



WORLD
METEOROLOGICAL
ORGANIZATION



Assimilação de Dados

Determinação da Condição Inicial, R2O e Atividades Operacionais

CARLOS FREDERICO BASTARZ

TRAINING COURSE ON WEATHER FORECASTING - AND BEYOND

17 de Novembro de 2025

Sumário

1. Assimilação de Dados

- 1.1 O que é Assimilação de Dados?
- 1.2 Motivação
- 1.3 Intuição matemática

2. Determinação da Condição Inicial

- 2.1 Evolução do Skill da Assimilação de Dados
- 2.2 Sistema de Modelagem Numérica e Assimilação de dados do CPTEC
- 2.3 Gridpoint Statistical Interpolation

3. R2O – Research to Operations

- 3.1 O que é e por que é necessário R2O?
- 3.2 Ferramentas de Apoio
- 3.3 Fluxo de Transição

4. Atividades Operacionais

- 4.1 Custo Operacional
- 4.2 Monitoramento
- 4.3 Comparações com outros produtos numéricos

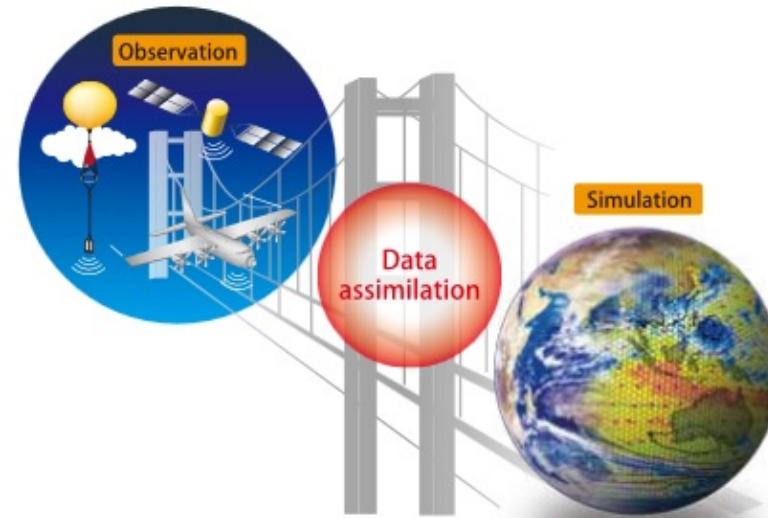
5. Conclusões

1. Assimilação de Dados

1.1 O que é Assimilação de Dados?

- Assimilação de Dados compreende um conjunto de técnicas que permitem a combinação, de forma ótima, de observações e previsões numéricas levando em consideração os seus respectivos erros
 - As observações são espalhadas de forma irregular no espaço
 - As previsões são apresentadas em uma grade regular
 - A combinação de ambos acarreta na atualização/correção da previsão, e isso chamamos de análise

- A análise é a condição inicial dos modelos numéricos



Fonte: <https://www.data-assimilation.riken.jp/en/research/index.html>

1. Assimilação de Dados

1.2 Motivação

- 👉 Modelos e observações possuem incertezas
 - 🔴 Modelos
 - 💻 Discretização das equações, parametrizações físicas etc
 - 🔵 Observações
 - 🔧 Calibração dos instrumentos, local das medidas (e.g., proximidade a rios), erros de anotação etc
- ▣ A assimilação de dados precisa considerar estes fatores de forma que estas incertezas possam ponderar a contribuição das parcelas
 - 👉 Quanto maior o erro do modelo/observação, menor a sua precisão e, consequentemente, menor o seu peso

1. Assimilação de Dados

1.3 Intuição matemática

- Escrevendo esta combinação entre observação e previsão como uma combinação linear

$$x_a = \alpha y_o + (1 - \alpha)x_b$$

- Onde
 - x_a é a análise
 - x_b é a previsão
 - y_o é a observação
 - α é o peso atribuído à observação
 - $1 - \alpha$ é o peso atribuído à previsão
- Para que seja uma combinação linear, a soma dos pesos deve ser igual a 1 

- Como deve ser o peso α ?



1. Assimilação de Dados

1.3 Intuição matemática

- α é um parâmetro que relaciona as medidas das variâncias da observação e do modelo

$$\alpha = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2}$$

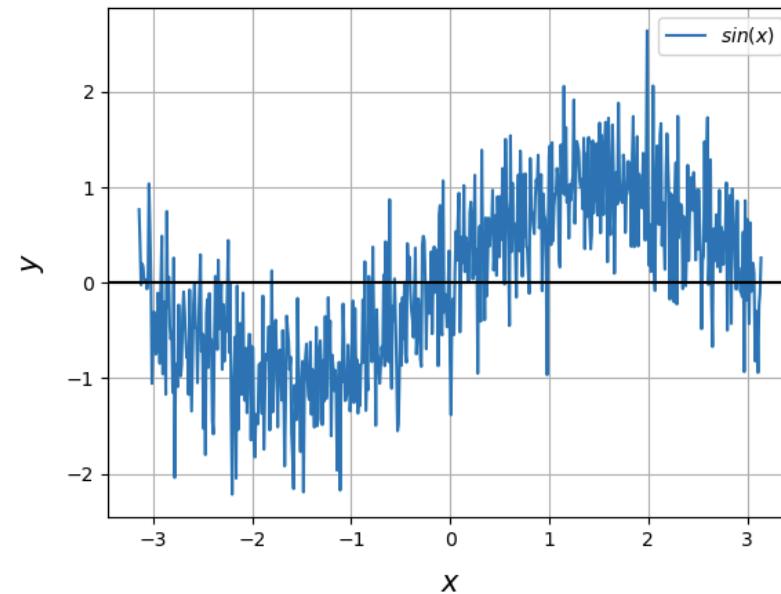
- Onde
 - σ_b^2 e σ_o^2 são as variâncias do background e das observações
- 👉 Portanto, podemos entender α como um parâmetro que representa a razão entre a variância do erro do modelo e a variância total do erro do sistema (modelo e observação)

1. Assimilação de Dados

1.3 Intuição matemática

- Considere um modelo matemático simples, a função seno com a adição de um ruído normalmente distribuído:

$$f(x) = \sin(x) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2), \quad -\pi \leq x \leq \pi$$

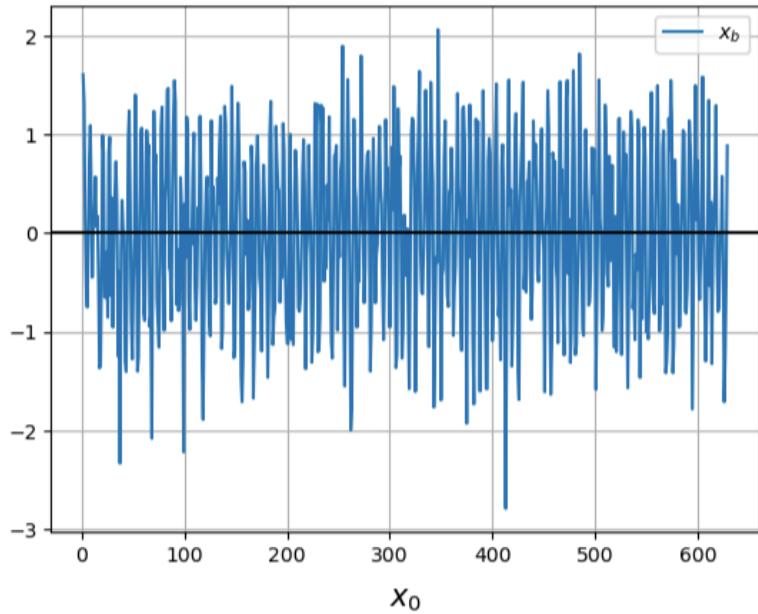


1. Assimilação de Dados

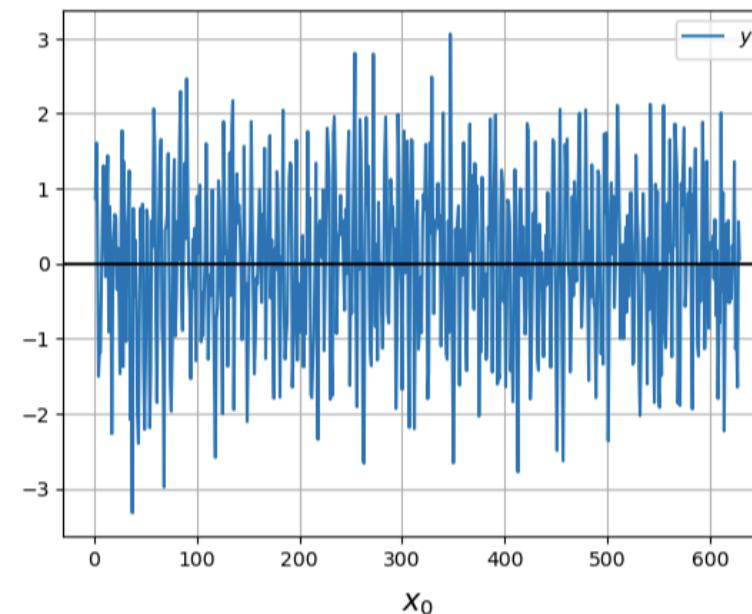
1.3 Intuição matemática

- Definimos um domínio inteiro onde aplicamos o modelo para extrair informações de uma "previsão" e "observações"

xb = função $\sin(x)$ aplicado ao domínio x_0



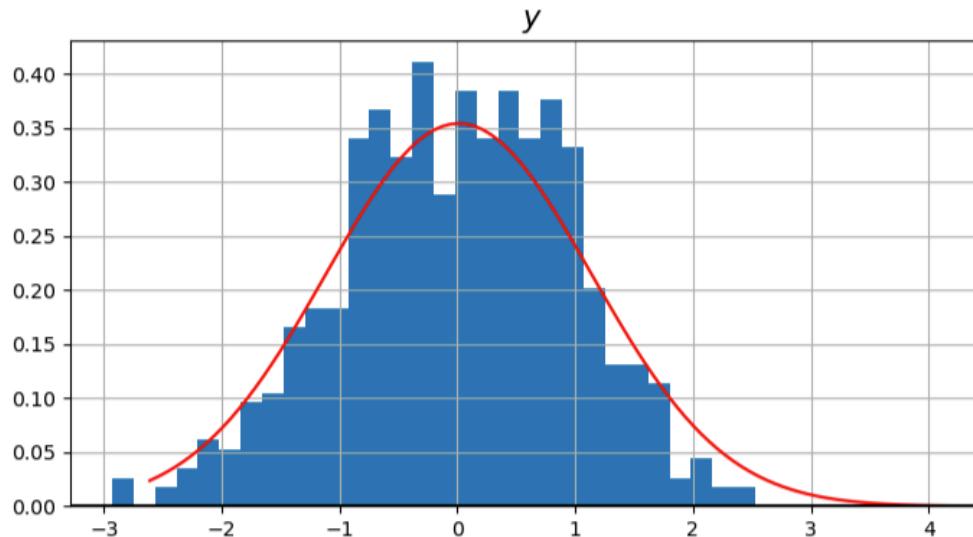
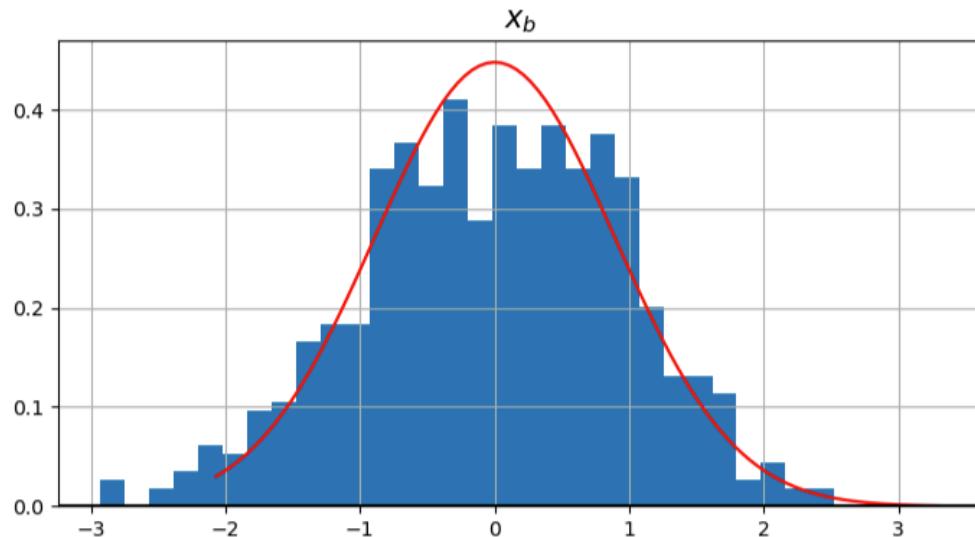
y = Observações (possuem a mesma natureza de xb)



