

AVALIAÇÃO DA PREVISIBILIDADE DO MODELO METEOROLÓGICO GFS PARA FINS DE ALERTA PRECOCE

Palavras-Chave: TEMPERATURAS EXTREMAS, CAMPINAS, DESEMPENHO DO MODELO

Autores(as):

BEATRIZ JULIÃO REIS, IMECC – UNICAMP

BRUNO KABKE BAINY, CEPAGRI – UNICAMP

Prof^(a). Dr^(a). ANA MARIA HEUMINSKI DE AVILA, CEPAGRI – UNICAMP

INTRODUÇÃO:

Alertas meteorológicos são um dos maiores desafios para os serviços de meteorologia, sobretudo em regiões tropicais. Este trabalho tem por objetivo principal investigar a potencialidade da utilização do modelo GFS para previsão de temperaturas extremas para fins de alerta. A pesquisa consiste avaliar estatística e comparativamente as previsões obtidas a partir do modelo GFS (*Global Forecast System*), com antecedência de até 7 dias, e os dados medidos na estação meteorológica do CEPAGRI, em Campinas, ao longo do período de 2 (dois) anos (2020 e 2021). Como resultado final, avalia-se o desempenho do modelo GFS, sobretudo nos quesitos da antecedência da previsão e da confiabilidade para emissão de alertas frente aos eventos de temperaturas extremas.

METODOLOGIA:

O primeiro passo do projeto foi a obtenção de dados do modelo GFS (*Global Forecast System*), que foram obtidos em <https://rda.ucar.edu/datasets/ds084.1/>. Foram solicitados dados de Análise (interpolação/ajuste de dados observacionais para o horário em questão) e dados de previsão de temperatura (para cada simulação do modelo, os dados são disponibilizados em intervalos de 6h em 6h, e este trabalho contemplou previsões para até 168h - ou seja, previsões para até 7 dias de antecedência – o “lead time”, ou LT).

Como as simulações do modelo GFS são feitas 4 vezes por dia, foram obtidos 84825 dados, que consistem aproximadamente em 731 dias (entre as 00UTC do dia 01/01/2020 e as 00UTC do dia 01/01/2022), com quatro simulações por dia (ou seja, 2924 simulações), cada qual com uma série de 29 horários (sendo um deles a Análise e as 28 remanescentes as previsões propriamente ditas).

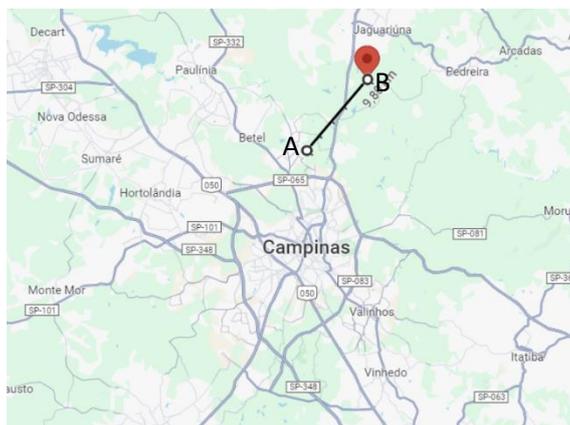


Figura 1. Localização da estação meteorológica do Cepagri (ponto "A") e o ponto de grade mais próximo do modelo (ponto "B"). Fonte: Google Maps.

Foi selecionado o ponto de grade mais próximo da estação meteorológica do Cepagri (uma vez que é essa a origem dos dados que utilizamos de referência para a avaliação do modelo), nas coordenadas 22°45'00.0"S 47°00'00.0"W, a aproximadamente 9,8km da estação meteorológica do Cepagri, que foi considerada uma distância satisfatória (Figura 1). Dados de temperatura do Cepagri (médias sobre 10 minutos), foram amostrados nos horários coincidentes com os fornecidos pelo modelo GFS.

RESULTADOS E DISCUSSÃO:

Na primeira etapa de análise de dados, foram calculados os limiares de temperaturas (percentis

Tabela 1. Extremos de temperatura em Campinas nos anos de 2020 e 2021, conforme dados do CEPAGRI.

