



WORK VIII 2025 Eta

WORKSHOP EM MODELAGEM NUMÉRICA DE TEMPO, CLIMA
E MUDANÇAS CLIMÁTICAS UTILIZANDO O MODELO ETA

Métricas de Avaliação de Previsões Numéricas

Daniela Carneiro Rodrigues
membros do Grupo Eta
Divisão de Impactos Adaptação e
Vulnerabilidade – DIIAV/INPE

Introdução

- ***Avaliar*** é testar o grau de correspondência entre as previsões numéricas e as observações/análises.
- ***A ideia de avaliar as previsões tempo surgiu em 1884 (Finley, 1884) e a partir da década 50 foi incorporada a previsão numérica.***
- ***A avaliação tornou-se uma das principais etapas durante o processo de desenvolvimento e aperfeiçoamento de modelos.***

Porque avaliar as previsões numéricas?

- Para monitorar a qualidade das previsões ao longo do tempo;
- Para melhorar a qualidade da previsão por meio de uma melhor compreensão dos erros;
- Comparar a qualidade de diferentes sistemas de previsão;
- Verificar o impacto de inserção de novas implementações nos modelos numéricos (física, dinâmica, resolução, etc).

Atributos que caracterizam a qualidade das previsões

- **Bias:** correspondência entre a média da previsão e a média da observação.
- **Acurácia:** nível de concordância entre a previsão e a observação. Erros baixos indicam grande acurácia.
- **Skill:** é acurácia relativa de uma previsão sobre alguma previsão de referência. Previsão de tempo podem ter maior acurácia simplesmente porque o tempo é mais fácil de prever. Skill leva isso em conta.
- **Confiabilidade:** concordância média entre os valores previstos e valores observados. Existe uma diferença no valor para cada categoria. Indica quão bem calibradas são as probabilidades emitidas.
- **Resolução:** é a capacidade da previsão de produzir frequências relativas que se destacam do nível da climatologia geral da amostra, ou seja, como a frequência de ocorrência do evento difere à medida que a probabilidade emitida muda.
- **Nitidez:** tendência da previsão de prever valores extremos. Um previsão climatológica não tem nitidez.
- **Discriminação:** capacidade de distinguir com sucesso eventos de não eventos.
- **Incerteza:** a variabilidade das observações. Quanto maior a incerteza, mais difícil a previsão tende a ser.

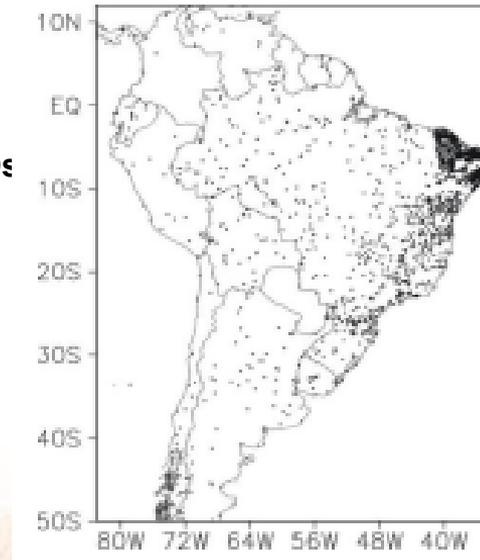
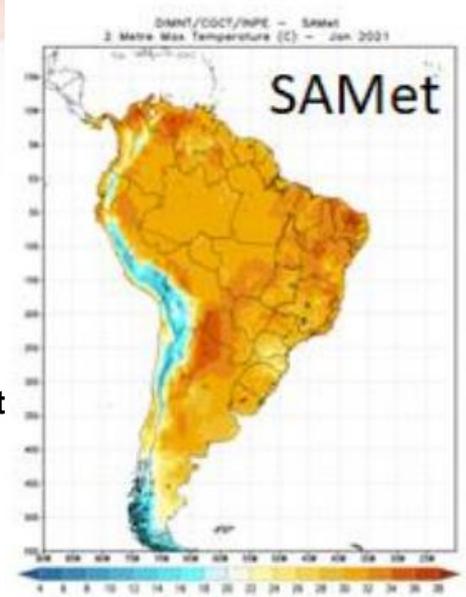
Formas de avaliar as previsões numéricas:

1. Dados espacializados: a verificação deve ser feita contra observações em uma grade comum de latitude/longitude

- Análises
- Reanálises (CFSR, ERA5, ERA-Interim)
- Dados observados espacializados em grade (GPCP, CRU, MERGE, SAMet)
- Estimativas a partir de dados de satélites (CMORPH, TRMM)

2. Avaliar contra dados pontuais:

- Dados de estações meteorológicas (PCDs:
- Synop, Metar
- Radiossondas
- Bóias, Ship (oceanos)



Tipos de Avaliações:

- 1. Avaliação Objetiva** - padrões de comportamento entre as previsões e as observações podem ser medidos através de **números/índices**.
Variáveis contínuas (P, T, RH,..)
Variáveis discretas (PREC, NEVE,..)
- 2. Avaliação Subjetiva** - padrões de comportamento entre as previsões e as observações são verificados através de fatos observáveis por meio da **interferência humana**.
- 3. Avaliação Semi-Objetiva** - padrões de comportamento entre as previsões e as observações que podem ser verificados através de **algoritmos numéricos juntamente com a interferência humana**.

