WORLD METEOROLOGICAL ORGANIZATION

Data Assimilation

Determination of the Initial Condition, R2O and Operational Activities

CARLOS FREDERICO BASTARZ

TRAINING COURSE ON WEATHER FORECASTING - AND BEYOND

17 November 2025

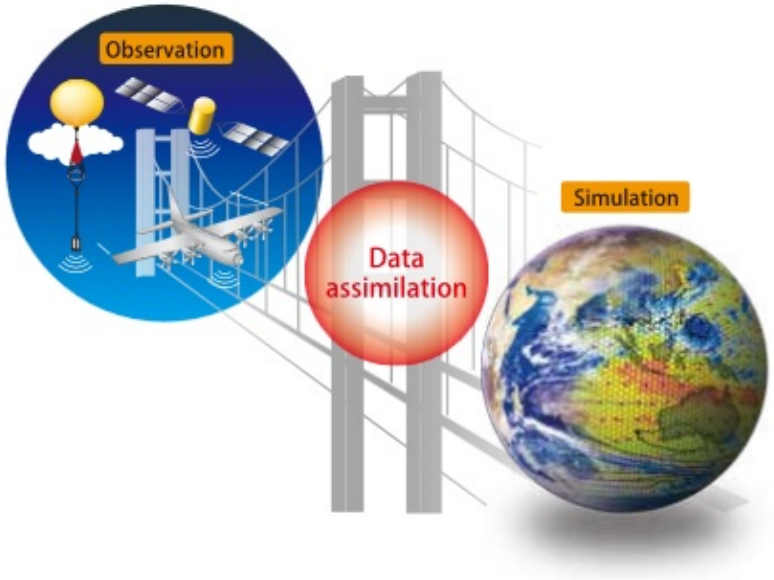
Summary

1. Data Assimilation

1.1 What is Data Assimilation?

- Data Assimilation comprises a set of techniques that allow the optimal combination of observations and numerical forecasts while taking their respective errors into account
  - Observations are irregularly distributed in space
  - Forecasts are presented on a regular grid
  - Combining both results in an updated/corrected forecast, which we call the analysis

- The analysis is the initial condition of numerical models



Source: https://www.data-assimilation.riken.jp/en/research/index.html

1. Data Assimilation

1.2 Motivation

- 👉 Models and observations have uncertainties
  - 🔴 Models
    - 🖥️ Discretization of equations, physical parameterizations, etc.
  - 🔵 Observations
    - 🔪 Instrument calibration, measurement location (e.g., proximity to rivers), recording errors, etc.
- 🛑 Data assimilation needs to account for these factors so that these uncertainties can weight their contributions
  - 👉 The higher the model/observation error, the lower its precision and, consequently, the lower its weight

1. Data Assimilation

1.3 Mathematical Intuition

- Writing this combination between observation and forecast as a linear combination:

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

- Where:
  - x_a is the analysis
  - x_b is the forecast
  - y_o is the observation
  - α is the weight given to the observation
  - $1 - \alpha$ is the weight given to the forecast
- For it to be a linear combination, the sum of the weights must equal 1 👍

- How should the weight α be determined?

1. Data Assimilation

1.3 Mathematical Intuition

- α is a parameter that relates the variance of the observation and the model

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

- Where:
  - σ_b^2 and σ_o^2 are the variances of the background and the observations, respectively

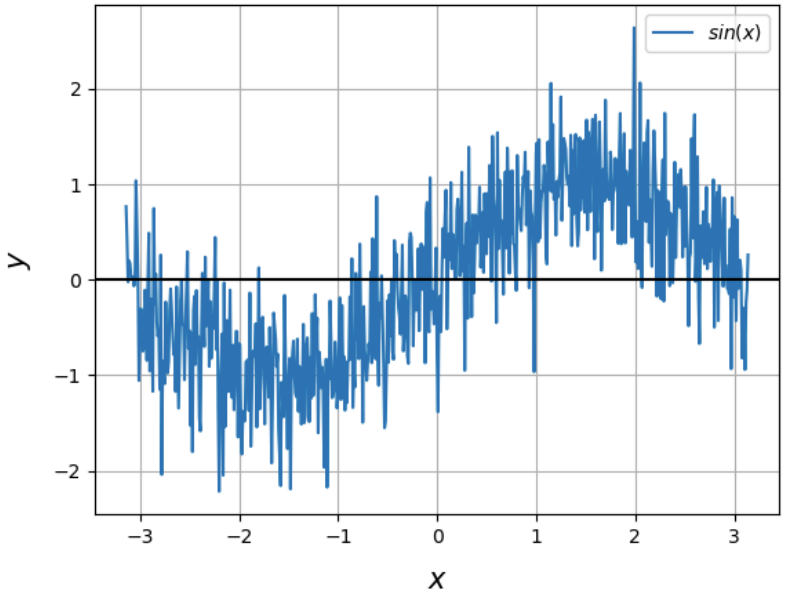
- 👉 Therefore, α can be understood as a parameter representing the ratio between the model error variance and the total system error variance (model + observation)

1. Data Assimilation

1.3 Mathematical Intuition

- Consider a simple mathematical model, the sine function with added normally distributed noise:

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

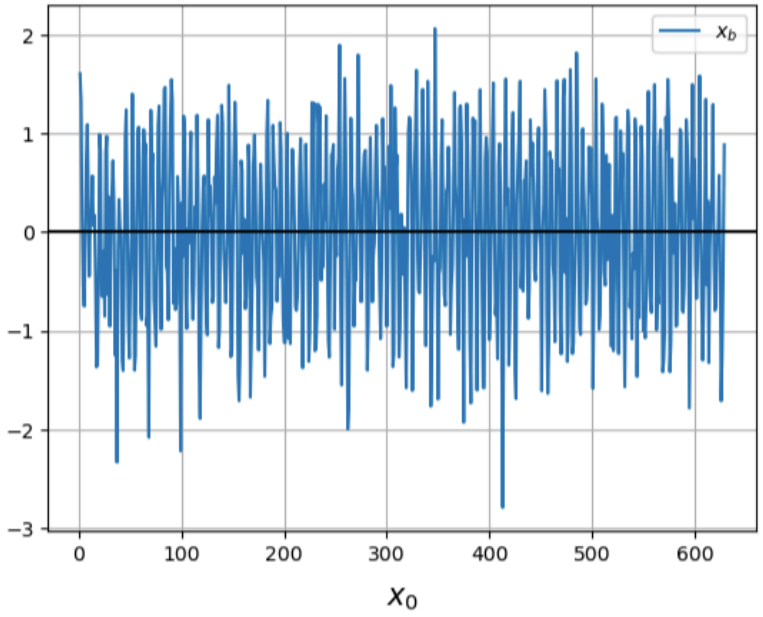


1. Data Assimilation

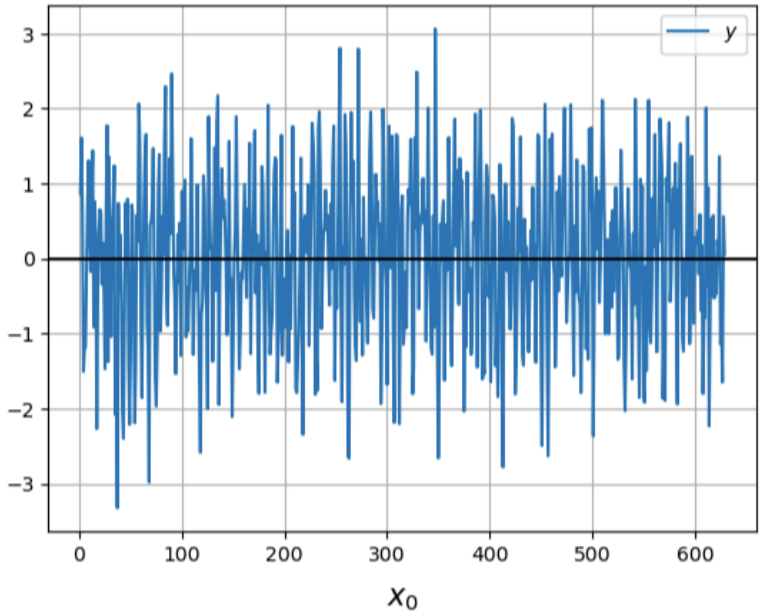
1.3 Mathematical Intuition

- We define a full domain where we apply the model to extract information as a "forecast" and "observations"

xb = sine function applied to domain x0



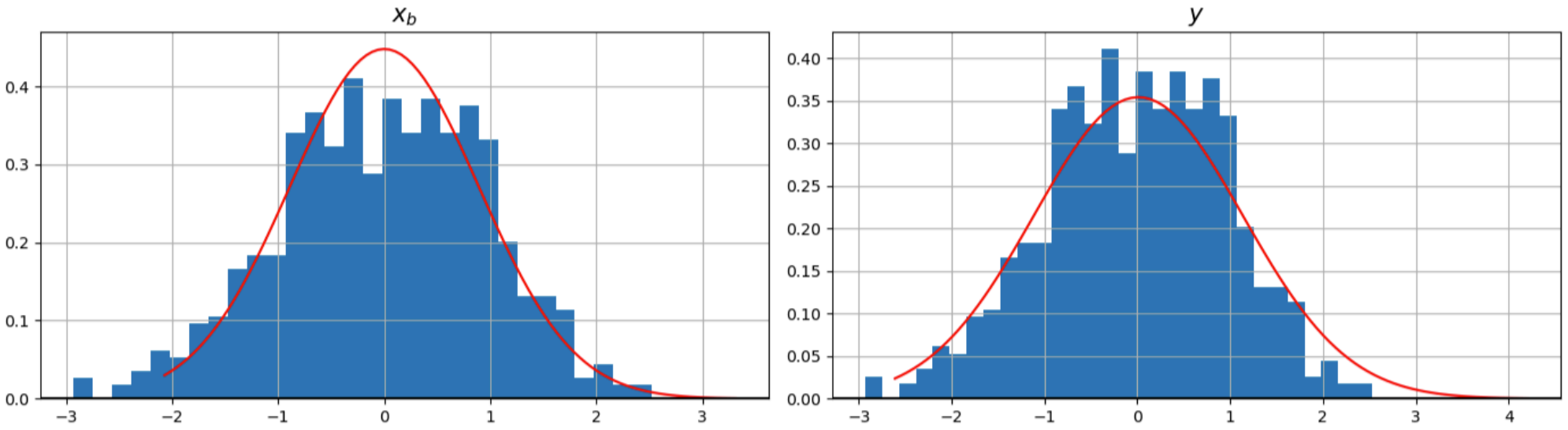
y = Observations (same nature as xb)



