Métricas de Avaliação dos Modelos de Circulação Geral

Disciplina: PNTC

https://subsazonal.cptec.inpe.br/

http://funceme.br/dashboard/subsaz_forecast

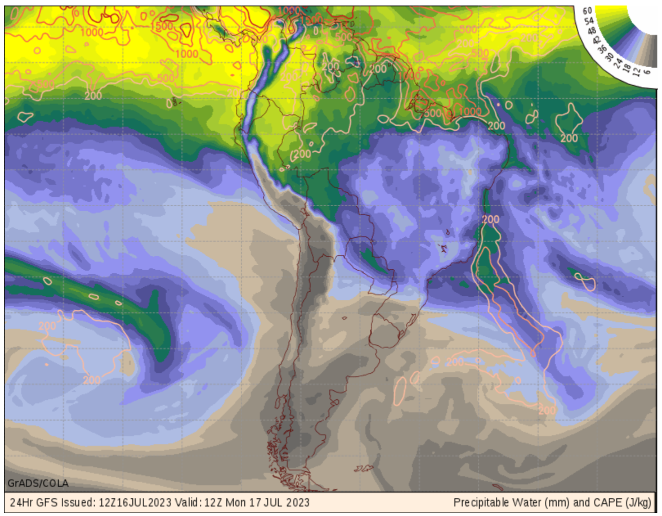
https://www.wmolc.org/seasonPmmeUI/plot_PMME

Qual é a Finalidade do Modelo?

- Previsão
  - Previsão em tempo real;
  - Previsão retrospectiva.
- Simulação
  - AMIP (Model Intercomparison Project)
  - Pesquisa

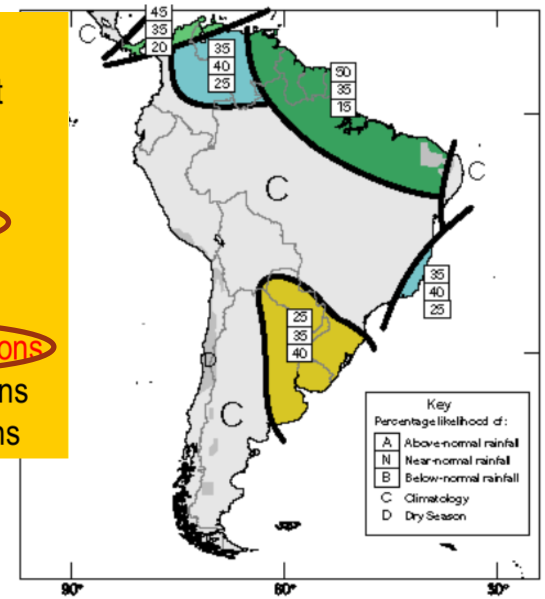
Previsão em Tempo Real

Previsão de Tempo



Present	Future
Weather →	Weather forecast
• hours	Nowcasting
• days	Medium-range
• weeks	Sub-seasonal
Climate →	Climate forecast
• months	seasonal predictions
• decades	decadal predictions
• centuries	secular projections

Previsão climática sazonal

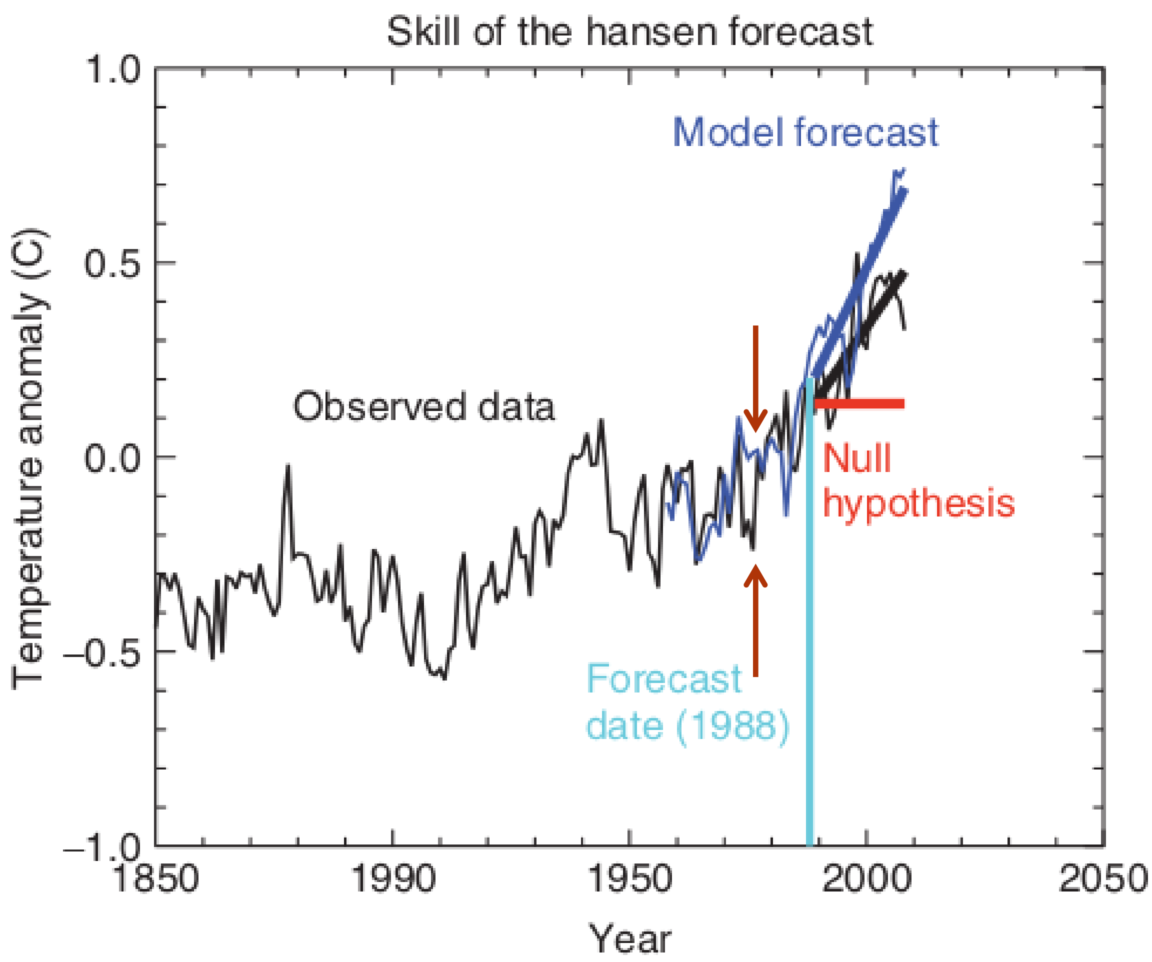


Previsão Retrospectiva

- Mesmo que a previsão em tempo real. Porém, inicializadas no presente com datas do passado

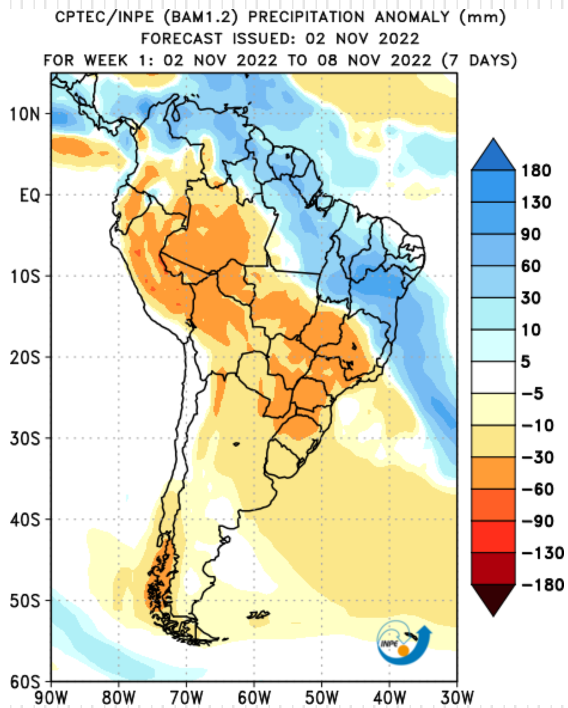
Projeção

Hargreaves (2010)

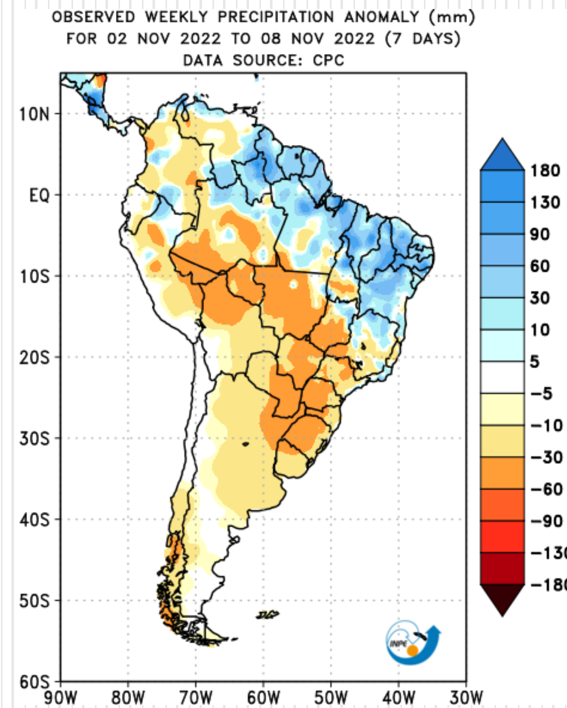


Avaliação das Previsões Determinísticas

Previsão

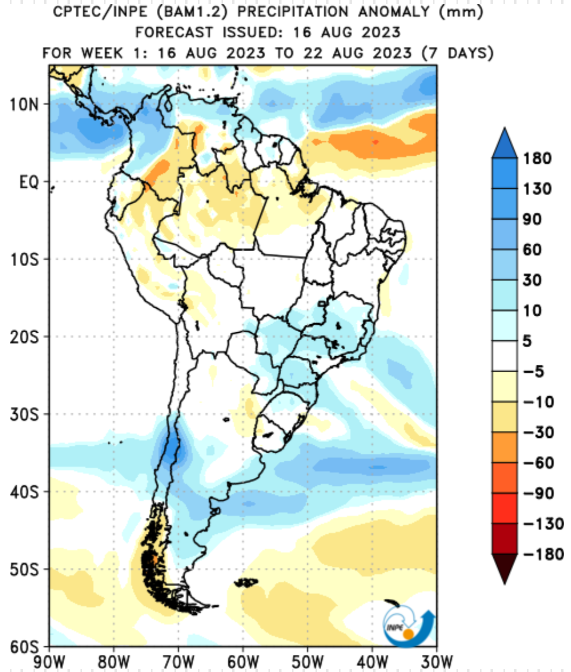


Dado de Referência (observação)



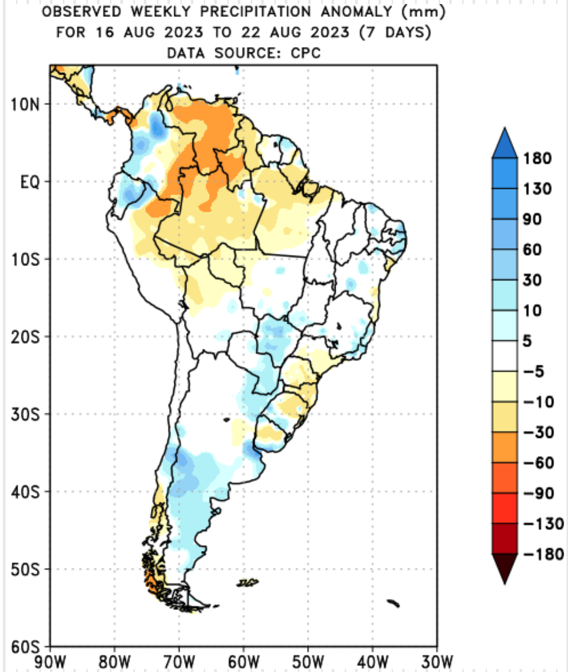
Avaliação das Previsões Determinísticas

Previsão



VS

Dado de Referência (observação)

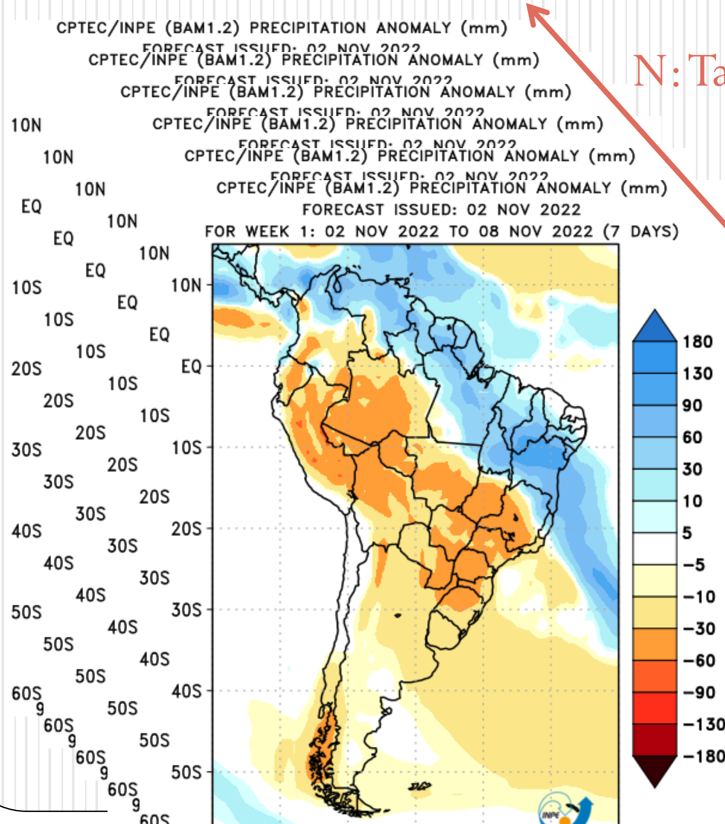


Avaliação das Previsões Determinísticas

Previsão

N: Tamanho da amostra

Dado de Referência (observação)