Métricas de avaliação objetiva de previsões determinísticas

- **Variância amostral (s^2):**

Descreve a variabilidade da variável.

$$s_O^2 = \frac{1}{N-1} \sum_{i=1}^N (O_i - \bar{O})^2$$

$$s_F^2 = \frac{1}{N-1} \sum_{i=1}^N (F_i - \bar{F})^2$$

N = número de dados
 F = previsão
 O = observação/análise

- **Erro médio (ME):**

É a diferença entre a média da previsão e a média das observações/análises.
Indica tendências de superestimativas ou subestimativas.

$$ME = \frac{1}{N} \sum_{i=1}^N (F_i - O_i) = \bar{F} - \bar{O}$$

- **Erro médio absoluto (MAE):**

Mede a magnitude do erro médio. Utilizamos o módulo de cada erro para evitar a subestimação, isso porque, o valor é menos afetado por pontos especialmente extremos.

$$MAE = \frac{1}{N} \sum_{i=1}^N |F_i - O_i|$$

- **Erro quadrático médio (MSE):**

Mede a magnitude do erro quadrático médio e é frequentemente usado na construção de pontuações de habilidades.

$$MSE = \frac{1}{N} \sum_{i=1}^N (F_i - O_i)^2$$

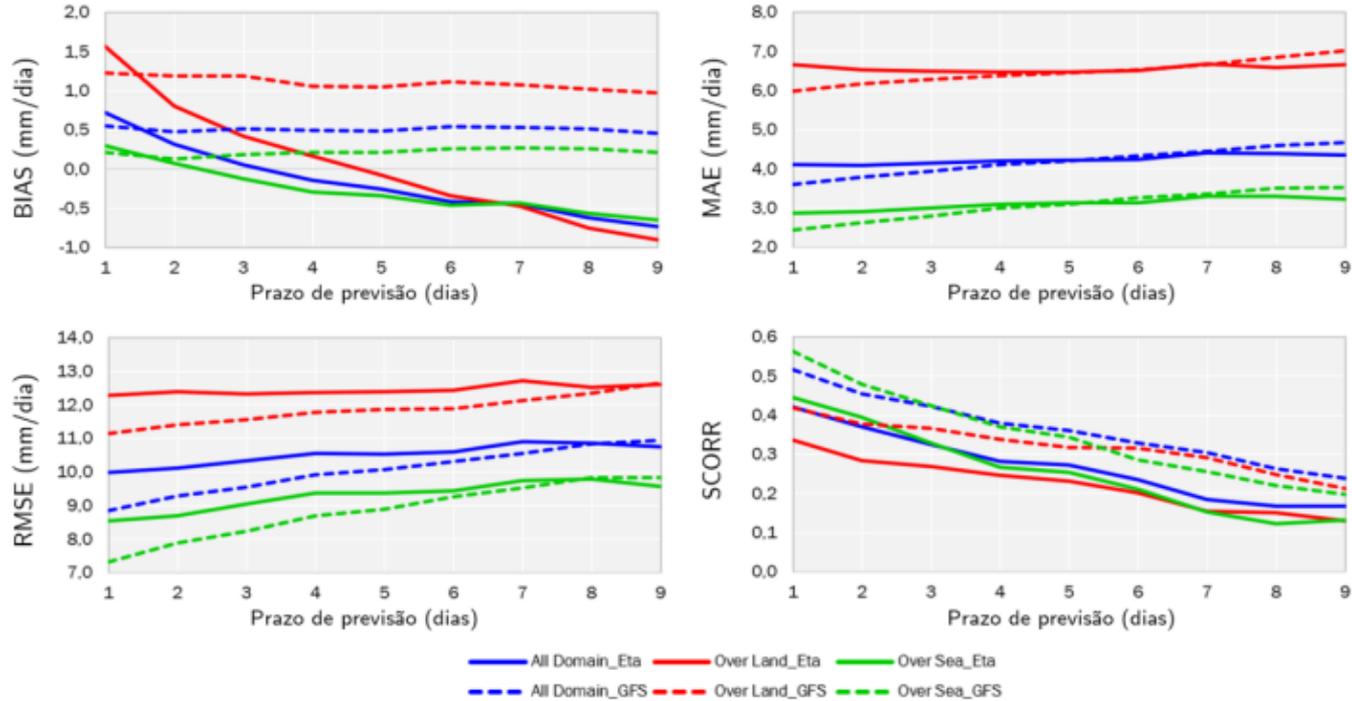
- **Raiz do erro quadrático médio (RMSE):**

É a raiz quadrada do valor médio ao quadrado das diferenças entre a previsão e a observação. É uma medida de erro absoluto que eleva os desvios ao quadrado para impedir que os desvios positivos e negativos se cancelem. Indica a magnitude dos erros.

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

Exemplo de avaliações objetivas – Previsões determinísticas

Eta x GFS



Erro médio (BIAS), erro médio absoluto (MAE), raiz quadrada do erro quadrático médio (RMSE) E correlação espacial (SCORR), calculados entre a precipitação do modelo Eta 8km (linhas contínuas) e o GFS (linhas tracejadas) e a base observada do MERGE. Previsões para o 1º ao 9º dia. Valores médios para todo o domínio (Linhas azuis), continente (linhas vermelhas) e oceano (linhas verdes).