1. Data Assimilation

1.3 Mathematical Intuition

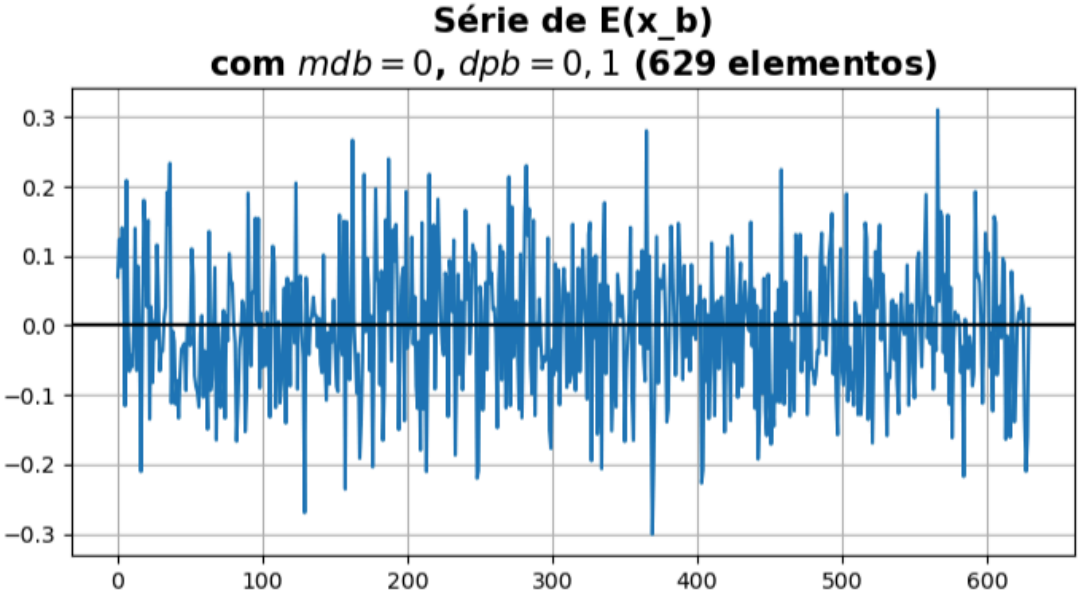
- x_b and y_o follow a normal distribution, both represented by random values over a normal curve with $\mu_{x_b} = 0.0019$, $\sigma_{x_b} = 0.8909$ and $\mu_{y_o} = -0.011$, $\sigma_{y_o} = 0.8563$

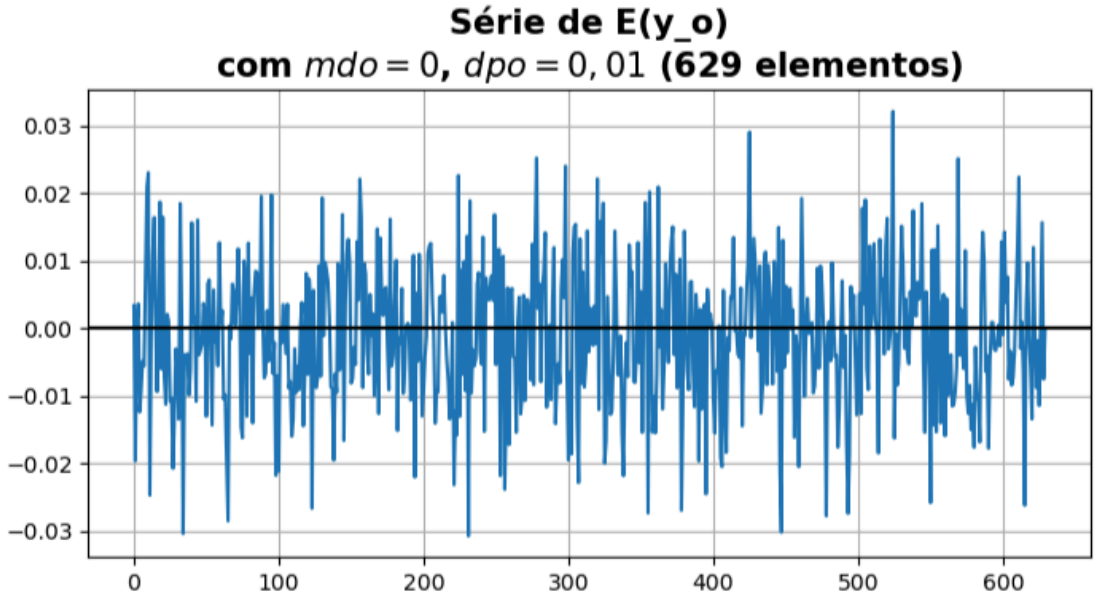


1. Data Assimilation

1.3 Mathematical Intuition

- Define the error series for forecasts and observations

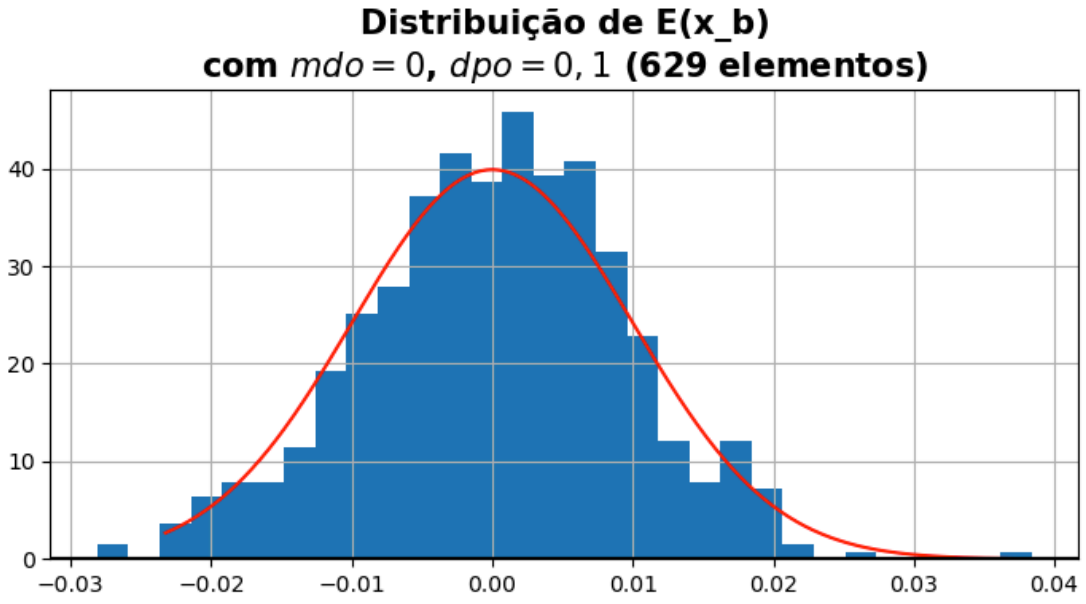


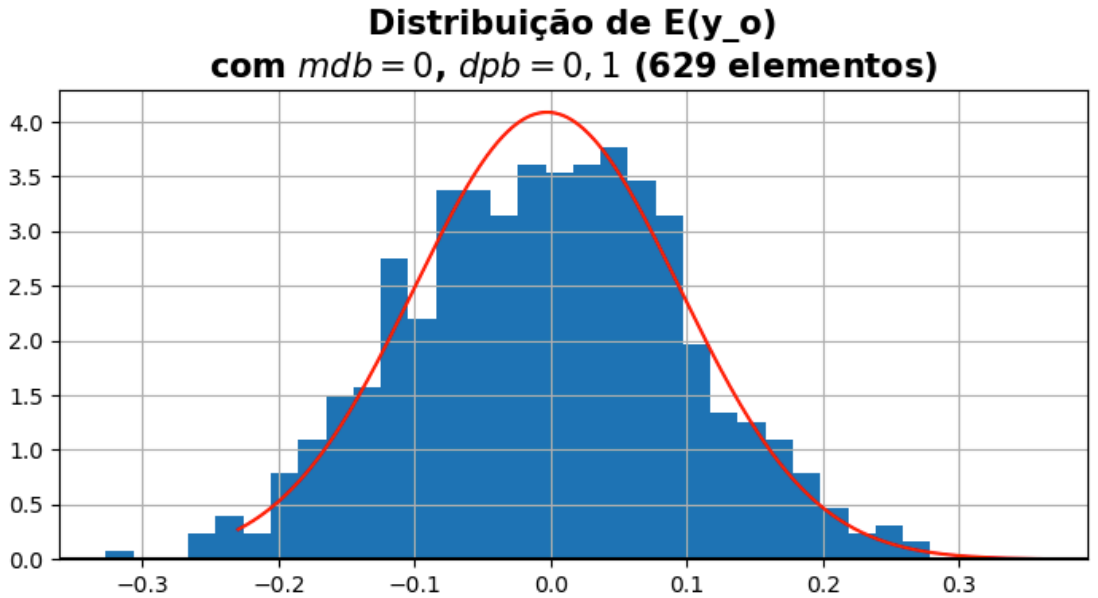


1. Data Assimilation

1.3 Mathematical Intuition

- Checking the error distributions of "forecast" and "observation"





1. Data Assimilation

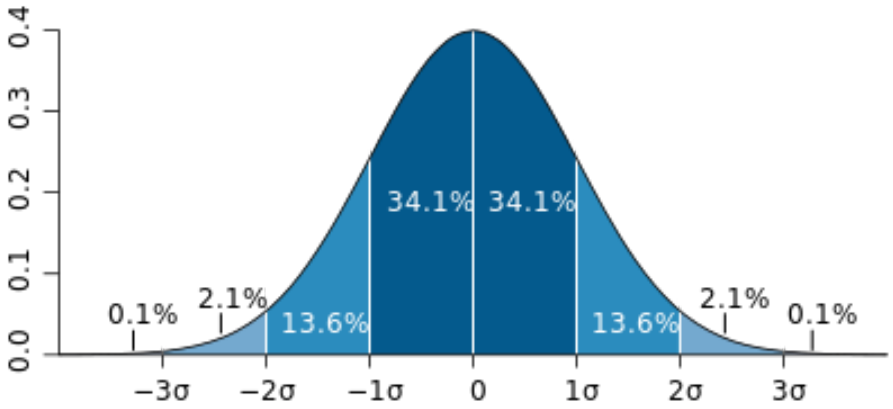
1.3 Mathematical Intuition

Why the Normal Distribution?

- We keep the distributions of x_b and y_o close to a normal distribution because it has the following properties:

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

- ~68% of values lie within 1σ of the mean
- ~95% of values lie within 2σ of the mean
- ~99.7% of values lie within 3σ of the mean



1. Data Assimilation

1.3 Mathematical Intuition

- From this information, we calculate
  - The variance of forecast and observation errors

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

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

- The variance of observation errors is much smaller than the variance of background errors
- The value of $\alpha = \frac{\sigma_b^2}{\sigma_b^2+\sigma_o^2}$