1. Assimilação de Dados

1.3 Intuição matemática

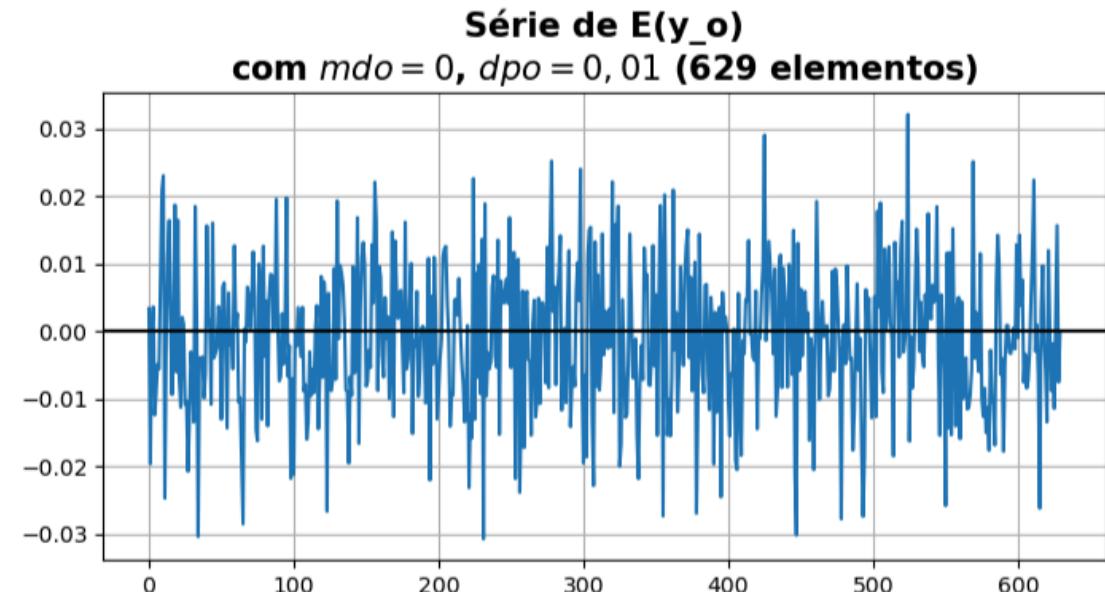
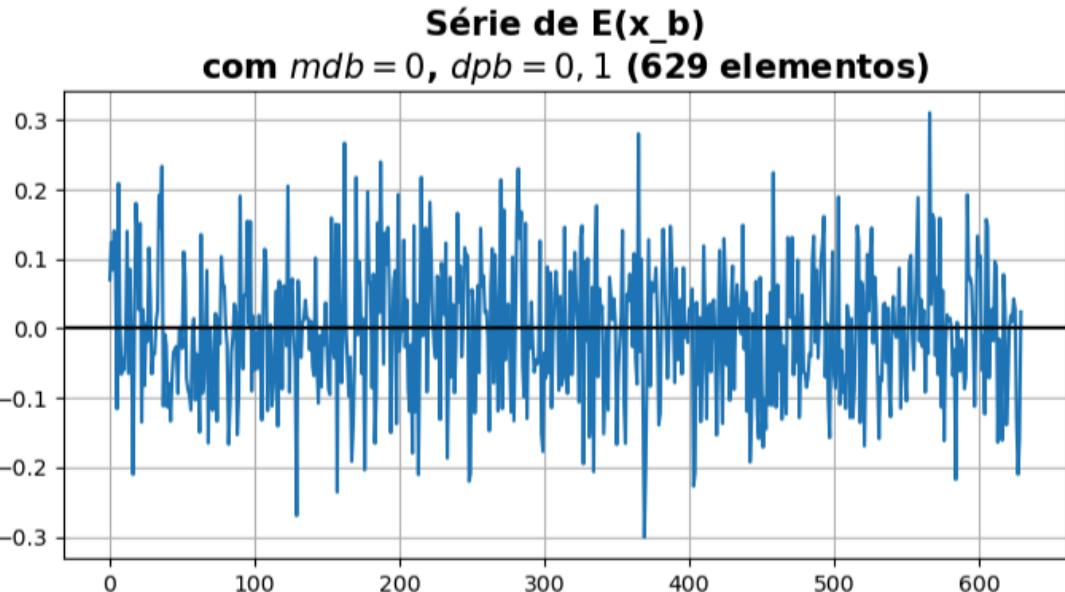
- x_b e y_o possuem distribuição normal, ambos são representados por valores aleatórios distribuídos sobre uma curva normal com $\mu_{x_b} = 0.0019$ e $\sigma_{x_b} = 0.8909$ e $\mu_{y_o} = -0.011$ e $\sigma_{y_o} = 0.8563$



1. Assimilação de Dados

1.3 Intuição matemática

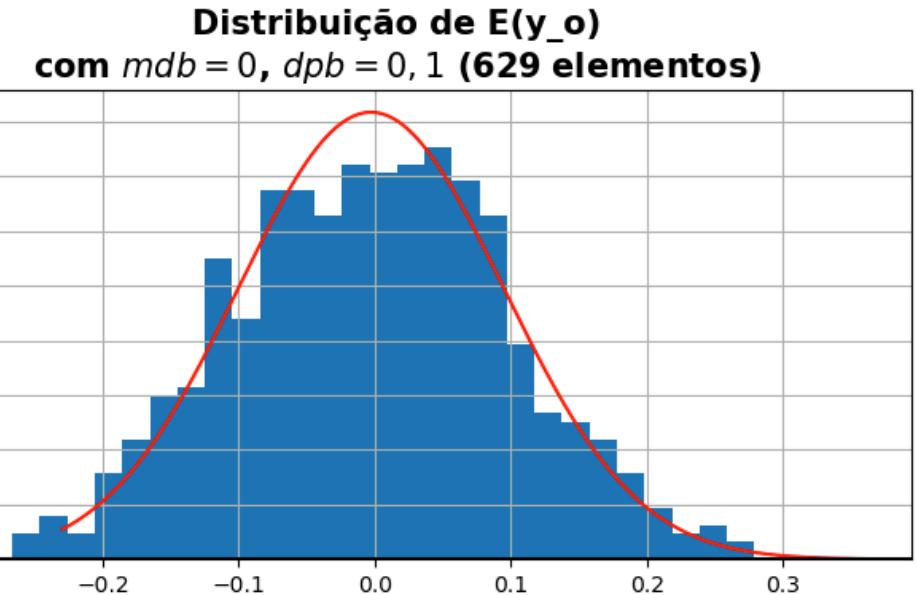
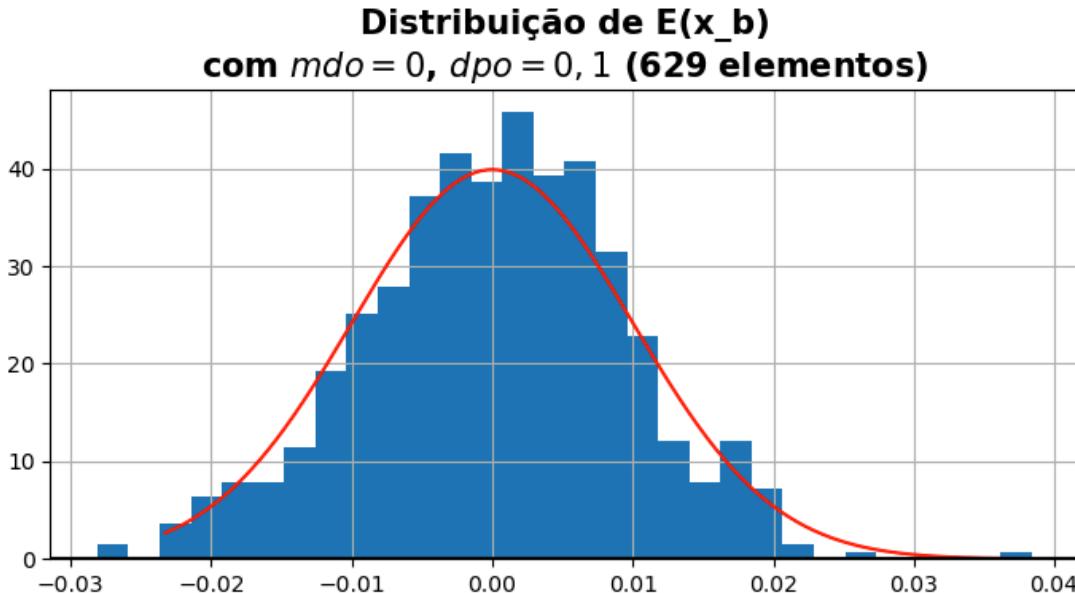
- Definimos as séries de erros das previsões e das observações



1. Assimilação de Dados

1.3 Intuição matemática

- Verificação da distribuição dos erros de "previsão" e "observação"



1. Assimilação de Dados

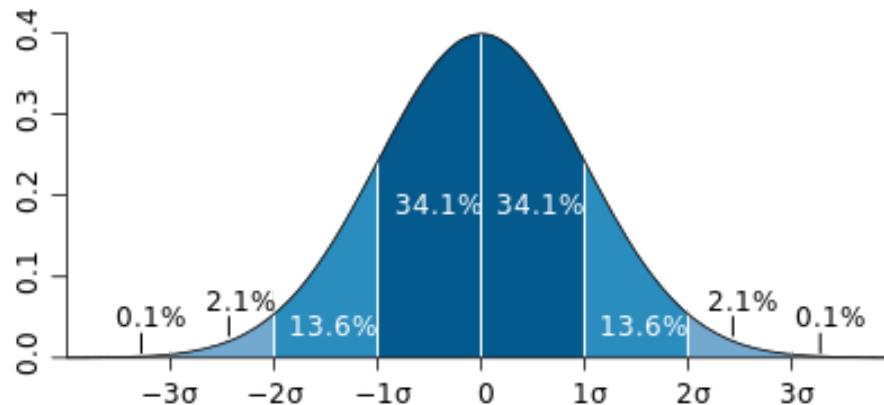
1.3 Intuição matemática

Por que a Distribuição Normal?

- Estamos mantendo as distribuições de x_b e y_o próximas à distribuição normal, porque esta distribuição possui as seguintes propriedades:

$$f(\psi) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\psi-\mu)^2}{2\sigma^2}}$$

- ~68% dos valores encontram-se a uma distância da média inferior a 1σ
- ~95% dos valores encontram-se a uma distância da média inferior a 2σ
- ~99,7% dos valores encontram-se a uma distância da média inferior a 3σ



1. Assimilação de Dados

1.3 Intuição matemática

- A partir destas informações, calculamos
 - A variância dos erros de previsão e observação

```
sigmab2 = np.var(errb)  
sigmao2 = np.var(erro)
```

```
sigmab2 = 0.0095226361060977  
sigmao2 = 0.00011333207595536619
```

- A variância dos erros de observação é muito menor do que a variância dos erros de background
- O valor de $\alpha = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2}$

```
alpha = sigmab2 / (sigmab2 + sigmao2)  
alpha = 0.9882386415340758
```