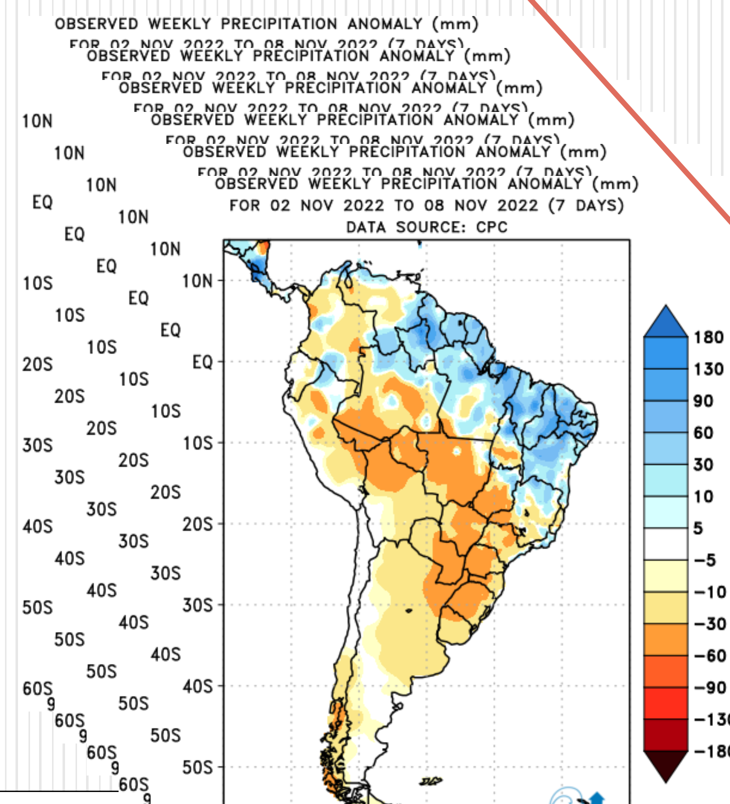


TABLE 1 Common Metrics (or Scores) for Verifying Continuous Deterministic Forecasts (F_i) Against the Observations (O_i)

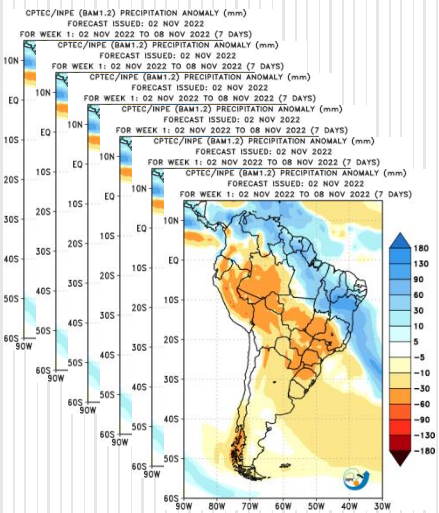
Metric	Equation	Attribute Measured	Characteristics
Bias (linear bias, B)	$B = \frac{1}{N}\left[\sum_{i=1}^{N}(F_i - O_i)\right]$	Accuracy (average error)	Estimates the persistent or average error based on a specific data set; negative orientation (best when $B = 0$)
MAE	$\mathrm{MAE} = \frac{1}{N}\left[\sum_{i=1}^{N}\lvert F_i - O_i\rvert\right]$	Accuracy	Average error magnitude, negative orientation (best when MAE = 0)
RMSE	$\mathrm{RMSE} = \left[\frac{1}{N}\sum_{i=1}^{N}(F_i - O_i)^2\right]^{1/2}$	Accuracy	Average error magnitude weighted to larger errors; Negative orientation (best when RMSE = 0)
SS (Skill Score)	$\mathrm{SS} = \frac{S_f - S_r}{S_p - S_r} = \frac{S_r - S_f}{S_r} = 1 - \frac{S_f}{S_r}$	Skill (general format) For negatively oriented scores, perfect score $S_p = 0$)	Fractional improvement of the forecast over an unskilled reference. Range: $-\infty$ to 1.
Pearson correlation coefficient (r)	$r = \frac{\sum_{i=1}^{N}(F_i - \overline{F})(O_i - \overline{O})}{\sqrt{\sum_{i=1}^{N}(F_i - \overline{F})^2}\sqrt{\sum_{i=1}^{N}(O_i - \overline{O})^2}}$	Association	Strength of the linear relationship between forecasts and observations Range: -1 to 1.

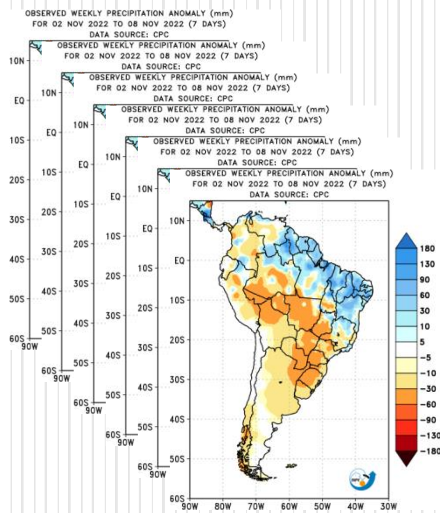
Note: Subscripts i refer to the ith case of the verification sample; the sample of forecast and observation pairs is of size N; the overbar indicates sample averaging. S_f (usually) refers to either the MAE or RMSE scores computed for the N pairs of F_i and O_i according to the equations in the table; S_r refers to the same score computed using a unskilled reference forecast such as the variable mean (climatology) or the latest observed value of the variable (persistence); S_p refers to the score for the perfect forecast. For perfect forecasts where $F_i = O_i$ for all N pairs, both MAE and RMSE equal zero (i.e., $S_p = 0$).

Coelho et al. (2018)

Avaliação das Previsões Determinísticas

- BIAS (VIÉS)

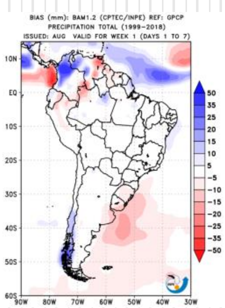




$$\mathrm{v} = \frac{1}{n}\sum_{i=1}^{n}(P_i - o_i)$$

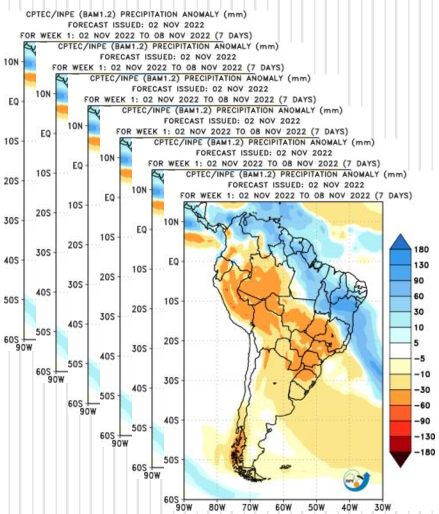
$$\mathrm{v} = \bar{p} - \bar{o}$$

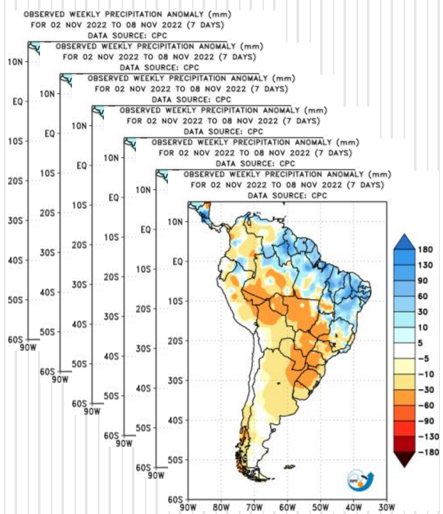
$$\mathrm{MAR} = \frac{1}{n}\sum_{i=1}^{n}(|p_i - o_i|)$$



Avaliação das Previsões Determinísticas

- RMSE | MSE



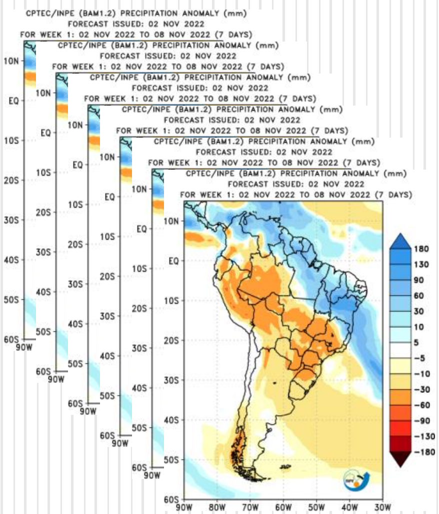


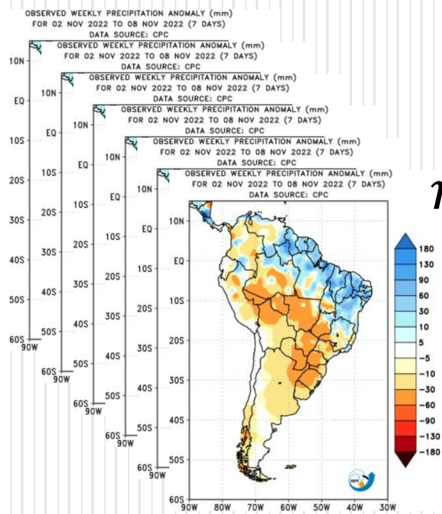
$$\mathrm{RMSE} = \sqrt[2]{\frac{1}{n}\sum_{i=1}^{n}(p_i - o_i)^2}$$

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(p_i - o_i)^2$$

Avaliação das Previsões Determinísticas

- Correlação

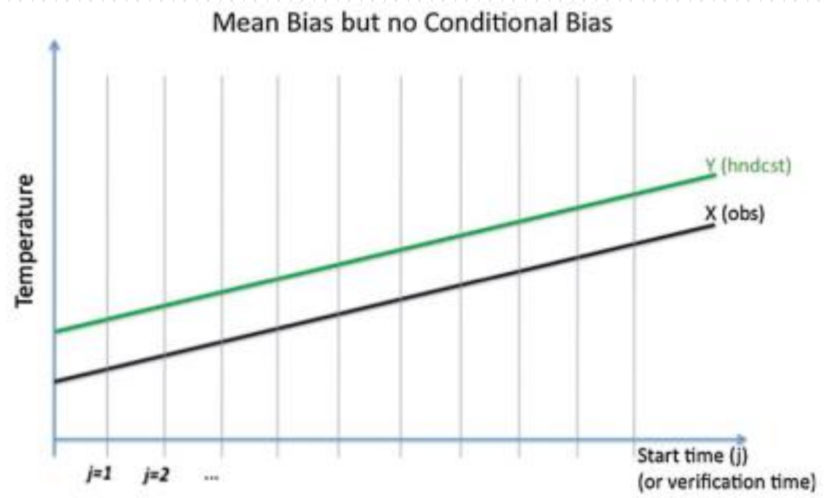


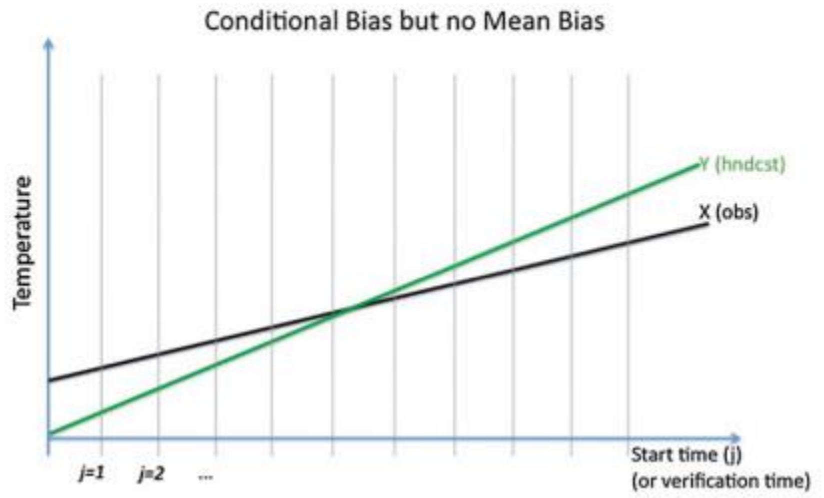


$$r = \frac{\sum_{i=1}^{n}(p_i - \bar{p})(o_i - \bar{o})}{\sqrt[2]{\sum_{i=1}^{n}(p_i - \bar{p})^2}\ \sqrt[2]{\sum_{i=1}^{n}(o_i - \bar{o})^2}}$$

Avaliação das Previsões Determinísticas

- Correlação não é skill. Ela é um potencial skill.





