal de Campina Grande.