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

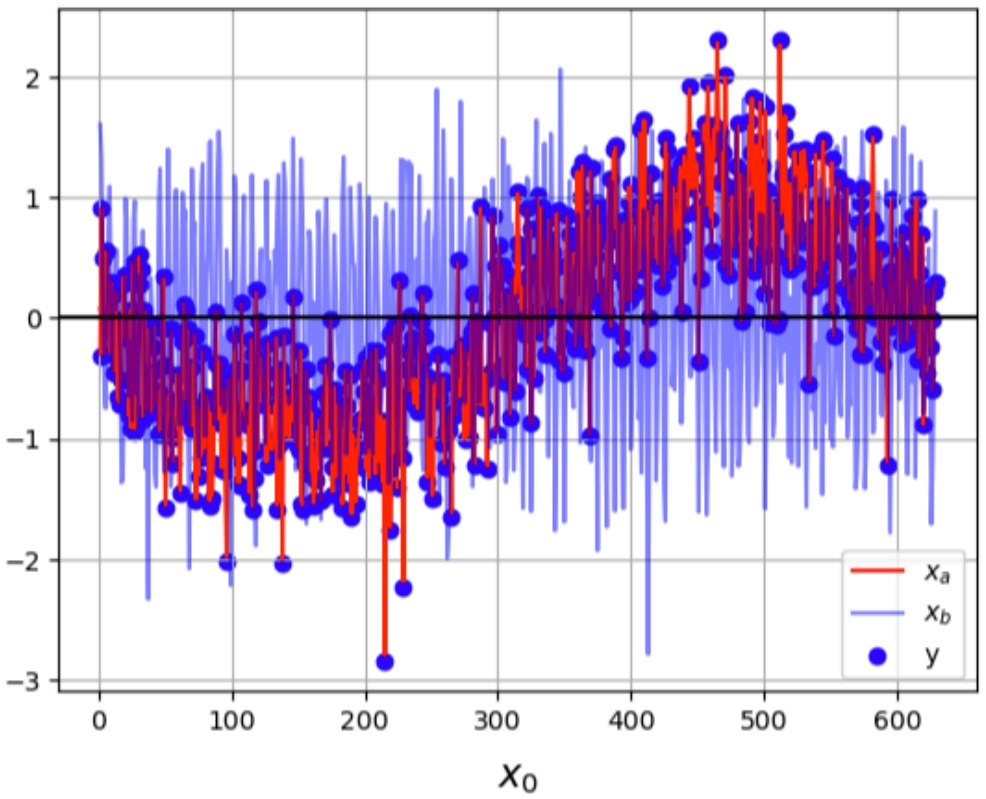
1. Data Assimilation

1.3 Mathematical Intuition

- The value $\alpha \approx 0.99$ indicates that 99% of the weight in the linear combination between x_b and y_o is given to the observations, while 1% of the weight is given to the background

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

- For further exploration
  - Jupyter notebook with univariate empirical analysis Open in Colab
  - Jupyter notebook with multivariate empirical analysis Open in Colab



1. Data Assimilation

1.3 Mathematical Intuition

- In real-world, multivariate, and multidimensional problems, the weight α is represented by error covariance matrices, which require modeling and approximations for their representation 👉 we have limited control or influence over these errors
- Optimal Interpolation

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

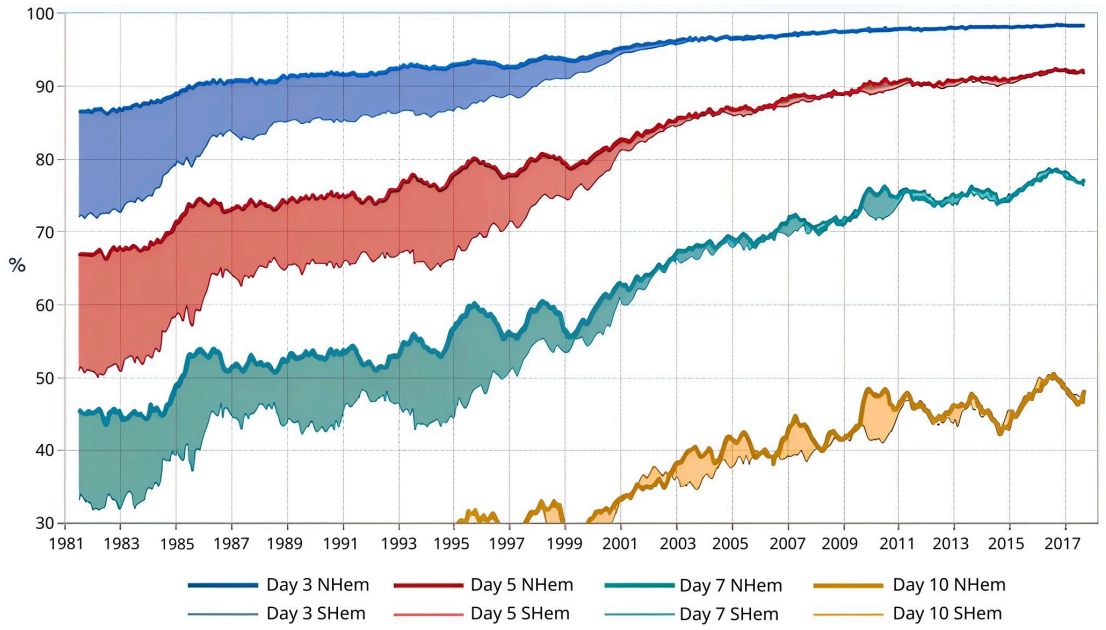
- Where
  - $\mathbf{x}_a$ is the analysis vector (estimated state)
  - $\mathbf{x}_b$ is the background vector (first guess)
  - $\mathbf{y}$ is the observation vector
  - $\mathbf{W}$ is the weight (or gain) matrix
  - H is the nonlinear observation operator (maps model space to observation space)

2. Determining the Initial Condition

2.1 Evolution of Data Assimilation Skill

- 📈 Evolution of forecast skill for 500 hPa geopotential height
  - In the early 1980s, the 7-day forecast skill for the Northern Hemisphere did not exceed 50%, and was below 40% in the Southern Hemisphere
  - Over time, the skill difference between hemispheres decreased drastically, becoming very similar by the 2000s
  - Only from the mid-1990s did 10-day forecasts start reaching some skill (~30%)...
  - Currently, 10-day forecasts already reach 50% skill for both hemispheres

- Although improvements were significant, it seems short-term forecast skill is approaching its limit - **why?**



2. Determining the Initial Condition

2.2 CPTEC Numerical Modeling and Data Assimilation System

- SMNA is CPTEC's data assimilation system
  - Global spectral model BAM (Brazilian Atmospheric Model)
  - Data assimilation framework GSI (Gridpoint Statistical Interpolation)
  - Provides analysis for the BAM model at spatial resolution TQ0299L064
    - 👉 TQ0299 = triangular spectral truncation of order 299, using quadratic Gaussian grid
    - 👉 L064 = 64 vertical levels in hybrid sigma-pressure coordinates
- At CPTEC, the combination of BAM model and GSI has been applied since 2012
  - With updates to the atmospheric model (surface model, convective parameterization, vertical coordinate)
  - With updates to the GSI version (including new observation types, covariance matrix updates, etc.)

2. Determining the Initial Condition

2.3 Gridpoint Statistical Interpolation

- GSI is a data assimilation framework developed by NCEP
  - Provides software implementation for all components related to data assimilation
    - Variational methods (3D/4DVar, FGAT, hybrid-variational, 3D/4DEnVar)
    - Ensemble-based methods (EnKF, EnSRF, LETKF)
    - Minimization of the variational cost function
    - Observation operator H (Radiative Transfer Model)
    - Support for global (spectral) and regional (grid-point) models
- Maintained by DTC/NCAR
  - Centralizes contributions, manages the code, distributes releases, and provides tutorials for the user community
  - 🔗 https://ral.ucar.edu/solutions/products/gridpoint-statistical-interpolation-gsi

2. Determining the Initial Condition

2.3 Gridpoint Statistical Interpolation

⚙ 3DVar

- 3DVar is a variational data assimilation method that finds the best initial state of the atmosphere (or ocean) by adjusting an analysis that minimizes the difference between observations and the background, weighted by their errors (covariance matrices)
- **Cost Function**

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

- **Gradient**

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

- **Exact Analytical Solution**

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

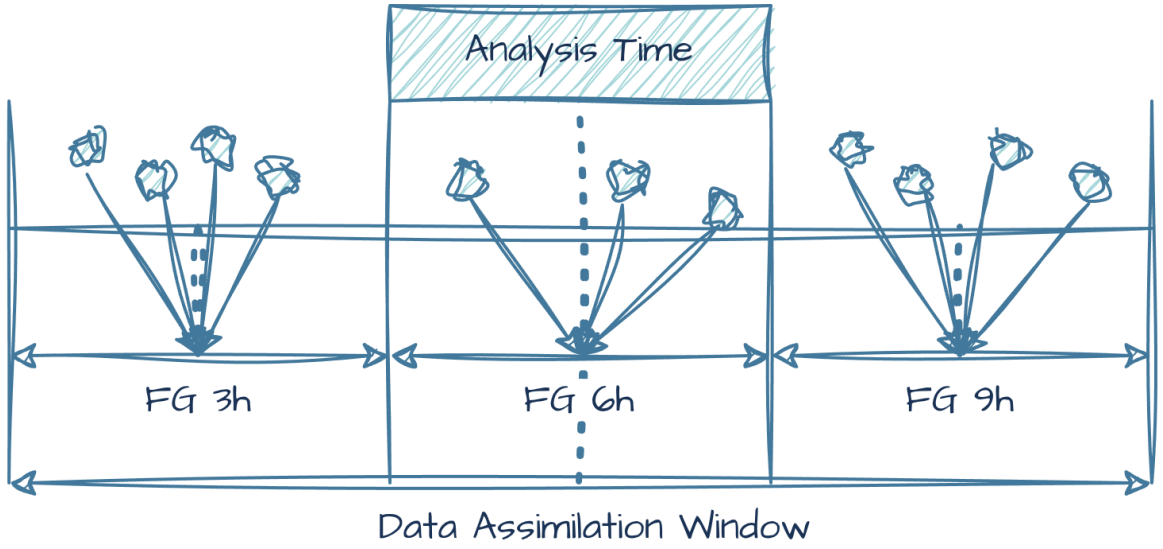
2. Determining the Initial Condition

2.3 Gridpoint Statistical Interpolation

⚙ FGAT

- FGAT (First Guess at Appropriate Time) uses the background at the observation time to improve temporal consistency in 3DVar
  - Cost function remains 3D, as the background correction does not evolve over time
  - Improves temporal alignment of observations that are not at the analysis time (e.g., non-conventional observations)
  - Requires the first guess to be partitioned over the assimilation window

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



2. Determining the Initial Condition

2.3 Gridpoint Statistical Interpolation

⚙ Observation Operator $\mathbf{H}$

- Responsible for mapping the model background to the observation space
  - If y_o and x_b are equivalent quantities (e.g., temperatures), then $\mathbf{H}$ performs only an interpolation and the innovation is calculated as $y_o - H(x_b)$
  - If y_o is a radiance, then $\mathbf{H}$ needs to compute a radiance profile from x_b to calculate the innovation
    - In this case, $\mathbf{H}$ is a radiative transfer model (in the SMNA, it is CRTM - Community Radiative Transfer Model)

2. Determining the Initial Condition

2.3 Gridpoint Statistical Interpolation

⚙ Covariance Matrix $\mathbf{B}$

- Sources of uncertainty in the modeling process are represented by:
  - Numerical model (e.g., dynamics and physics)
  - Observations (e.g., measurement, instrument, processing level)
  - Data assimilation system (e.g., observation operators, adjoint and tangent-linear models, ensemble size)
- The forecast error covariance matrix ($\mathbf{B}$) represents the covariance of the model "error" (an estimate)
- In data assimilation, these errors are modeled in covariance matrices that account for spatiotemporal relationships between observed and diagnosed/prognosed quantities
- 3DVar cost function:

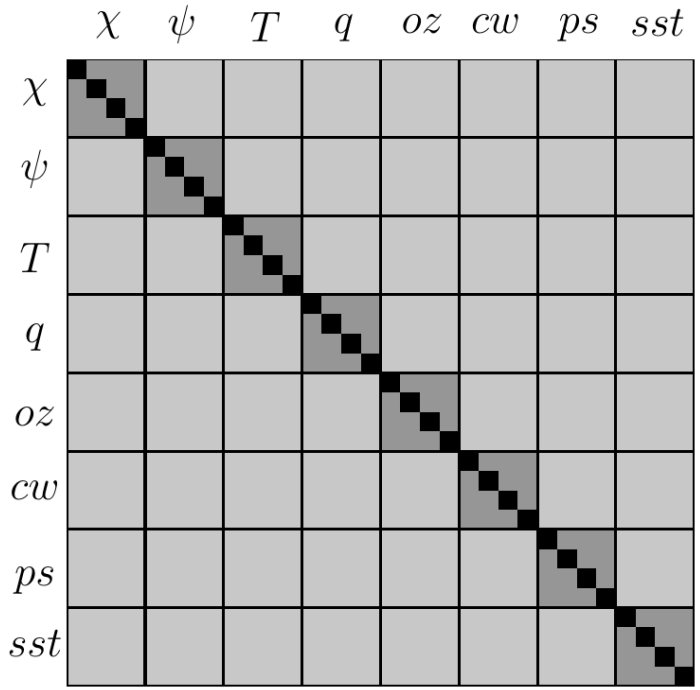
$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) + \frac{1}{2}[\mathbf{y}^o - \mathbf{H}(\mathbf{x})]^T \mathbf{R}^{-1}[\mathbf{y}^o - \mathbf{H}(\mathbf{x})]$$

2. Determining the Initial Condition

2.3 Gridpoint Statistical Interpolation

⚙ Covariance Matrix $\mathbf{B}$

- **NMC Method (National Modeling Center)**
  - The NMC method assumes that the spatial correlation of model errors is similar to the spatial correlation of differences between 48- and 24-hour forecasts
  - **Assumption:** linear growth of forecast errors during the first hours of prediction
- Example of a valid forecast pair (BAM model)
  - 2013122418-2013122618 (48-hour forecast)
  - 2013122518-2013122618 (24-hour forecast)



2. Determining the Initial Condition

2.3 Gridpoint Statistical Interpolation

⚙ Observation Quality Control

- This step occurs before and during data assimilation
  - Before, it involves preparing observation data (at CPTEC, this means creating their own prepBUFR files for conventional and radiance data)
  - During assimilation, GSI performs multiple checks and tests
- Generally, there is a type of quality control for each observation type

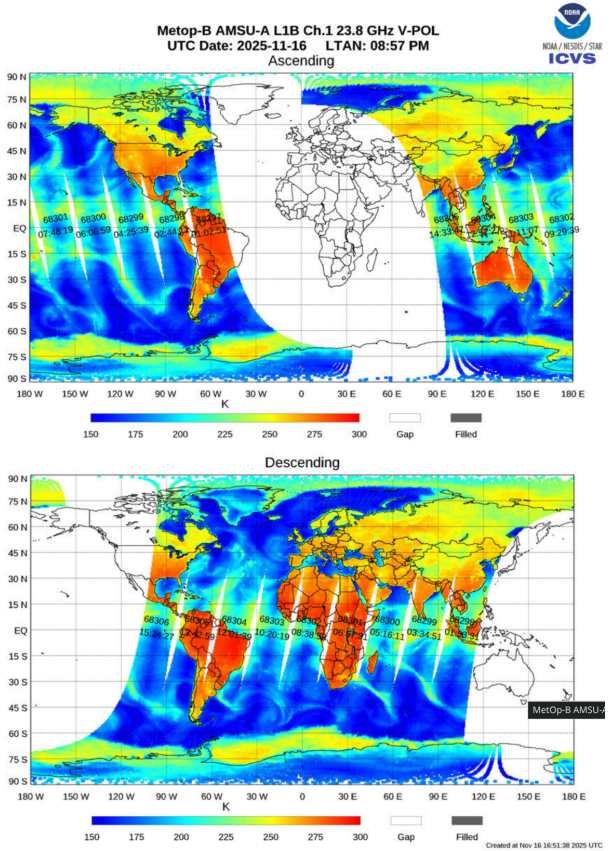
- 👨‍🔬 **Pre-QC:** for radiosondes, checks altitude, pressure, temperature, humidity, and removes duplicate data; for radiances, applies provider flags, cloud cover, and viewing angle checks
- 👨‍🔬 **OMF:** calculates innovations $\mathbf{y}_o - H(\mathbf{x}_b)$ and compares them with $\mathbf{B}$ and $\mathbf{R}$ variances
- 👨‍🔬 **Buddy Check:** compares observations with neighbors (can reject or reduce weights of observations)
- 👨‍🔬 **Adaptive/Variational QC:** adjusts observation weights instead of immediately discarding them

2. Determining the Initial Condition

2.3 Gridpoint Statistical Interpolation

⚙ Observations Assimilated by SMNA

- Currently, the following observation datasets are assimilated in CPTEC's SMNA
  - 🌡️ Conventional observations: u, v, t, q, and ps
  - 🛰️ Non-conventional satellite/instrument observations:
    - AMSUA/METOP-B (microwave, infers t)
    - SATWND (infers u and v)
    - GPSRO (GPS signal refraction, bending angle, infers t, p, and q)

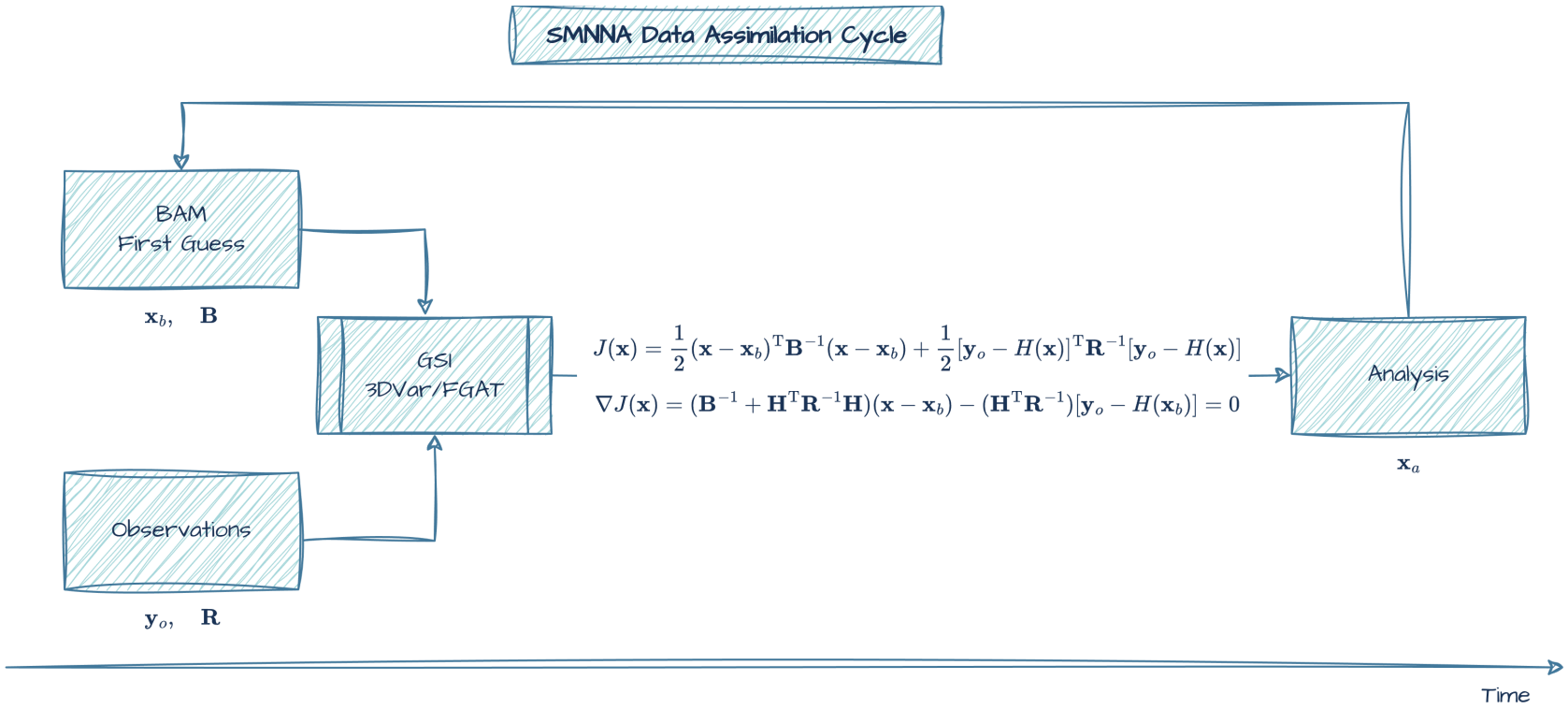


Source: https://www.star.nesdis.noaa.gov/icvs/status_MetOPB_AMSUA.php

2. Determining the Initial Condition

2.3 Gridpoint Statistical Interpolation

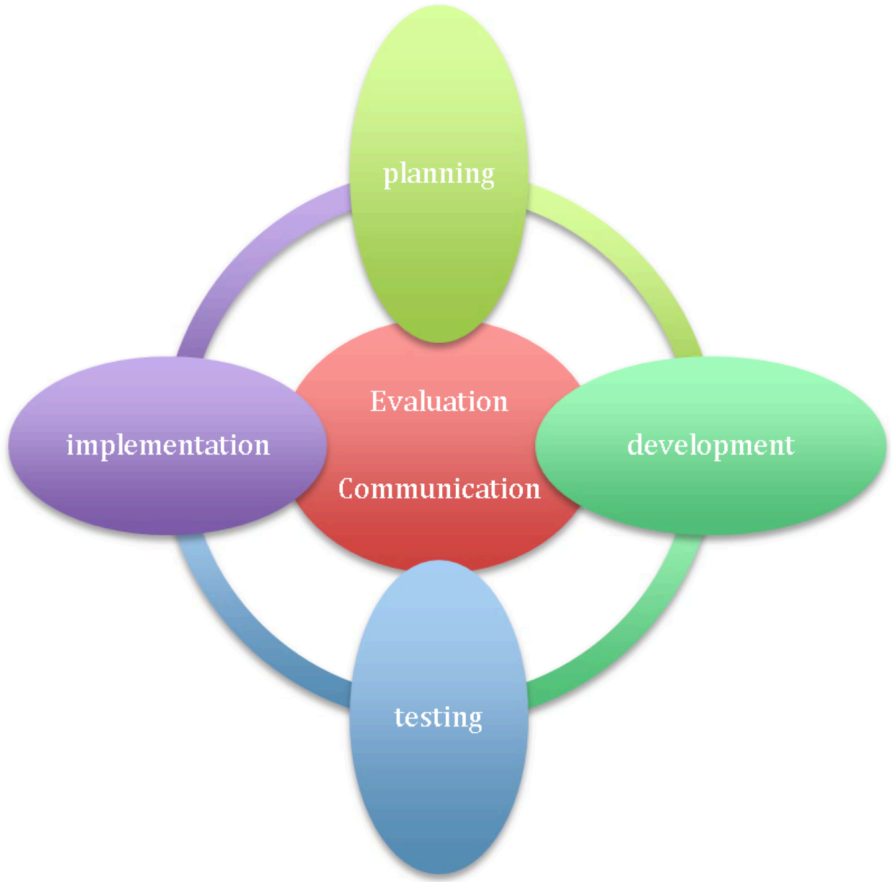
⚙ Data Assimilation Cycle



3. R2O - Research to Operations

3.1 What is R2O and Why is it Necessary?

- 👉 A crucial step for maintaining and developing the operational data assimilation suite
  - This is when research results are transferred along with new developments to the operational environment
  - Computational artifacts produced in research can also serve as operational diagnostic tools
- 🤹 Challenges:
  - Technical limitations often linked to computational capacity (processing and storage)
  - Proper validation, outperforming the previous version, and comparison with other products

