Estimating and Reducing Precipitation Uncertainty in Atmospheric Systems using Data Assimilation: A Local-Scale Forecasting Study in The Andes

Danilo[2], **OLucia**[2], **ManuelZ**[2]

[2]School of Applied Sciences and Engineering, Universidad EAFIT

*For correspondence:
dasuarezh@eafit.edu.com (EAFIT)

Present address:
[†]Medellín,Antioquia, Colombia

Abstract   The entire document is under construction and subject to modification

1   Introduction

Numerical Weather Prediction (NWP) is a forecasting method based on a set of equations that describe the behavior and interaction of fluids in the atmosphere. These equations are implemented within a spatial discretization scheme, where it is possible to make predictions based on the system's temporal evolution. Since the equations of atmospheric dynamics (e.g., the Navier–Stokes equations) are fundamentally a system of nonlinear and nonlocal partial differential equations, the time integration of an atmospheric model constitutes an initial value problem. Consequently, producing skillful forecasts requires both a realistic representation of the atmosphere and an accurate specification of the initial conditions (*Kalnay, 2003*).

According to *Lackmann* (*2011*), the main sources of error in a model can be classified as errors in the initial conditions, errors in physics or dynamics and intrinsic predictability limit. Errors in the initial conditions arise from inaccuracies in the observations themselves and from shortcomings in the DA systems, errors in physics or dynamics depend largely on the scale and the phenomenon being simulated and also includes errors such as inadequate spatial resolution or a poor representation of topography. Finally the predictability limit is a consequence of the chaotic nature of atmospheric dynamics (*Lorenz and Haman, 1996*). In general, it is not straightforward to determine how long it takes for infinitesimal differences in the initial condition to amplify sufficiently to eliminate the model's forecasting skill.

Given the nature of the errors that arise in models, the challenges associated with improving forecasts are considerable. Nevertheless, progress in this area has been remarkable *Kalnay* (*2003*). The improvement is the result of increased computing power, better representation of physical processes, greater availability of observations and the use of more accurate data assimilation methods, with the latter being the focus of our discussion here.

DA can be classically approached in two ways: variational and statistical (filtering), in both we seek an optimal solution. Statistically we will, for example, seek a solution with minimum variance, whereas variationally we will seek a solution that minimizes a suitable cost function. Also we can refer to several flavors in DA: nudging methods, reduced methods, ensemble methods and hybrid methods (*Asch et al., 2016*).

data assimilation review...

2   Methodology

2.1   Study Area and target precipitation event

The Medellín–Aburrá River is located in Antioquia, Colombia over The Andes region. At present, this watercourse flows through a highly urbanized basin and has posed multiple challenges over the years in terms of sanitation, land use planning, disaster risk management, among others. For this study, the watershed was delineated with an area of 711.7 km^2, a perimeter of 191.5 km, an average slope of 7.58%, a mean elevation of 2434 m a.s.l., a maximum elevation of 3125 m a.s.l., and a minimum elevation of 1424 m a.s.l. The main channel has a length of 52.4 km and an average slope of 3.25%.

Between november 25 and 26 of 2018, a precipitation event was recorded that triggered several emergency reports in the Aburrá Valley, the most catastrophic of which occurred in the municipality of Copacabana. According to official data from the National Unit for Disaster Risk Management (*UNGRD, 2018*) and various media reports, flooding took place as a result of the overflow of the Aburrá–Medellín River, which caused the collapse of a retaining wall. Although the disaster caused no loss of human life, it affected approximately 600 people in multiple ways.


```
IMG/01_maptotal.png
```

Figure 1. Study area localization, basin elevation and Digital Elevation Model of watershed

In *Figure 2* the discharge and precipitation time series can be observed. The maximum river level was recorded on November 26 around 01 UTC, reaching nearly 600 cm. It is important to note that rainfall events had also occurred in the preceding days, which increased the river level, but in those cases the level returned to a reference value (horizontal black line in 2). On the day prior to the maximum river level (vertical red line in 2) , a precipitation event was recorded; however, on this occasion the river level did not return to the reference value before the next precipitation event occurred. This subse-

quent event became the trigger of the emergency, as it produced a river response of great magnitude.

This situation highlights two important aspects: first, that antecedent soil moisture in a watershed is a key factor when assessing its response to rainfall events; and second, that the ability to accurately anticipate precipitation events plays a decisive role.


```
IMG/01_evento_nozo.png
```
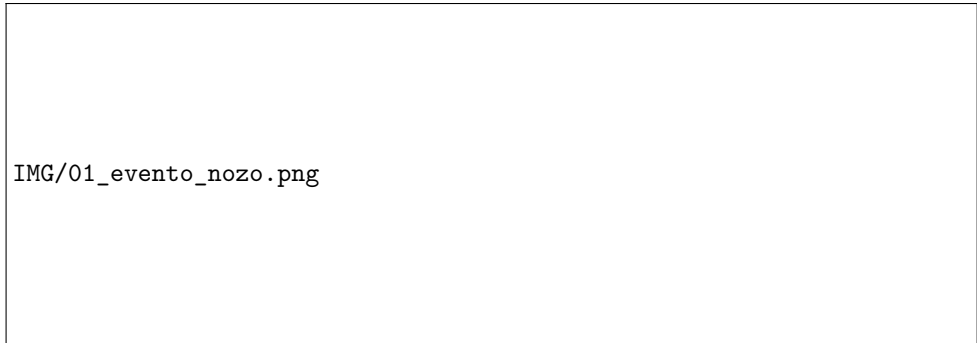
Figure 2. Series of precipitation and water level in the watershed around days of target precipitation event.

2.2   Observations

2.2.1   Weather radar

A dual-polarization C-band Doppler weather radar from Enterprise Electronics Corporation (ECC) and operated by the Early Warning System of Medellín and the Aburrá Valley (SIATA, for its acronym in Spanish). The radar is located at an altitude of 2800 m a.s.l. and performs quasi-horizontal scans, also known as Plan Position Indicator (PPI), approximately every 5 minutes at four elevation angles (0.5°, 1.0°, 2.0°, and 4.0°). In addition, the radar alternates PPIs with vertical scans, or Range Height Indicators (RHI), at azimuths of 0°, 90°, 180°, and 270°. Depending on the elevation angle, the radar provides horizontal coverage with a range of 100 to 250 km (the red circle in 1 has 120 km at 1.0° elevation). The scanning strategy of the instrument collects information through circular sweeps around the radar coverage area. Consequently, the data are stored in a polar coordinate system (range, elevation, azimuth), which is the native format for collection and storage. To enable a simpler use of this information it is necessary to interpolate the data into a Cartesian coordinate system (x, y, z). The interpolation makes use of the RadX computational routines (*Dixon and Javornik, 2016*), in particular the method REORDER (*Oye and Case, 1995*).