Goddard et al. (2012)

Avaliação das Previsões Determinísticas

- Skill Score (índice de destreza do MSE)

$$SS = 1 - \left(\frac{MSE}{MSE_R}\right)$$

Se Ref. for a climatologia

$$SS = 2\left(\frac{s_p}{s_o}\right)r - \left(\frac{s_p}{s_o}\right)^2 - \left[\frac{(\bar{p} - \bar{o})}{s_o}\right]^2$$

$$SS = r^2 - \left[r - \left(\frac{s_p}{s_o}\right)\right]^2 - \left[\frac{(\bar{p} - \bar{o})}{s_o}\right]^2$$

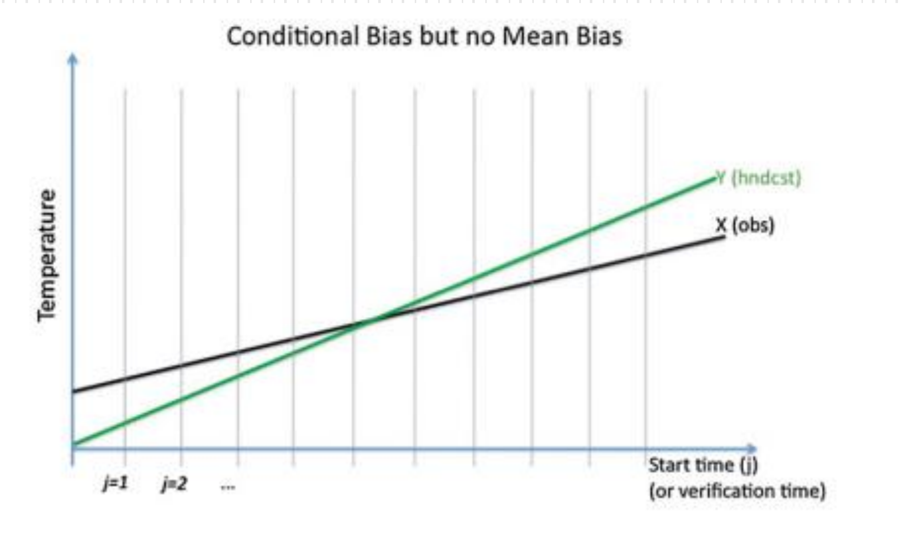
$$s_p = \frac{1}{n}\sqrt{\sum_{i=1}^{n}(p_i - \bar{p})^2}$$

$$s_o = \frac{1}{n}\sqrt{\sum_{i=1}^{n}(o_i - \bar{o})^2}$$

Avaliação das Previsões Determinísticas

- BIAS (viés) condicional

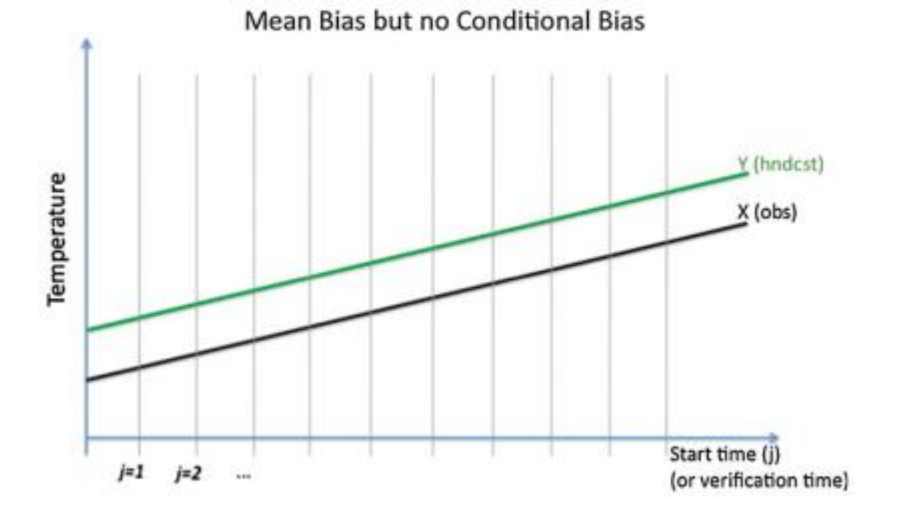
$$BC = \left[r - \left(\frac{s_p}{s_o} \right) \right]^2$$



Avaliação das Previsões Determinísticas

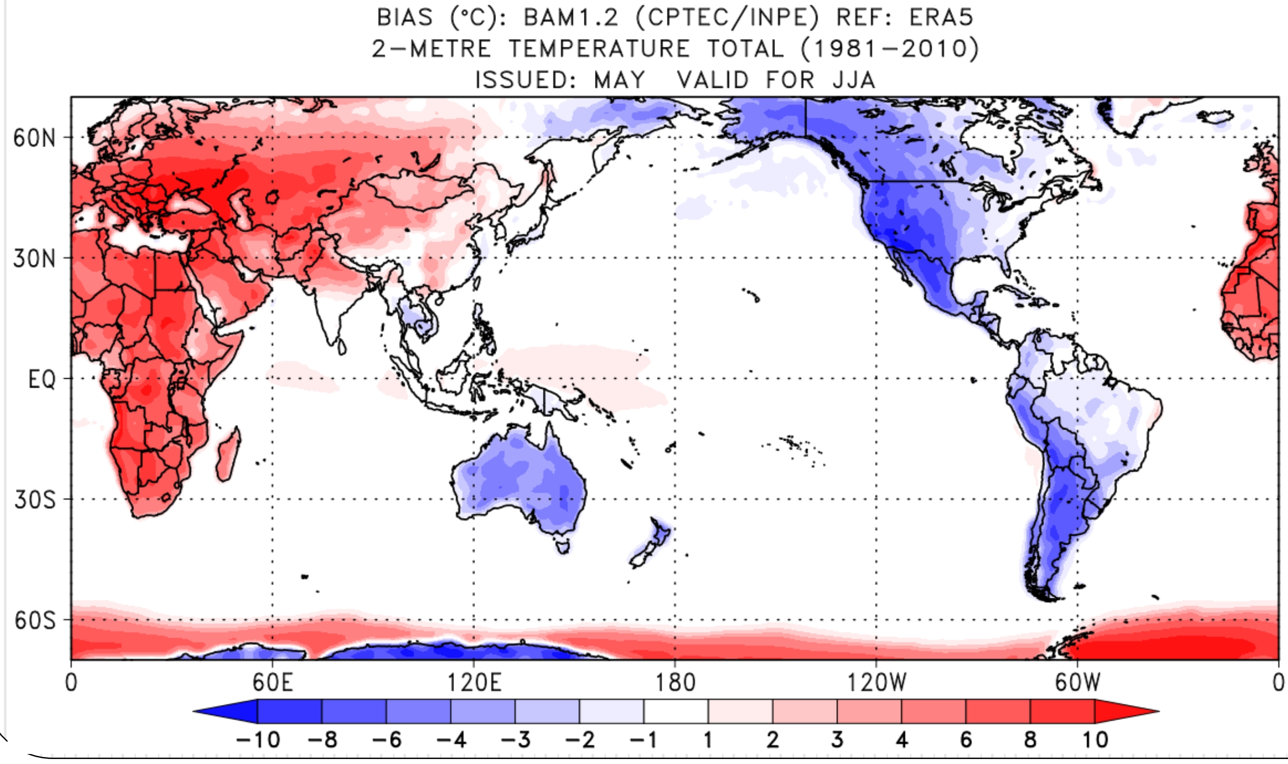
- BIAS (viés) incondicional

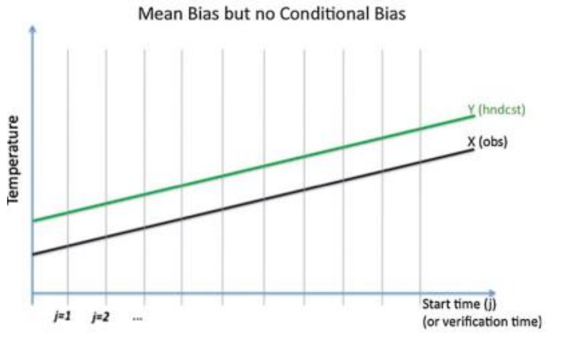
$$BI = \left[\frac{(\bar{p} - \bar{o})}{s_o}\right]^2$$



Avaliação das Previsões Determinísticas

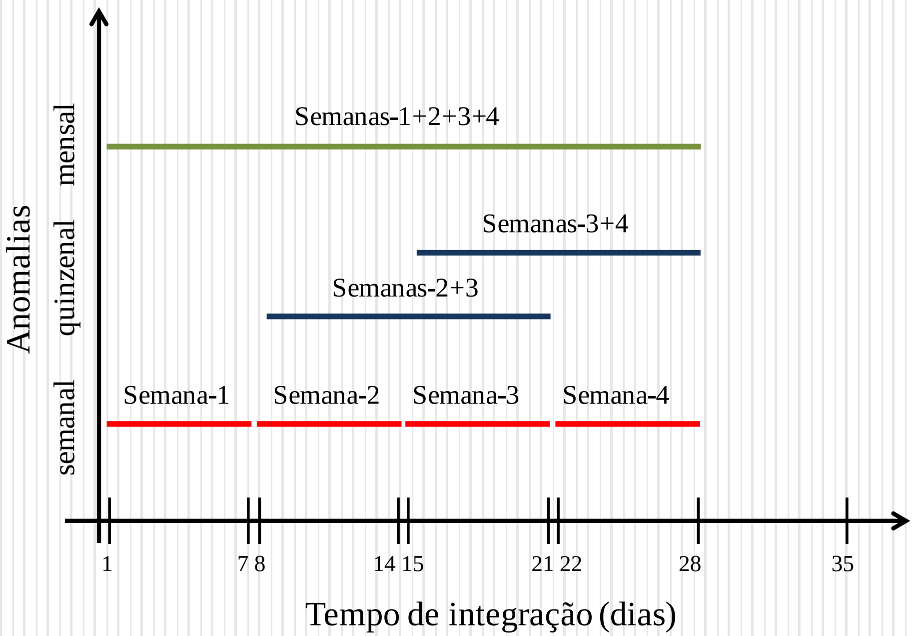
- BIAS (viés) incondicional





Interpretação das Métricas Determinísticas

- Ver scripts



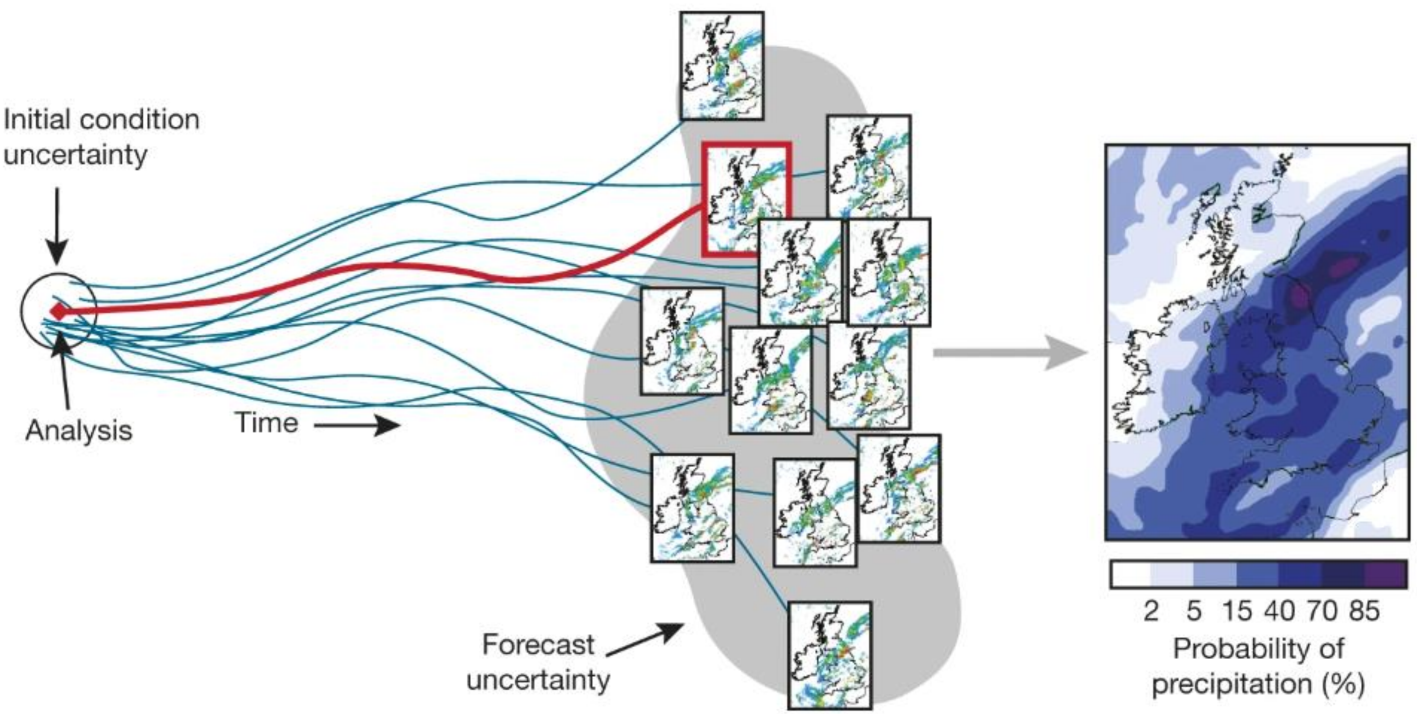
Datas das inicializações das previsões retrospectivas do BAM-1.2