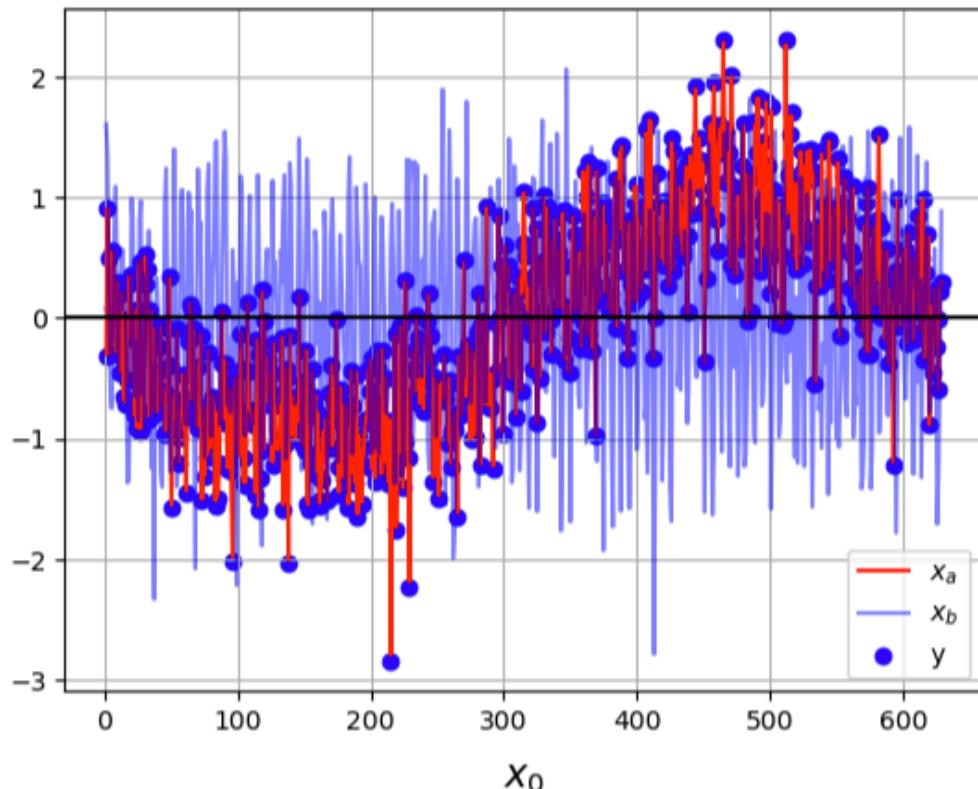
1. Assimilação de Dados

1.3 Intuição matemática

- O valor de $\alpha \approx 0.99$ indica que 99% do peso na combinação linear entre x_b e y_o é dado para as observações, enquanto que 1% do peso é dado para o background

$$x_a = \alpha y_o + (1 - \alpha)x_b$$

- Para quem tiver mais curiosidade
 - 🎲 Jupyter notebook com análise empírica univariada [Open in Colab](#)
 - 🎲 Jupyter notebook com análise empírica multivariada [Open in Colab](#)



1. Assimilação de Dados

1.3 Intuição matemática

- Em problemas reais, multivariados e multidimensionais, o peso α é representado por matrizes de covariâncias de erros que requerem modelagem e aproximações para a sua representação ➡ temos pouco controle ou influência sobre estes erros
- Interpolação Ótima

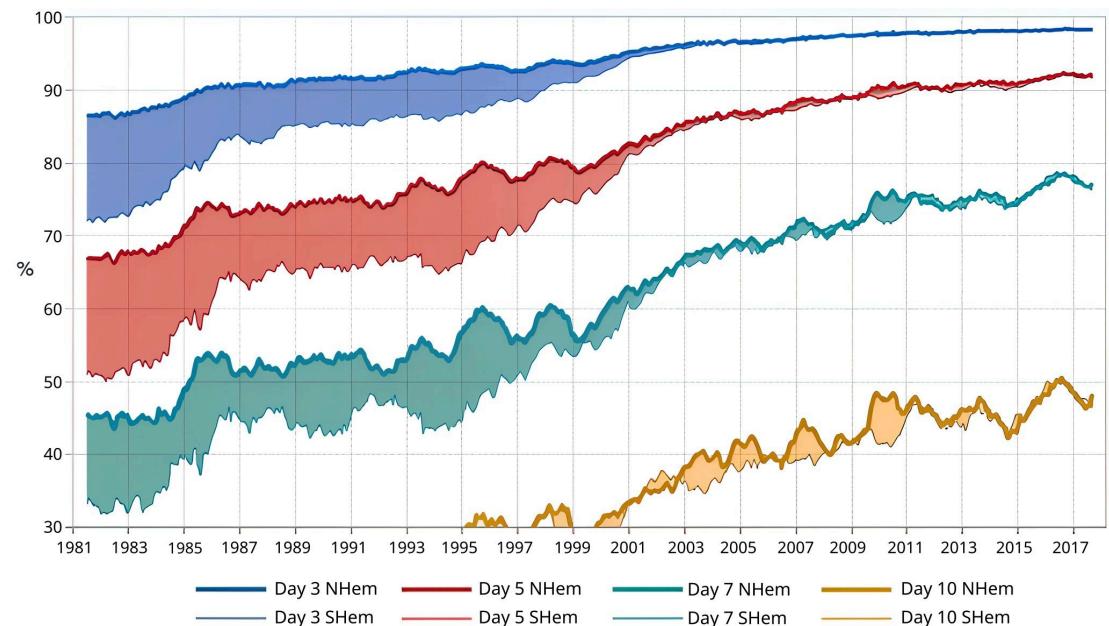
$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{W}[\mathbf{y} - H(\mathbf{x}_b)], \quad \mathbf{W} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}$$

- Onde
 - \mathbf{x}_a é o vetor de análise (estado estimado)
 - \mathbf{x}_b é o vetor de *background* ou *first guess*
 - \mathbf{y} é o vetor de observações
 - \mathbf{W} é a matriz de peso (ou ganho)
 - H é o operador observação não linear (transforma o espaço do modelo para o espaço físico das observações)

2. Determinação da Condição Inicial

2.1 Evolução do Skill da Assimilação de Dados

- Evolução do skill da previsão da altura geopotencial em 500 hPa
 - No início dos anos 1980, a previsão de 7 dias para o Hemisfério Norte, o skill da previsão não chegava a 50%, sendo inferior a 40% no Hemisfério Sul
 - Com o tempo, a diferença do skill entre os hemisférios diminuiu drasticamente, sendo muito próximos a partir dos anos 2000
 - Apenas a partir da metade dos anos 1990, a previsão de 10 dias começa a atingir algum skill (~30%)...
 - Atualmente, a previsão de 10 dias já alcança skill de 50% para ambos os hemisférios
- Embora a melhoria tenha sido importante, parece que o skill das previsões mais curtas está alcançando o seu limite - **por que?**



2. Determinação da Condição Inicial

2.2 Sistema de Modelagem Numérica e Assimilação de dados do CPTEC

- O SMNA é o sistema de assimilação de dados do CPTEC
 - Modelo global espectral BAM (Brazilian Atmospheric Model)
 - Framework de assimilação de dados GSI (Gridpoint Statistical Interpolation)
 - Fornece análise para o modelo BAM na resolução espacial TQ0299L064
 - ⚡ TQ0299 = truncamento espectral triangular de ordem 299, usando grade Gaussiana quadrática
 - ⚡ L064 = 64 níveis verticais em coordenadas híbridas sigma-pressão
- No CPTEC, a combinação entre modelo BAM e GSI tem sido aplicado desde 2012
 - Com atualizações na versão do modelo atmosférico (modelo de superfície, parametrização convectiva, coordenada vertical)
 - Com atualizações na versão do GSI (incluindo novos tipo de dados de observação, matriz de covariâncias entre outros)

2. Determinação da Condição Inicial

2.3 Gridpoint Statistical Interpolation

- O GSI é um framework de assimilação de dados desenvolvido pelo NCEP
 - Fornece a implementação de software para todas as componentes relacionadas à assimilação de dados
 - Métodos variacional (3D/4DVar, FGAT, híbrido-variacional e 3D/4DEnVar)
 - Métodos baseados em conjuntos (EnKF, EnSRF, LETKF)
 - Métodos de minimização da função custo variacional
 - Operador H (Modelo de Transferência Radiativa)
 - Suporte para modelos globais (espectrais) e regionais (em ponto de grade)
- Mantido pelo DTC/NCAR
 - Centraliza as contribuições, faz o gerenciamento do código, distribui releases e realiza tutoriais para a comunidade de usuários
 -  <https://ral.ucar.edu/solutions/products/gridpoint-statistical-interpolation-gsi>

2. Determinação da Condição Inicial

2.3 Gridpoint Statistical Interpolation

⚙️ 3DVar

- 3DVar é um método variacional de assimilação de dados que busca o melhor estado inicial da atmosfera (ou oceano) ajustando uma análise que minimiza a diferença entre as observações e o background, ponderada pelos seus erros (matrizes de covariâncias)
- **Função Custo**

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) + \frac{1}{2} [\mathbf{y}_o - H(\mathbf{x})]^T \mathbf{R}^{-1} [\mathbf{y}_o - H(\mathbf{x})]$$

- **Gradiente**

$$\nabla J(\mathbf{x}) = (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})(\mathbf{x} - \mathbf{x}_b) - (\mathbf{H}^T \mathbf{R}^{-1})[\mathbf{y}_o - H(\mathbf{x}_b)] = 0$$

- **Solução Analítica Exata**

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{W}[\mathbf{y}_o - H(\mathbf{x}_b)], \quad \mathbf{W} = \mathbf{B} \mathbf{H}^T (\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R})^{-1}$$

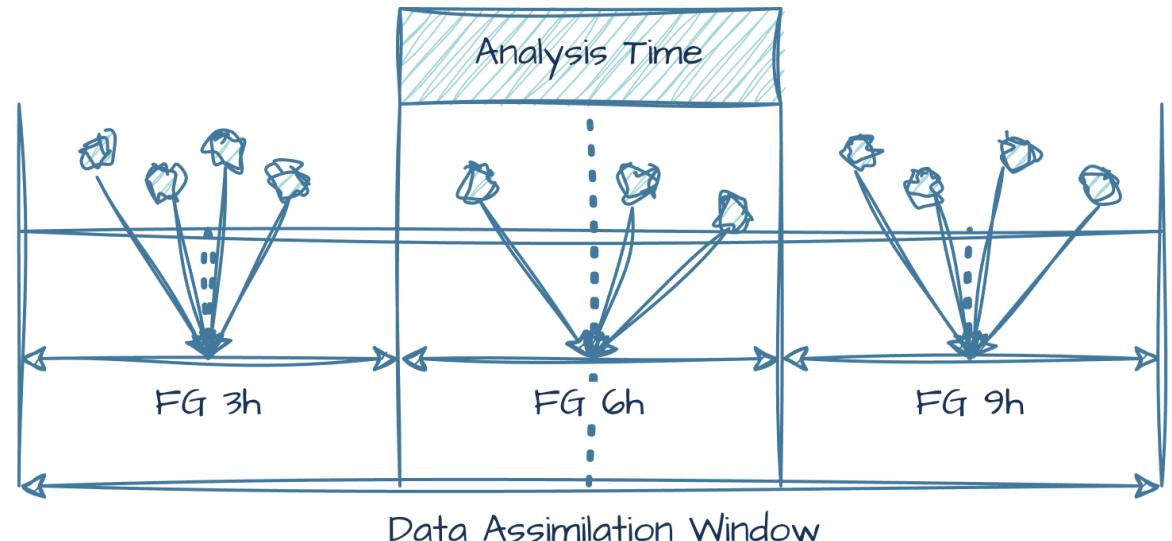
2. Determinação da Condição Inicial

2.3 Gridpoint Statistical Interpolation

⚙️ FGAT

- FGAT (First Guess at Appropriate Time) utiliza o background no horário da observação para melhorar a consistência temporal no 3DVar
 - Função custo continua 3D, pois não evolui a correção do background ao longo do tempo
 - Melhora a sincronia temporal das observações que não estão no tempo da análise (e.g., observações não convencionais)
 - Exige que o first guess seja particionado na janela de assimilação

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}_b(t_0))^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b(t_0)) + \frac{1}{2} \sum_i [\mathbf{y}_i - H_i(\mathbf{x}(t_i))]^T \mathbf{R}_i^{-1} [\mathbf{y}_i - H_i(\mathbf{x}(t_i))]$$



2. Determinação da Condição Inicial

2.3 Gridpoint Statistical Interpolation

⚙ Operador de Observação \mathbf{H}

- Responsável por levar o background do modelo até o espaço físico das observações
 - Se y_o e x_b forem quantidades equivalentes (e.g., temperaturas), então \mathbf{H} realiza apenas uma interpolação e o cálculo da inovação é feito $y_o - H(x_b)$
 - Se y_o for uma radiância, então \mathbf{H} precisa calcular um perfil de radiância a partir de x_b para calcular a inovação
 - Neste caso, \mathbf{H} é um modelo de transferência radiativa (no caso do SMNA, é o CRTM - Community Radiative Transfer Model)