2.2.2   Weather stations

IDEAM + SIATA

(a) D01 **(b)** D03 **(c)** D04
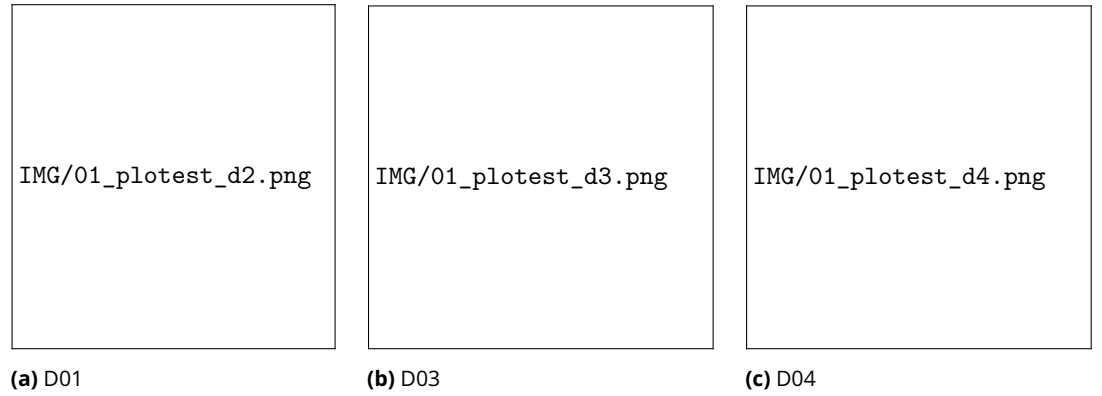
Figure 3. 2018-11-26 00:30:00

2.2.3 Precipitation

Inverse Distance Weighted (IDW) deterministic interpolation method developed by the U.S. National Weather Service *Workneh et al.* (*2024*) and it is a method for estimating the value of a variable at a location where no information is available, by using surrounding points where measurements exist. It is carried on the assumption that the attribute value of unknown point is weighted average of the nearby known values and its one of the more popular methods adopted by geoscientists *Lu and Wong* (*2008*).

$$Z_j = \frac{\sum_{i=1}^{n} \frac{Z_i}{d_{ij}^{p}}}{\sum_{i=1}^{n} \frac{1}{d_{ij}^{p}}} \tag{1}$$

Where Z_j is the value at the unknown point, Z_i the value at the known point, d_{ij} the distance to the known point, and finally p a method sensitivity parameter. In this work, IDW interpolation is used to create a precipitation field on a regular grid from ground measurement stations, with a sensitivity parameter set to 6.
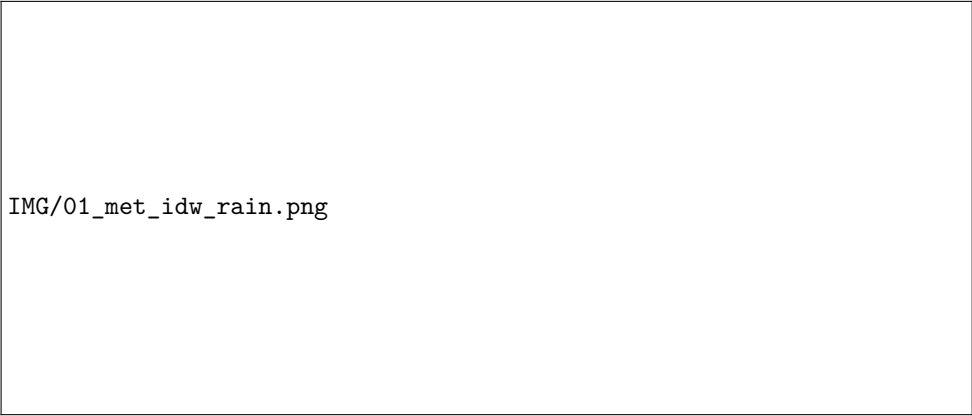
For this study, the rainfall network of SIATA (*Figure 4*) was used. At the time of data revision, a total of 145 stations were identified as part of the measurement network, 126 stations were located within the IDW interpolation domain (black dashed rectangle), and 110 stations were located precisely within the study watershed. In adittion, between 109 and 111 stations provided valid information for interpolation during the validation period (2018-11-25 12:00 to 2018-11-26 12:00) The grid of interpolation domain is 500 x 500 m.

Figure 4. basin ppt from ground stations for 2018-11-25 23:00:00

The IDW interpolation is performed over all points located within the dashed-line box. Then, using weather radar information, an overlay is applied to the IDW rainfall field, where values in locations without radar measurements are set to zero, while in other cases the original values are retained. Finally, the field is clipped so that only the precipitation values within the watershed are preserved (*Figure 5*).

Figure 5. basin ppt from ground stations for one step

Following the procedure described above, a precipitation time series over the basin can be obtained at 10-minute intervals. This time series serves as the reference measurement and is compared with the model precipitation over the same region for validation purposes.

2.3 NWP

The computational implementation of NWP models can be quite complex, as it usually involves a numerical discretization scheme of equations. Therefore, numerical convergence criteria must be satisfied according to the spatiotemporal scale of the model. Nevertheless, the foundations of any such model lie in the governing equations of the evolution of the atmosphere, which can be summarized as follows:

- Conservation of momentum:

$$\frac{d\mathbf{v}}{dt} = -\alpha\nabla p - \nabla\phi + \mathbf{F} - 2\mathbf{\Omega}\times\mathbf{v} \tag{2}$$

- Conservation of mass:

$$\frac{\partial\rho}{\partial t} = -\nabla(\rho\mathbf{v}) \tag{3}$$

- State for perfect gases:

$$p\alpha = RT \tag{4}$$

- Conservation of energy

$$Q = C_p\frac{dT}{dt} - \alpha\frac{dp}{dt} \tag{5}$$

- Conservation of water vapor mixing ratio

$$\frac{\partial\rho q}{\partial t} = -\nabla(\rho\mathbf{v}q) + \rho(E - C) \tag{6}$$

Where $\mathbf{v} = (u, v, w)$, $\frac{d\mathbf{v}}{dt}$ is the relative acceleration of an air parcel, $\alpha\nabla p$ is the pressure gradient force, $\nabla\phi$ is the apparent gravity, $\mathbf{F}$ is the frictional force, and $2\mathbf{\Omega}\times\mathbf{v}$ is the Coriolis force. ρ is the density of an air parcel, p is the pressure, α is the specific volume, T is the temperature and R is the gas constant for air. Q is the heat, C_p is the specific heat at constant pressure, q is the water vapor mixing ratio, E is the evaporation, and C is the condensation.

In addition to the aforementioned considerations, it is important to note that there are processes that cannot be resolved at the spatial scale of the grid defined for the model and therefore cannot be explicitly resolved (*subgrid-scale processes*), but their contributions cannot be ignored, because these subgrid-scale processes depend on and in turn affect the large-scale fields and processes that are explicitly resolved by numerical models (*Pu and Kalnay, 2018*). These include turbulent motions that may range from a few centimeters to several kilometers, as well as molecular-scale processes such as condensation, evaporation, friction, and radiation. Consequently, parameterizations based on physical or statistical approximations are required in order to estimate their effects. Among the most relevant of these processes are: hydrological processes, cloud processes, boundary layer processes, chemical and radiative processes, surface processes (land or ocean).