DIAS	MESES	PERÍODO
03 e 14	NOVEMBRO	1999 a 2010
01 e 15	DEZEMBRO	1999 a 2010
04 e 14	JANEIRO	2000 a 2011
01 e 15	FEVEREIRO	2000 a 2011
03 e 14	MARÇO	2000 a 2011

Interpretação das Métricas Determinísticas

- Ver site subsazonal

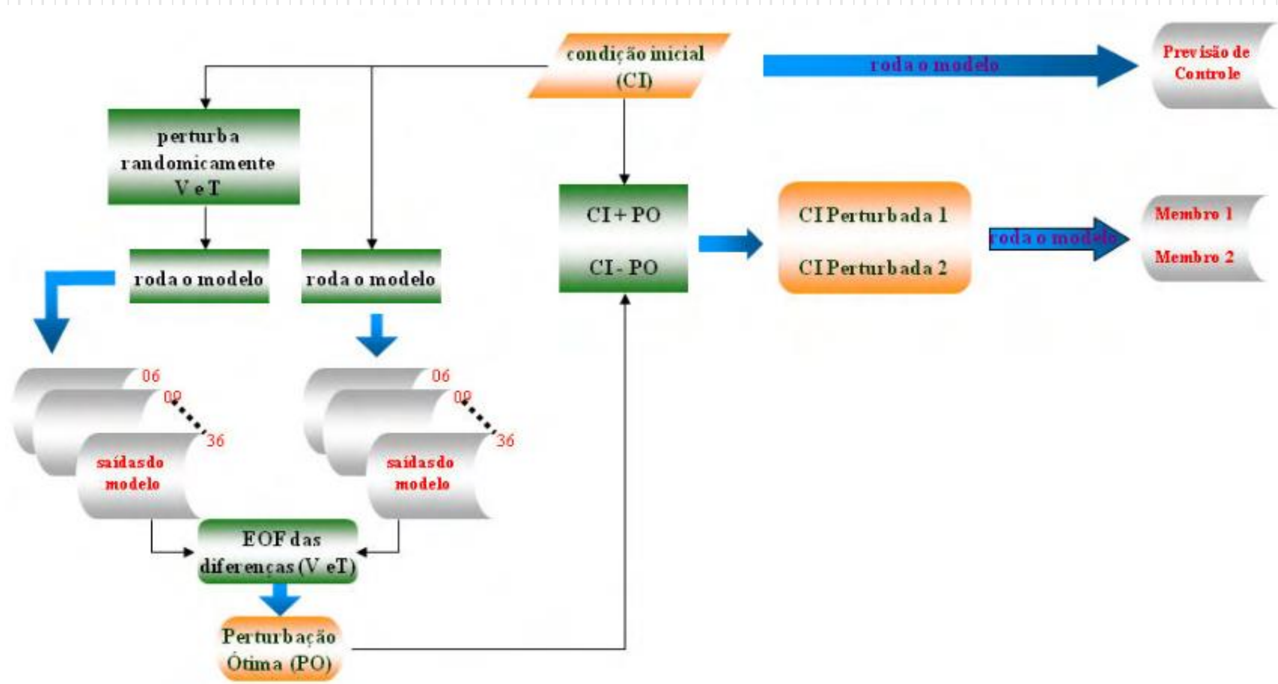
Ensemble



Initial condition uncertainty
Analysis
Time
Forecast uncertainty
2 5 15 40 70 85
Probability of precipitation (%)

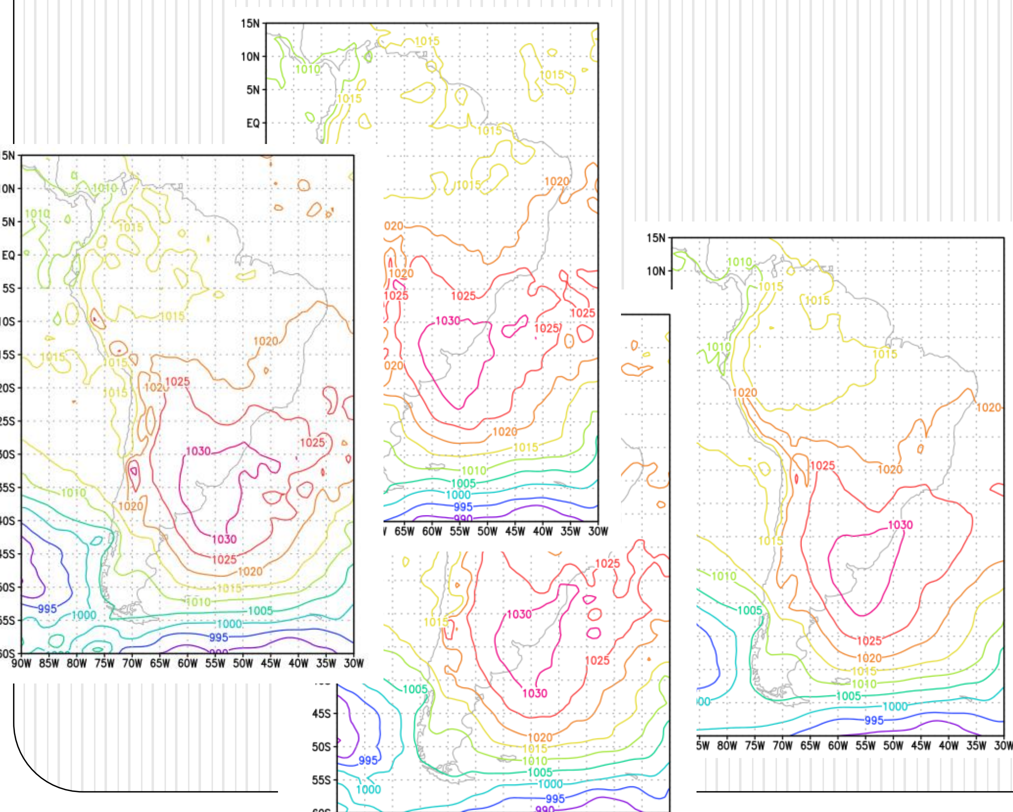
Ensemble

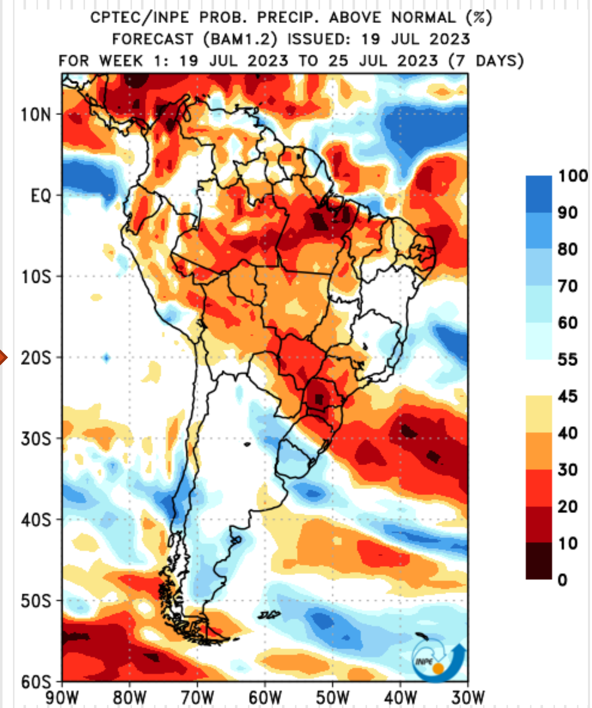
- Mudança na condição inicial



Ensemble

- Perturbações na condição inicial





Avaliação das Previsões Probabilísticas

- Definição do evento
  - Precipitação acima de um limiar (e.g., 50 mm/dia)
  - Ciclone tropical
  - Anomalia positiva de precipitação
  - Temperatura do ar no tercil superior

TABLE 2 Contingency Table Format and Associated Scores

<table>
<tr><th>Measure</th><th>Equation/Format</th><th>Range-Orientation</th><th>Characteristics</th></tr>
<tr><td>Contingency table</td><td>
<table>
<tr><td></td><td></td><td colspan="2">Observed</td><td></td></tr>
<tr><td></td><td></td><td>Yes</td><td>No</td><td>Total fcst</td></tr>
<tr><td rowspan="2">Forecast</td><td>Yes</td><td>a
(Hits)</td><td>b
(False alarms)</td><td>a + b
(total events forecast)</td></tr>
<tr><td>No</td><td>c
(Missed events)</td><td>d
(Correct negatives)</td><td>c + d
(total non-events forecast)</td></tr>
<tr><td></td><td>Total obs</td><td>a + c
(total events obs)</td><td>b + d
(total non-events obs)</td><td>N = a + b + c + d
(sample size)</td></tr>
</table>
</td><td>Normally, as shown, columns are conditional observation totals, and rows are conditional forecast totals.</td><td>Equivalent to a scatterplot for categorized variable; 2 × 2 table most common—two categories, one threshold.</td></tr>
<tr><td>FB</td><td>$\mathrm{FB} = \frac{a+b}{a+c}; \frac{c+d}{b+d}$
Ratio between the total number of events forecast (or not forecast) and the total number of events observed (or not observed).</td><td>0 to ∞</td><td>Best score = 1. Simple comparison of forecast frequency to observed frequency.</td></tr>
<tr><td>H
(Probability of detection)</td><td>$H = \frac{a}{a+c}$</td><td>0 to 1</td><td>Best = 1. Incomplete score—does not account for false alarms.</td></tr>
<tr><td>F
(probability of false detection)</td><td>$F = \frac{b}{b+d}$</td><td>1 to 0</td><td>Best = 0. Can be improved by forecasting the event less often to reduce false alarms.</td></tr>
<tr><td>FAR</td><td>$\mathrm{FAR} = \frac{b}{a+b}$</td><td>1 to 0</td><td>Best = 0. Sensitive to false alarms but ignores misses. Use with H.</td></tr>
<tr><td>TS
(critical success index)</td><td>$\mathrm{TS} = \frac{a}{a+b+c}$</td><td>0 to 1</td><td>Best = 1. Sensitive to both false alarms and misses; ignores correct negatives.</td></tr>
</table>

Continued

Coelho et al. (2018)

TABLE 2 Contingency Table Format andAssociated Scores—cont'd

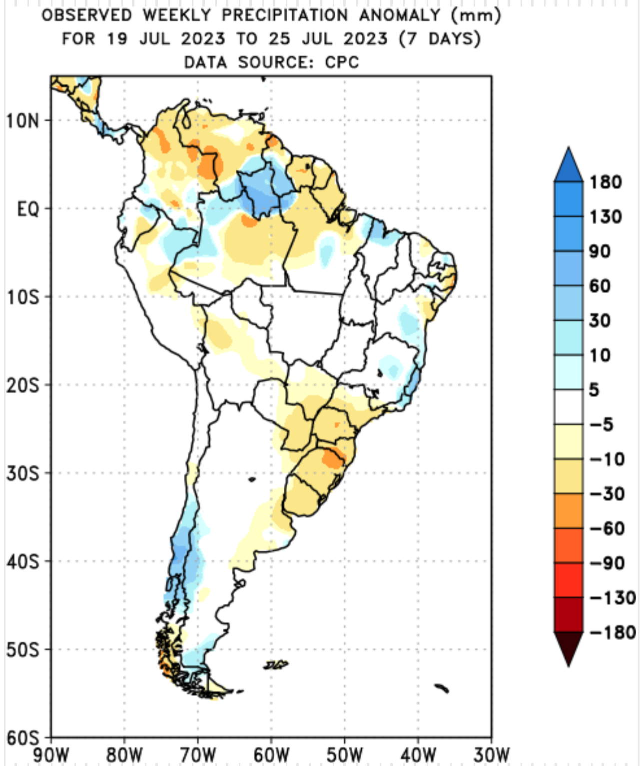
Measure	Equation/Format	Range-Orientation	Characteristics
ETS (Gilbert skill score)	$\mathrm{ETS} = \frac{a - a_r}{a + b + c - a_r}$ where $a_r = \frac{(a+b)(a+c)}{N}$	−1/3 to 1; 0 indicates no skill over chance.	Best = 1. TS adjusted for the number correct by chance (guessing), a form of SS. Always <TS.
KSS (also true skill statistic TSS or Pierce skill score	$\mathrm{KSS} = \frac{a}{a+c} - \frac{b}{b+d} = H - F$	−1 to 1; 0 indicates no discriminant ability	Best = 1. Related to the ROC area and EDI/SEDI scores. Indicates the ability of the forecast to discriminate between events and nonevents, as a basis for decision-making.
HSS	$\mathrm{HSS} = \frac{(a+d) - E_r}{N - E_r}$ where $E_r = \frac{1}{N}[(a+c)(a+b) + (c+d)(b+d)]$	−∞ to 1	Best = 1. SS in the general format, with "chance" as the reference forecast.
EDI	$\mathrm{EDI} = \frac{\ln F - \ln H}{\ln F + \ln H}$	−1 to 1; 0 indicates no accuracy.	Best = 1. Designed to avoid convergence to 0 or 1 for low frequency (rare) events. Most often used for verifying extreme event forecasts.
SEDI	$\mathrm{SEDI} = \frac{\ln F - \ln H + \ln(1-H) - \ln(1-F)}{\ln F + \ln H + \ln(1-H) + \ln(1-F)}$	−1 to 1; 0 indicates no accuracy.	Best = 1. Similar to EDI, but approaches 1 only for unbiased forecasts.