Métricas para avaliar previsão determinística

- **Skill score:**

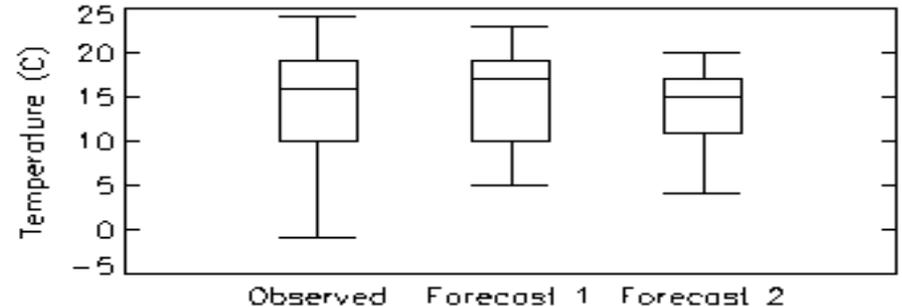
O skill score mede a melhoria fracionária do sistema de previsão em relação a uma previsão de referência.

Os escores mais utilizados são o MAE e o MSE. A estimativa de referência é persistência para previsões de 24h ou menos, e climatologia para previsões mais longas.

$$MAE_{SS} = \frac{MAE_{forecast} - MAE_{reference}}{MAE_{perfect} - MAE_{reference}} = 1 - \frac{MAE_{forecast}}{MAE_{reference}}$$
$$MSE_{SS} = \frac{MSE_{forecast} - MSE_{reference}}{MSE_{perfect} - MSE_{reference}} = 1 - \frac{MSE_{forecast}}{MSE_{reference}}$$

- **Box Plot:**

Caixas de plotagem mostram o intervalo de dados entre os percentis 25 e 75. A linha horizontal dentro da caixa mostra o valor mediano e os bigodes mostram o intervalo completo dos dados.



Índices quantitativos: avalia variáveis discretas (prec, neve)

		Observado		
		Sim	Não	
Previsão	Sim	Acertos a	Alarme falso b	Previsto sim
	Não	Em falta c	Rejeições corretas d	Não previsto
		<i>observed yes</i>	<i>observed no</i>	$N = total$

a = previsto e observado
b = previsto e não observado
c = não previsto e observado
d = não previsto e não observado

- **Bias categórico (BIAS):**

fornece a razão entre a frequência da variável prevista e a frequência da variável observada. (Ex.: chuva)

$$Bias = \frac{Acertos + Alarmes falsos}{Acertos + Em falta}$$

- **Probabilidade de detecção (POD (HR)):**

mede a fração de eventos observados que foram corretamente previstos.

$$POD = \frac{Acertos}{Acertos + Em falta}$$

- **Taxa de alarme falso (FAR):**

fornece a fração de eventos previstos que foram observados como não eventos.

$$FAR = \frac{Alarmes falsos}{Acertos + Em falta}$$

- **Equitable threat score (ETS):**

mede a fração de todos os eventos previstos e/ou observados que foram diagnosticados corretamente, contabilizando os acertos que ocorreriam puramente devido ao acaso.

$$ETS = \frac{Acertos - Acertos_{aleatórios}}{Acertos + Em falta - Acertos_{aleatórios}}$$