2.3.1 WRF

The Weather Research and Forecasting (WRF) model is a three-dimensional, limited-area model that allows for dynamic downscaling. It is widely used in atmospheric research and

numerical weather prediction, as it is often employed as a tool for operational forecasting. The model is jointly developed by the National Center for Atmospheric Research (NCAR) and the National Centers for Environmental Prediction (NCEP). The dynamical core used here is ARW, WRF-ARW it is portable, efficient, and can be executed on High-Performance Computing (HPC) platforms.

The version of the model used for this work corresponds to WRF-4.5.2, released on 22 December 2023. This version includes the preprocessing package (WPS), data assimilation (WRF-DA) module, and the ARW solver core, which among its main features incorporates: a complete nonhydrostatic equation set with; both regional and global applications; full Coriolis and curvature terms; one-way and two-way nesting; moving nests; hybrid sigma–pressure vertical coordinates based on mass; vertically variable grid spacing; map scale factors for four projections; Arakawa C-grid; second and third order Runge–Kutta time integration options; scalar conservation form for prognostic variables; second and sixth order advection options; upper-boundary absorption and Rayleigh damping; coupled ocean models; and a comprehensive set of physical parameterizations. (*Skamarock et al., 2019*). The WRF-DA (WRF Data Assimilation) system is framework that Supports variational, statistical and hybrid data assimilation techniques, allowing users to integrate observations dynamically.

(a) WRF Domains setup (D01, D02, D03, D04)   **(b)** WRF Domains setup (D03, D04)
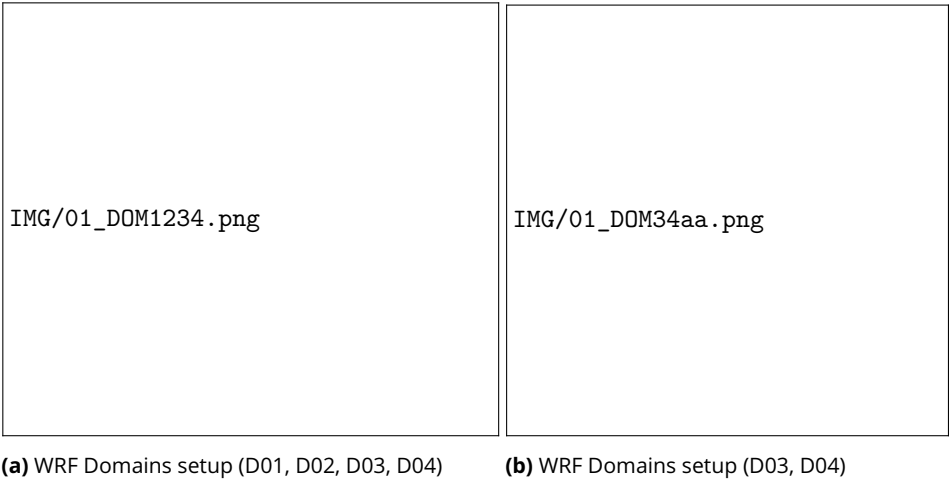
Figure 6. WRF Domains setup. Radar Cover (orange circle), watershed limits (red contour), topography isoline 1800m (black contour)

	D01	**D02**	**D03**	**D04**
Δ_{xy} (m)	18000	3600	900	300
Nx X Ny	240 X 200	391 X 421	437 X 437	331 X 331
Nz	80	80	80	80
Time step (s)	54	10	2	0.5
Nudging*	True	False	False	False

Table 1. Model dimensional setup for every domain. Analysis nudging*

	D01	D02	D03	D04
Microphysics	Thompson (*Thompson et al., 2008*)	Thompson	Thompson	Thompson
LW Radiation	RRTMG (*Iacono et al., 2008*)	RRTMG	RRTMG	RRTMG
SW Radiation	Dudhia (*Dudhia, 1989*)	Dudhia	Dudhia	Dudhia
Surface layer	MM5 (*Jiménez et al., 2012*)	MM5	MM5	MM5
Land surface	Noah (*Tewari et al., 2004*)	Noah	Noah	Noah
Cumulus	Kain-Fritsch (*Kain, 2004*)	Kain-Fritsch	None	None
PBL	YSU (*Hong et al., 2006*)	YSU	YSU	SMS-3DTKE (*Zhang et al., 2018*)
Land Use	Modis (15" ~ 500 m) (*Friedl et al., 2002*)	Modis (15" ~ 500 m)	Modis (15" ~ 500 m)	Modis (15" ~ 500 m)
Topography	GTOPO30 (30" ~ 1 km) (*USGS, 1996*)	GTOPO30 (30" ~ 1 km)	ALOS PALSAR* (0.4" ~ 50 m) (*ASF, 2015*)	ALOS PALSAR* (0.4" ~ 50 m)

Table 2. Model parametrization setup and geographical database in every domain. The original ALOS PALSAR resolution of 12.5 m was resampled to 50 m*.

A comprehensive literature review was conducted to support and define the parameterizations employed..

The initial and boundary conditions for D01 come from NCEP operational Global Forecast System *NCEP* (*2015*), the datasets contains analysis and forecast grids on a 0.25° x 0.25° global latitude longitude grid. Grids include analysis and forecast time steps at a 3 hourly interval from 0 to 240, and a 12 hourly interval from 240 to 384. Model forecast runs occur at 00, 06, 12, and 18 UTC daily. The simulation for this study starts at 2018-11-25 12:00:00 UTC.

2.3.2 Model precipitation and DA

Surface precipitation rate cannot be resolved directly at the scale of the grid, consequently, rainfall must be parameterized. The need for parameterization arises because moist processes that control precipitation (autoconversion, accretion, deposition, riming, melting, evaporation) occur on scales far below typical grid spacings and are strongly nonlinear. In this study, a double-moment microphysics scheme (Thompson microphysics) is used (*Thompson et al., 2008*), which includes calculations for five hydrometeor classes: cloud water, rain, graupel, snow, and ice.

In operational NWP, the standard practice is to assimilate observations that directly constrain the model's prognostic state (winds, temperature, humidity, pressure, hydrometeor mixing ratios), and then verify precipitation as a diagnostic outcome of moist processes. Direct precipitation is also possible, but general difficulties in precipitation assimilation are reported mainly due to the strong nonlinearity of moist processes and the non-Gaussian statistics of precipitation variables *Kotsuki et al.* (*2017*)

2.3.3 Reflectivity estimation

Reflectivity is an indirect measure of the amount and size of hydrometeors in the atmosphere provided by radar. In practice, comparison between radar measurements and the results of an atmospheric model is required, so a transformation must be applied to convert the model output into reflectivity values. The approach relies on the mixing ratio values of different types of water particles in the atmosphere and assumes certain conditions regarding the shape, density, and size distribution of the particles. the following assumptions are considered:

- The particles in the atmosphere are spherical and have a constant density according to each particle.
- The distribution of particle sizes, in terms of their diameter, follows an exponential form given by $N(D) = N_0 \exp(-\lambda D)$, where the intercept parameter is taken as a constant value according to each particle.