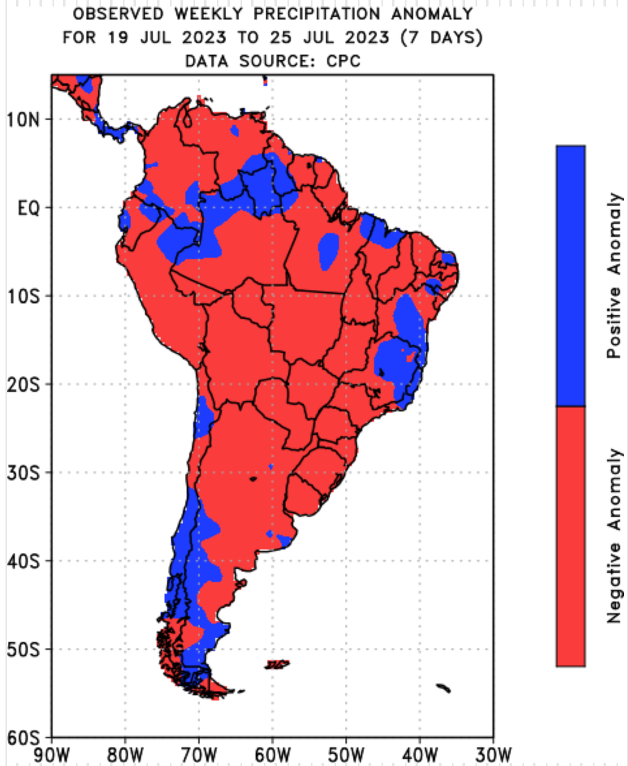
Note: The letters *a*, *b*, *c*, and *d* refer to total counts of cases with the corresponding pairing of forecast and observation. Sample size is denoted as *N*.

Coelho et al. (2018)

Avaliação das Previsões Probabilísticas

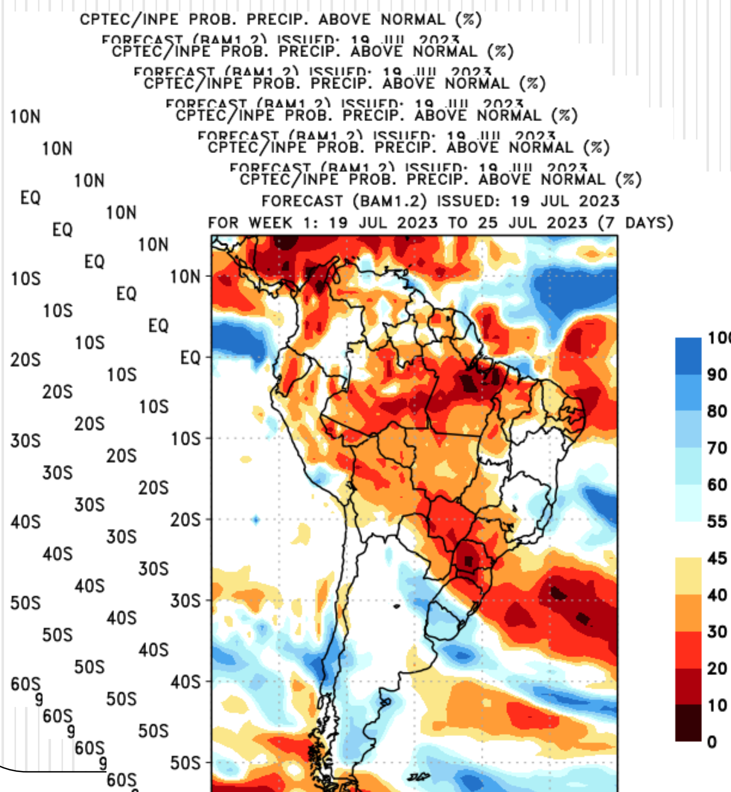
- Definição do evento precipitação positiva



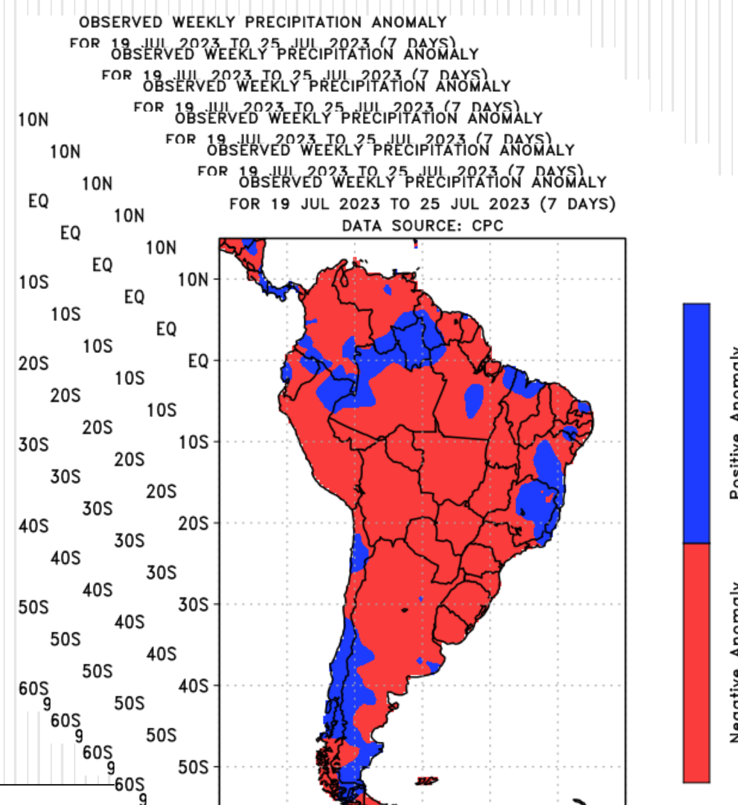


Avaliação das Previsões Probabilísticas

Previsão Probabilística



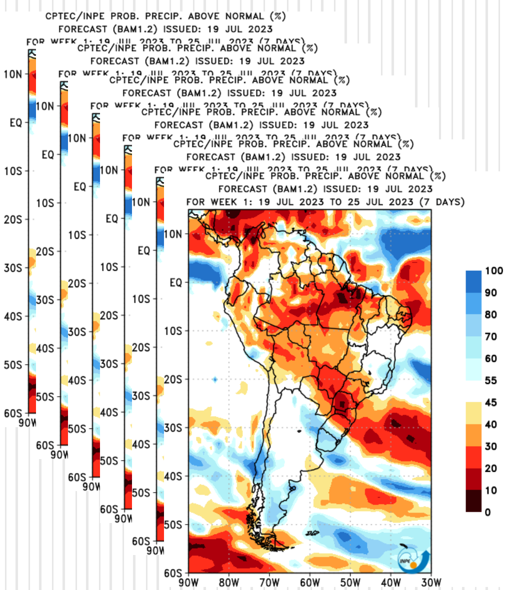
Ocorrência ou não do evento(observação)



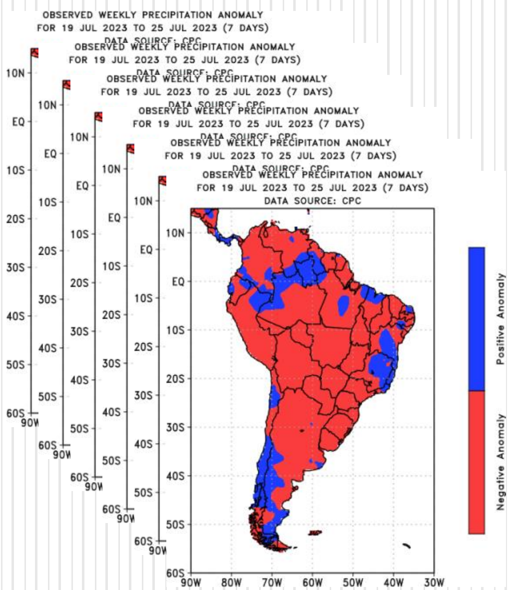
Avaliação das Previsões Probabilísticas

Previsão Probabilística

Ocorrência ou não do evento(observação)



VS



P: 0,4 | 0,3 |0,5 |0,1 | 0,6 | 0,2

O: 1 | 1 | 0 | 1 | 0 | 0

Avaliação das Previsões Probabilísticas

- Brier Score

$$\mathrm{BS} = \frac{1}{n}\sum_{i=1}^{n}(p_i - o_i)^2$$

$$0 \leq \mathrm{BS} \leq 1$$

Avaliação das Previsões Probabilísticas

- Decomposição do Brier Score

$$\mathrm{BS} = \frac{1}{n}\sum_{i=1}^{n}(p_i - o_i)^2 \qquad 0 \le \mathrm{BS} \le 1$$

$$\mathrm{BS} = \underbrace{\frac{1}{N}\sum_{k=1}^{j} I_k(p_k - \bar{O}_k)^2}_{\text{Confiabilidade}} - \underbrace{\frac{1}{N}\sum_{k=1}^{j} I_k(\bar{O}_k - \bar{o})^2}_{\text{Resolução}} + \underbrace{\bar{o}(1-\bar{o})}_{\text{Incerteza}}$$

$$\bar{O}_k = p(O_k|p_k) = \frac{1}{N_k}\sum_{t \in N_k} O_t \qquad \bar{o} = \frac{1}{n}\sum_{i=1}^{n} o_i$$

$k = 1, \ldots, j=11$: $p_1=0$, $p_2=0{,}1 \ldots p_{11}=1$

Avaliação das Previsões Probabilísticas

$$BS = Con - Res + Inc$$

$$BSS = 1 - \left(\frac{BS}{BS_R}\right)$$

Se Ref for a climatologia

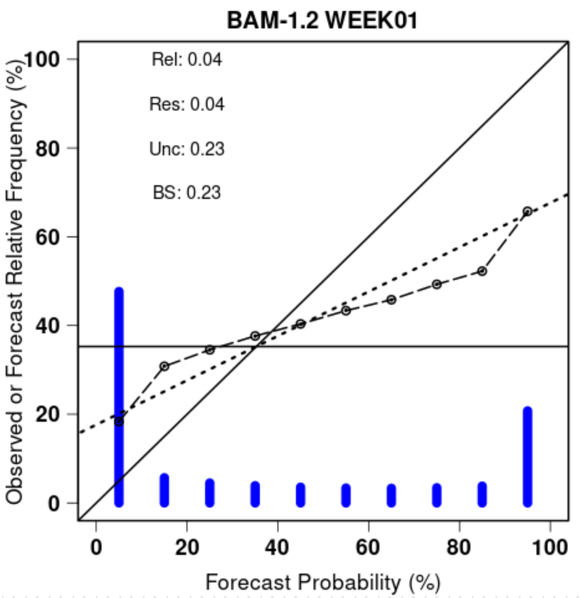
$$BS_R = 0 - 0 + Inc$$

$$BSS = 1 - \left(\frac{Con - Res + Inc}{Inc}\right)$$

$$BSS = \frac{(Res - Con)}{Inc}$$

Avaliação das Previsões Probabilísticas

- Diagrama de Confiabilidade (Reliability Diagram)

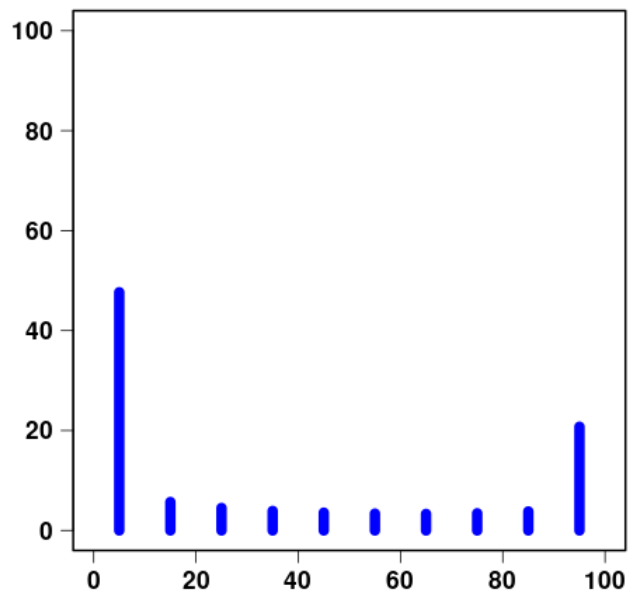


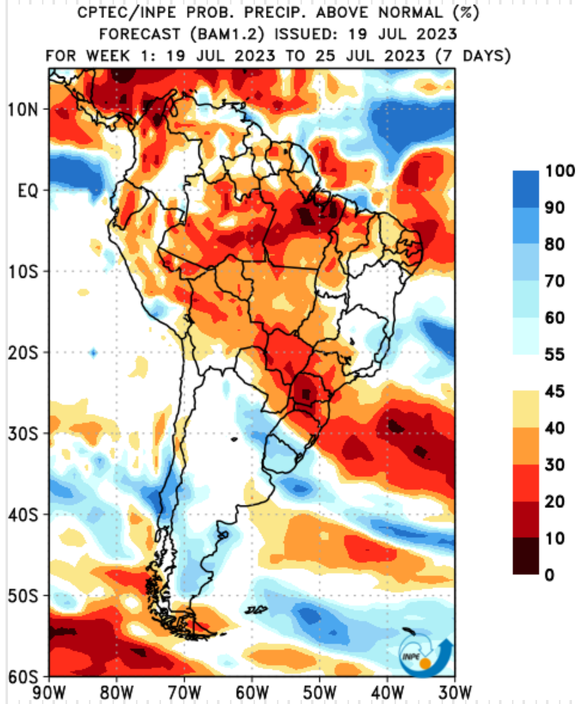
Atributos do diagrama de confiabilidade:

- Confiabilidade;
- Resolução;
- Nitidez.