Percentis de temperatura	Limiares de temperatura (°C)		
	5%	1%	Absoluto
Mais frios	14,1	10,7	3,6
Mais quentes	30,5	33,2	39,0

de 5%, 1% e absoluto - mais quentes e mais frios) dos dados observacionais do Cepagri, para um panorama geral do comportamento da temperatura nesse período (Tabela 1). Além disso, nota-se que o maior valor

registrado (39°C) foi uma das maiores temperaturas já registradas pelo CEPAGRI (desde 1989) e o menor valor (3,6°C) não havia sido registrado desde o início dos anos 2000.

O conjunto dos dados do modelo foi organizado em 29 séries temporais (temperaturas a cada 6h), baseado no critério de LT. Posteriormente, esses dados foram analisados comparativamente e estatisticamente com os dados do Cepagri para cada horário em questão. Foram calculadas estatísticas, apresentadas graficamente na Figura 2: Coeficiente de correlação (r), Fator de Dois (FAC2), Erro Médio Absoluto (MGE), Erro Médio Quadrático (RMSE) e Viés Médio (MB), para cada LT. Pode-se observar que, como seria esperado, a Correlação entre os dados observacionais e as previsões decai quanto maior o LT, enquanto o Erro Médio Quadrático (RMSE) e o Erro Médio Absoluto (MGE) aumentam – em ambos casos, o valor ideal é zero, para um modelo perfeito.

Contudo, o Viés Médio (MB) se aproxima de 0.0, ou seja, melhora – a principal hipótese para isso é que a previsão do modelo oscila mais para períodos mais longos (maiores LTs), e, dessa forma, super e subestimativas dessas oscilações se compensam. O Fator de Dois (FAC2) o valor ideal é 1.0), o que era esperado – ele indica a fração das previsões que ficaram entre a metade e o dobro do valor observado; ou seja, num contexto de previsão de temperaturas em graus Celsius para a região de Campinas, é muito raro que isso não aconteça. Outro importante ponto a se notar é a discrepância da Análise em comparação com as primeiras previsões: espera-se que a Análise seja a representação mais próxima da realidade, sendo inesperado que ela possua uma Correlação menor, e Erros Médio

Quadrático (RMSE) e Médio Absoluto (MGE) maiores do que as primeiras previsões (até 72h, aproximadamente).

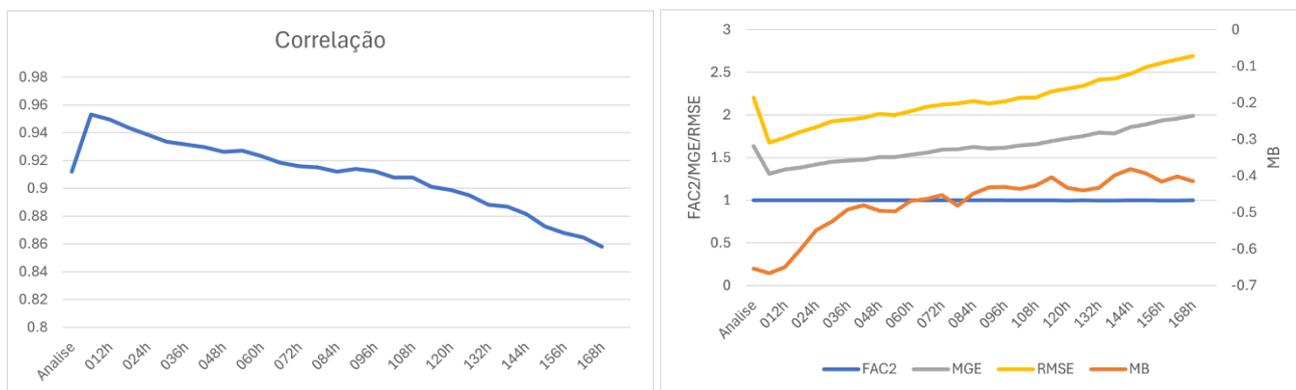


Figura 2. Coeficientes estatísticos para as previsões com base na antecedência com que a previsão é feita.