The number of hydrometeor types used in WRF depends on the implementation of each microphysics parameterization. However, the method employed considers only three types of particles (rain, snow, and graupel), the equivalence in reflectivity factor depends on the state of the particle:

$$Ze = \begin{cases} \Gamma(7) \, N_0 \, \lambda^{-7}, & \text{if liquid particle (rain)} \\[2ex] \Gamma(7) \, N_0 \, \lambda^{-7} \left(\dfrac{\rho_s}{\rho_i} \right)^2 \left(\dfrac{|K|_i^2}{|K|_l^2} \right), & \text{if solid particle (snow, graupel)} \end{cases} \tag{7}$$

Where Γ denotes the gamma function and λ is the slope parameter, ρ represents the densities of snow and ice, and K denotes the dielectric factor of ice and liquid water.

From the preceding expressions, the reflectivity factors for the three hydrometeor types mentioned above can be derived. The implementation of this methodology is available in *Ladwig* (*2017*). For further details on the equations, assumptions, and parameter values used, the reader is referred to *Stoelinga* (*2005*). Given the implications involved in converting model moisture field outputs to reflectivity, this approach does not allow a direct comparison with radar measurements. Nevertheless, it is highly useful for identifying storm patterns and obtaining an approximation of the reflectivity values that the model is able to represent.

2.4 Data Assimilation

According to *Carrassi et al.* (*2018*), whose proposed notation will be adopted in this section, the observations and the model are complementary sources of information, but both incomplete and inaccurate. Data assimilation provides the conceptual and methodological tools to tackle the problem by extracting synergies between model and observations by exploiting their respective informational content.

Assume that a model of the natural processes of interest is available as a discrete stochastic-dynamical system and fixed model parameters:

$$\mathbf{x}_t = \mathcal{M}_{k:k-1}(\mathbf{x}_{k-1}) + \boldsymbol{\eta}_k \tag{8}$$

Where $\mathbf{x}_t \in \mathbb{R}^m$ is state vector, $\mathcal{M}_{k:k-1}(\mathbf{x}_{k-1}) : \mathbb{R}^m \to \mathbb{R}^m$ is a function, typically nonlinear and possibly chaotic in nature, that relates the system state from time t_{k-1} to time t_k, and finally, $\boldsymbol{\eta}_k \in \mathbb{R}^m$ is the model error represented as an additive stochastic term. In this study $\mathbf{x}_t = \mathcal{M}_{k:k-1}(\mathbf{x}_{k-1}) = \mathcal{WRF}_{k:k-1}(\mathbf{x}_{k-1})$.

Additionally, observations of $x_t \in \mathbb{R}^d$ are available and can be expressed as components of an observation vector y_t:

$$\mathbf{y}_k = \mathcal{H}_t(\mathbf{x}_k) + \epsilon_k \tag{9}$$

Equation *Equation 9* defines the observational operator $\mathcal{H} : \mathbb{R}^m \to \mathbb{R}^d$, which is could be nonlinear and may involve interpolations, convolutions, or spectral transformations. Similarly, ϵ_k is an additive stochastic term that accounts for measurement instrument errors, deficiencies in the formulation of the observational operator and representativeness error.

Once the model and the data or observations are available as complementary sources of information, the main goal is the estimation of the system state and the outcome of the estimation process, according to Bayes' theorem, is the *a posteriori* distribution of the process $\mathbf{x}$ conditioned on the data $\mathbf{y}$:

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})} \tag{10}$$

Both the model ($\boldsymbol{\eta}_k$) and observational error sequences (ϵ_k) are assumed to be independent in time, mutually independent and distributed according to $p(\boldsymbol{\eta}_k)$ y $p(\epsilon_k)$. Also the probability density functions are related to prior and likelihood terms of Bayes's rule as follows:

$$p(\mathbf{x}_k|\mathbf{x}_{k-1}) = p(\boldsymbol{\eta}_k) = p\big[\mathbf{x}_k - \mathcal{M}_{k:k-1}(\mathbf{x}_{k-1})\big] \tag{11}$$

$$p(\mathbf{y}_k|\mathbf{x}_k) = p(\epsilon_k) = p\big[\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)\big] \tag{12}$$

If we define the sequences of system states $\mathbf{x}_{K:0}$ and observations $\mathbf{y}_{K:1}$ within the interval $[t_o, t_k]$. Since the observational errors area assumed to be independent in time:

$$p(\mathbf{y}_{K:1} \mid \mathbf{x}_{K:0}) = \prod_{k=1}^{K} p(\mathbf{y}_k \mid \mathbf{x}_k) = \prod_{k=1}^{K} p\big[\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)\big] \tag{13}$$

If we presume $\mathcal{M}_{k:k-1}$ as markovian process:

$$p(\mathbf{x}_{K:0}) = p(\mathbf{x}_0)\, p(\mathbf{x}_1|\mathbf{x}_0)\, p(\mathbf{x}_2|\mathbf{x}_1\mathbf{x}_0) \dots p(\mathbf{x}_k|\mathbf{x}_{k-1}, ..., \mathbf{x}_0) \tag{14}$$

$$p(\mathbf{x}_{K:0}) = p(\mathbf{x}_0) \prod_{k=1}^{K} p(\mathbf{x}_k \mid \mathbf{x}_{k-1}) = p(\mathbf{x}_0) \prod_{k=1}^{K} p\big[\mathbf{x}_k - \mathcal{M}_{k:k-1}(\mathbf{x}_{k-1})\big] \tag{15}$$

$p(\mathbf{y})$ is the marginal distribution of the observation and is independent of $\mathbf{x}$, so it will be treated as normalization coefficient.

$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) \tag{16}$$

Considering *Equation 16*, we can use *Equation 13* and *Equation 15*:

$$p(\mathbf{x}_{K:0}|\mathbf{y}_{K:1}) \propto p(\mathbf{x}_0)\prod_{k=1}^{K} p\big[\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)\big] \prod_{k=1}^{K} p\big[\mathbf{x}_k - \mathcal{M}_{k:k-1}(\mathbf{x}_{k-1})\big] \tag{17}$$

One of the most widely used approaches, due to its simplicity and computational efficiency, is that the observational and model noises are assumed to be Gaussian distributed, furthermore assume the existence of a prior, commonly referred to as the *background* ($\mathbf{x}^b \in \mathbb{R}^m$), for the initial state at the beginning of the window:

$$\boldsymbol{\eta}_t \sim \mathcal{N}(0,\mathbf{Q}), \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}(0,\mathbf{R}), \quad \boldsymbol{\tau}_t = \mathbf{x}_0 - \mathbf{x}^b \sim \mathcal{N}(0,\mathbf{B}) \tag{18}$$