Avaliação das Previsões Probabilísticas

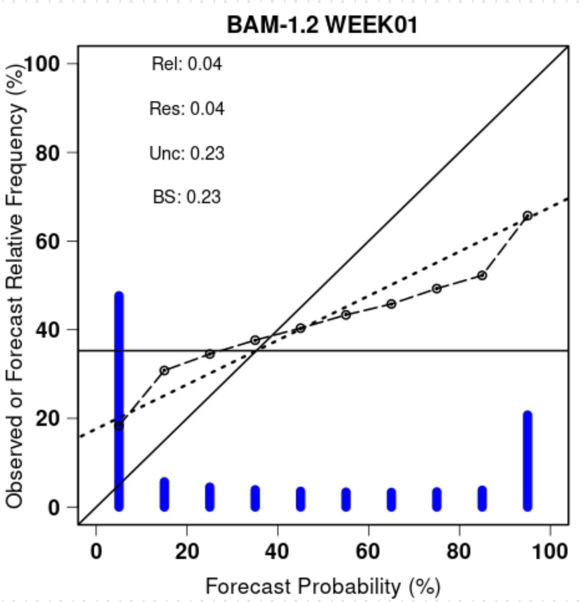
- Diagrama de Confiabilidade (Reliability Diagram)

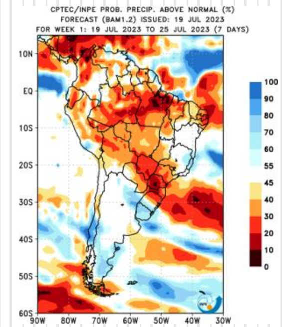


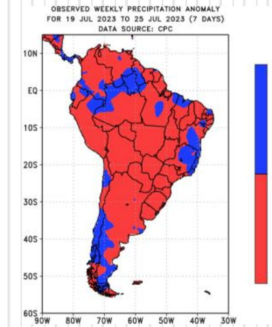


Avaliação das Previsões Probabilísticas

- Diagrama de Confiabilidade (Reliability Diagram)



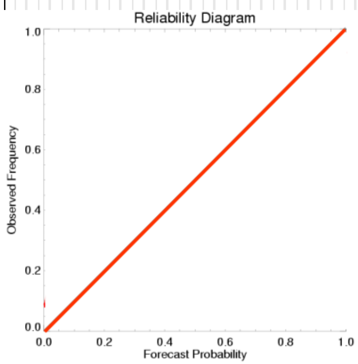




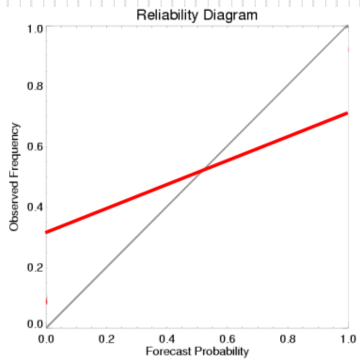
Probabilidades previstas	N° Previsões	Previsão perfeita	Previsão real
100%	8000	8000 (100%)	7200 (90%)
90%	5000	4500 (90%)	4000 (80%)
80%	4500	3600 (80%)	3000 (66%)
…	…	…	…
10%	5500	550 (10%)	800 (15%)
0%	7000	0 (0%)	700 (10%)

Avaliação das Previsões Probabilísticas

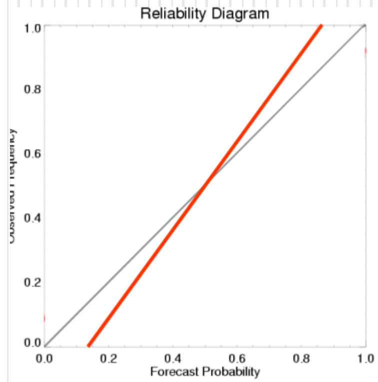
- Diagrama de Confiabilidade (Reliability Diagram)



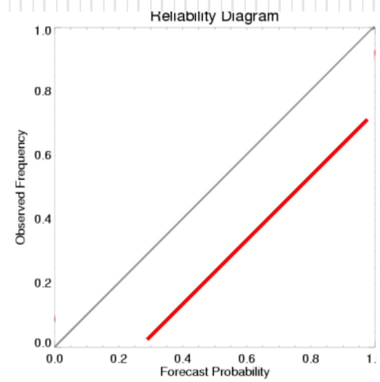
Previsões confiáveis



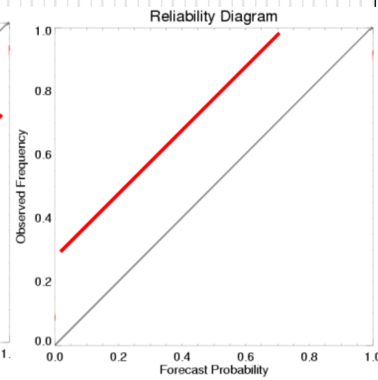
Previsões superconfiantes com pouca resolução



Previsões sub confiantes com resolução



Viés positivo



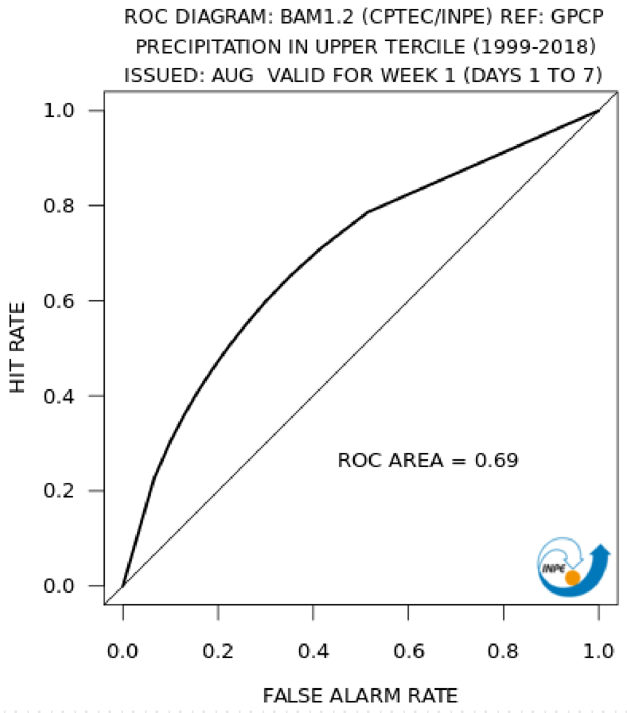
Viés negativo

Avaliação das Previsões Probabilísticas

- Área sob a curva ROC (Relative operating characteristics)

		Observed		
		Yes	No	Total fcst
Forecast	Yes	*a* (Hits)	*b* (False alarms)	*a* + *b* (total events forecast)
	No	*c* (Missed events)	*d* (Correct negatives)	*c* + *d* (total non-events forecast)
	Total obs	*a* + *c* (total events obs)	*b* + *d* (total non-events obs)	*N* = *a* + *b* + *c* + *d* (sample size)

Área sob a curva ROC é uma medida de discriminação



Interpretação das Métricas Probabilísticas

- Ver scripts

Interpretação das Métricas Probabilísticas

- Ver site subsazonal

Avaliação de Simulações

- Simulações com modelos de circulação geral são realizadas para diversos fins. Dessa forma, não existe um protocolo de avaliação definido para todos os tipos de avaliação (assim como mostrado no caso das previsões).

Avaliação de Simulações

MJO Simulation Diagnostics

CLIVAR Madden–Julian Oscillation Working Group:

D. Waliser,* *Jet Propulsion Laboratory, Pasadena, California;* K. Sperber,* *Lawrence Livermore National Laboratory, PCMDI, Livermore, California;* H. Hendon, *Centre for Australian Weather and Climate Research, Melbourne, Australia;* D. Kim, *Seoul National University, Seoul, South Korea;* E. Maloney, *Colorado State University, Fort Collins, Colorado;* M. Wheeler, *Centre for Australian Weather and Climate Research, Melbourne, Australia;* K. Weickmann, *NOAA/Earth System Research Laboratory, Boulder, Colorado;* C. Zhang, *Rosenstiel School of Marine and Atmospheric Science, Miami, Florida;* L. Donner, *NOAA/GFDL, Princeton, New Jersey;* J. Gottschalck, W. Higgins, *NOAA/NCEP, Camp Springs, Maryland;* I.-S. Kang, *Seoul National University, Seoul, South Korea;* D. Legler, *U.S. CLIVAR Office, Washington, D.C.;* M. Moncrieff, *NCAR, Boulder, Colorado;* S. Schubert, *NASAGSFC, Greenbelt, Maryland;* W. Stern, *NOAA/GFDL, Princeton, New Jersey;* F. Vitart, *European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom;* B. Wang, *IPRC, University of Hawaii at Manoa, Honolulu, Hawaii;* W. Wang, *NOAA/NCEP, Camp Springs, Maryland;* S. Woolnough, *University of Reading, Reading, United Kingdom*

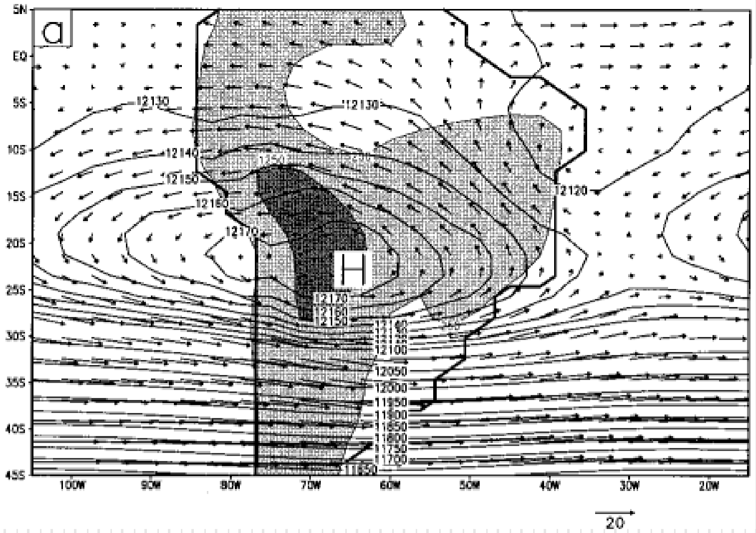
(Manuscript received 17 July 2008, in final form 13 November 2008)

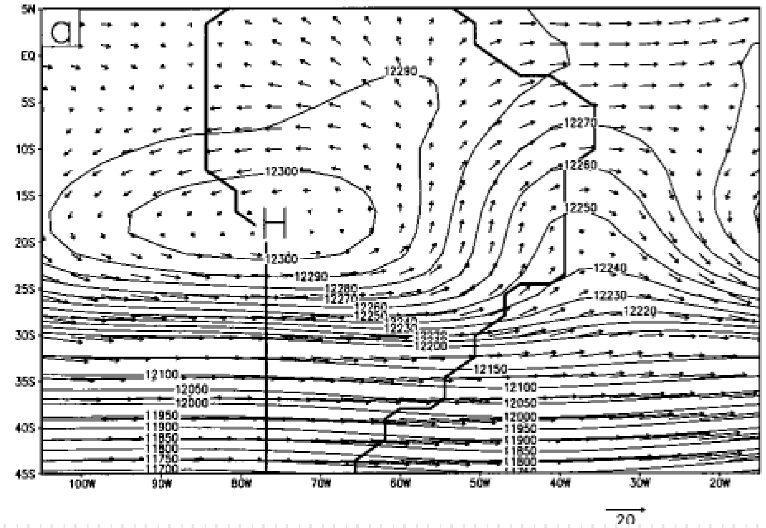
ABSTRACT

The Madden–Julian oscillation (MJO) interacts with and influences a wide range of weather and climate phenomena (e.g., monsoons, ENSO, tropical storms, midlatitude weather), and represents an important, and as yet unexploited, source of predictability at the subseasonal time scale. Despite the important role of the MJO in climate and weather systems, current global circulation models (GCMs) exhibit considerable shortcomings in representing this phenomenon. These shortcomings have been documented in a number of multimodel comparison studies over the last decade. However, diagnosis of model performance has been challenging, and model progress has been difficult to track, because of the lack of a coherent and standardized set of MJO diagnostics. One of the chief objectives of the U.S. Climate Variability and Predictability (CLIVAR) MJO Working Group is the development of observation-based diagnostics for objectively evaluating global model simulations of the MJO in a consistent framework. Motivation for this activity is reviewed, and the intent and justification for a set of diagnostics is provided, along with specification for their calculation, and illustrations of their application. The diagnostics range from relatively simple analyses of variance and correlation to more sophisticated space–time spectral and empirical orthogonal function analyses. These diagnostic techniques are used to detect MJO signals, to construct composite life cycles, to identify associations of MJO activity with the mean state, and to describe interannual variability of the MJO.