2. Determinação da Condição Inicial

2.3 Gridpoint Statistical Interpolation

⚙️ Matriz B

- Fontes de incerteza do processo de modelagem são representadas por:
 - Modelo numérico (e.g., dinâmica e física)
 - Observações (e.g., medição, instrumento, grau de processamento)
 - Sistema de assimilação de dados (e.g., operadores de observação, modelos adjunto e tangente linear, tamanho do conjunto de um ensemble)
- A matriz de covariâncias dos erros de previsão (**B**), representa a covariância do "erro" (uma estimativa) do modelo
- Na assimilação de dados, estes erros são modelados em matrizes de covariâncias que tratam das relações espaço-temporais entre as quantidades observadas e diagnosticadas/prognosticadas
- Função custo 3DVar:

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}^b) + \frac{1}{2} [\mathbf{y}^o - \mathbf{H}(\mathbf{x})]^T \mathbf{R}^{-1} [\mathbf{y}^o - \mathbf{H}(\mathbf{x})]$$

2. Determinação da Condição Inicial

2.3 Gridpoint Statistical Interpolation

Matriz B

- **Método NMC (National Modeling Center)**
 - O método NMC preconiza que a correlação espacial dos erros do modelo são semelhantes à correlação espacial das diferenças entre as previsões de 48 e 24 horas
 - **Suposição:** crescimento linear dos erros de previsão durante as primeiras horas de previsão
 - Exemplo de par de previsões válido (modelo BAM)
 - 2013122418-2013122618 (previsão 48 horas)
 - 2013122518-2013122618 (previsão 24 horas)

2. Determinação da Condição Inicial

2.3 Gridpoint Statistical Interpolation

⚙️ Controle de Qualidade das Observações

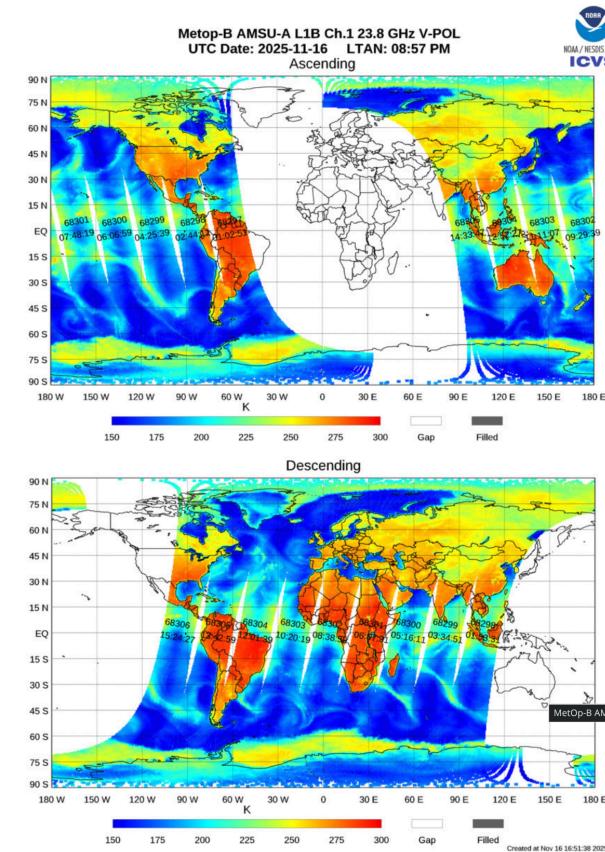
- Esta etapa ocorre antes e durante a assimilação de dados
 - Antes, envolve a preparação dos dados de observação (CPTEC está trabalhando neste assunto, i.e., preparar os seus próprios arquivos prepBUFR para dados convencionais e de radiâncias)
 - Durante a assimilação, o GSI realiza vários testes e checagens
 - Em geral, há um tipo de controle de qualidade para cada tipo de observação
- 💡 **Pré-controle de qualidade:** para as radiossondas, verifica a altitude, pressão, temperatura, umidade e descarta dados duplicados; para as radiâncias, aplica as flags de uso dos provedores dos dados, verifica a cobertura de nuvens e o ângulo de visada
 - 💡 **OMF:** cálculo das inovações $\mathbf{y}_o - H(\mathbf{x}_b)$ e compara valores com as variâncias de **B** e **R**
 - 💡 **Buddy Check:** compara as observações com a sua vizinhança (pode rejeitar observações ou reduzir o peso delas)
 - 💡 **Controle de Qualidade Adaptativo/Variacional:** ajusta o peso das observações e não as descarta imediatamente

2. Determinação da Condição Inicial

2.3 Gridpoint Statistical Interpolation

Observações Assimiladas pelo SMNA

- Atualmente, são assimiladas os seguintes conjuntos de dados de observações no SMNA do CPTEC
 - Observações convencionais u , v , t , q e ps
 - Observações não convencionais dos seguintes satélites/instrumentos
 - AMSUA/METOP-B (micro-ondas, infere t)
 - SATWND (infere u e v)
 - GPSRO (refração do sinal GPS, ângulo de curvatura, infere t , p e q)

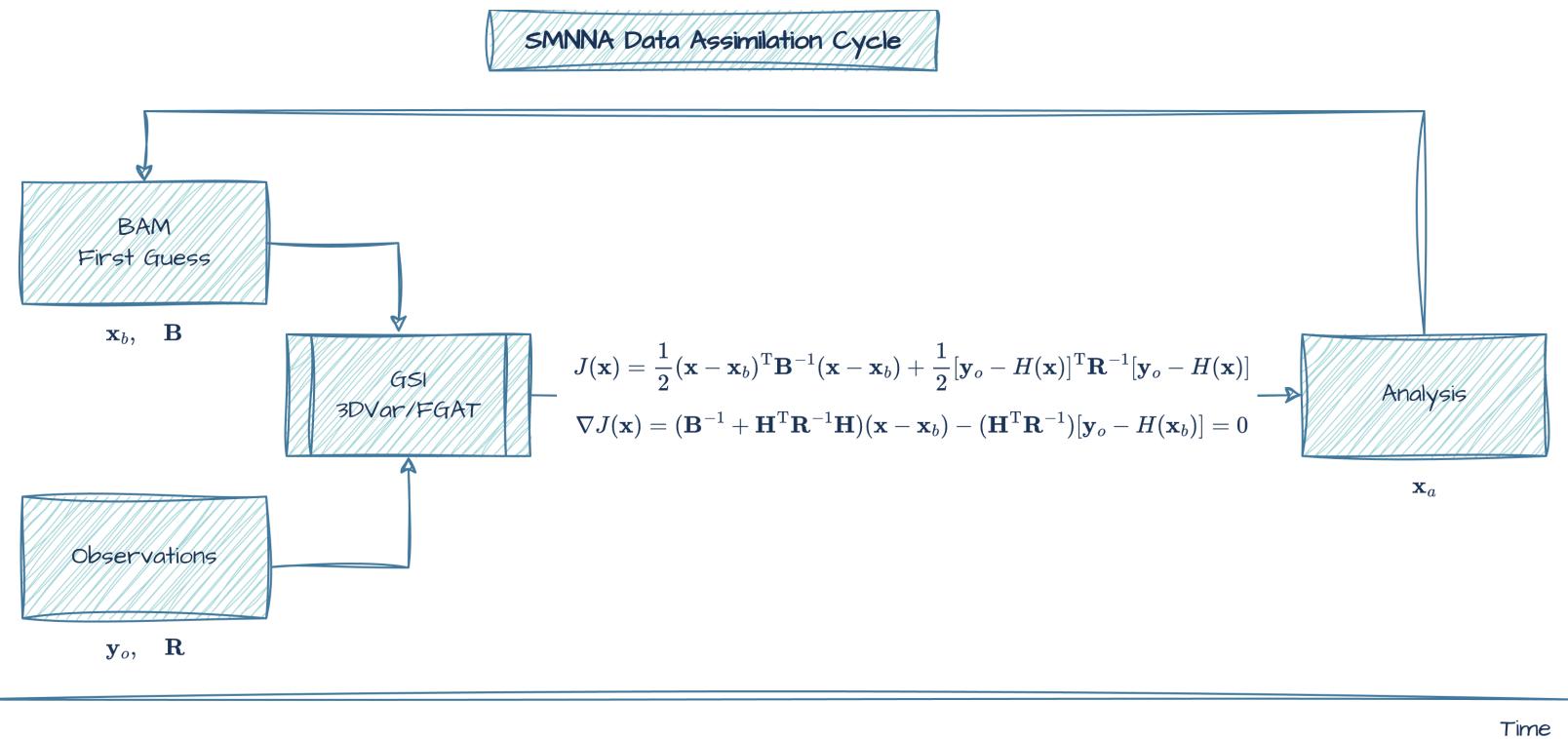


Fonte: https://www.star.nesdis.noaa.gov/icvs/status_MetOPB_AMSUA.php

2. Determinação da Condição Inicial

2.3 Gridpoint Statistical Interpolation

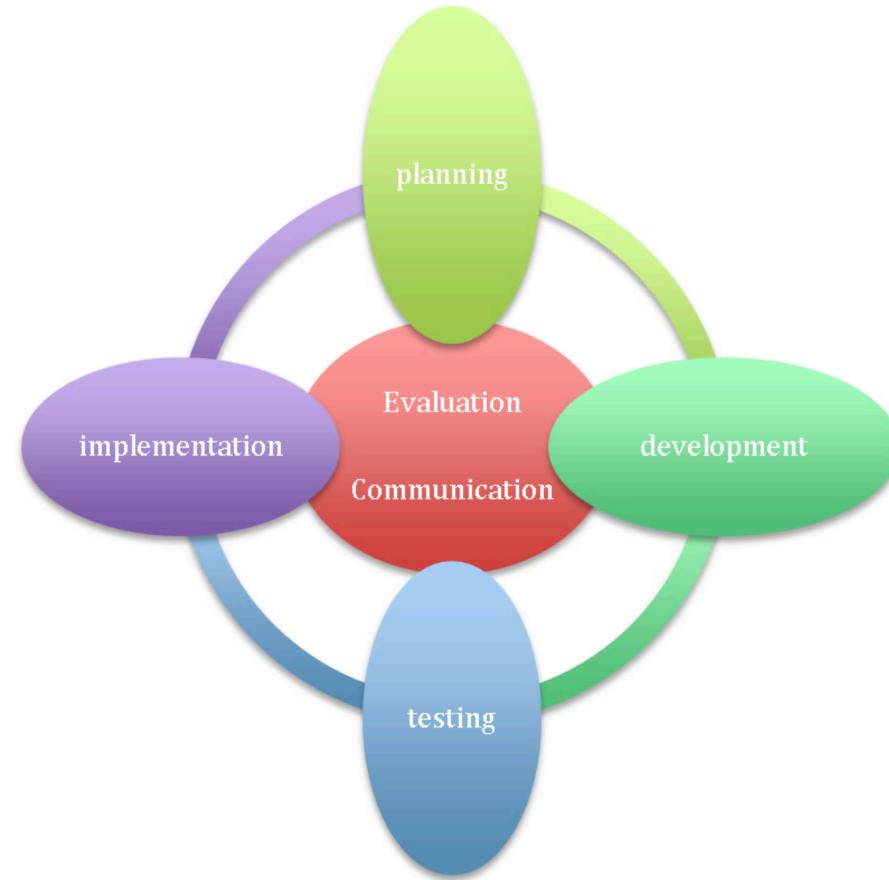
⚙ Ciclo de Assimilação de Dados



3. R2O - Research to Operations

3.1 O que é e por que é necessário R2O?

- 👉 Etapa crucial de manutenção e desenvolvimento da suíte de assimilação de dados operacional
 - É quando a pesquisa realizada é passada junto com os novos desenvolvimentos para a situação operacional
 - Consideram-se também os artefatos computacionais produzidos como apoio à pesquisa, os quais podem ser utilizados como ferramentas de diagnóstico operacionais
- 🧙 Desafios
 - Limitações técnicas frequentemente associadas à capacidade computacional (processamento e armazenamento)
 - Validação adequada, superar a versão anterior e comparar com os demais produtos



Fonte: <https://www.ecmwf.int/sites/default/files/elibrary/2017/17549-ecmwf-research-operations-r2o-process.pdf>

3. R2O - Research to Operations

3.2 Ferramentas de Apoio

- Grupo de Assimilação de Dados possui uma organização no GitHub
 - Permite a organização da equipe e dos desenvolvimentos
 -  Roadmaps
 -  Tags e releases para distribuição
 -  Registro de issues
 -  Discussões
 -  Wikis etc
 -  <https://github.com/GAD-DIMNT-CPTEC>

J(x) Grupo de Assimilação de Dados (GAD)
Grupo de Assimilação de Dados da CGCT/INPE
13 followers Brazil joao.gerd@inpe.br

README.md

Grupo de Assimilação de Dados (GAD)

Este repositório hospeda os projetos desenvolvidos e mantidos pelo Grupo de Assimilação de Dados (GAD) na [Divisão de Modelagem Numérica do Sistema Terrestre \(DIMNT\)](#) do [Centro de Previsão de Tempo e Estudos Climáticos \(CPTEC\)](#).