Also we are looking for maximizing the *a posteriori* probability:

$$\arg\max_{\mathbf{x}}\ p(\mathbf{x}_{K:0}|\mathbf{y}_{K:1}) = \arg\max_{\mathbf{x}}\ p(\mathbf{x}_0)\prod_{k=1}^{K} p\big[\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)\big] \prod_{k=1}^{K} p\big[\mathbf{x}_k - \mathcal{M}_{k:k-1}(\mathbf{x}_{k-1})\big] \tag{19}$$

$$\arg\max_{\mathbf{x}}\ p(\mathbf{x}_{K:0}|\mathbf{y}_{K:1}) = \arg\min_{\mathbf{x}}\ -\Bigg[\log p(\mathbf{x}_0) + \sum_{k=1}^{K} \log p\big[\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)\big] + \sum_{k=1}^{K} \log p\big[\mathbf{x}_k - \mathcal{M}_{k:k-1}(\mathbf{x}_{k-1})\big]\Bigg] \tag{20}$$

$$\arg\max_{\mathbf{x}}\ p(\mathbf{x}_{K:0}|\mathbf{y}_{K:1}) = \arg\min_{\mathbf{x}}$$

$$\tfrac{1}{2}(\mathbf{x} - \mathbf{x}^b)^{\mathrm{T}}\mathbf{B}_0^{-1}(\mathbf{x} - \mathbf{x}^b)$$

$$+ \sum_{k=0}^{K} \tfrac{1}{2}\big[\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)\big]^{\mathrm{T}}\mathbf{R}^{-1}\big[\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)\big] \tag{21}$$

$$+ \sum_{k=0}^{K} \tfrac{1}{2}\big[\mathbf{x}_k - \mathcal{M}_{k:k-1}(\mathbf{x}_{k-1})\big]^{\mathrm{T}}\mathbf{Q}^{-1}\big[\mathbf{x}_k - \mathcal{M}_{k:k-1}(\mathbf{x}_{k-1})\big]$$

The result of maximizing the *a posteriori* probability will be called *analysis* and it is denoted by $\mathbf{x}^a$.

$$\mathbf{x}^a = \arg\max_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}) = \arg\min_{\mathbf{x}} \mathcal{J}(\mathbf{x}) \tag{22}$$

The cost function contains three components: errors associated with initial conditions, observations, and dynamical model.:

$$\mathcal{J}(\mathbf{x}) \;=\; \mathcal{J}^b(\mathbf{x}) + \mathcal{J}^o(\mathbf{x}) + \mathcal{J}^q(\mathbf{x}) \tag{23}$$

$$\mathcal{J}(\mathbf{x}) \;=\; \frac{1}{2}\left\|\mathbf{x} - \mathbf{x}^b\right\|^2_{\mathbf{B}_0^{-1}} \;+\; \frac{1}{2}\sum_{k=0}^{K}\left\|\mathbf{y}_k - \mathcal{H}(\mathbf{x}_{t-1})\right\|^2_{\mathbf{R}_t^{-1}} \;+$$
$$\frac{1}{2}\sum_{k=0}^{K}\left\|\mathbf{x}_k - \mathcal{M}_{k:k-1}(\mathbf{x}_{k-1})\right\|^2_{\mathbf{Q}_t^{-1}} \tag{24}$$

Since this study focuses on the estimation of the model's initial condition, several simplifications can be introduced. First, the model is not evolved in time during the optimization process. Furthermore, we assume the strong constraint considering perfect the model dinamic $\mathbf{x}_t = \mathcal{M}_{t-1,t}(\mathbf{x}_{t-1})$, so $\mathcal{J}^q$ term is zero. These assumptions allow for certain simplifications and the resulting expression is also known as 3DVAR:

$$\mathcal{J}(\mathbf{x}) \;=\; \frac{1}{2}\left\|\mathbf{x} - \mathbf{x}^b\right\|^2_{\mathbf{B}^{-1}} \;+\; \frac{1}{2}\left\|\mathbf{y} - \mathcal{H}\{\mathbf{x}\}\right\|^2_{\mathbf{R}^{-1}} \tag{25}$$

Using incremental variational formulation *Courtier et al.* (*1994*) as used in WRFDA is one of the most common approach. In general, variational DA problems are badly conditioned and the rate of convergence of the minimization algorithms depends on the conditioning of the Hessian of the cost function (*Asch et al., 2016*). Preconditioning is a technique for improving the condition number, for that we make a change of variable:

$$\mathbf{x} = \mathbf{x}^b + \delta\mathbf{x} \tag{26}$$

Considering the linearization of $\mathcal{H}$ (**H**) and the approximation:

$$\mathcal{H}\{\mathbf{x}\} \approx \mathcal{H}\{\mathbf{x}\} + \mathbf{H}\delta\mathbf{x} \tag{27}$$

and define:
$$\mathbf{d} = \mathbf{y} - \mathcal{H}\{\mathbf{x}\} \tag{28}$$

finally:

$$\mathcal{J}(\mathbf{x}) \;=\; \frac{1}{2}\left\|\delta\mathbf{x}\right\|^2_{\mathbf{B}^{-1}} \;+\; \frac{1}{2}\left\|\mathbf{d} - \mathbf{H}\delta\mathbf{x}\}\right\|^2_{\mathbf{R}^{-1}} \tag{29}$$

For optimization we start with $\delta\mathbf{x} = \mathbf{0}$ and $\mathbf{x} = \mathbf{x}^b$, at the end of the cost function minimization we get the analysis $\mathbf{x}^a = \mathbf{x}$.

According to *Bannister* (*2017*), the background error covariance matrix **B** is never explicitly computed in the model space **x** (u, v, T, q, p_s), even modern computers are incapable of dealing with such large matrices. Instead, it is represented through matrix decomposition $\mathbf{B} = \mathbf{U}\mathbf{U}^T$, and the cost function is minimized in the control variable space **v**, which is related to the model variables via the control variable transform:

$$\delta\mathbf{x} = \mathbf{U}\mathbf{v} \tag{30}$$

$$\mathcal{J}(x) = \frac{1}{2}(\delta \mathbf{x}^T \mathbf{B}^{-1} \delta \mathbf{x}) + \frac{1}{2} \|\mathbf{d} - \mathbf{H}\delta \mathbf{x}\}\|^2_{\mathbf{R}^{-1}} \tag{31}$$

$$\mathcal{J}(\mathbf{v}) = \frac{1}{2} \|\mathbf{v}\|^2_{\mathbf{U}^T\mathbf{B}^{-1}\mathbf{U}} + \frac{1}{2} \|\mathbf{d} - \mathbf{H}\mathbf{U}\mathbf{v}\}\|^2_{\mathbf{R}^{-1}} \tag{32}$$

$$\mathcal{J}(\mathbf{v}) = \frac{1}{2} \|\mathbf{v}\|^2_{\mathbf{I}} + \frac{1}{2} \|\mathbf{d} - \mathbf{H}\mathbf{U}\mathbf{v}\}\|^2_{\mathbf{R}^{-1}} \tag{33}$$