$$Acertos_{aleatórios} = \frac{1}{N} (Observado\ sim \times\ previsto\ sim)$$

Exemplo de Índices quantitativos – Previsões determinísticas

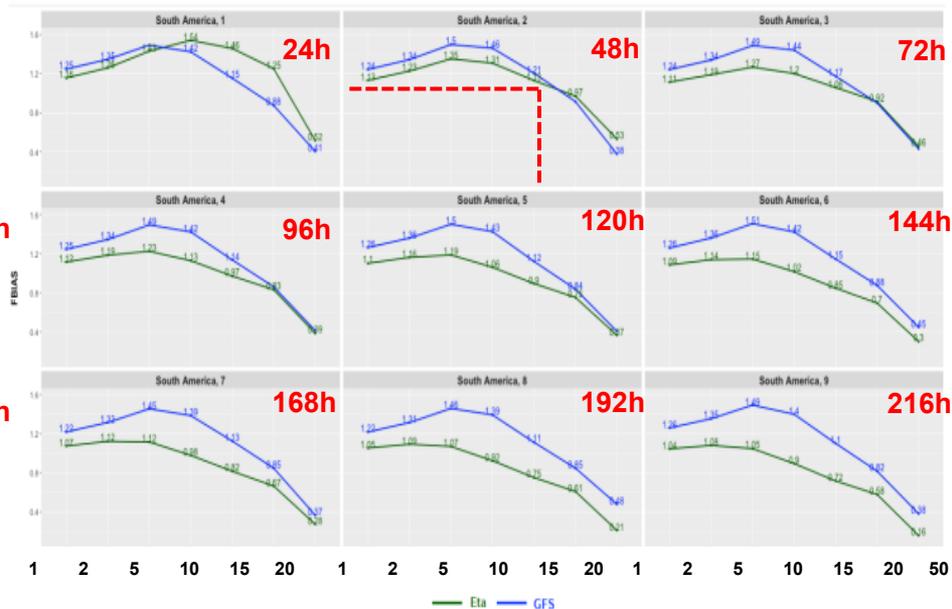
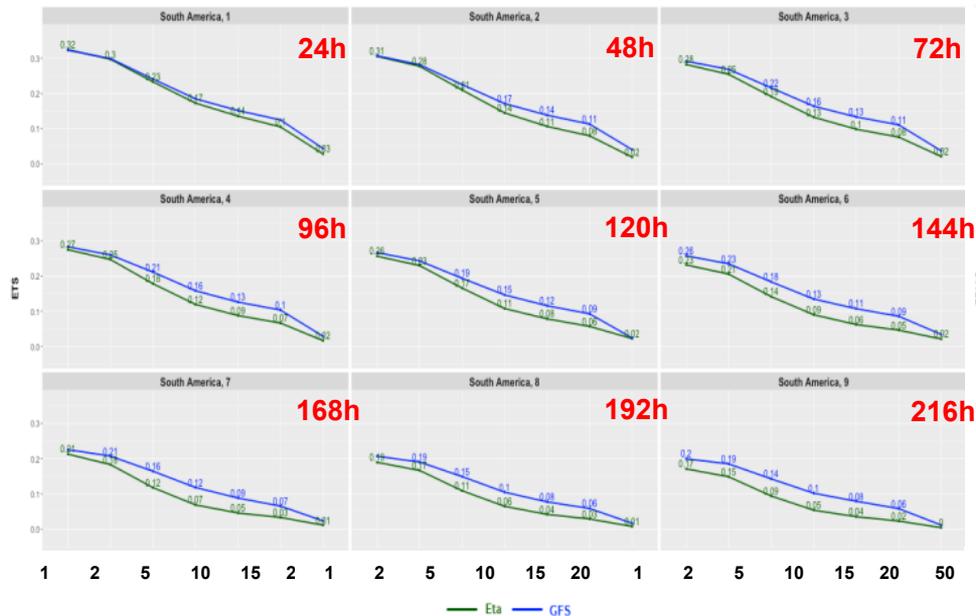
- Previsões de precipitação de 9 dias de antecedência dos **modelos Eta 8km (verde)** e **GFS (azul)**, referentes ao período de janeiro de 2021. O eixo vertical representa os índices e o eixo horizontal o limiar da chuva (fraca, modera e intensas).

Eta x GFS

ETS

BIAS CATEGÓRICO

*score perfeito teria valor = 1.