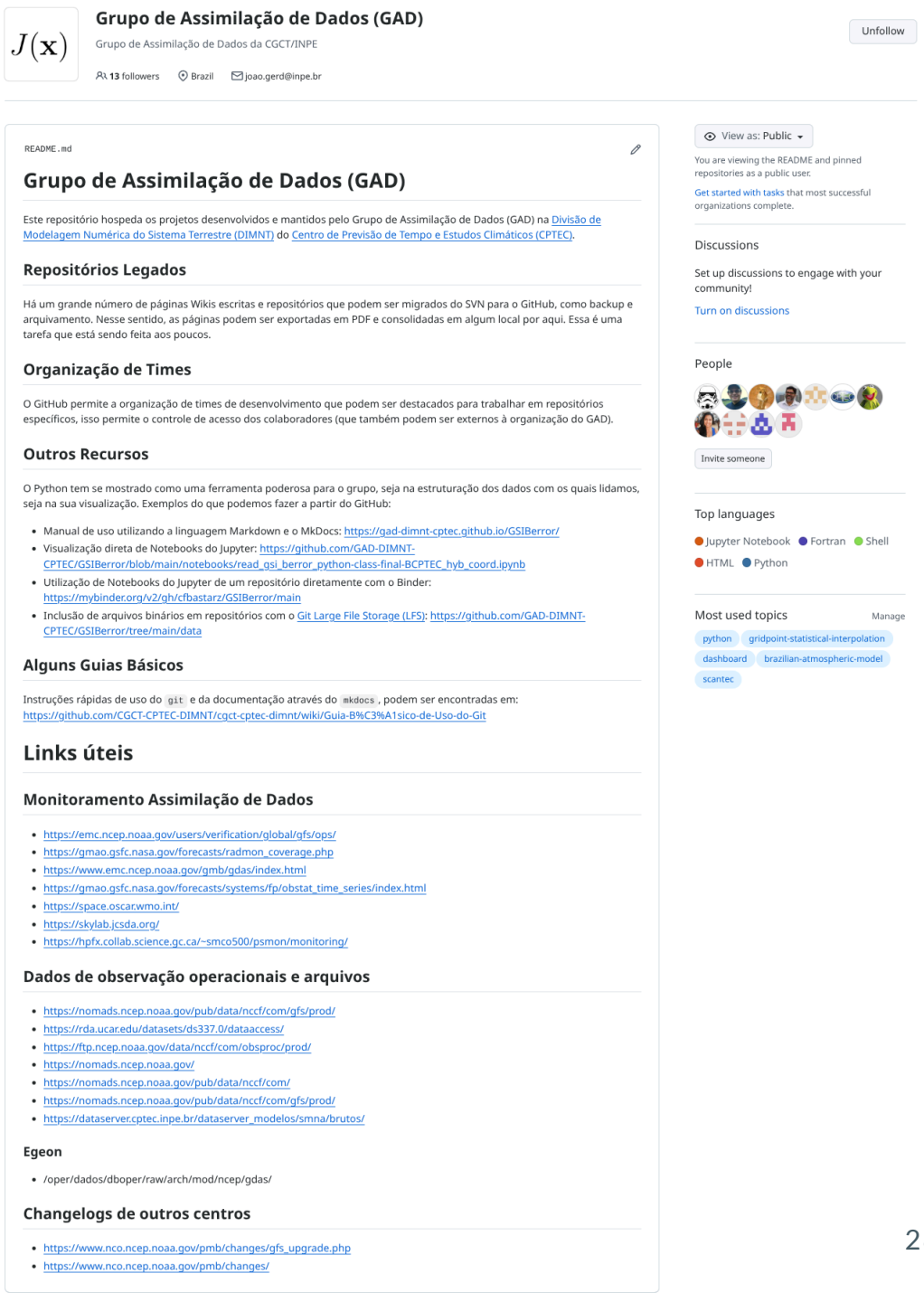


Source: https://www.ecmwf.int/sites/default/files/elibrary/2017/17549-ecmwf-research-operations-r20-process.pdf

3. R2O - Research to Operations

3.2 Support Tools

- The Data Assimilation Group maintains an organization on GitHub
  - Facilitates team organization and development tracking
    - 🗺️ Roadmaps
    - 🏷️ Tags and releases for distribution
    - 📋 Issue tracking
    - 🗣️ Discussions
    - 📄 Wikis, etc.
    - 🔗 https://github.com/GAD-DIMNT-CPTEC



Grupo de Assimilação de Dados (GAD)

Grupo de Assimilação de Dados da CGCT/INPE

Unfollow

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_gsi_berror_python-class-final-BCPTEC_hyb_coord.ipynb
- Utilização de Notebooks do Jupyter de um repositório diretamente com o Binder: https://mybinder.org/v2/gh/cfbastarz/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.jcsda.org/
- https://hpfx.collab.science.gc.ca/~smco500/psmon/monitoring/

Dados de observação operacionais e arquivos

- https://nomads.ncep.noaa.gov/pub/data/nccf/com/gfs/prod/
- https://rda.ucar.edu/datasets/ds337.0/dataaccess/
- https://ftp.ncep.noaa.gov/data/nccf/com/obsproc/prod/
- https://nomads.ncep.noaa.gov/
- https://nomads.ncep.noaa.gov/pub/data/nccf/com/
- https://nomads.ncep.noaa.gov/pub/data/nccf/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/

View as: Public

You are viewing the README and pinned repositories as a public user.

Get started with tasks that most successful organizations complete.

Discussions

Set up discussions to engage with your community!

Turn on discussions

People

Invite someone

Top languages

Jupyter Notebook · Fortran · Shell · HTML · Python

Most used topics Manage

python · gridpoint-statistical-interpolation · dashboard · brazilian-atmospheric-model · scantec

3. R2O - Research to Operations

3.2 Support Tools

Examples of Support Tools

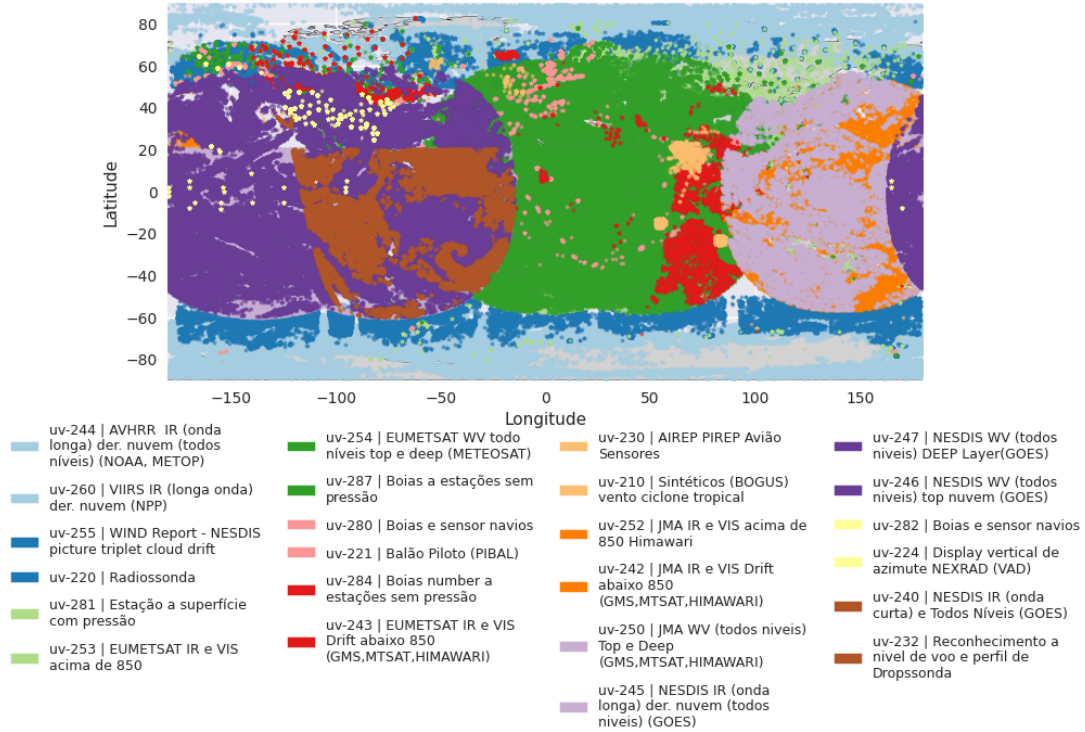
- **readDiag**
  - Tool for diagnosing radiance assimilation in GSI (Python)
- **GSIBerror**
  - Tool for diagnosing the GSI background error covariance matrix (Python)
- **pyBAM**
  - Tool to read BAM forecast fields (recomposes spectral coefficients to physical space, Python)
- **SCANTEC**
  - Community System for Evaluation of Numerical Weather and Climate Models (Fortran)
- **SCANPLOT**
  - Plotting system for SCANTEC (Python)
- **SMNAMonitoringApp**
  - Tool for monitoring operational SMNA simulations (under development, Python)
- **Observation Impact and Observing System Experiments**
  - Diagnostic tools to study the impact and contribution of different observation types on the analysis

3. R2O - Research to Operations

3.2 Support Tools

Examples of Support Tools

- **readDiag**
  - 🔪 Tool for diagnosing radiance assimilation in GSI (Python)
  - 🔗 https://gad-dimnt-cptec.github.io/readDiag
  - 🐙 https://github.com/GAD-DIMNT-CPTEC/readDiag
  - 🔳 Jupyter notebooks available for testing
  - 🐍 Available on PyPi
  - 📄 Documentation in Portuguese and English

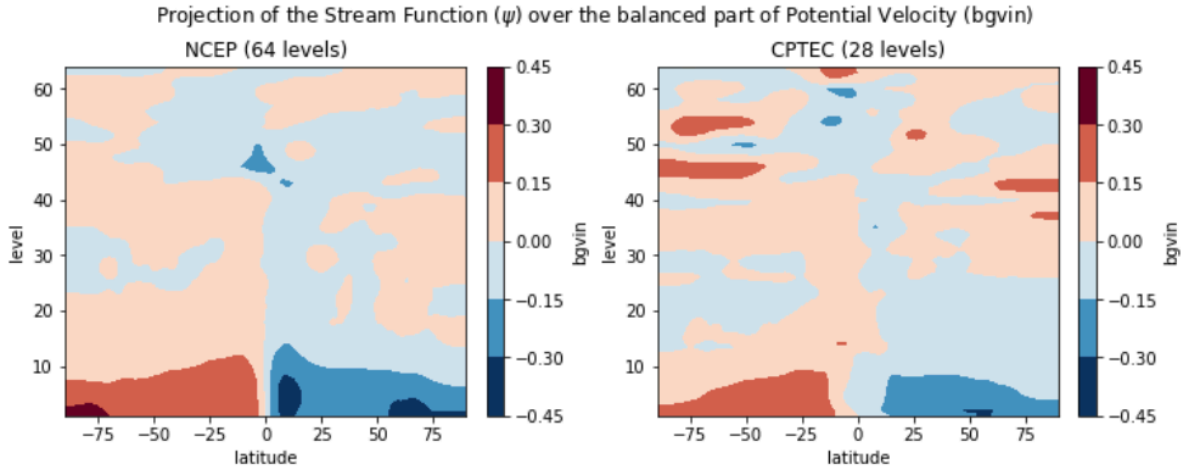


3. R2O - Research to Operations

3.2 Support Tools

Examples of Support Tools

- **GSIBerror**
  - 🔪 Tool for diagnosing the GSI background error covariance matrix (Python)
  - 🔗 https://gad-dimnt-cptec.github.io/GSIBerror
  - 🐙 https://github.com/GAD-DIMNT-CPTEC/GSIBerror
  - 🏁 Jupyter notebooks available for testing
  - 🐍 Available on PyPi
  - 📄 Documentation in Portuguese and English