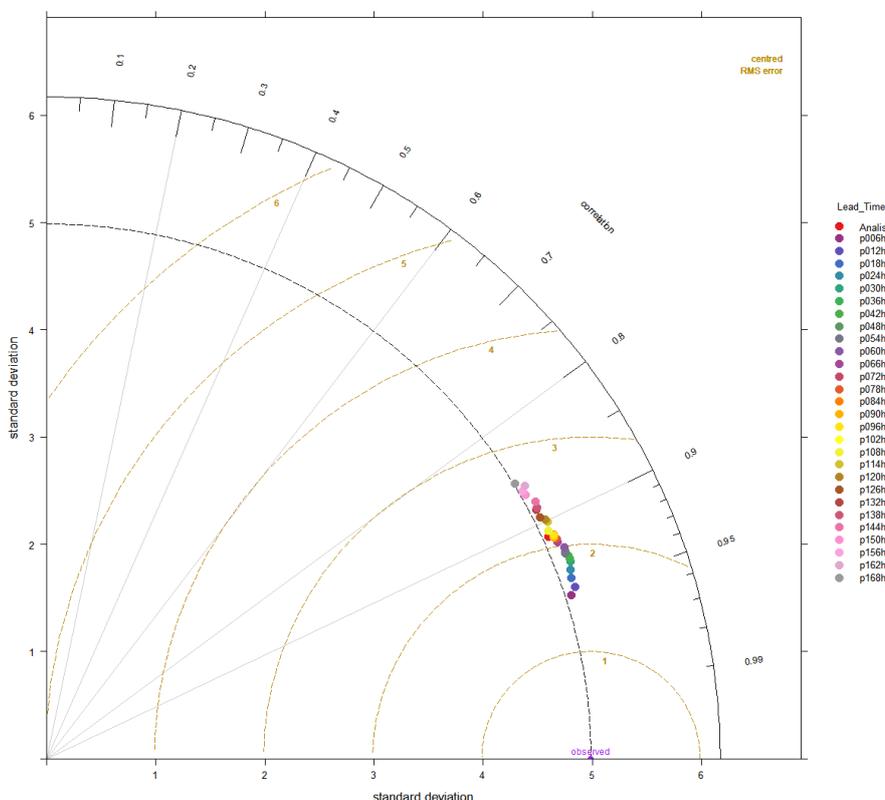


Figura 3. Diagrama de Taylor, concentrando informações sobre correlação, desvio padrão e erro médio quadrático centralizado, para os diferentes “lead times”.

Além disso, foi construído um Diagrama de Taylor (Figura 3), que mostra a correlação (ângulo), o desvio padrão (raio), e o Erro Quadrático Médio Centrado (distância da “observação”), para cada LT. Pode-se perceber que os desvios padrões das previsões e da análise foram todos muito semelhantes ao da observação. A correlação foi, de maneira geral, satisfatória, mantendo-se numa faixa entre 0.85 e 0.96. O Erro Quadrático Médio Centrado permaneceu entre 1,5 e 3,0°C. Vê-se, novamente, que

ambas essas estatísticas (Correlação e Erro Quadrático Médio Centralizado) tiveram a tendência de menor previsibilidade quanto maior a antecedência da previsão - com exceção da Análise, como já mencionado, que apresentou um desempenho intermediário.

Foi feito um gráfico de Quantis Condicionais (Figura 4), que dispõe, para cada valor previsto, os correspondentes observados. Nota-se que o modelo contemplou todas as faixas de temperatura observadas, e que os melhores desempenhos gerais (faixas correspondentes aos quantis) se concentram nas previsões até 96h e na análise. Em previsões com antecedência maior do que 96h, a

faixa correspondente aos quantis fica mais larga, sobretudo nas temperaturas mais altas, indicando