Avaliação de Simulações

- Influência dos Andes na formação da Alta da Bolívia

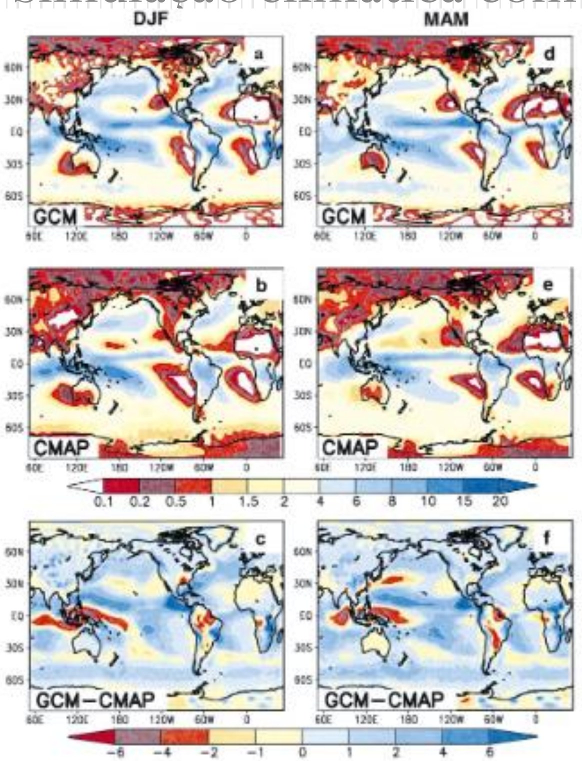




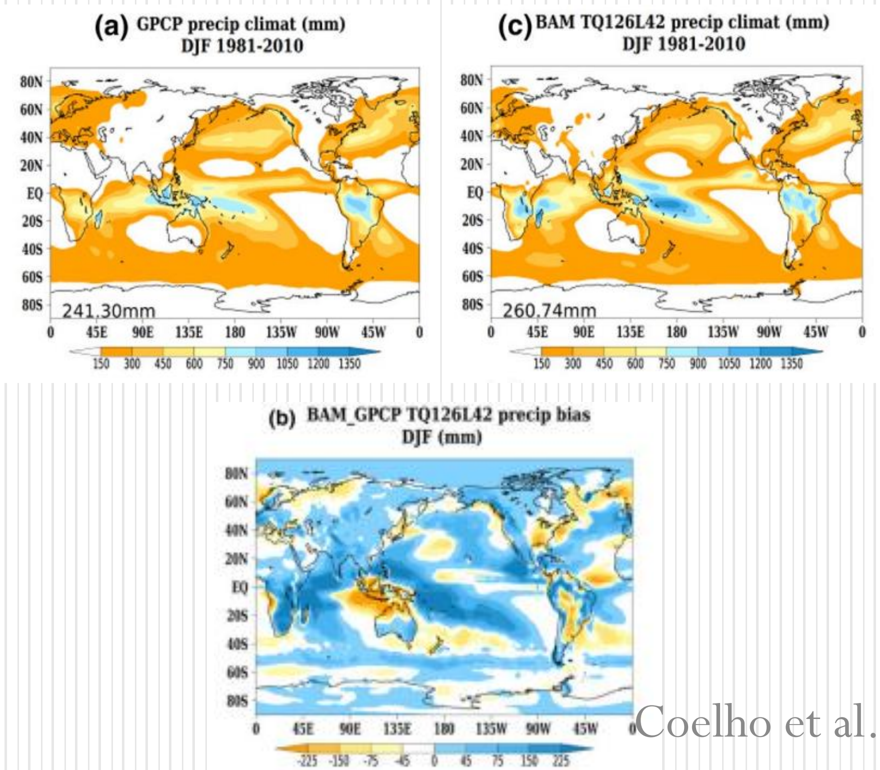
Lenters e Cook (1997)

Avaliação de Simulações

- Simulação climática com os modelos globais do CPTEC



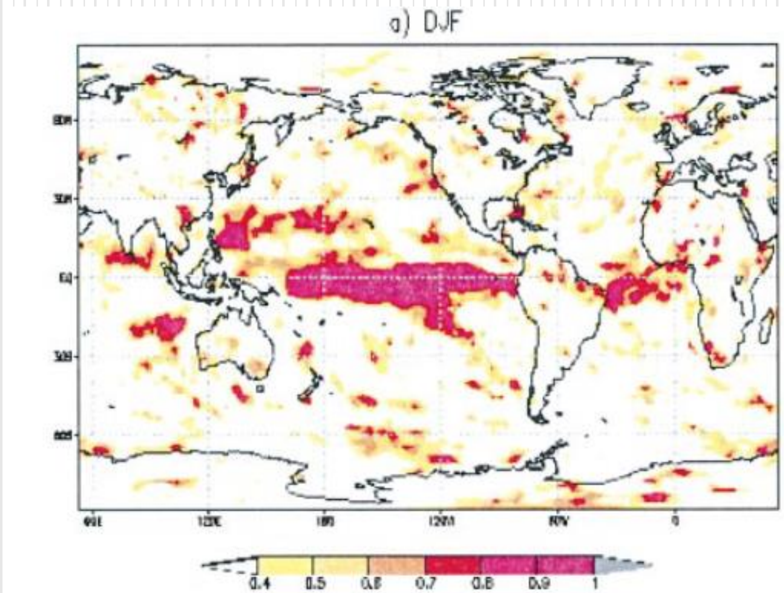
Cavalcanti et al. (2002)



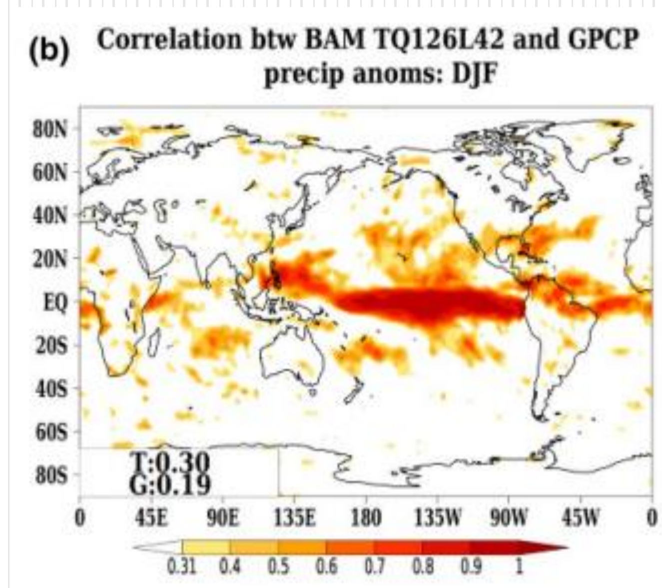
Coelho et al. (2020)

Avaliação de Simulações

- Simulações climáticas com os modelos globais do CPTEC

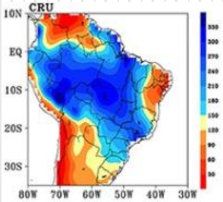


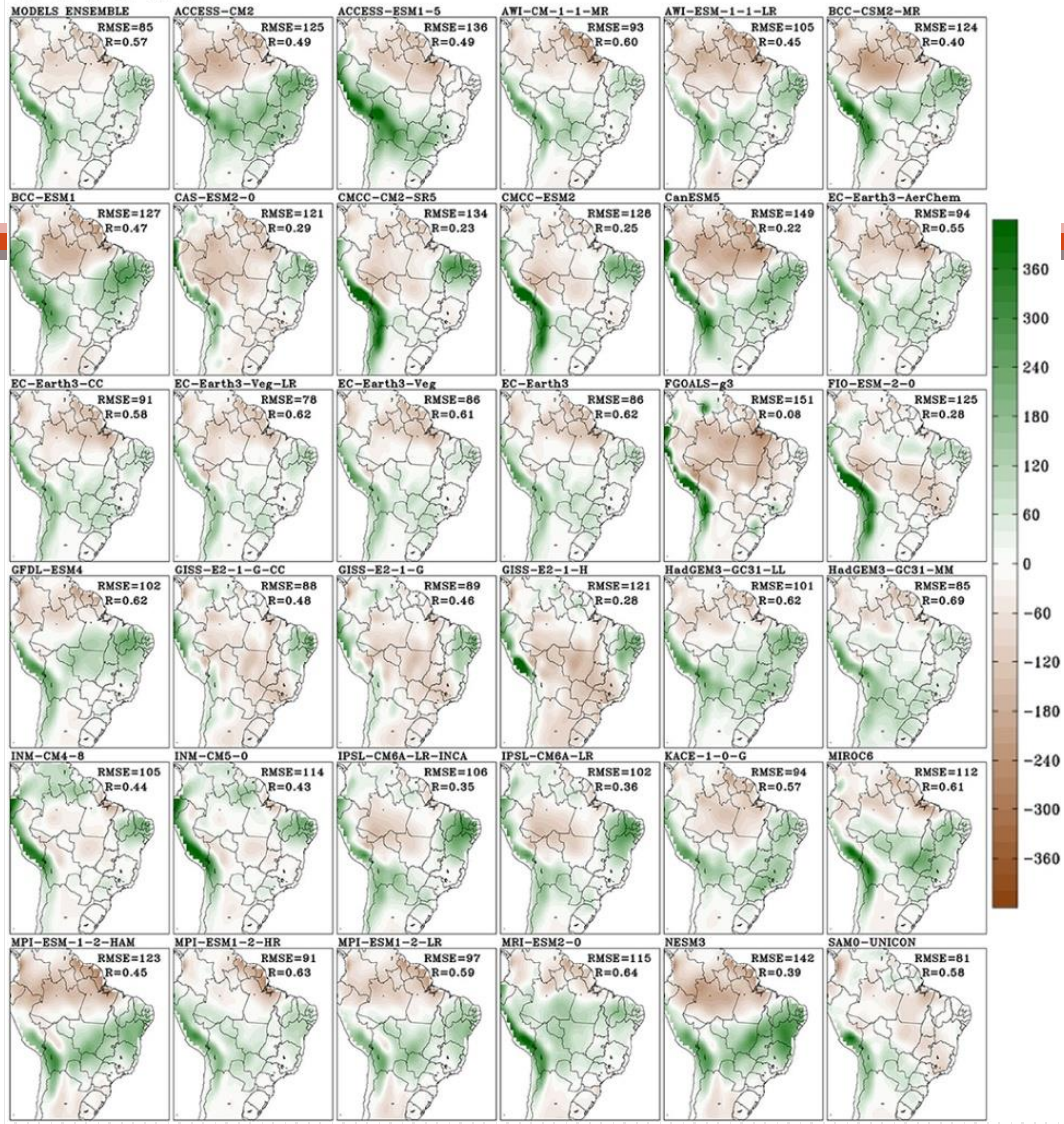
Cavalcanti et al. (2002)



Coelho et al. (2020)

Avaliação de Simulações



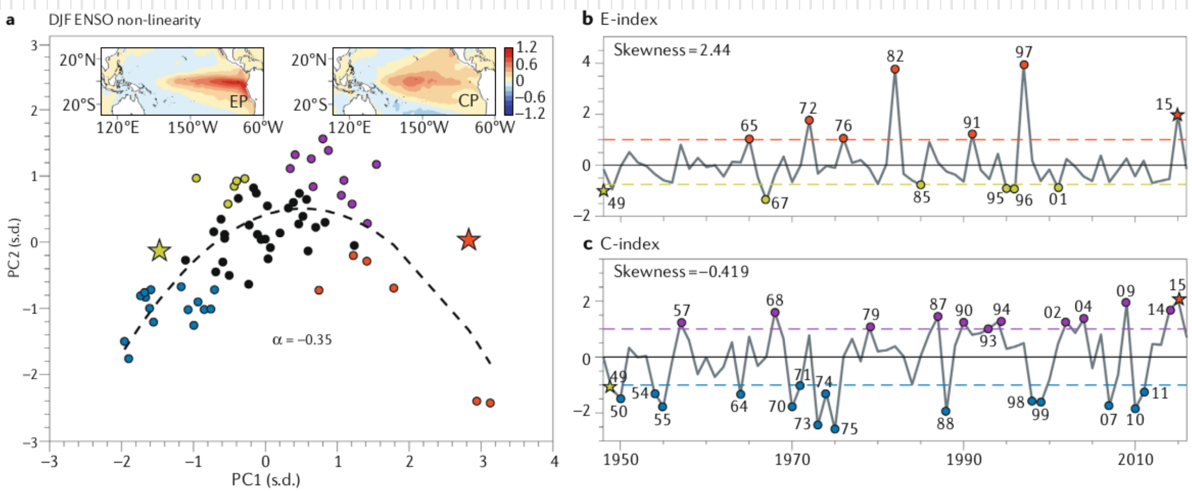


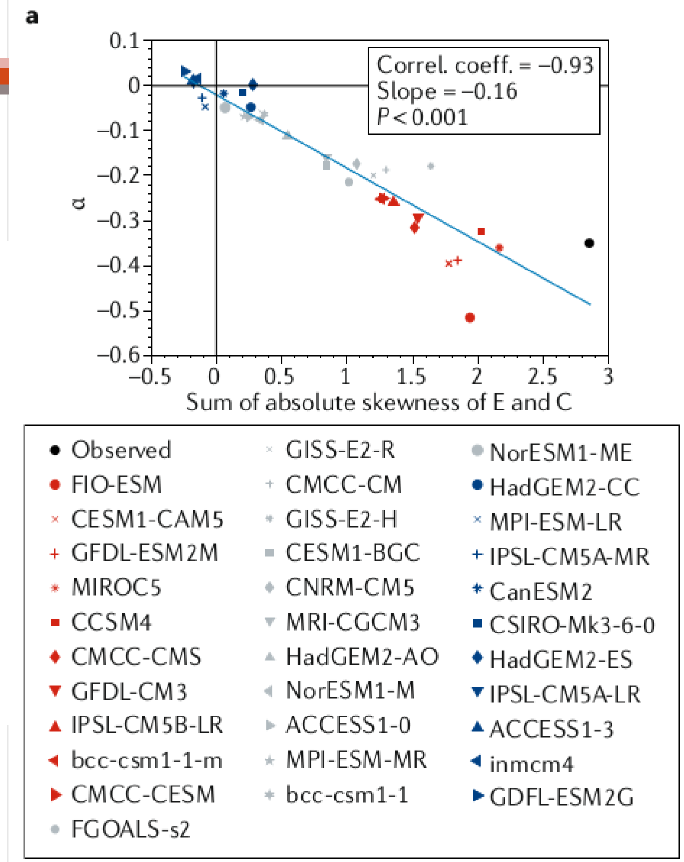
- CMIP6

Firpo et al. (2022)