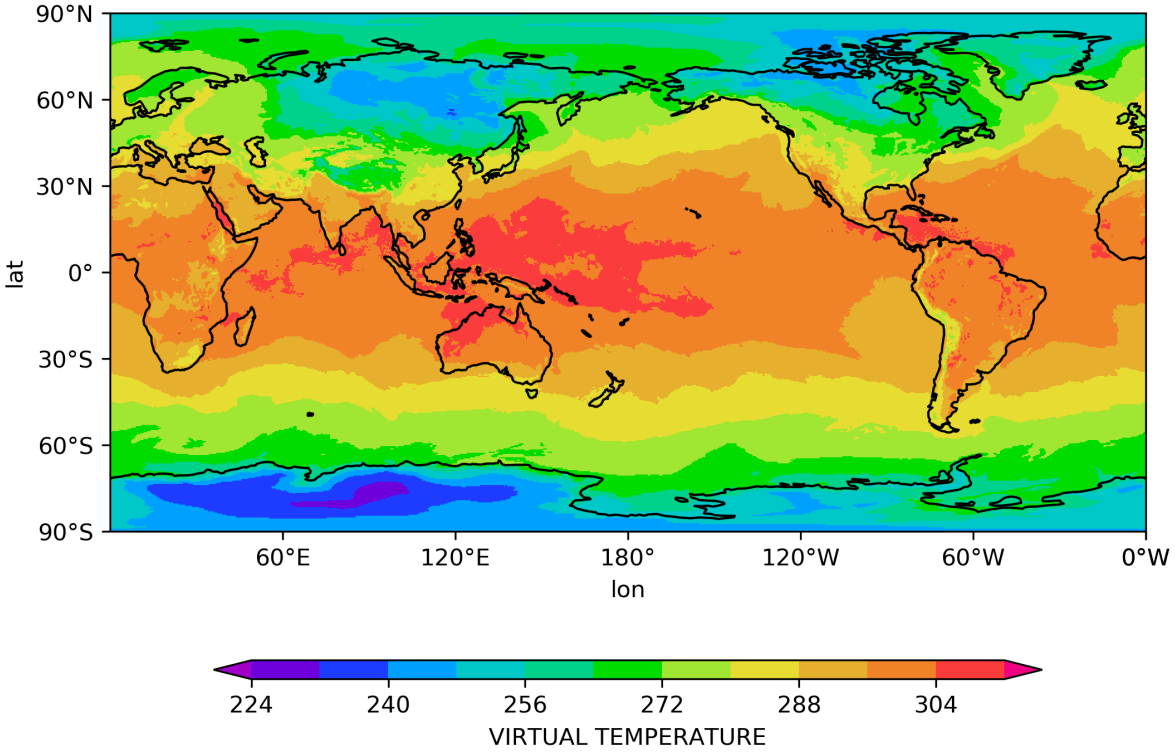


3. R2O - Research to Operations

3.2 Support Tools

Examples of Support Tools

- **pyBAM**
  - 🪛 Tool for reading BAM forecast fields (Python)
  - 🔗 https://gad-dimnt-cptec.github.io/pyBAM
  - 🐙 https://github.com/GAD-DIMNT-CPTEC/pyBAM
  - 📄 Documentation only in Portuguese



3. R2O - Research to Operations

3.2 Support Tools

Examples of Support Tools

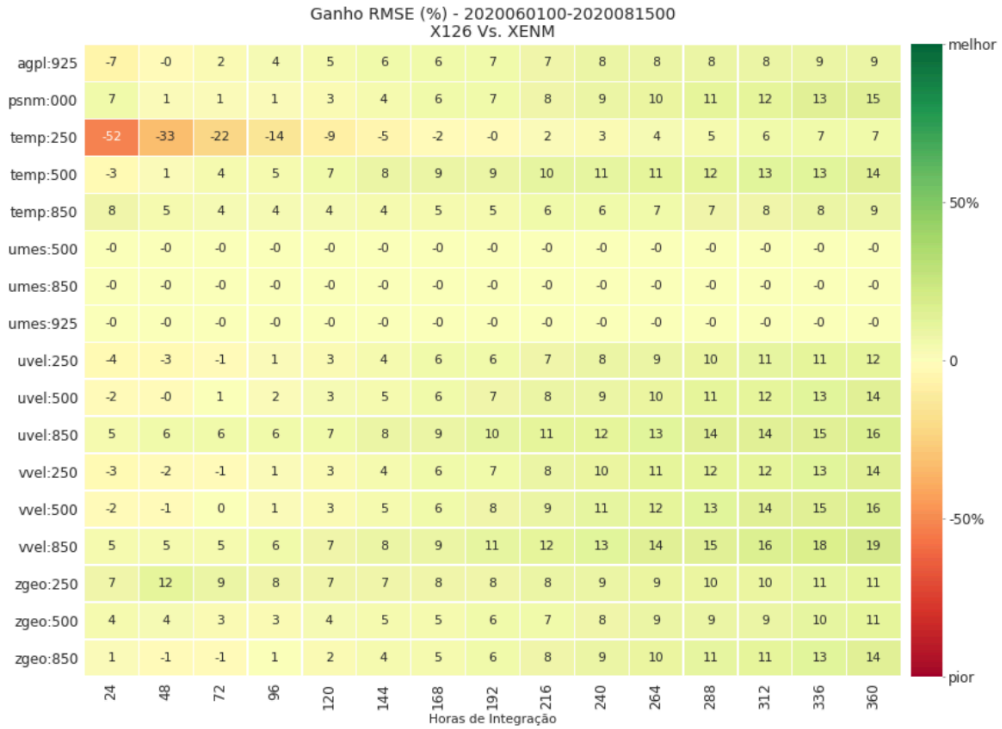
- **SCANTEC**
  - 🔪 Community System for Evaluation of Numerical Weather and Climate Models (Fortran)
  - 🔗 https://gad-dimnt-cptec.github.io/SCANTEC
  - 🐙 https://github.com/GAD-DIMNT-CPTEC/SCANTEC
  - 📄 Documentation only in Portuguese

3. R2O - Research to Operations

3.2 Support Tools

Examples of Support Tools

- **SCANPLOT**
  - 🪛 Plotting system for SCANTEC (Python)
  - 🔗 https://gad-dimnt-cptec.github.io/SCANPLOT
  - 🐙 https://github.com/GAD-DIMNT-CPTEC/SCANPLOT
  - 🏁 Jupyter notebooks available to test the CLI
  - ⚙️ Demo GUI available at https://huggingface.co/spaces/cfbastarz/SCANPLOT
  - 📄 Documentation only in Portuguese

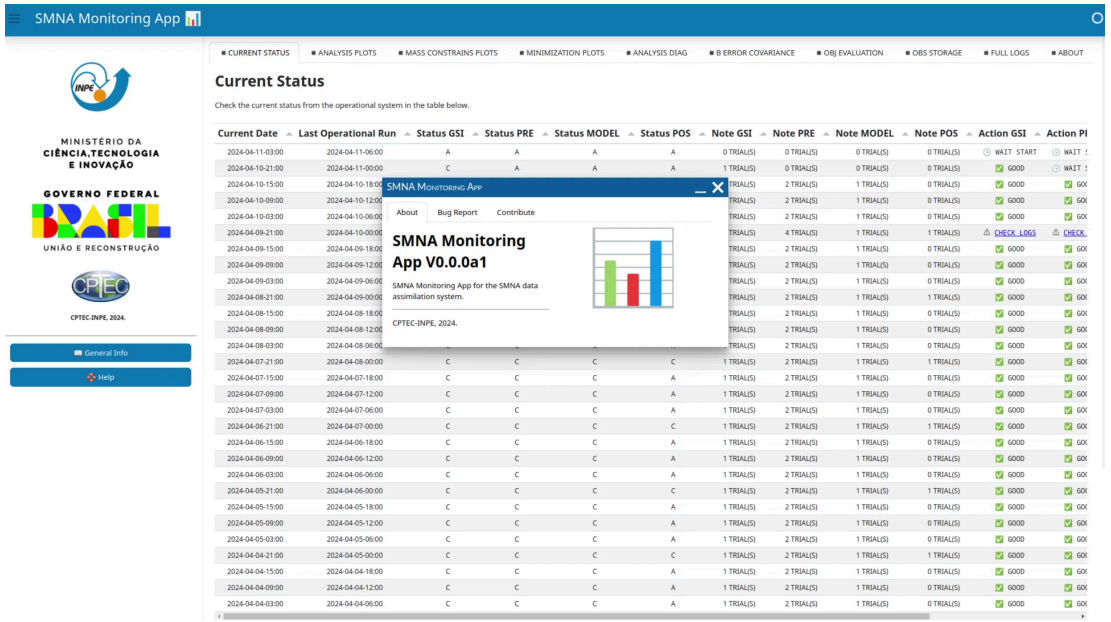


3. R2O - Research to Operations

3.2 Support Tools

Examples of Support Tools

- **SMNAMonitoringApp**
  - 🔪 Tool for monitoring operational SMNA simulations (under development, Python)
  - 🔗 https://gad-dimnt-cptec.github.io/SMNAMonitoringApp
  - 🐙 https://github.com/GAD-DIMNT-CPTEC/SMNAMonitoringApp
  - ⚙️ Demo available at https://huggingface.co/spaces/cfbastarz/SMNAMonitoringApp
  - 📄 Documentation only in Portuguese

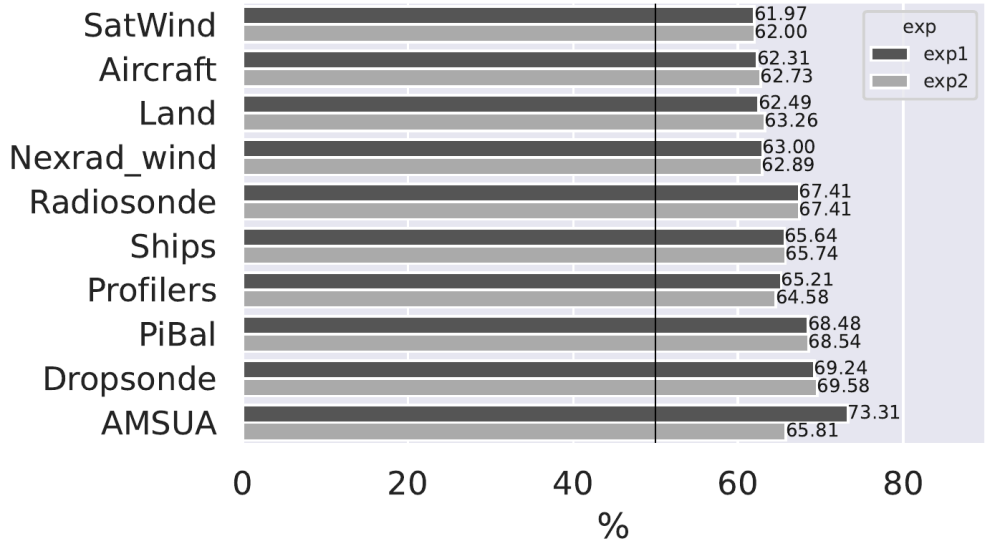


3. R2O - Research to Operations

3.2 Support Tools

Examples of Support Tools

- **Fractional Observation Impact (Viana & de Mattos, 2024)**



meteorology

MDPI

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 ($-3.54/-3.45$ J kg^{-1})and the profilers ($-2.13/-1.97$ J kg^{-1}), and the smallest contributions came from the land ($-0.28/-0.29$ J kg^{-1}) and sea ($-0.44/-0.44$ 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.