Repositórios Legados

Há um grande número de páginas Wikis escritas e repositórios que podem ser migrados do SVN para o GitHub, como backup e arquivamento. Nesse sentido, as páginas podem ser exportadas em PDF e consolidadas em algum local por aqui. Essa é uma tarefa que está sendo feita aos poucos.

Organização de Times

O GitHub permite a organização de times de desenvolvimento que podem ser destacados para trabalhar em repositórios específicos, isso permite o controle de acesso dos colaboradores (que também podem ser externos à organização do GAD).

Outros Recursos

O Python tem se mostrado como uma ferramenta poderosa para o grupo, seja na estruturação dos dados com os quais lidamos, seja na sua visualização. Exemplos do que podemos fazer a partir do GitHub:

- Manual de uso utilizando a linguagem Markdown e o MkDocs: <https://gad-dimnt-cptec.github.io/GSIBerror/>
- Visualização direta de Notebooks do Jupyter: https://github.com/GAD-DIMNT-CPTEC/GSIBerror/blob/main/notebooks/read_gsib_error_python-class-final-BCPTEC_hyb_coord.ipynb
- Utilização de Notebooks de um repositório diretamente com o Binder: <https://mybinder.org/v2/gh/cbastarz/GSIBerror/main>
- Inclusão de arquivos binários em repositórios com o Git Large File Storage (LFS): <https://github.com/GAD-DIMNT-CPTEC/GSIBerror/tree/main/data>

Alguns Guias Básicos

Instruções rápidas de uso do git e da documentação através do mkdocs, podem ser encontradas em:
<https://github.com/CGCT-CPTEC-DIMNT/cgct-cptec-dimnt/wiki/Guia-B%C3%A1sico-de-Uso-do-Git>

Links úteis

Monitoramento Assimilação de Dados

- <https://emc.ncep.noaa.gov/users/verification/global/gfs/ops/>
- https://gmao.gsfc.nasa.gov/forecasts/radmon_coverage.php
- <https://www.emc.ncep.noaa.gov/gmb/gdas/index.html>
- https://gmao.gsfc.nasa.gov/forecasts/systems/fp/obstat_time_series/index.html
- <https://space.oscar.wmo.int/>
- <https://skylab.csda.org/>
- <https://hpfx.collab.science.gc.ca/~smco500/psmon/monitoring/>

Dados de observação operacionais e arquivos

- <https://nomads.ncep.noaa.gov/pub/data/ncfc/com/gfs/prod/>
- <https://rda.ucar.edu/datasets/ds337.0/dataaccess/>
- <https://ftp.ncep.noaa.gov/data/ncfc/com/obsproc/prod/>
- <https://nomads.ncep.noaa.gov/>
- <https://nomads.ncep.noaa.gov/pub/data/ncfc/com/>
- <https://nomads.ncep.noaa.gov/pub/data/ncfc/com/gfs/prod/>
- https://dataserver.cptec.inpe.br/dataserver_modelos/smna;brutos/

Egeon

- /oper/dados/dboper/raw/arch/mod/ncep/gdas/

Changelogs de outros centros

- https://www.nco.ncep.noaa.gov/pmb/changes/gfs_upgrade.php
- <https://www.nco.ncep.noaa.gov/pmb/changes/>

3. R2O - Research to Operations

3.2 Ferramentas de Apoio

Alguns exemplos de ferramentas de apoio

- **readDiag**
 - Ferramenta para o diagnóstico da assimilação das radiâncias no GSI (Python)
- **GSIError**
 - Ferramenta para o diagnóstico da matriz de covariâncias dos erros de previsão do GSI (Python)
- **pyBAM**
 - Ferramenta para leitura dos campos de previsão do BAM (recompõe para o espaço físico os coeficientes espectrais, Python)
- **SCANTEC**
 - Sistema Comunitário de Avaliação de modelos Numéricos de Tempo E Clima (Fortran)
- **SCANPLOT**
 - Sistema de plotagem para o SCANTEC (Python)
- **SMNAMonitoringApp**
 - Ferramenta para o monitoramento das simulações operacionais do SMNA (em construção, Python)
- **Impacto das Observações e Observing System Experiments**
 - Ferramentas de diagnóstico para estudar o impacto e a contribuição dos diferentes tipos de observações na análise



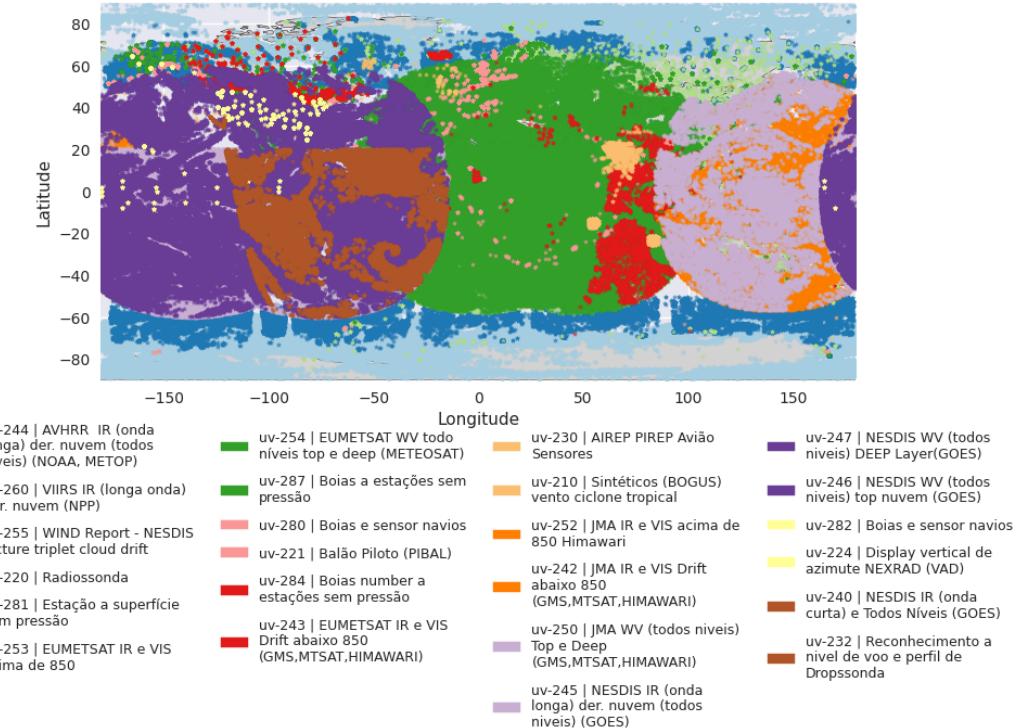
3. R2O - Research to Operations

3.2 Ferramentas de Apoio



Alguns exemplos de ferramentas de apoio

- **readDiag**
 - 🔐 Ferramenta para o diagnóstico da assimilação das radiâncias no GSI (Python)
 - 🌐 <https://gad-dimnt-cptec.github.io/readDiag>
 - 🐙 <https://github.com/GAD-DIMNT-CPTEC/readDiag>
 - 📝 Notebooks do Jupyter disponíveis para testar
 - 🌐 Disponível no PyPi
 - 📄 Documentação em Português e Inglês



3. R2O - Research to Operations

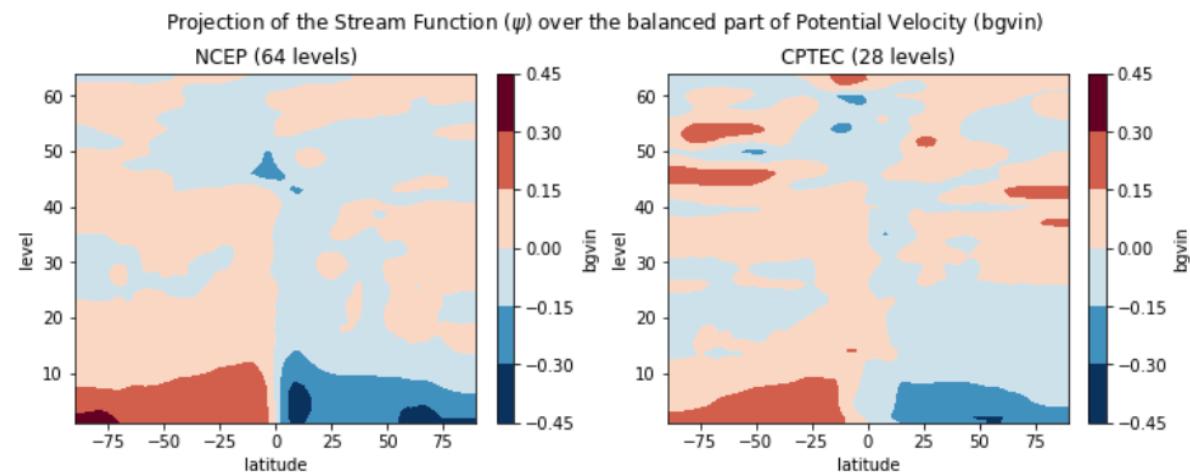
3.2 Ferramentas de Apoio



Alguns exemplos de ferramentas de apoio

- **GSIBerror**

- 🔧 Ferramenta para o diagnóstico da matriz de covariâncias dos erros de previsão do GSI (Python)
- 🌐 <https://gad-dimnt-cptec.github.io/GSIBerror>
- 🐙 <https://github.com/GAD-DIMNT-CPTEC/GSIBerror>
- 📝 Notebooks do Jupyter disponíveis para testar
- 🐍 Disponível no PyPi
- 📄 Documentação em Português e Inglês



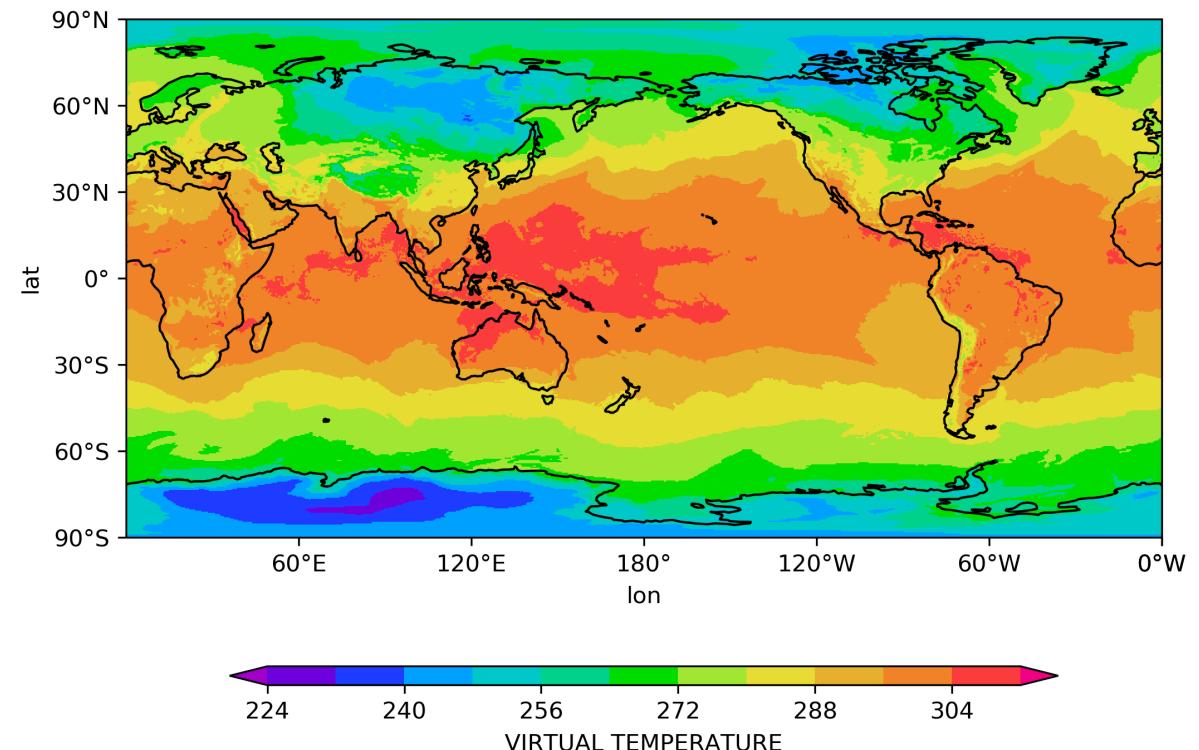
3. R2O - Research to Operations

3.2 Ferramentas de Apoio



Alguns exemplos de ferramentas de apoio

- **pyBAM**
 - Ferramenta para leitura dos campos de previsão do BAM (Python)
 - [🔗 https://gad-dimnt-cptec.github.io/pyBAM](https://gad-dimnt-cptec.github.io/pyBAM)
 - [🔗 https://github.com/GAD-DIMNT-CPTEC/pyBAM](https://github.com/GAD-DIMNT-CPTEC/pyBAM)
 - Documentação apenas em Português