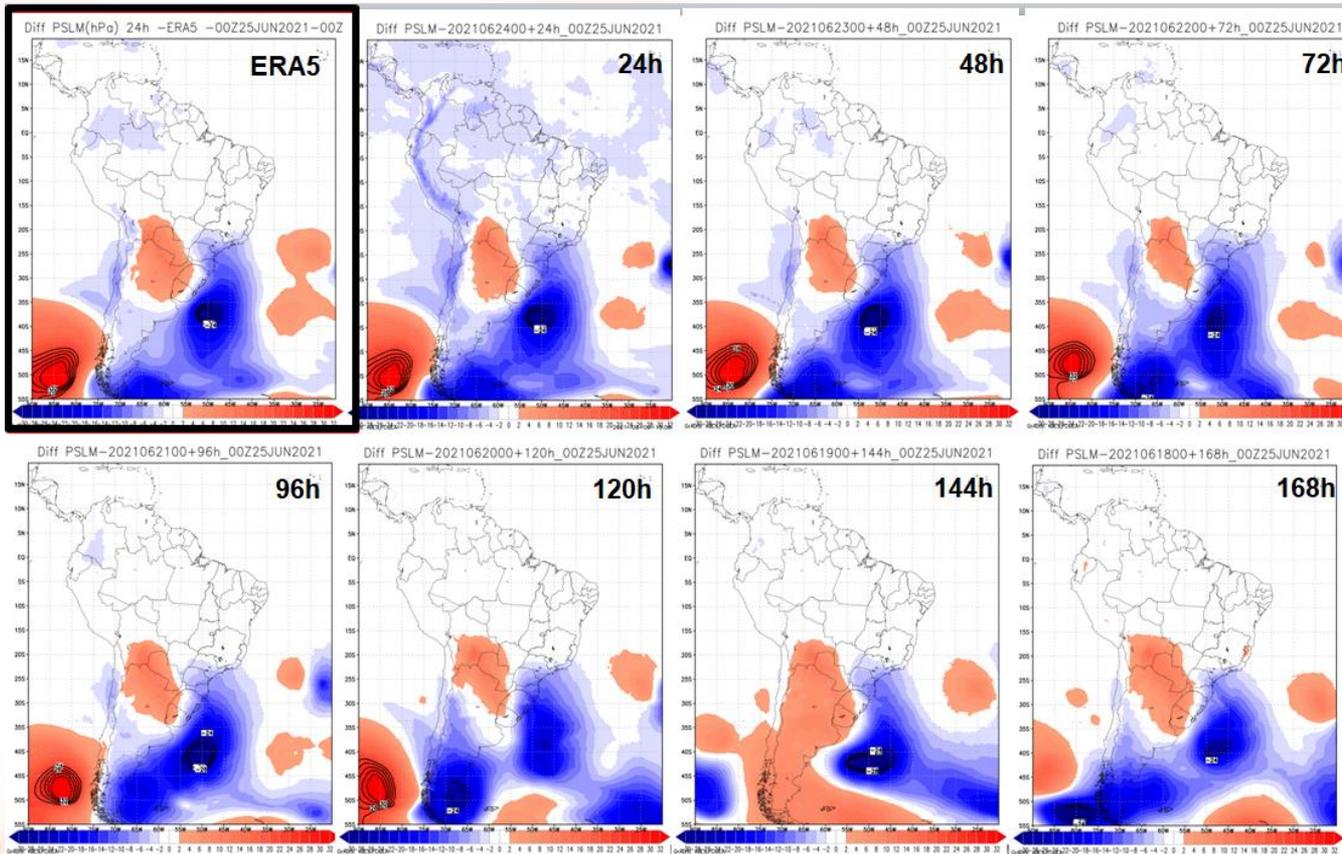


Índice ETS (Equitable Threat Score) para as previsões de precipitação dos Modelos Eta (verde) e GFS (azul) referentes ao período de estudo. Acumulados de 24 em 24h para os 9 dias de previsão.

BIAS categórico para as previsões de precipitação dos Modelos Eta (verde) e GFS (azul) referentes ao período de estudo. Acumulados de 24 em 24h para os 9 dias de previsão.

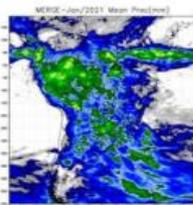
Exemplo de avaliações subjetivas – Previsões determinísticas

Evento: Queda na PSLM > 24 (hPa) em 24h para detecção de “ciclone bomba”:



- O objetivo aqui é por meio da interferência humana, avaliar em comparação a reanálise ERA5, se o modelo Eta 8km foi capaz de prever a queda na pslm em 24h > 24 hPa, configurando um ciclone bomba, conforme verificado na reanálise.
- Além disso, identificar a partir de quantas horas de antecedência o modelo era capaz de prever o sistema e se posicionou corretamente sua localização.

Exemplo de avaliações subjetivas – Previsões determinísticas

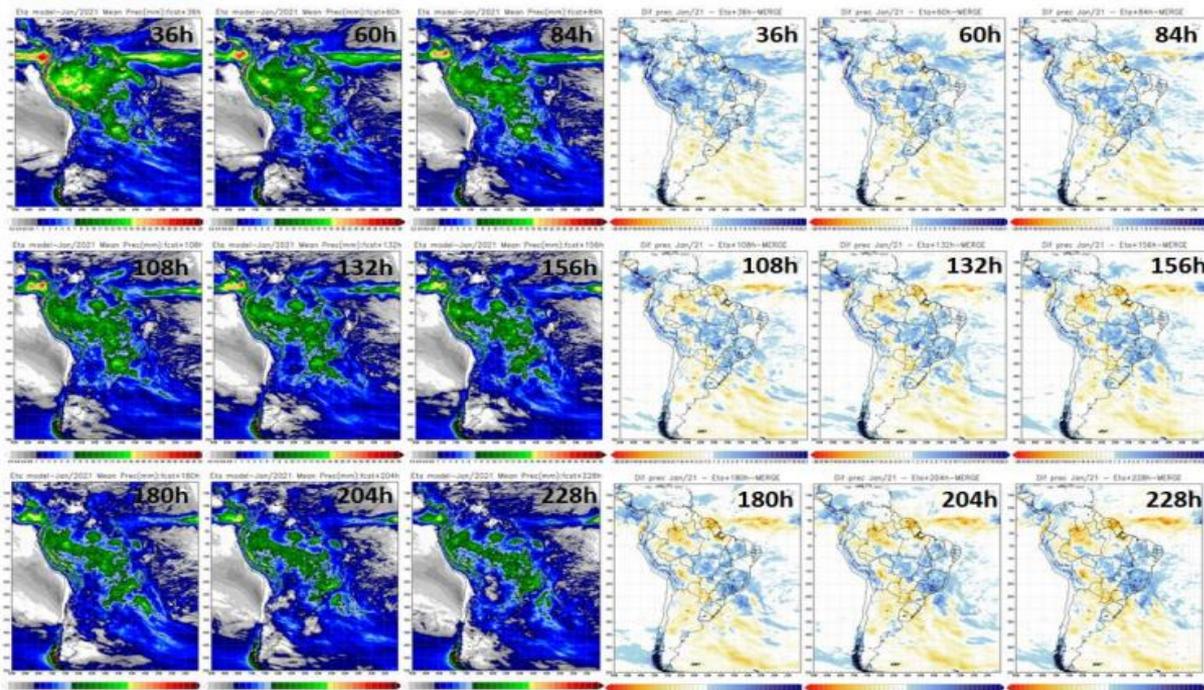


MERGE

Precipitação média
de Jan/2021

Previsões de chuva Eta 8km

Diferenças: Eta 8km - MERGE



Precipitação do Modelo Eta 8km para os prazos de previsão de 36 a 228 horas, médias para o mês de janeiro e diferenças médias entre Eta e MERGE em cada prazo de previsão.

The background features a warm, orange-toned sky with soft, white clouds. In the top-left corner, a portion of a globe with a grid pattern is visible. The bottom of the image shows a textured, wavy surface in shades of orange and white, resembling sand dunes or a stylized landscape.