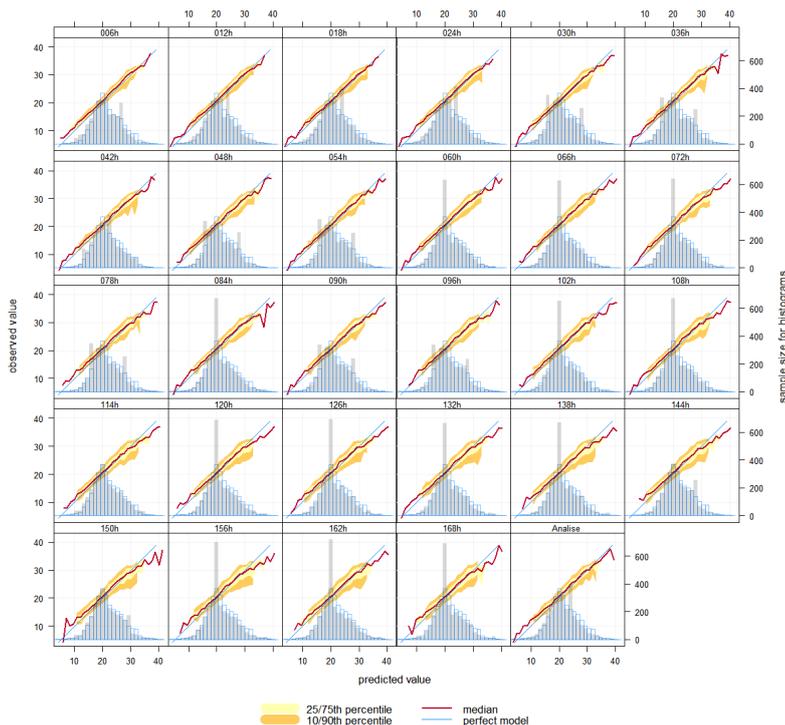


Figura 4. Quantis Condicionais. A linha vermelha representa a mediana dos valores previstos e observações correspondentes, e os hachurados correspondem às faixas que compreendem os quantis de 25/75 (hachurado claro) e 10/90 (hachurado escuro). Um modelo perfeito coincide com a linha diagonal azul e quando as faixas hachuradas são estreitas. No plano de fundo, estão os histogramas de frequências das temperaturas observadas (linhas azuis) e das temperaturas simulada (hachurado cinza).

menor precisão da previsão. Apesar do bom desempenho geral, o modelo tende a superestimar os valores nas temperaturas mais baixas, e subestimá-los nas temperaturas mais altas.

A seguir, foi feita uma divisão por faixas de temperatura, que foram analisadas separadamente. Na Tabela 2 pode-se ver os índices estatísticos encontrados em cada faixa, em comparação àquela para o modelo como um todo. De forma geral, as duas faixas mais “extremas” tiveram desempenho inferior. Para temperaturas menores ou iguais a 10°C, pode-se ver que o Fator de Dois foi inferior a 1 (possivelmente porque para temperaturas mais baixas o critério de “metade ou dobro” do índice necessita de um erro absoluto menor para ser violado), e o Viés Médio (MB) foi

positivo (ou seja, o modelo tende a superestimar as temperaturas - diferente do que ocorre nas outras

faixas). Vê-se também que os Erros Médio Absoluto (MGE) e Médio Quadrático (RMSE) foram maiores nas 2 faixas extremas. O Viés Médio (MB) foi ligeiramente negativo, o que significa que, de forma geral, o modelo tende a subestimar as temperaturas.

Então, esses índices foram

calculados separadamente para cada LT e cada faixa de temperatura observada, representados graficamente na Figura 5.

CONCLUSÕES:

Tabela 2. Índices estatísticos para avaliar o desempenho das previsões com base nas faixas de temperatura observadas. "n" representa o número de amostras sobre as quais os índices foram calculados, e os "Critérios" são os valores ideais

Faixas de Temperatura (T)	n	FAC2	MB	MGE	NMB	NMGE	RMSE
Geral	83867	1.000	-0.475	1.63	-0.02168	0.0745	2.19
T ≤ 10 °C	493	0.978	1.159	2.05	0.14373	0.2542	2.88
10 °C < T ≤ 20 °C	31552	1.000	-0.171	1.32	-0.00993	0.0767	1.80
20 °C < T ≤ 30 °C	45993	1.000	-0.604	1.77	-0.02521	0.0741	2.35
30 °C < T	5829	1.000	-1.236	2.14	-0.03856	0.0669	2.71
Critérios	NA	1.000	0.000	0.00	0.00000	0.0000	0.00

O modelo GFS demonstrou ser, como esperado, um sistema de modelagem numérica do tempo de alta confiabilidade no quesito de previsão de temperaturas para o município de Campinas. Nas faixas de temperatura que representam a maior proporção das ocorrências observadas no município (de 10 a 30°C), o modelo apresentou um excelente desempenho, mesmo para previsões com maior antecipação. Tratando-se de eventos denominados “extremos” (neste trabalho, consideradas temperaturas iguais ou inferiores a 10°C e temperaturas acima de 30°C), as limitações são maiores e há necessidade de uso mais cauteloso das informações geradas pelo modelo. Uma vez que alertas meteorológicos requerem uma maior precisão, considerando que seu objetivo é preparar a população e tomadores de decisão para eventos, é importante que haja um alto grau de confiabilidade nos dados. Por análise comparativa das estatísticas para os diferentes LTs em diferentes faixas de temperatura, verifica-se que, para antecipação de extremos frios/quentes, previsões com até 84h/60h de antecedência possuem um desempenho satisfatório para a finalidade de avisos meteorológicos e preparo de eventuais ações de enfrentamento. Informações com maior antecedência podem ser empregadas na previsão operacional do tempo, mas não com o rigor técnico e autoritativo de um alerta.