3. R2O - Research to Operations

3.2 Ferramentas de Apoio

Alguns exemplos de ferramentas de apoio

- **SCANTEC**

- 🔐 Sistema Comunitário de Avaliação de modelos Numéricos de Tempo E Clima (Fortran)
- 🌐 <https://gad-dimnt-cptec.github.io/SCANTEC>
- 💻 <https://github.com/GAD-DIMNT-CPTEC/SCANTEC>
- 📄 Documentação apenas em Português



3. R2O - Research to Operations

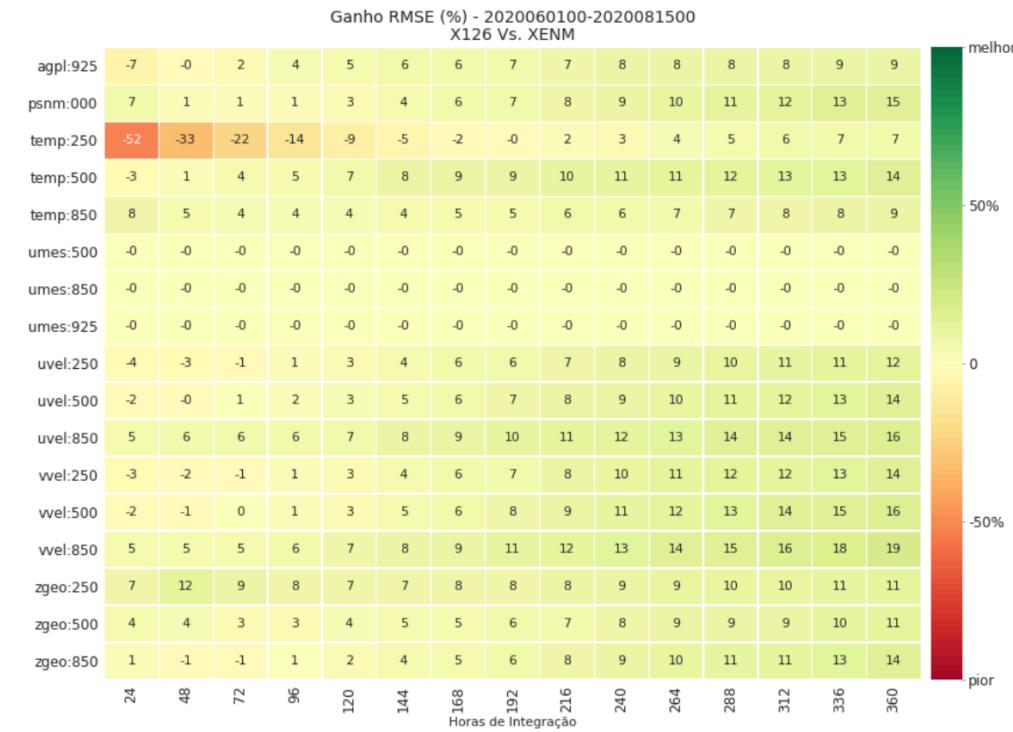
3.2 Ferramentas de Apoio



Alguns exemplos de ferramentas de apoio

- **SCANPLOT**

- 🔧 Sistema de plotagem para o SCANTEC (Python)
- 🔗 <https://gad-dimnt-cptec.github.io/SCANPLOT>
- 🐒 <https://github.com/GAD-DIMNT-CPTEC/SCANPLOT>
- 📝 Notebook do Jupyter disponíveis para testar a CLI
- 🛠 GUI de demonstração disponível em <https://huggingface.co/spaces/cfbastarz/SCANPLOT>
- 📄 Documentação apenas em Português



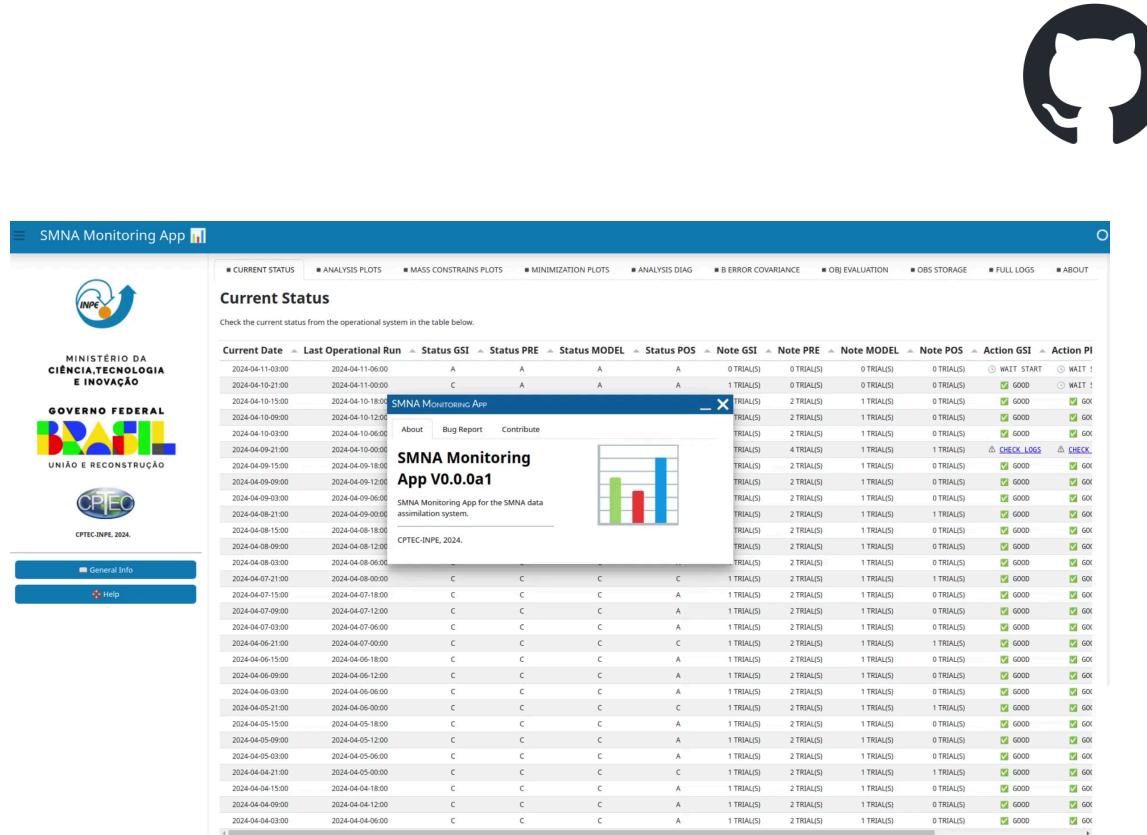
3. R2O - Research to Operations

3.2 Ferramentas de Apoio

Alguns exemplos de ferramentas de apoio

- **SMNAMonitoringApp**

- Ferramenta para o monitoramento das simulações operacionais do SMNA (em construção, Python)
- <https://gad-dimnt-cptec.github.io/SMNAMonitoringApp>
- <https://github.com/GAD-DIMNT-CPTEC/SMNAMonitoringApp>
- Demonstração disponível em <https://huggingface.co/spaces/cfbastarz/SMNAMonitoringApp>
- Documentação apenas em Português

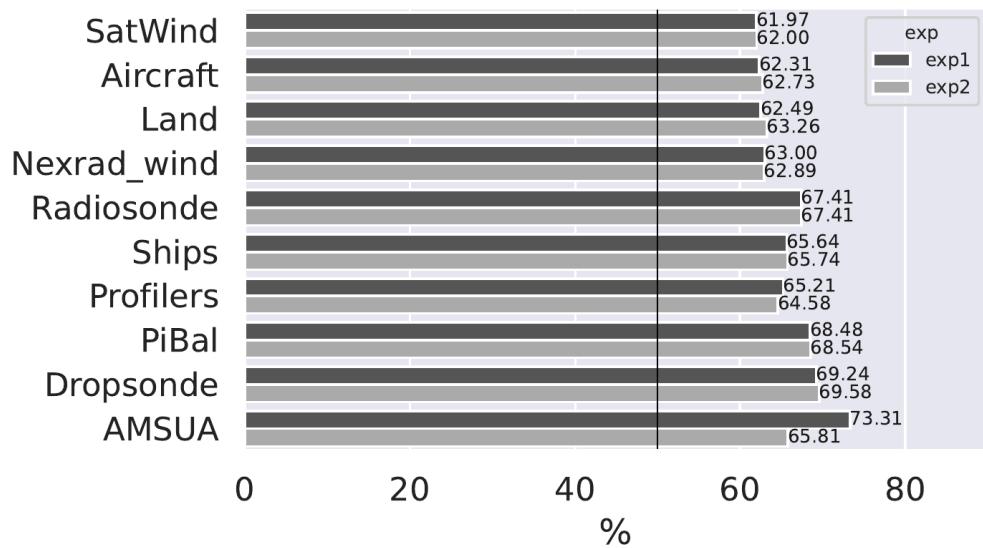


3. R2O - Research to Operations

3.2 Ferramentas de Apoio

Alguns exemplos de ferramentas de apoio

- Impacto fracionário das observações (Viana e de Mattos, 2024)



Article

Assessing the Impact of Observations on the Brazilian Global Atmospheric Model (BAM) Using Gridpoint Statistical Interpolation (GSI) System

Liviany Pereira Viana * and João Gerd Zell de Mattos 

National Institute for Space Research, Cachoeira Paulista, São Paulo 12630-000, Brazil; joao.gerd@inpe.br
* Correspondence: liviany.viana@inpe.br

Abstract: This article describes the main features of the impacts of global observations on the reduction of errors in the data assimilation (DA) cycle carried out in the Brazilian Global Atmospheric Model (BAM) at Center for Weather Forecast and Climate Studies [Centro de Previsão de Tempo e Estudos Climáticos (CPTEC)] at the Brazilian National Institute for Space Research [Instituto Nacional de Pesquisas Espaciais (INPE)]. These results show the importance of studying and evaluating the contribution of each observation to the DA system, therefore, two experiments (exp1/exp2) were performed with different configurations of the BAM model, with exp2 presenting the best fit between the Gridpoint Statistical Interpolation (GSI) and BAM systems. The BAM model was validated by the statistical metrics of root mean-square error and correlation anomaly, but this validation is not explored in this paper. A metric was applied that does not depend on the adjoint-based method, but only on the residuals that are made available in the GSI system for the observation space, given by the total impact, the fractional impact and the fractional beneficial impact. In general, the average daily showed that the observations of the global system that contribute most to the reduction of errors in the DA cycle are from the pilot balloon data ($-3.54/-3.45 \text{ J kg}^{-1}$) and the profilers ($-2.13/-1.97 \text{ J kg}^{-1}$), and the smallest contributions came from the land ($-0.28/-0.29 \text{ J kg}^{-1}$) and sea ($-0.44/-0.44 \text{ J kg}^{-1}$) surfaces. The same pattern was observed for the synoptic times presented. However, when verifying the fraction of the impact by each type of observation, it was found that the radiance data (64.88/30.30%), followed by radiosondes (14.85/27.42%) and satellite winds (11.03/22.70%), are the most important fractions for both experiments. These results show that the DA system is working to generate the best analyses at the research center and that the deficiencies found in some observations can be adjusted to improve the development of the GSI and the BAM model, since together, the entire database used is evaluated, as well as the forecast model itself, indicating the relationship between the assertiveness of the atmospheric model and the DA system used at the research center.