The matrix $\mathbf{U}$ consists of a series of operations $\mathbf{U} = \mathbf{U}_p \mathbf{U}_v \mathbf{U}_h$ and represents the stages of covariance modeling: $\mathbf{U}_h$ through recursive filters and $\mathbf{U}_v$ through the EOF decomposition stand for horizontal and vertical transforms, respectively. $\mathbf{U}_p$ via linear regression is the physical transform and converts the increment in the control variable space to model variable space. As evidenced in *Sun et al.* (*2016*) and *Gopalakrishnan* (*2019*), the choice of control variables for this study is a formulation that uses zonal velocity (u), meridional velocity (v), temperature (T), surface pressure (P_s) and pseudo relative humidity (RH_s), that is, in this case the transform $\mathbf{U}_p$ only involves the conversion analysis increments of relative humidity to mixing ratio and there is no cross-corelation between variables:

$$\begin{pmatrix} u \\ v \\ T \\ P_s \\ q \end{pmatrix} = \begin{pmatrix} I & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 \\ 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & C_{q,rh} \end{pmatrix} \begin{pmatrix} u \\ v \\ T \\ P_s \\ RH_s \end{pmatrix} \tag{34}$$

2.4.1 Background error modeling

$\mathbf{B}$ plays an important role because it ensures a balance-preserving transfer of observational information to the unobserved parts of the model domain. A common method to model the BE covariance matrix is National Meteorological Center (NMC) method (*Parrish and Derber, 1992*). In this method the model errors are approximated as the difference between two different forecasts that are valid at same time, typically with 12h and 24h forecast lead time.

$$\mathbf{B} = \overline{\left(\varepsilon\varepsilon^T\right)} \approx \overline{\left(\mathbf{x}^{24} - \mathbf{x}^{12}\right)\left(\mathbf{x}^{24} - \mathbf{x}^{12}\right)^T} \tag{35}$$

$$\mathbf{B} = \overline{\left(\varepsilon\varepsilon^T\right)} \approx \overline{\left(\mathbf{x}^{24+3} - \mathbf{x}^{12+3}\right)\left(\mathbf{x}^{24+3} - \mathbf{x}^{12+3}\right)^T} \tag{36}$$

It is important to include forecast differences valid at different parts of the day to remove contributions from the diurnal cycle (*Skamarock et al., 2019*), so in this study we used 03 UTC and 15 UTC from set of 30 hour simulations initialized at 00 UTC and 12 UTC performed from 2017-11-01 to 2017-12-10, the same month of the target precipitation event of the previous year. So $\mathbf{B}$ was estimated using total of 79 forecast differences for every data assimilation domain (D02, D03, D04).

explore B

2.4.2 Nudging

Nudging is a continuous DA method (*Lei and Hacker, 2015*) that augments the prognostic equations with a Newtonian–relaxation term to constrain the model state $\mathbf{x}(t)$ toward a time–interpolated 3D external analysis $\mathbf{x}^{an}(t)$ throughout the domain:
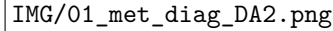
$$\frac{d\mathbf{x}}{dt} \;=\; \mathcal{M}(\mathbf{x}) \;+\; \mathbf{G}\,\mathbf{W}\left(\mathbf{x}^{an}(t) - \mathbf{x}(t)\right),$$

where $\mathcal{M}$ denotes resolved dynamics/physics, $\mathbf{G}$ sets the inverse relaxation timescale (gain), and $\mathbf{W}$ applies vertical/variable weights.

Properly tuned, analysis nudging anchors the large-scale flow while allowing mesoscale features to evolve (*Spero et al., 2018*; *Routray et al., 2012*). Nudging is especially recommended for large domains (*Miguez-Macho et al., 2004*). In the present study, we only implement the method in Domain D01 over the whole forecast time window and in this case $\mathbf{x}^{an}(t)$ corresponds to the NCEP operational GFS used as initial/boundary conditions.

2.5 DA Experiment

A 24-hour simulation is conducted [2025-11-25 12:00:00, 2025-11-26 12:00:00], after the initialization, a spin-up period of 3 hours is followed. Then data assimilation (DA) method is applied at hour 2025-11-25 15:00:00. Note that domain D01 is not used for assimilating observations, although nudging is applied. In contrast, domains D02, D03, and D04 are used to assimilate observations from *in situ* stations, a radiometer, and a wind profiler.

IMG/01_met_diag_DA2.png

Figure 7. DA Experiment proposed.

The observations and variables used for assimilation can be classified by the source as IDE, SWS, SCC, SRD and SPR:

- **IDE**: IDEAM Weather Stations (surface) where are u from?
- **SWS**: SIATA Weather Stations (surface). Composed mainly Vaisala and Thiess
- **SCC**: SIATA Citizen Scientist Stations (surface). Low-cost sensor (Davis 6830)
- **SRD**: SIATA Radiometer (vertical). From Radiometrics Corporation
- **SPR**: SIATA Profiler (vertical) From Radiometrics Corporation (Raptor)

The variables available from each type of observation source are summarized in *Table 3*, and the observations assimilated in each domain are summarized in *Table 4*.

Observations	Variable
IDE	PR, TM, RH, WS, WD
SWS	PR, TM, RH, WS, WD
SCC	PR*, TM, HR
SRD	TM
SPR	WS, WD

Table 3. Table of types of observations and variables. Pressure(PR), temperature (TM), relative humidity (RH), wind speed (WS), wind direction (WD). SCC pressure estimated from TM and HR*

Domain	Observations
D01	
D02	IDE, SWS, SRD, SPR
D03	IDE, SWS, SRD, SPR
D04	IDE, SWS, SCC, SRD, SPR

Table 4. WRF domains and observations (*Table 3*) for data assimilation

It should be noted that with respect to measurement uncertainty, to simplify certain procedures in the simulation, in cases IDE, SWS, SCC the measurement errors of the sensors are assumed to be identical for each variable, regardless of the measurement source.

2.6 Validation

After the assimilation step, an additional spin-up period is required; therefore, the validation window spans 20 hours [2025-11-25 12:00:00, 2025-11-26 12:00:00]. A description of methods and metrics for verification is presented here.

2.6.1 RMSE

The RMSE metric quantifies the average magnitude of the errors between observed and predicted values. Its main properties are: non-negativity, sensitivity to large errors and dimensional consistency.

Given an observed value y_i^{obs} and a predicted value y_i^{pred}, the individual error is defined as $e_i = y_i^{\text{obs}} - y_i^{\text{pred}}$. The RMSE is expressed as

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i^{\text{obs}} - y_i^{\text{pred}})^2} = \frac{1}{\sqrt{n}}\|\mathbf{e}\|_2. \tag{37}$$

2.6.2 Fractions skill score

Fractions Skill Score (FSS) is a verification method designed to measure how the skill of model varies with spatial scale (*Roberts and Lean, 2008*). Here we extend the original definition using convolutions for permitting odd and even windows length, that is the common implementation (*Pulkkinen et al., 2019*; *Imhoff et al., 2023*).