Keywords: data assimilation; observation impact; Brazilian atmospheric model; numerical weather prediction

check for updates

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

1. Introduction

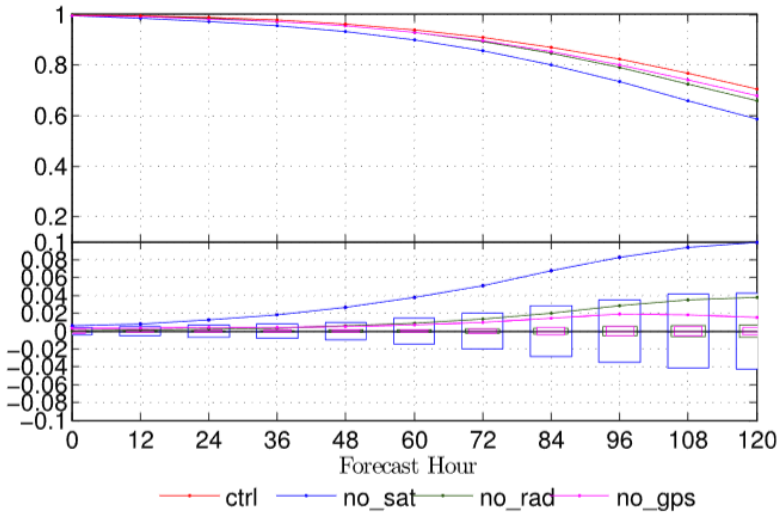
Initial conditions (ICs) are important components of numerical weather prediction (NWP) and represent the balance between collected data and numerical model estimates, these values are obtained through the data assimilation (DA) process. Observational data, together with short-term forecasts, can provide 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 Support Tools

Examples of Support Tools

- **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

3. R2O - Research to Operations

3.3 Transition Workflow

- 🤹 Project Management
  - 🗺️ Roadmap organization
    - 💡 Goals for each version?
    - 🚕 How to get there?
- 🐙 Development and Version Control
  - 📝 All changes are tracked in the repository (for SMNA, this repository is internal)
  - 👷 Operations report any issues found
  - 👩‍🔬 Development and research team investigate problems and propose code changes
    - 🧪 New artifacts are generated and used by operations
  - 📕 Important for operations to log occurrences
    - 🔬 History and system stability assessment
- 🗣️ Science Communication
  - 📄 Through reports and technical notes
  - 📜 Scientific articles

4. Operational Activities

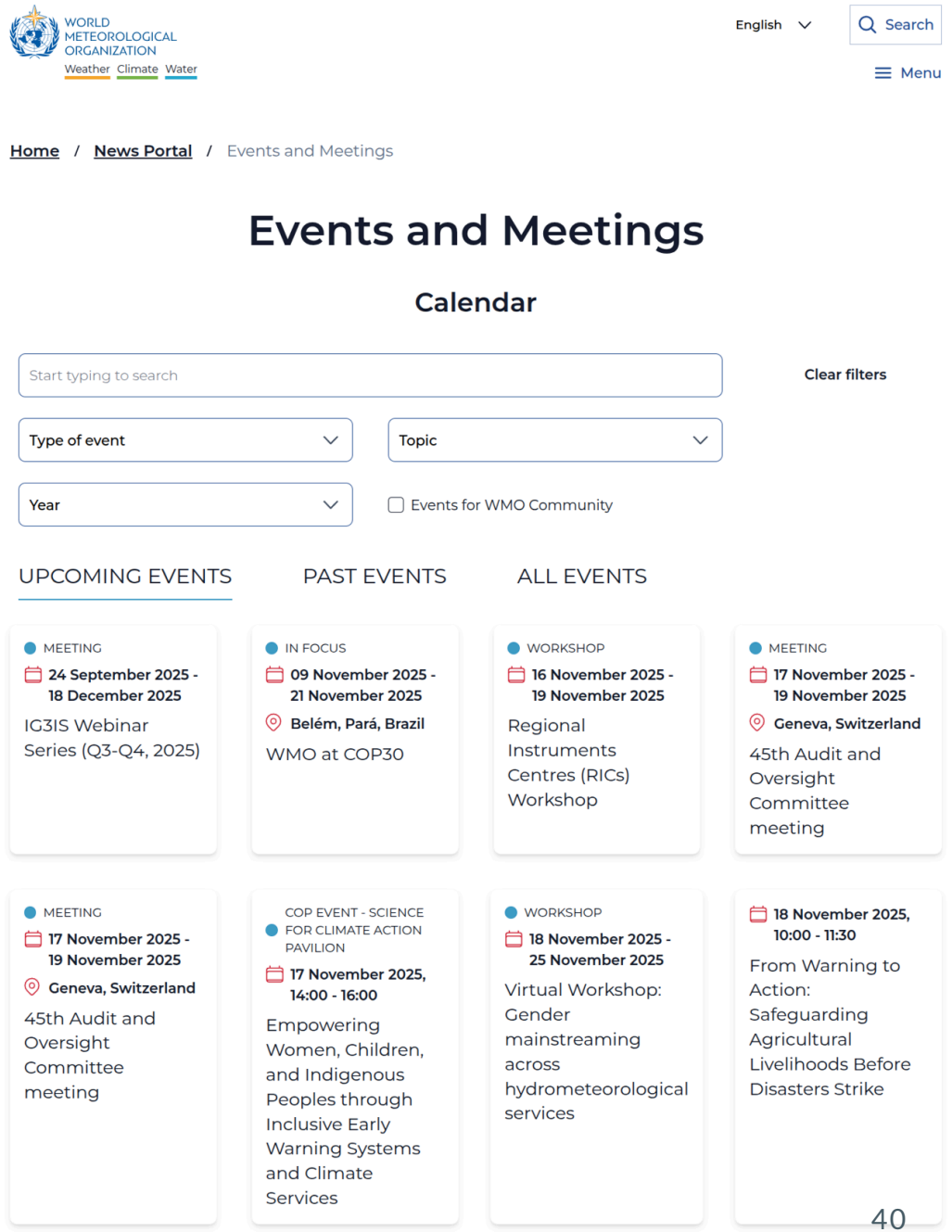
4.1 Operational Cost

- $ Storage cost per SMNA cycle (BAM+GSI)
  - GSI spectral analysis TQ0299L064: ~89MB
  - Spectral background (3x FGAT): ~2GB
  - GSI diagnostic files: ~1.5GB
  - Observation files (AMSUA, GNSS RO, Conventional): ~150MB
  - Other artifacts in the process (GSI): ~3GB
  - Post-processed forecasts for 11 days: ~13GB
- Total: ~20GB
  - 👉 4 cycles per day: ~80GB
  - 👉 1 month: ~10TB
  - 👉 1 year: ~115TB

4. Operational Activities

4.2 Monitoring

- ⚙️ Once the data assimilation system is operational, it is necessary to
  - 🔬 Track daily system simulations in terms of computational performance and analysis/forecast quality (objective evaluation)
  - 🔬 Monitor the operational status of satellite sensors with other centers
  - 🔬 Monitor the dissemination of observation data with other centers (especially at CPTEC, which still does not generate its own observation data)
  - 🤝 WMO Events and Meetings 🔗 link



4. Operational Activities

4.3 Comparison with Other Numerical Products

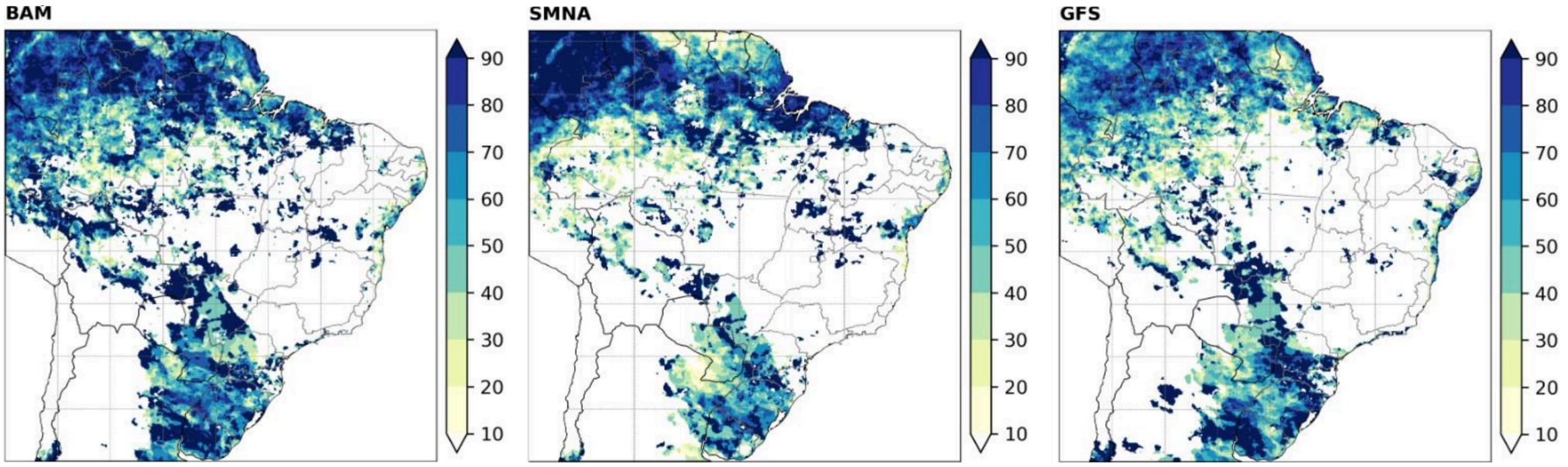
- Monthly evaluation of CPTEC models
  - 📊 The model evaluation team issues technical reports on the skill of global and regional weather prediction models from INPE, alongside the NCEP GFS model
  - ☔ Significant accumulated precipitation events (e.g., >20mm in 24h) are selected to evaluate model performance
  - 💦 Comparison with MERGE (Rozante et al., 2010: Combining TRMM and Surface Observations of Precipitation: Technique and Validation over South America - 🔗 link)

- Models evaluated
  - 👉 BAM (global)
  - 👉 BAM/SMNA (global)
  - 👉 MPAS/MONAN (global)
  - 👉 WRF/CPTEC (regional)
  - 👉 Eta (regional)
  - 👉 BRAMS (regional)

4. Operational Activities

4.3 Comparison with Other Numerical Products

- Hit Percentage (BAM X SMNA X GFS) – May 2025
  - Number of predicted events that occurred ($\frac{hits}{hits+misses}$)