Avaliação de Simulações

- CMIP e ENOS





Cai et al. (2020)

Diagrama de Taylor

Taylor (2001)

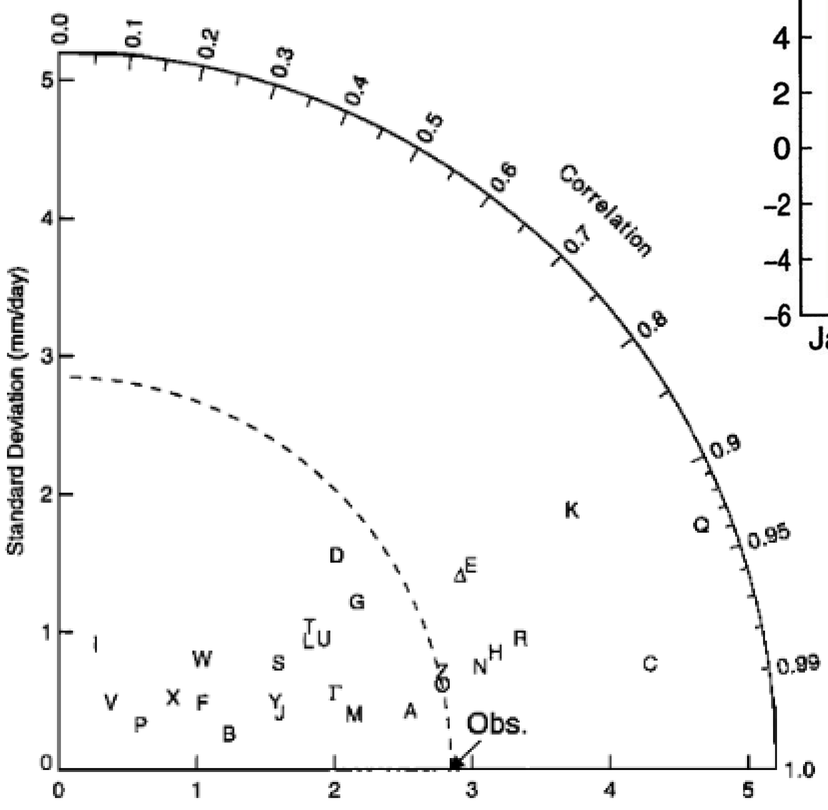
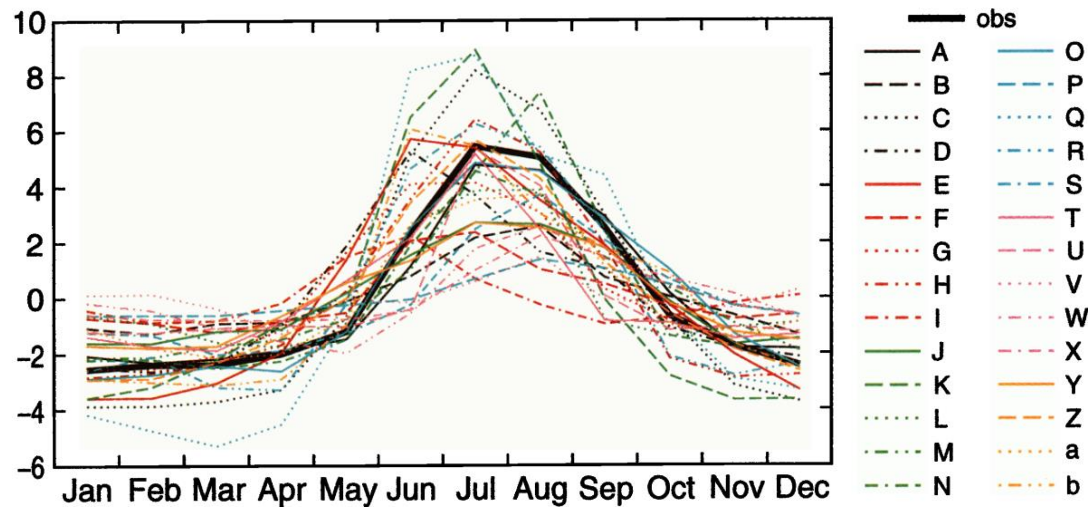
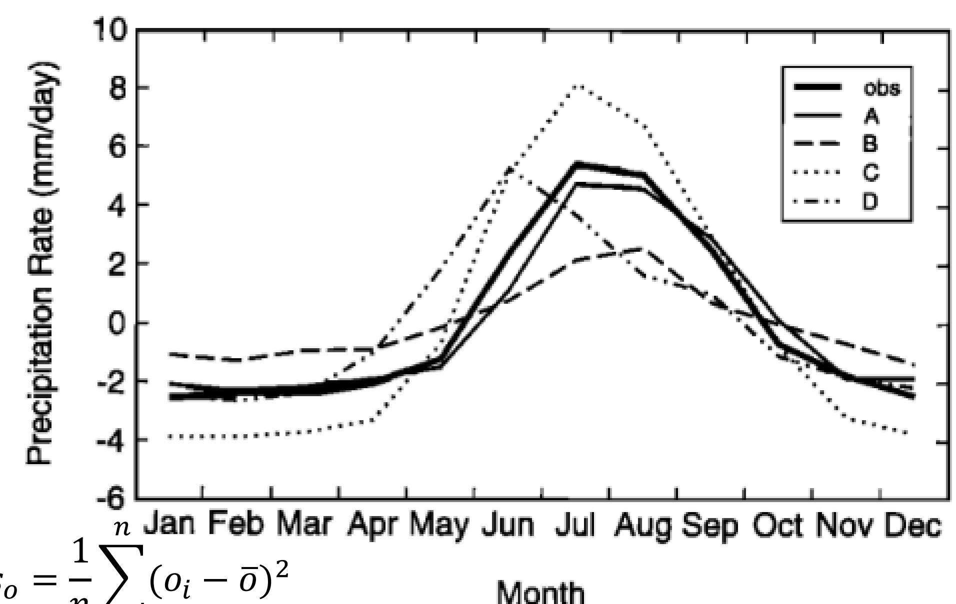


Figure 3. Pattern statistics describing the climatological annual cycle of precipitation over India simulated by 28 models compared with the observed [*Parthasarathy et al.*, 1994]. To simplify the plot, the isolines indicating correlation, standard deviation, and RMS error have been omitted.





$$s_o = \frac{1}{n}\sum_{i=1}^{n}(o_i - \bar{o})^2$$

Diagrama de Taylor

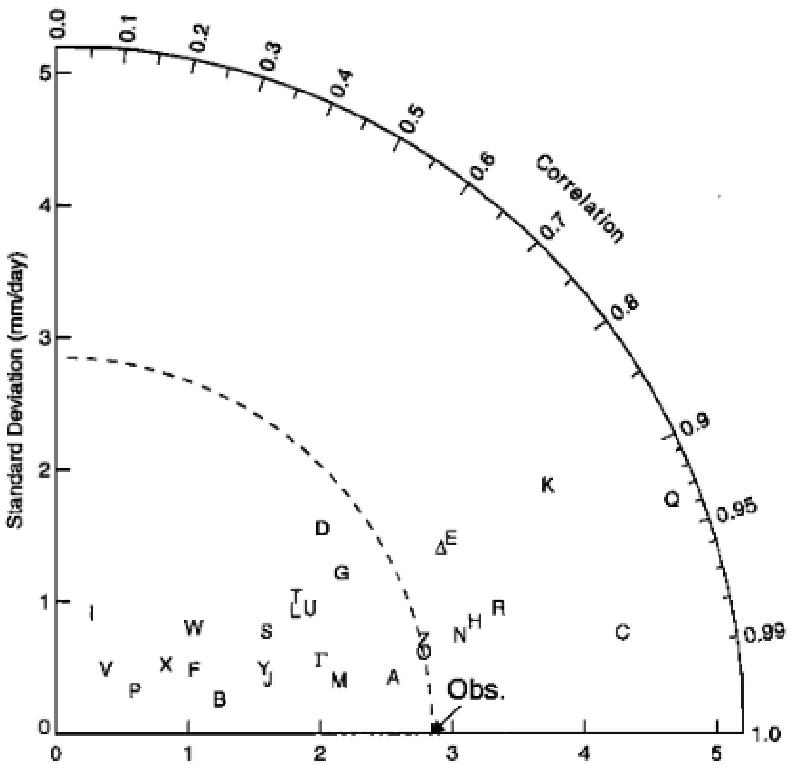
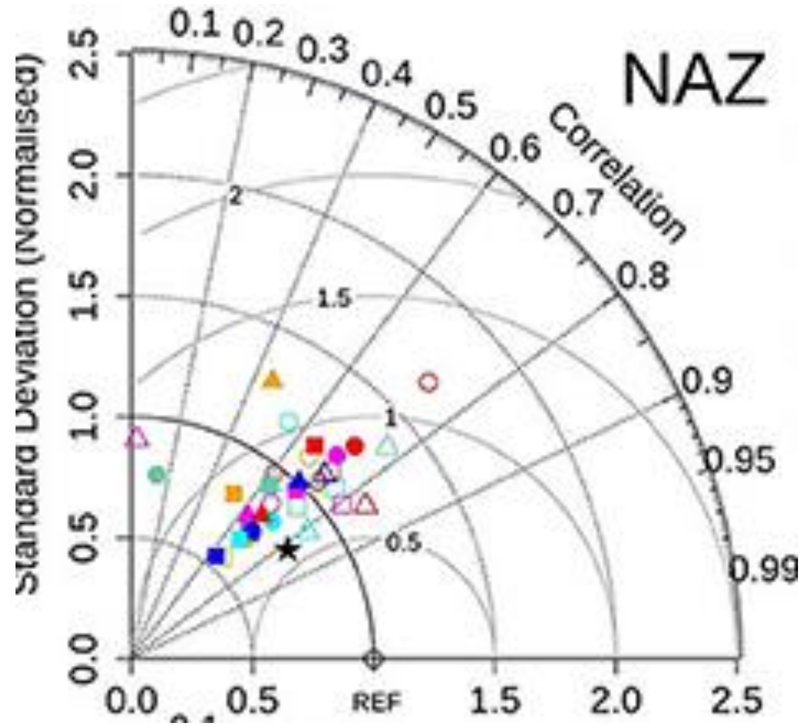


Figure 3. Pattern statistics describing the climatological annual cycle of precipitation over India simulated by 28 models compared with the observed [*Parthasarathy et al.*, 1994]. To simplify the plot, the isolines indicating correlation, standard deviation, and RMS error have been omitted.



$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(p_i - o_i)^2$$

$$\mathrm{CMSE} = \frac{1}{n}\sum_{i=1}^{n}(p_i - \bar{p}) - (o_i - \bar{o}))^2$$

Pontos destacados

- Definição das principais métricas estatísticas de verificação;
- Atributos de cada métrica;
- Diferenças entre previsão e simulação;
- Aplicação das métricas estatísticas na avaliação de previsão e simulação;
- Prática.

Referências

- COELHO, Caio AS et al. Forecast verification for S2S timescales. In: **Sub-Seasonal to Seasonal Prediction**. Elsevier, 2019. p. 337-361.
- GODDARD, Lisa et al. A Verification Framework for Interannual-to-Decadal Prediction Experiments. **Climate Prediction S&T Digest**, p. 95, 2011.
- MURPHY, Allan H. A new vector partition of the probability score. **Journal of Applied Meteorology and Climatology**, v. 12, n. 4, p. 595-600, 1973.
- MURPHY, Allan H. Skill scores based on the mean square error and their relationships to the correlation coefficient. **Monthly weather review**, v. 116, n. 12, p. 2417-2424, 1988.