Métricas de avaliações de previsões probabilísticas - Ensemble

Aspectos da qualidade de previsões probabilísticas

- **Diagrama de confiabilidade:** sintetiza a confiabilidade, Nitidez e a resolução, indicando o grau de precisão do conjunto. **Para um conjunto de previsão ser preciso, ele precisa ter:**

Confiabilidade - concordância entre a probabilidade de previsão e a frequência média observada.

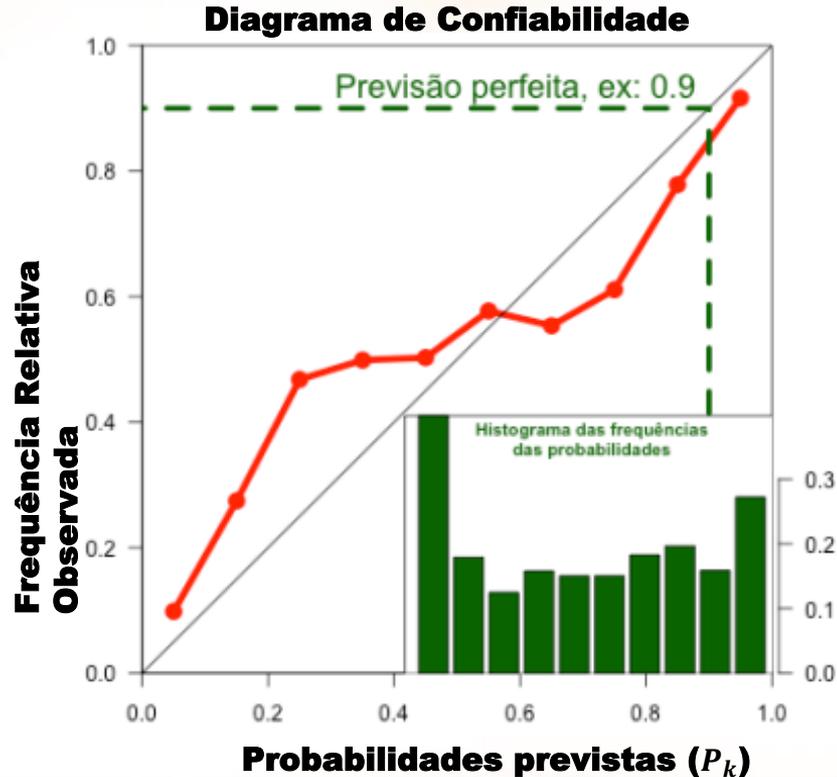
Nitidez - tendência a prever probabilidades próximas de 0 ou 1, em oposição a valores agrupados em torno da média.

Resolução - capacidade da previsão de resolver o conjunto de eventos de amostra em subconjuntos com resultados caracteristicamente diferentes.

- **Brier score (BS):** Score geral da **precisão** das previsões probabilísticas. Resume previsão probabilística vs. ocorrência (ou não) do evento.
- **Histograma ranqueado:** mede o **espalhamento** do conjunto de previsões. Verifica onde a observação verificadora geralmente cai em relação aos dados do conjunto de previsão.

Aspectos de qualidade das previsões probabilísticas

- Diagrama de Confiabilidade:



- A linha diagonal dá o viés condicional.
- Curva abaixo da linha, indica probabilidades muito altas; pontos acima da linha indicam probabilidades muito baixas. Quanto menor a distância da curva em relação à linha, mais bem calibrada (**confiável**) é a previsão.
- Quanto mais plana for a curva no diagrama de confiabilidade, menor será a **resolução**.
- A frequência das previsões em cada bin de probabilidade (histograma) mostra a **nitidez** da previsão.
- **O diagrama de confiabilidade está condicionado às previsões e fornece informações sobre o real significado da previsão.**

→ O intervalo de probabilidades de previsão é dividido em K bins (0-10%, 10-20%, 20-30%, etc.).

Aspectos de qualidade de previsões probabilísticas

- Brier score (BS):**

Score geral da **precisão** das previsões probabilísticas. Resume previsão probabilística vs. ocorrência (ou não) do evento.

$$\text{Brier score} - BS = \frac{1}{N} \sum_{i=1}^N (p_i - o_i)^2 = \frac{1}{N} \sum_{k=1}^K n_k (p_k - \bar{o}_k)^2 - \frac{1}{N} \sum_{k=1}^K n_k (\bar{o}_k - \bar{o})^2 + \bar{o}(1 - \bar{o})$$

valor observado binário (pointing to o_i)
Probabilidades revistas (pointing to p_i)
confiabilidade (pointing to n_k)
resolução (pointing to \bar{o}_k)
incerteza (pointing to \bar{o})

medidas de **precisão de uma previsão.*

*N= número de eventos/previsões
 i= indexa o eventos de 1 a N.
 Pi = previsão para o evento i
 (probabilidade de 0 a 1) .
 Oi=observação do evento t.*

PROB	OBS_VAL	OBS_BIN
0.8	-0.66	1
0.4	3.19	0
0.6	-3.69	1
0.4	5.89	0
0.2	1.37	0
0.4	7.80	0

**O valor do BS varia de 0 a 1. Como é uma medida de erro, quanto menor, melhor.*

O valor observado binário é 1 quando a observação corresponde à probabilidade e 0 quando a observação não corresponde à probabilidades.