Let $F, O \in \mathbb{R}^{M \times N}$ denote the forecast field and the corresponding observation field. Let $q \in \mathbb{R}$ be the threshold. Define the binary fields:

$$B_O^q(i,j) = \begin{cases} 1 & \text{if } O(i,j) \geq q \\ 0 & \text{if } O(i,j) < q \end{cases}, \qquad B_M^q(i,j) = \begin{cases} 1 & \text{if } M(i,j) \geq q \\ 0 & \text{if } M(i,j) < q \end{cases} \tag{38}$$

$$\text{for } i = 1, \ldots, Nx, \ j = 1, \ldots, N_y$$

Let $n \in \mathbb{N}$ be the spatial scale (window side length). Define the box (uniform) kernel:

$$K_n[u,v] = \frac{1}{n^2}, \quad u, v = 0, 1, 2, \ldots, n-1 \tag{39}$$

The kernel K_n has size $n \times n$ and is normalized so that $\sum_{u,v} K_n[u,v] = 1$.

$$a = \left\lfloor \frac{n}{2} \right\rfloor, \qquad b = n - 1 - a. \tag{40}$$

Define a boundary-handling operator E_{ref} that extends indices outside the domain $\{1, \ldots, N_x\} \times \{1, \ldots, N_y\}$ by reflection and define the neighborhood-fraction fields are, then for any possibly-outside index (i,j):

$$E_{\text{ref}}(B)(i,j) = B\big(\phi_{\text{ref}}(i; M), \ \phi_{\text{ref}}(j; N) \big). \tag{41}$$

Where $\phi_{\text{ref}}(k; L)$ is defined as:

$$\phi_{\text{ref}}(k; L) = \begin{cases} m, & 0 \leq m < L, \\ 2(L-1) - m, & L \leq m < 2(L-1), \end{cases} \quad \text{where } m = k \bmod (2(L-1)). \tag{42}$$

The fractions of model and observations:

$$P_O^q(i,j;n) = \sum_{u=0}^{n-1} \sum_{v=0}^{n-1} K_n[u,v] \ E_{\text{ref}}\big(B_O^q \big)(i+u-a, \ j+v-a) \tag{43}$$

$$P_M^q(i,j;n) = \sum_{u=0}^{n-1} \sum_{v=0}^{n-1} K_n[u,v] \ E_{\text{ref}}\big(B_M^q \big)(i+u-a, \ j+v-a) \tag{44}$$

Using P_O^q and P_M^q:

$$\mathrm{MSE}(n, q) = \frac{1}{N_x \, N_y} \sum_{i=1}^{M} \sum_{j=1}^{N} \left(P_M^q(i, j; n) - P_O^q(i, j; n) \right)^2, \tag{45}$$

$$\mathrm{MSE}_{\mathrm{ref}}(n, q) = \frac{1}{N_x \, N_y} \sum_{i=1}^{M} \sum_{j=1}^{N} \left(P_M^q(i, j; n)^2 + P_O^q(i, j; n)^2 \right) \tag{46}$$

We can Define FSS as:

$$\mathrm{FSS}(n, q) = 1 - \frac{\mathrm{MSE}(n, q)}{\mathrm{MSE}_{\mathrm{ref}}(n, q)}. \tag{47}$$

If n is odd, then $a = b = (n-1)/2$, so the $n \times n$ neighborhood is symmetrically centered at (i, j). If n is even, then $a = \lfloor n/2 \rfloor$, $b = n - 1 - a$, so the window is half-shifted (one side has a offset and the other b). The boundary extension ensures that, near edges of the domain, the convolution always gets valid values by reflecting data across the boundary.

2.6.3 Funtional Data Analiysis

The basic philosophy of functional data analysis (FDA) is to think of observed data functions as single entities rather than merely as a sequence of individual observations. In practice, functional data are usually observed and recorded discretely as n pairs (t_j, y_j). In FDA the data instead of being a set of vectors, as in the classical multivariate analysis, is a set of curves. That is, a functional data $f(t)$, $t \in T \subset \mathbb{R}$, it is represented as a finite set of pairs (t_i, x_i), $t_i \in T$, $i = 1, 2, \ldots, N$, where N represents the observed points of the function. In most applications, the sample curves come from the observation of a stochastic process in continuous time.

Normalization

Normalization allows for a better manipulation of the data, adds resistance to outliers, and improves comparability across scales. The *Median Absolute Deviation (MAD)* is a robust measure of variability, defined for a dataset $X = \{x_1, \ldots, x_n\}$ as

$$MAD(X) := \mathrm{med}\left(|X - \mathrm{med}(X)| \right), \tag{48}$$

Where $\mathrm{med}(\cdot)$ denotes the median. A normalized version of the data is then given by

$$X_n = \log\left(|X| + MAD(X) \right). \tag{49}$$

Smoothing

To represent a series of points as functional data, linear combinations of basic functions will be used. The smoothing functional $S(f)$ balances fidelity to the data with smoothness of the estimated curve. Given observations $\{(x_i, y_i)\}$ and a candidate function $f(x)$, the functional is defined as

$$S(f) = p \sum_i w_i (y_i - f(x_i))^2 + (1+p) \int \left(\frac{d^2 f}{dx^2} \right)^2 dx, \tag{50}$$

where w_i are weights, and $p > 0$ is a smoothing parameter. The first term penalizes squared deviations from the data and the second term penalizes curvature, enforcing smoothness. Minimizing $S(f)$ yields a compromise: small p produces smoother functions, while large p prioritizes data fidelity. This is the basis of smoothing splines, widely used in statistics, functional data analysis, and numerical modeling.

Depth measurements

Let $C(I_x)$ the space of continuous functions defined on a compact interval I_x. Consider a stochastic process X with sample paths in $C(I_x)$ with distribution p_x. Let $x_1(t), x_2(t), \ldots, x_n(t)$ be a sample of curves from p_x. The graph of a function x in $C(I_x)$ will be denoted as $G(x)$, thus

$$G(x) = \{(t, x(t)), \quad t \in I_x\} \tag{51}$$

Define the hypograph (hyp) and the epigraph (epi) of a function x in $C(I)$ as:

$$hyp(x) = \{(t, y) \in I_x \times \mathbb{R} \ : \ y \leq x(t)\} \tag{52}$$

$$epi(x) = \{(t, y) \in I_x \times \mathbb{R} \ : \ y \geq x(t)\} \tag{53}$$

And also considering $GHyp_n$ and $GEpi_n$ as:

$$GHyp_n(x) = \frac{1}{n} \sum_{i=1}^{n} I\left(G(x_i) \subset hyp(x)\right) = \frac{1}{n} \sum_{i=1}^{n} I\left(x_i(t) \leq x(t), \ t \in I_x\right), \tag{54}$$

$$GEpi_n(x) = \frac{1}{n} \sum_{i=1}^{n} I\left(G(x_i) \subset epi(x)\right) = \frac{1}{n} \sum_{i=1}^{n} I\left(x_i(t) \geq x(t), \ t \in I_x\right), \tag{55}$$