Citation: Viana, L.P.; de Mattos, J.G.Z. Assessing the Impact of Observations on the Brazilian Global Atmospheric Model (BAM) Using Gridpoint Statistical Interpolation (GSI) System. *Meteorology* **2024**, *3*, 447–463. <https://doi.org/10.3390/meteorology3040021>

Academic Editor: Paul D. Williams

Received: 1 July 2024

Revised: 3 September 2024

Accepted: 9 September 2024

Published: 16 December 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

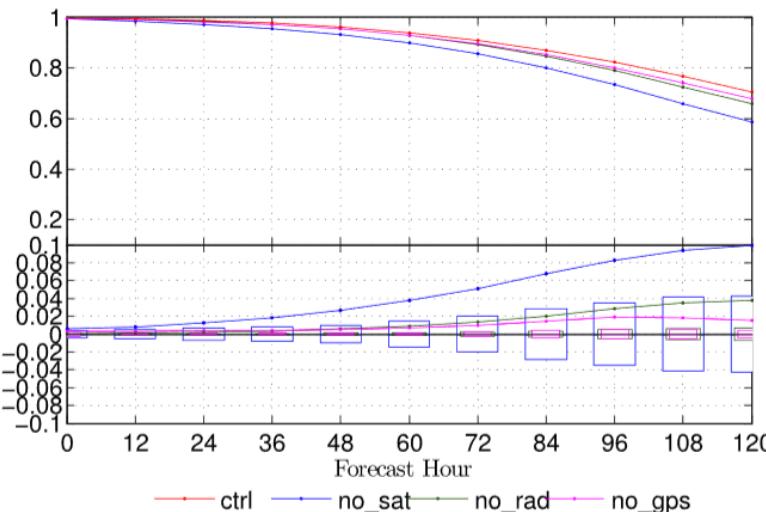
Initial conditions (ICs) are important components of numerical weather prediction (NWP) and represent the balance between collected data and numerical model estimates, these values are obtained through the data assimilation (DA) process. Observational data, together with short-term forecasts, can provide considerations and diagnostics for the initial calculation of the “true” state of the atmosphere and thus can be used as ICs in NWP models. In this way, the quality of these analyses can reflect the particularity of the forecast [1]. Furthermore, the forecasting ability of the model is significantly favored by the available observations [2]. In recent years, NWP has improved its suitability due to a number of factors, such as computational efficiency, good representation of smaller-scale

3. R2O - Research to Operations

3.2 Ferramentas de Apoio

Alguns exemplos de ferramentas de apoio

- **Observing System Experiments (de Azevedo et al., 2016)**



Observing System Experiments in a 3DVAR Data Assimilation System at CPTEC/INPE

HELENA BARBIERI DE AZEVEDO, LUIS GUSTAVO GONÇALVES DE GONÇALVES,
CARLOS FREDERICO BASTARZ, AND BRUNA BARBOSA SILVEIRA

Instituto Nacional de Pesquisas Espaciais, Cachoeira Paulista, São Paulo, Brazil

(Manuscript received 1 December 2015, in final form 15 December 2016)

ABSTRACT

The Center for Weather Forecast and Climate Studies [Centro de Previsão e Tempo e Estudos Climáticos (CPTEC)] at the Brazilian National Institute for Space Research [Instituto Nacional de Pesquisas Espaciais (INPE)] has recently operationally implemented a three-dimensional variational data assimilation (3DVAR) scheme based on the Gridpoint Statistical Interpolation analysis system (GSI). Implementation of the GSI system within the atmospheric global circulation model from CPTEC/INPE (AGCM-CPTEC/INPE) is hereafter referred to as the Global 3DVAR (G3DVAR) system. The results of an observing system experiment (OSE) measuring the impacts of radiosonde, satellite radiance, and GPS radio occultation (RO) data on the new G3DVAR system are presented here. The observational impact of each of these platforms was evaluated by measuring the degradation of the geopotential height anomaly correlation and the amplification of the RMSE of the wind. Losing the radiosonde, GPS RO, and satellite radiance data in the OSE resulted in negative impacts on the geopotential height anomaly correlations globally. Nevertheless, the strongest impacts were found over the Southern Hemisphere and South America when satellite radiance data were withheld from the data assimilation system.

1. Introduction

The Center for Weather Forecast and Climate Studies [Centro de Previsão de Tempo e Estudos Climáticos (CPTEC)] at the Brazilian National Institute for Space Research [Instituto Nacional de Pesquisas Espaciais (INPE)] recently implemented the Gridpoint Statistical Interpolation analysis system (GSI; Wu et al. 2002; Kleist et al. 2009) [with a three-dimensional variational data assimilation (3DVAR) approach] in the CPTEC/INPE atmospheric global circulation model (AGCM-CPTEC/INPE). This implementation of the GSI system, known as the Global 3DVAR (G3DVAR) system, has been operational since January 2013 and initializes AGCM-CPTEC/INPE forecasts on a global grid every 6 h. This implementation of the GSI system has replaced the Physical-space Statistical Analysis System (PSAS; Cohn et al. 1998), which was previously used to initialize the AGCM-CPTEC/INPE. The transition to the GSI system has increased the maximum number of observations we can

assimilate into our model and has provided the ability to assimilate satellite radiance data.

Since numerical weather prediction (NWP) is an initial value problem, the data assimilation process used to initialize forecasting models can have a significant impact on the quality of forecasts. Data assimilation is the process of combining observed data with short-range forecasts, therein considering the errors in the observations and errors associated with the numerical model, to generate an optimal estimate of the current state of the atmosphere (Talagrand 1997; Tsuyuki and Miyoshi 2007; Herdies et al. 2008). The information in the observing systems (i.e., the quantity and quality of the observations) plays a key role in the data assimilation process; it impacts the resulting analysis and consequently affects the quality of the forecasts. The resulting forecasts should benefit from a careful evaluation of how the different observing systems impact the NWP system since the inclusion of certain observations may degrade the forecasts. Furthermore, knowledge of which datasets provide better estimates of weather conditions can be used to optimize data assimilation systems by improving the process of selecting observations that contribute positively to the analysis.

Corresponding author e-mail: Helena de Azevedo, helenabdeazevedo@gmail.com

DOI: 10.1175/WAF-D-15-0168.1

© 2017 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](http://www.ametsoc.org/PUBSReuseLicenses) (www.ametsoc.org/PUBSReuseLicenses).

Unauthenticated | Downloaded 11/17/25 01:37 AM UTC

3. R2O - Research to Operations

3.3 Fluxo de Transição

-  Gerenciamento do projeto
 -  Organização do roadmap
 -  O que se deseja para cada versão?
 -  Como chegamos lá?
-  Desenvolvimento e Controle de versões
 -  Todas as alterações são registradas no repositório (para o SMNA, este repositório é interno)
 -  Operação reporta as issues que encontrar
 -  A equipe de desenvolvimento e pesquisa investiga os problemas e propõe alterações no código
 -  Novos artefatos são gerados e utilizados pela operação
 -  Importante que a operação faça o registro das ocorrências
 -  Histórico e avaliação da estabilidade do sistema
-  Comunicação da Ciência
 -  Por meio de relatórios e notas técnicas
 -  Artigos científicos

4. Atividades Operacionais

4.1 Custo Operacional

- \$ Custo de armazenamento de 1 ciclo do SMNA (BAM+GSI)
 - Análise espectral GSI TQ0299L064: ~89MB
 - Background espectral (3x FGAT): ~2GB
 - Arquivos de diagnóstico GSI: ~1.5GB
 - Arquivos de observações (AMSUA, GNSS RO, Convencionais): ~150MB
 - Outros artefatos que fazem parte do processo (GSI): ~3GB
 - Previsões pós-processadas para 11 dias: ~13GB
- Total: ~20GB
 - ➡ 4 ciclos por dia: ~80GB
 - ➡ 1 mês: ~10TB
 - ➡ 1 ano: ~115TB

4. Atividades Operacionais

4.2 Monitoramento

-  Uma vez que o sistema de assimilação de dados encontra-se em ambiente operacional, faz-se necessário
 -  Acompanhar as simulações diárias do sistema, em termos de desempenho computacional e qualidade das análises e previsões (avaliação objetiva)
 -  Acompanhar junto a outros centros a situação operacional dos sensores a bordo dos diversos satélites
 -  Acompanhar junto a outros centros a disseminação dos dados de observações utilizados (principalmente no caso do CPTEC, que ainda não gera os seus próprios dados de observação)
 -  WMO Events and Meetings 

[Home](#) / [News Portal](#) / [Events and Meetings](#)

Events and Meetings

Calendar

[Clear filters](#)

Type of event
Topic

Year
 Events for WMO Community

UPCOMING EVENTS
PAST EVENTS
ALL EVENTS

 MEETING

 24 September 2025 - 18 December 2025

IG3IS Webinar Series (Q3-Q4, 2025)

 IN FOCUS

 09 November 2025 - 21 November 2025

 Belém, Pará, Brazil

WMO at COP30

 WORKSHOP

 16 November 2025 - 19 November 2025

Regional Instruments Centres (RICs) Workshop

 MEETING

 17 November 2025 - 19 November 2025

 Geneva, Switzerland

45th Audit and Oversight Committee meeting

 MEETING

 17 November 2025 - 19 November 2025

 Geneva, Switzerland

45th Audit and Oversight Committee meeting

 COP EVENT - SCIENCE FOR CLIMATE ACTION PAVILION

 17 November 2025, 14:00 - 16:00

Empowering Women, Children, and Indigenous Peoples through Inclusive Early Warning Systems and Climate Services

 WORKSHOP

 18 November 2025 - 25 November 2025

Virtual Workshop: Gender mainstreaming across hydrometeorological services

 18 November 2025, 10:00 - 11:30

From Warning to Action: Safeguarding Agricultural Livelihoods Before Disasters Strike

4. Atividades Operacionais

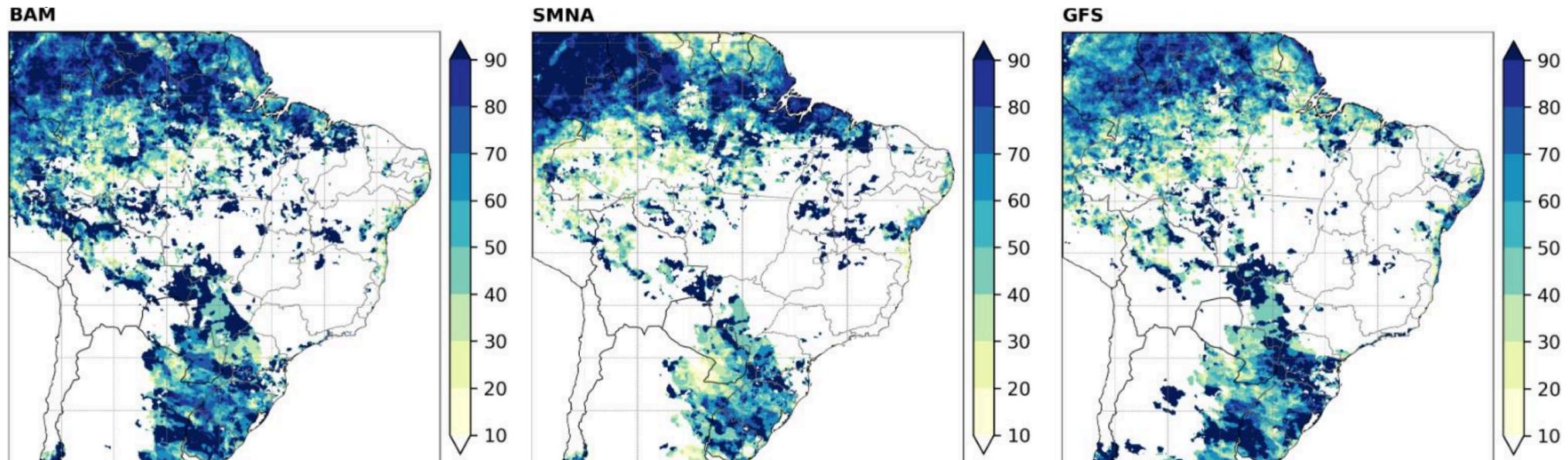
4.3 Comparações com outros produtos numéricos

- Avaliação mensal dos modelos do CPTEC
 -  Equipe de avaliação de modelos emite pareceres técnicos sobre o skill dos modelos globais e regionais de previsão de tempo do INPE, junto com o modelo GFS do NCEP
 -  São escolhidos os eventos de precipitação acumulada mais significativos (e.g., superior a 20mm em 24 horas), para os quais é avaliada a destreza dos modelos
 -  Comparação com o MERGE (Rozante et al., 2010: Combining TRMM and Surface Observations of Precipitation: Technique and Validation over South America -  [link](#))
- Modelos avaliados
 -  Modelo BAM (global)
 -  Modelo BAM/SMNA (global)
 -  MPAS/MONAN (global)
 -  WRF/CPTEC (regional)
 -  Eta (regional)
 -  BRAMS (regional)