Preparo dos dados de previsões por conjunto

Evento: Vai haver temperatura abaixo de 0° em Campos do Jordão?

Dataset de tempo e espaço de acordo com a avaliação

Valores dos membros do ensemble					Valor observado		
M1	M2	M3	M4	M5	PROB	OBS_VAL	OBS_BIN
-2.62	-8.48	-0.97	2.69	-5.54	0.8	-0.66	1
16.42	6.83	14.55	-3.05	-2.88	0.4	3.19	0
-11.14	5.55	-0.32	1.55	-2.50	0.6	-3.69	1
6.04	15.80	12.64	-2.28	-2.58	0.4	5.89	0
3.33	-10.58	15.48	3.78	6.10	0.2	1.37	0
-21.13	7.89	1.32	17.47	-0.63	0.4	7.80	0

Qual a probabilidade do evento ser previsto?
(fração dos membros do ensemble que prevêm o evento)

Valor observado binarizado
0 = não ocorreu
1 = ocorreu

Aspecto de espalhamento de previsões probabilísticas

- **Histograma ranqueado (Rank Histogram – RH):**

Indica quão bem a dispersão ou espalhamento do conjunto de previsão representa a **verdadeira variabilidade** das observações.

- ❑ Contabiliza (acumula, soma) a observação em um certo intervalo formado pelos membros do ensemble.
- ❑ Diagnostica o espalhamento médio do ensemble em relação a observação correspondente (toda previsão tem um observação).
- ❑ Assume que as observações deveriam ocorrer igualmente em todos os bins.

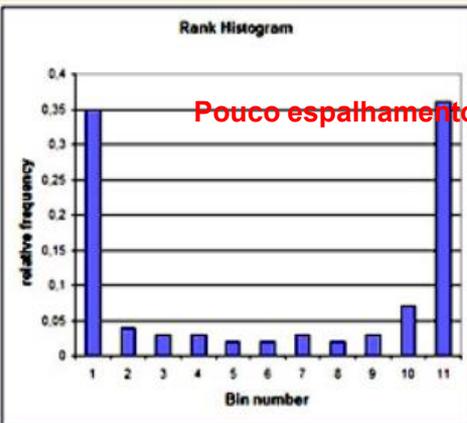
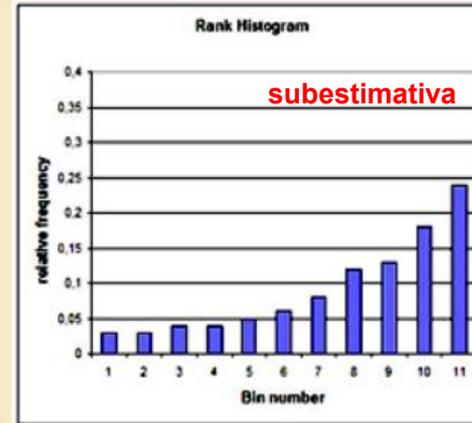
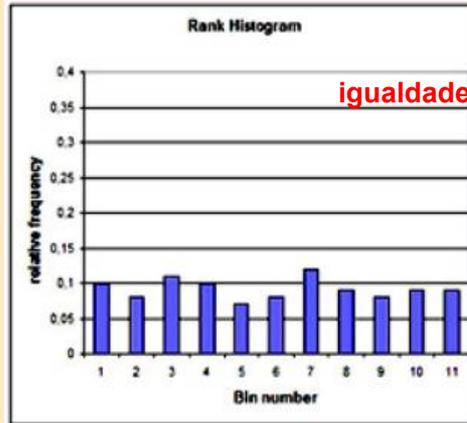
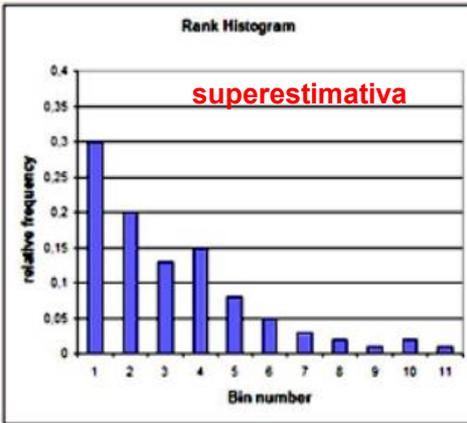
Rank histogram (Talagrand et al. 1997; Hamill, 2001)



- **O Histograma verifica onde a observação verificadora geralmente cai em relação aos dados do conjunto de previsão, que são organizados em ordem crescente em cada ponto da grade.**
- **Em um conjunto com espalhamento perfeito, cada membro representa um cenário igualmente provável, de modo que a observação tem a mesma probabilidade de cair entre quaisquer dos membros.**

Interpretação do Histograma Ranqueado

Fonte: Eumetcal, 2013



Plano - espalhamento do conjunto consegue representar a incerteza da previsão.

Em forma de U - conjunto muito pequeno, muitas observações fora dos extremos do conjunto.

Em forma de cúpula - o conjunto tem espalhamento muito grande, onde a maioria das observações cai perto do centro do conjunto.

Assimétrico - o conjunto contém viés (superestimativa ou subestimativa)

Nota: Um histograma de classificação plana não indica necessariamente uma boa previsão, apenas mede se a distribuição de probabilidade observada está bem representada pelo conjunto.