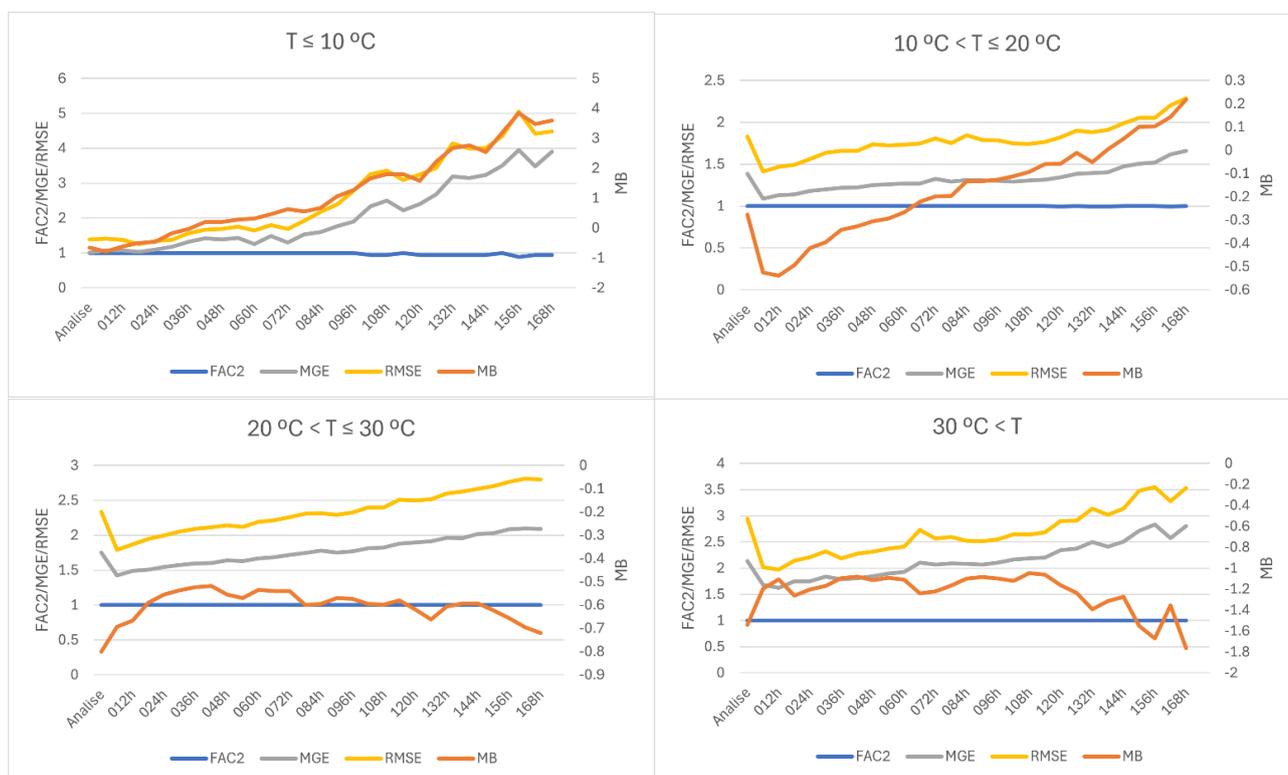


Figura 5. Índices estatísticos calculados para faixas de temperatura observada, em função da antecedência da previsão.

BIBLIOGRAFIA

Carslaw DC, Ropkins K. **openair — An R package for air quality data analysis**. Environmental Modelling & Software, 27–28(0), 52–61, doi: 10.1016/j.envsoft.2011.09.008, 2012.

Kendzierski, Sebastian, et al. **Air Temperature Forecasts' Accuracy of Selected Short-Term and Long-Term Numerical Weather Prediction Models over Poland**. Geofizika, vol. 35, 2018, pp. 19–37, <https://doi.org/10.15233/gfz.2018.35.5>, 2018.