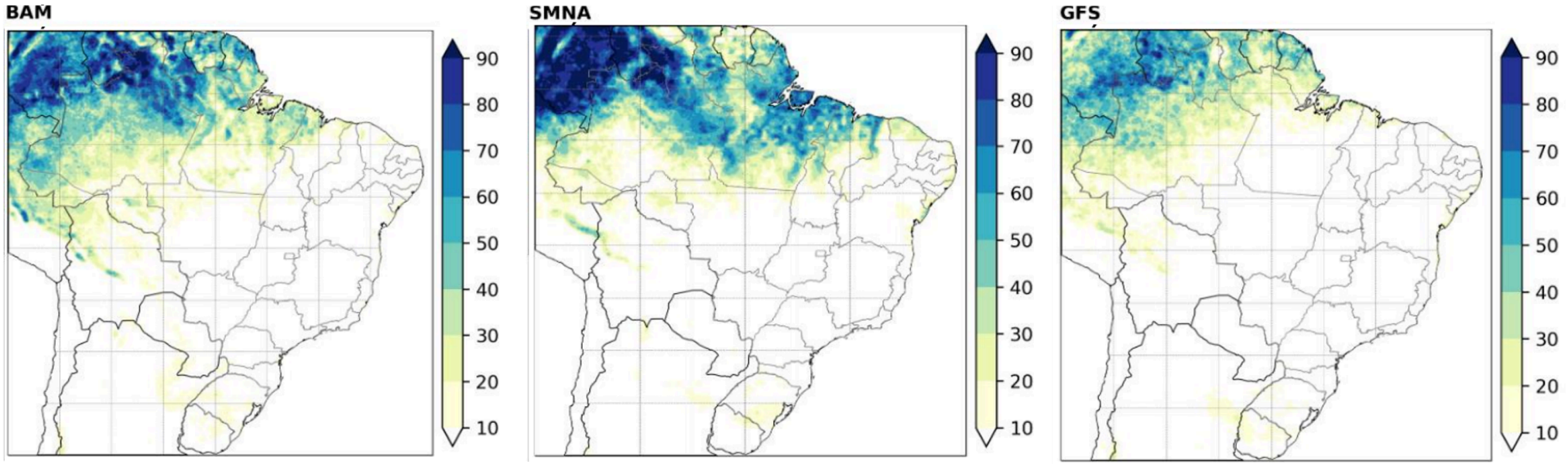


Source: Model Evaluation Group DIPTC (reproduced)

4. Operational Activities

4.3 Comparison with Other Numerical Products

- False Alarm Percentage (BAM X SMNA X GFS) – May 2025
  - Number of predicted events that did not occur ($\frac{fa}{fa+cr}$)

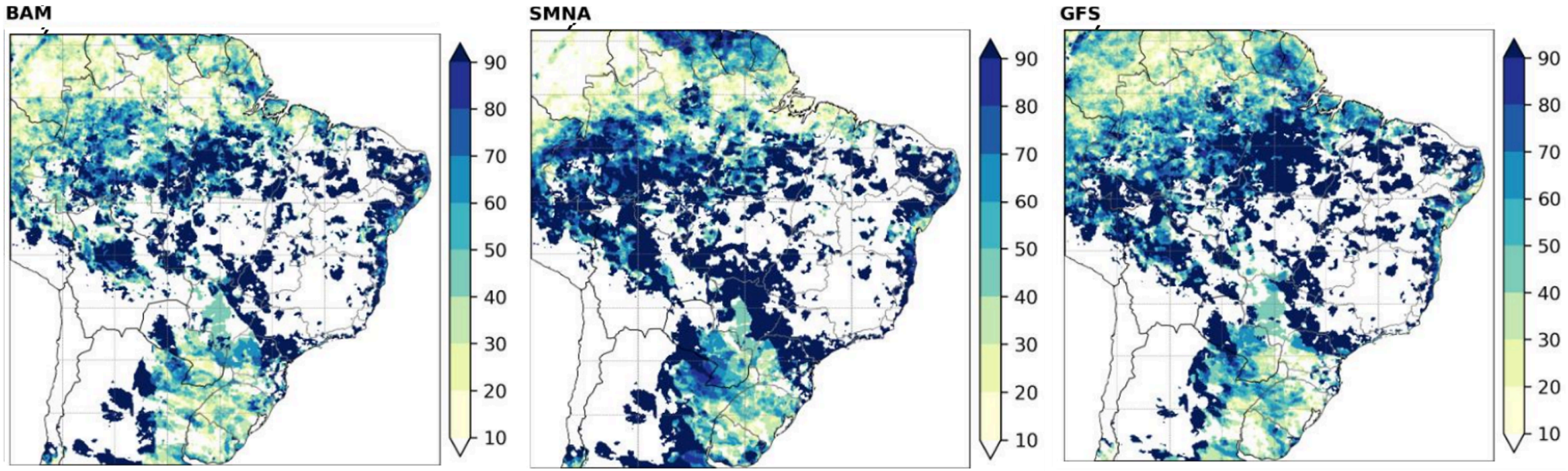


Source: Model Evaluation Group DIPTC (reproduced)

4. Operational Activities

4.3 Comparison with Other Numerical Products

- Miss Percentage (BAM X SMNA X GFS) – May 2025
  - Number of events that occurred but were not forecasted ($\frac{misses}{misses+hits}$)

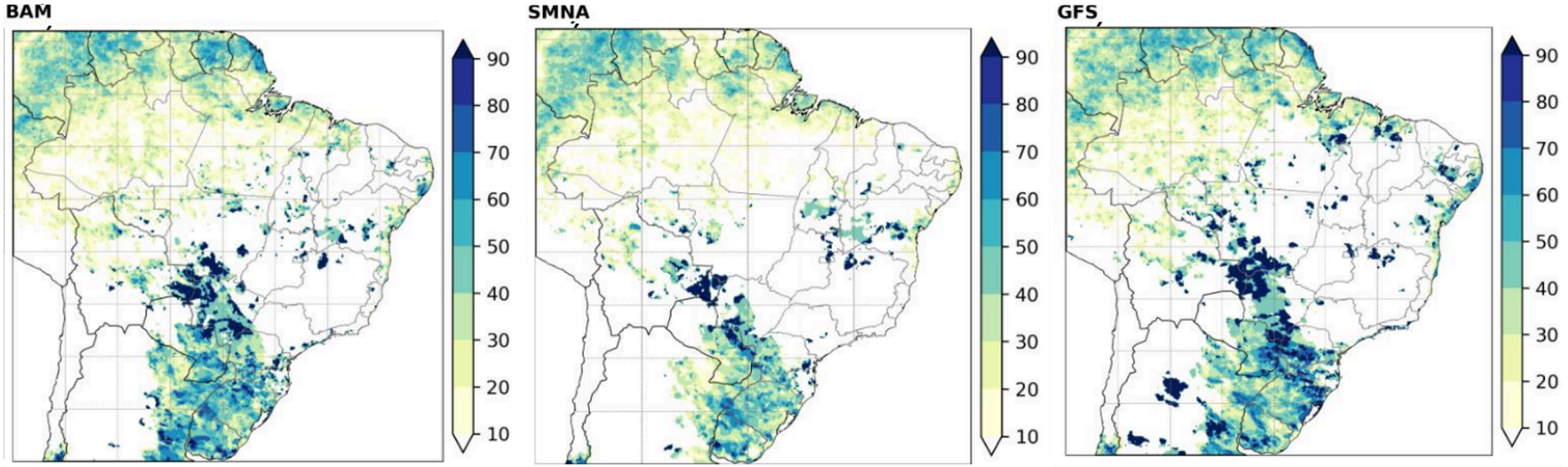


Source: Model Evaluation Group DIPTC (reproduced)

4. Operational Activities

4.3 Comparison with Other Numerical Products

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

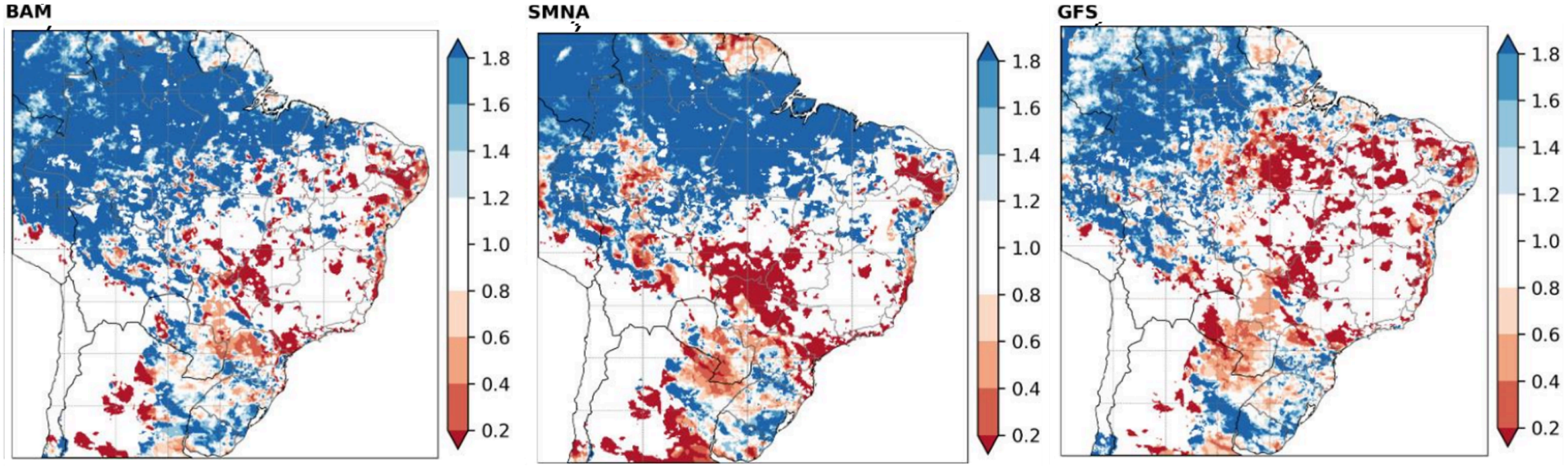


Source: Model Evaluation Group DIPTC (reproduced)

4. Operational Activities

4.3 Comparison with Other Numerical Products

- Bias Ratio (BAM X SMNA X GFS) – May 2025 ($\frac{hits+fa}{hits+misses}$)



Source: Model Evaluation Group DIPTC (reproduced)

5. Conclusions

- **Initial Condition Determination**
  - 🌉 Data Assimilation is the bridge between observations and the numerical model
  - ⚖️ Combines both sources to produce the best estimate of the optimal state of the atmosphere (or ocean, land surface, etc.)

- **Research to Operations**
  - 🛠️ Complexity of the data assimilation framework requires tools to diagnose issues in both operational and research environments
  - 🤝 Continuous cycles of development and fixes are applied in the operational system

- **Operational Activities**
  - 👷 Data assimilation in a NWP center requires collaboration among **modeling**, **scientific computing**, **satellite**, and **database** teams to ensure proper analysis generation
  - 🗣️ Continuous improvement relies on monitoring non-conventional observations and maintaining communication with international satellite groups

Thank You

https://github.com/GAD-DIMNT-CPTEC
https://cfbastarz.github.io
carlos.bastarz@inpe.br