Diagrama de Taylor:

- **Maneira de resumir graficamente o quão próximo um padrão (ou um conjunto de padrões) corresponde às observações.**

A semelhança entre dois padrões é quantificada em termos de:

- **Desvio padrão**
- **Coefficiente de correlação**
- **RMSE (raiz quadrática do erro médio)**

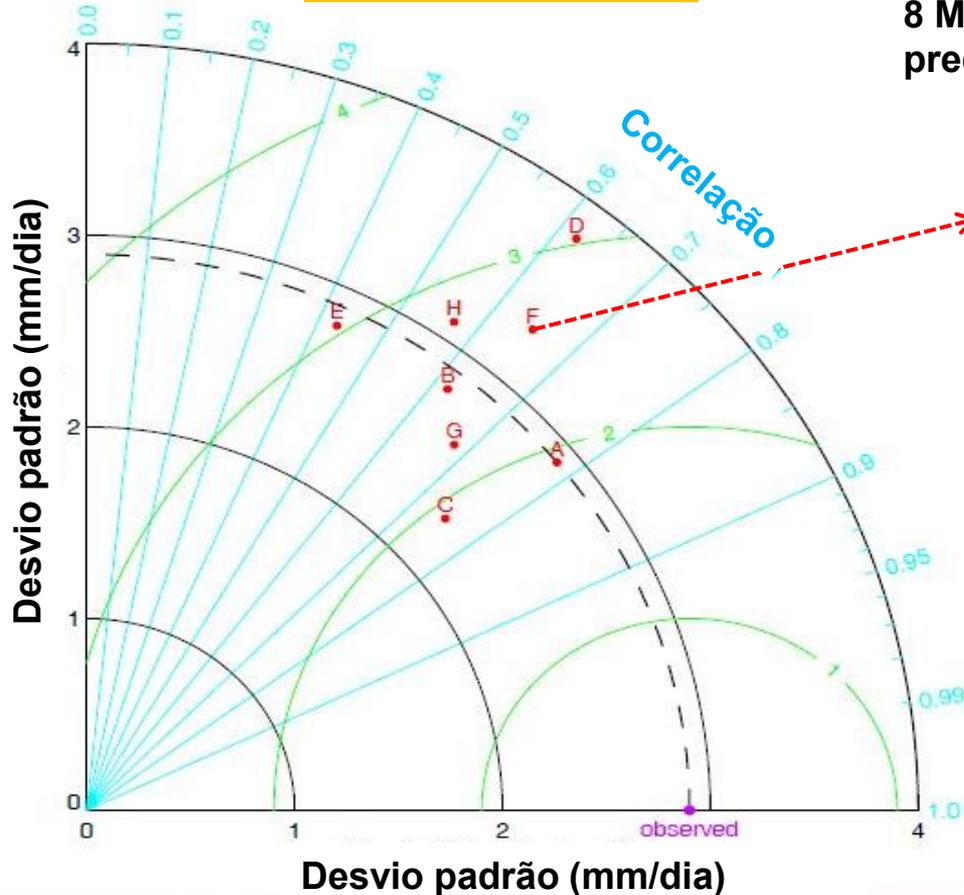
**Esses diagramas são especialmente úteis na avaliação de vários aspectos de modelos complexos ou na avaliação da habilidade relativa de muitos modelos diferentes (por exemplo, Modelos do CMIP/IPCC).*

- **As médias dos campos são subtraídas antes de computar essas estatísticas, então o diagrama não fornece informações sobre bias gerais, mas caracteriza unicamente o erro de padrão centrado.**

Taylor Diagram

A, B, C, D, E, F, G, H :

8 Modelos globais padrão espacial de precipitação média anual



Ex.: Modelo F:

Coeficiente de correlação : 0.65

RMSE : 2.6 mm/d

Desvio padrão : 3.3 mm/d

Diagrama de Taylor exibindo uma comparação estatística de padrão espacial de estimativas de oito modelos globais com observação de precipitação média anual.

Referências

WMO, 2002: Standardised Verification System (SVS) for Long-Range Forecasts (LRF). New attachment II-9 to the *Manual on the GPDS (WMO-No.485), Volume 1. Available on the internet at <http://www.wmo.ch/web/www/DPS/LRF-standardised-verif-sys-2002.doc>*

Wilks, 2005 - Statistical methods in the atmospheric science, 2ed.

Taylor, K.E., 2001: Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.*, 106, 7183-7192.

**JWGV, 2004: Forecast verification – Issues, methods, and FAQ.
http://www.bom.gov.au/bmrc/wefor/staff/eee/verif/verif_web_page.html**

Bougeault, P., 2002: WGNE survey of verification methods for numerical prediction of weather elements and severe weather events. *CAS/JSC WGNE Report No. 18, Appendix C. Available on the internet at <http://www.wmo.ch/web/wcrp/documents/wgne18rpt.pdf>.*

Atger, F., 2001: Verification of intense precipitation forecasts from single models and ensemble prediction systems. *Nonlin. Proc. Geophys.*, 8, 401-417.

Métricas de Avaliação de Previsões Numéricas (parte I e II). Apresentação Roberto Rozante e José Roberto no VI Worketa.



Obrigada!

daniela.rodrigues@inpe.br