4. Atividades Operacionais

4.3 Comparações com outros produtos numéricos

- Percentual de acertos (BAM X SMNA X GFS) - Maio de 2025
 - Número de previsões de ocorrência do evento que se confirmaram ($\frac{\text{hits}}{\text{hits} + \text{misses}}$)

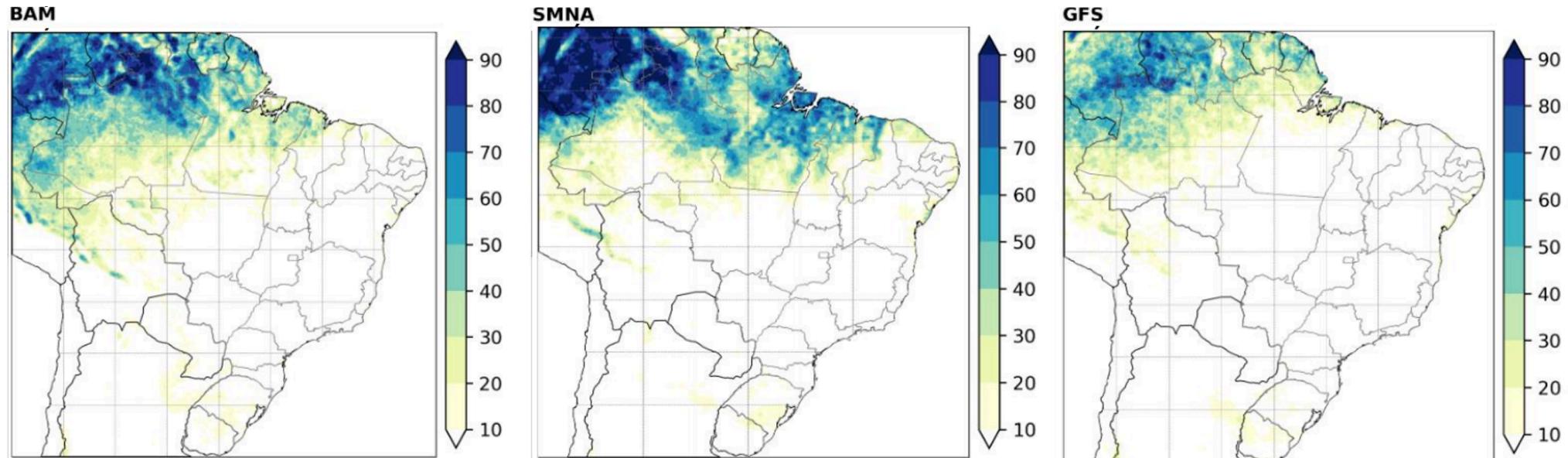


Fonte: Grupo de Avaliação de Modelos DIPTC (reprodução)

4. Atividades Operacionais

4.3 Comparações com outros produtos numéricos

- Percentual de falso alarme (BAM X SMNA X GFS) - Maio de 2025
 - Número de previsões de ocorrência do evento que não se confirmaram ($\frac{fa}{fa+cr}$)

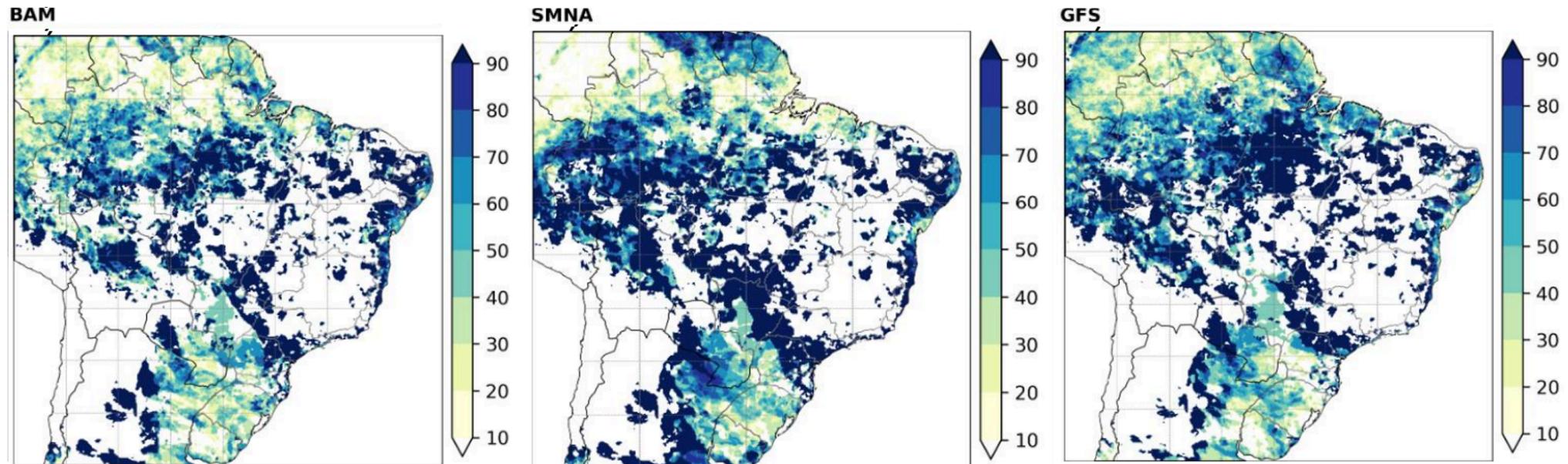


Fonte: Grupo de Avaliação de Modelos DIPTC (reprodução)

4. Atividades Operacionais

4.3 Comparações com outros produtos numéricos

- Percentual de perdas (BAM X SMNA X GFS) - Maio de 2025
 - Número de ocorrência de eventos sem previsão ($\frac{\text{misses}}{\text{misses} + \text{hits}}$)

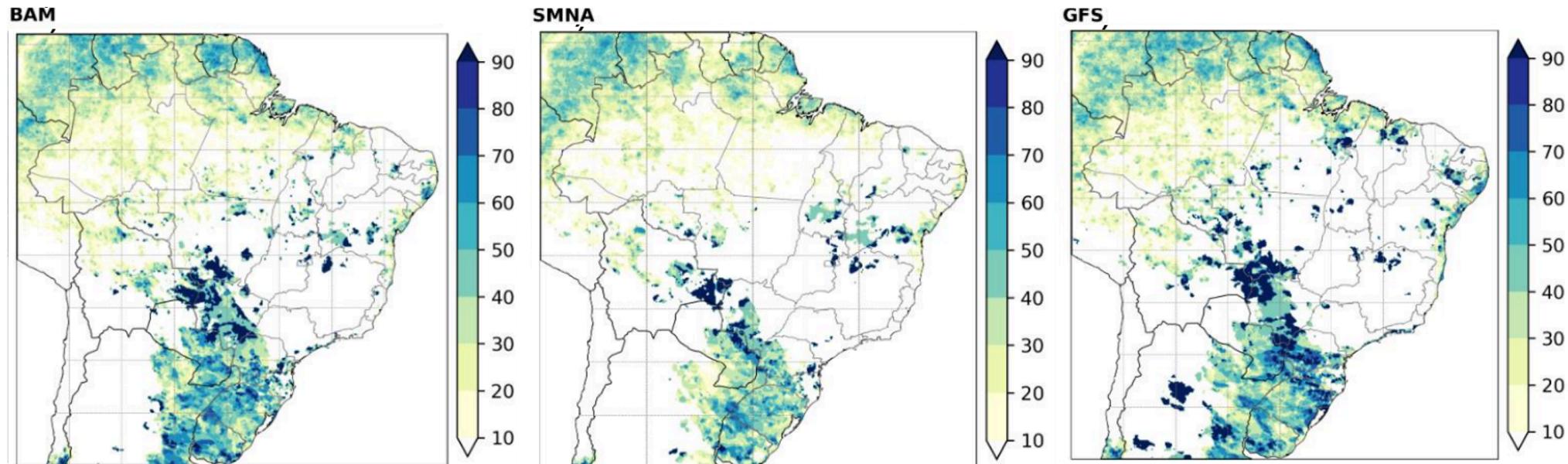


Fonte: Grupo de Avaliação de Modelos DIPTC (reprodução)

4. Atividades Operacionais

4.3 Comparações com outros produtos numéricos

- Threat Score (BAM X SMNA X GFS) - Maio de 2025 ($\frac{hits}{hits+misses+fa}$)

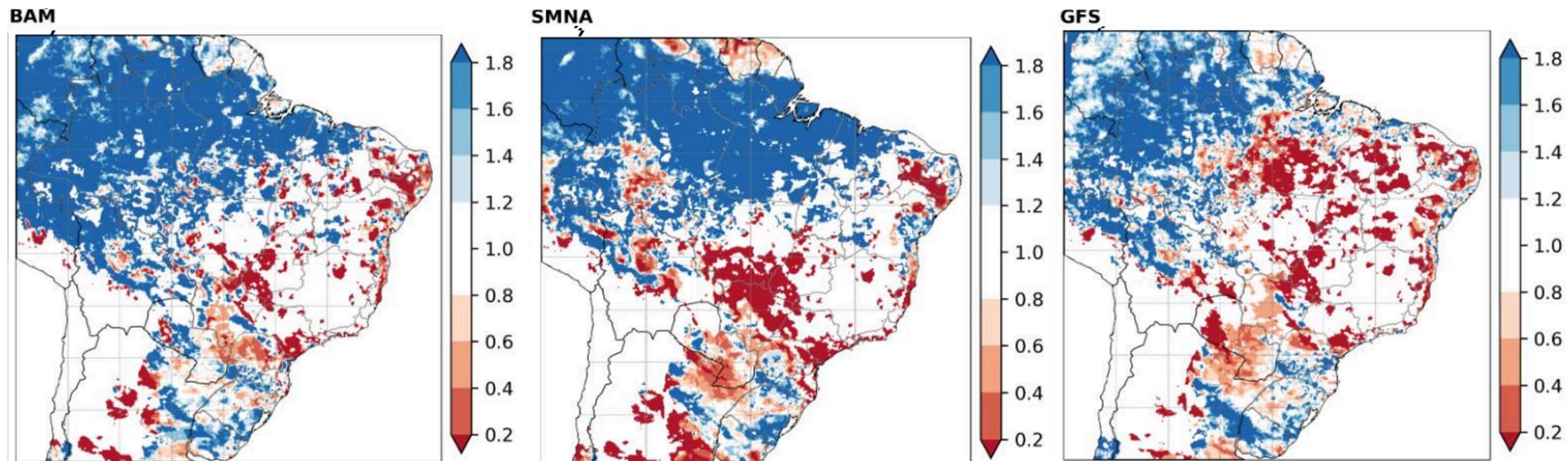


Fonte: Grupo de Avaliação de Modelos DIPTC (reprodução)

4. Atividades Operacionais

4.3 Comparações com outros produtos numéricos

- Razão de Viés (BAM X SMNA X GFS) - Maio de 2025 ($\frac{hits+fa}{hits+misses}$)



Fonte: Grupo de Avaliação de Modelos DIPTC (reprodução)

5. Conclusões

- **Determinação da Condição Inicial**

-  A Assimilação de Dados é a ponte entre as observações e o modelo numérico
-  É a técnica que combina ambas as informações para produzir a melhor estimativa do estado ótimo da atmosfera (ou do oceano, superfície continental etc)

- **Research to Operations**

-  Complexidade do framework de assimilação de dados demanda ferramentas para diagnosticar problemas no ambiente operacional e de pesquisa
-  Ciclos de entregas de novos desenvolvimentos e correções no ambiente operacional

- **Atividades Operacionais**

-  A atividade de assimilação de dados em um centro PNT requer a colaboração e o envolvimento das equipes de **modelagem, computação científica, satélites, banco de dados** para o correto estabelecimento dos processos envolvidos na determinação da análise
-  O contínuo aprimoramento desse tipo de produto demanda monitoramento constante, principalmente sobre a aplicação dos dados não convencionais e a comunicação com os grupos de satélites internacionais

Obrigado

 <https://github.com/GAD-DIMNT-CPTEC>

 <https://cfbastianz.github.io>

 carlos.bastarz@inpe.br