Where I(A) is the indicator function of a set A:

$$I(A)(x) = \begin{cases} 1, & \text{si } x \in A, \\ 0, & \text{si } x \notin A. \end{cases} \tag{56}$$

Then, The half-region depth at x with respect to a set of functions $x_1(t), \ldots, x_n(t)$ is

$$HRD(x) = S_{n,H}(x) = \min\left\{GHyp_n(x), GEpi_n(x)\right\}. \tag{57}$$

Hence, the half-region sample depth at x is the minimum between the proportion of functions of the sample whose graph is in the hypograph of x and the corresponding proportion for the epigraph of x. That is, a curve is considered deep if it achieves a balanced position with respect to the sample, having a comparable proportion of curves lying

above and below it. Conversely, if the curve lies too high (with most other curves entirely below) or too low (with most curves entirely above), the minimum proportion becomes small and its half-region depth attains a low value. The deepest curve, or S_h -sample median $\hat{\tau}_n$, is a curve from the sample which maximizes the half-region depth. This curve can be used to illustrate the representative pattern within the sample of curves.

$$\hat{\tau}_n = \underset{x \in \{x_1,\ldots,x_n\}}{\arg\max} S_{n,H}(x). \tag{58}$$

The concepts described before also can be adapted to finite-dimensional data and this can be consulted and expanded upon in *López-Pintado and Romo* (*2011*).

3   Results

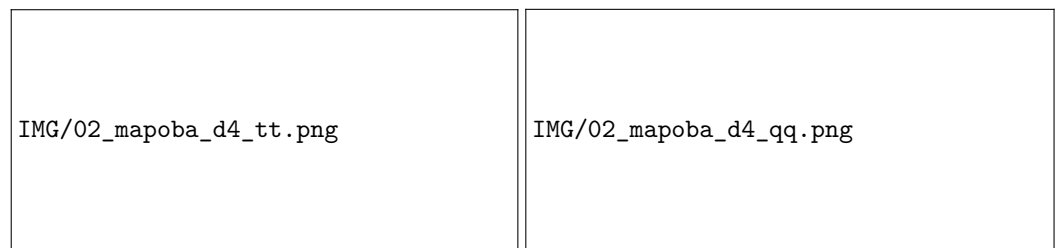### 3.1   Cost function optimization

IMG/02_jcost.png
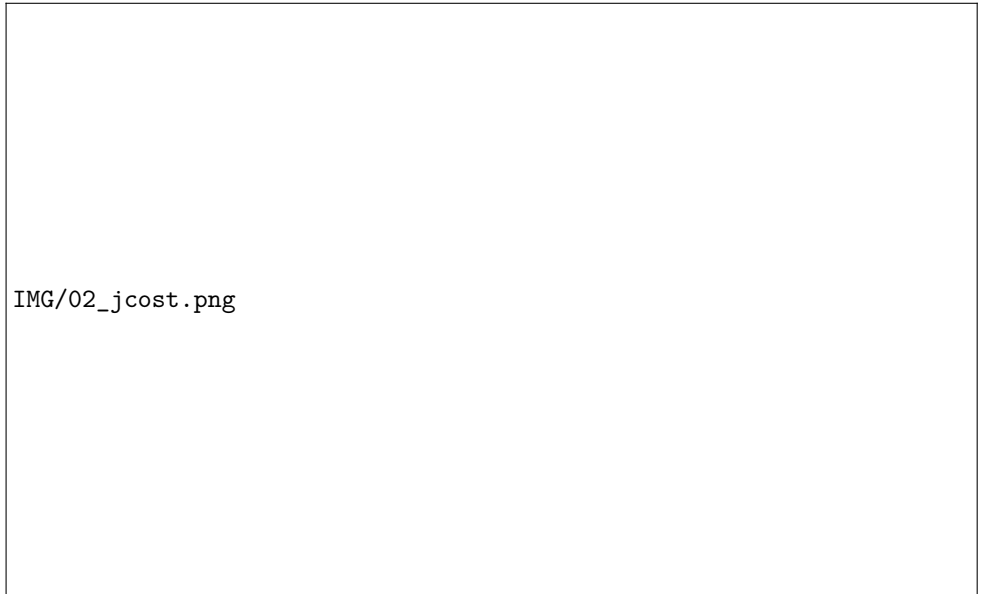
Figure 8. Cost function optimization for every domain
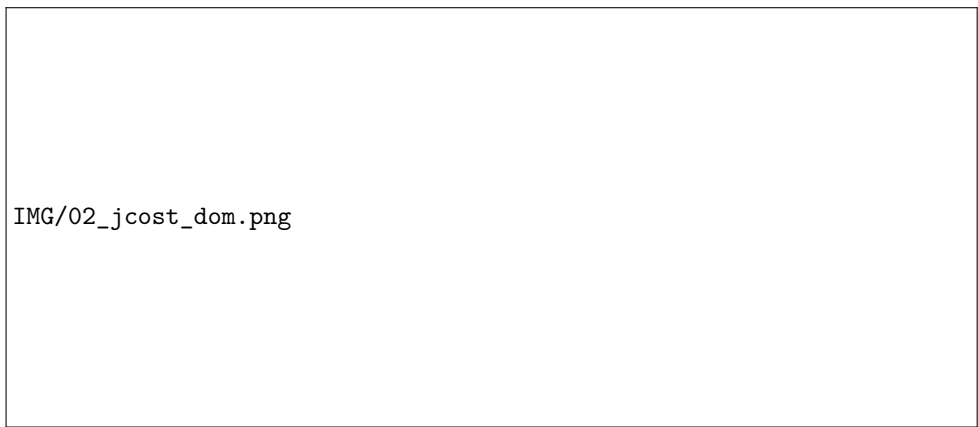
Figure 9. Cost function optimization and components ($\mathcal{J}_0$, $\mathcal{J}_b$) for every domain

check data for better render https://siata.gov.co/meteorologia/temporal/ot/
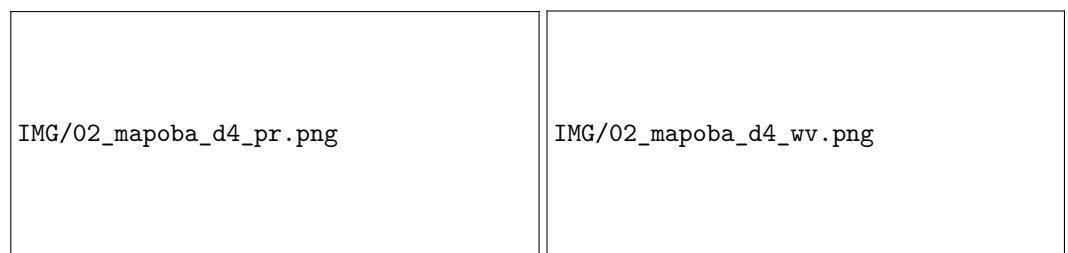

(a) |O-B| and |O-A| for temperature

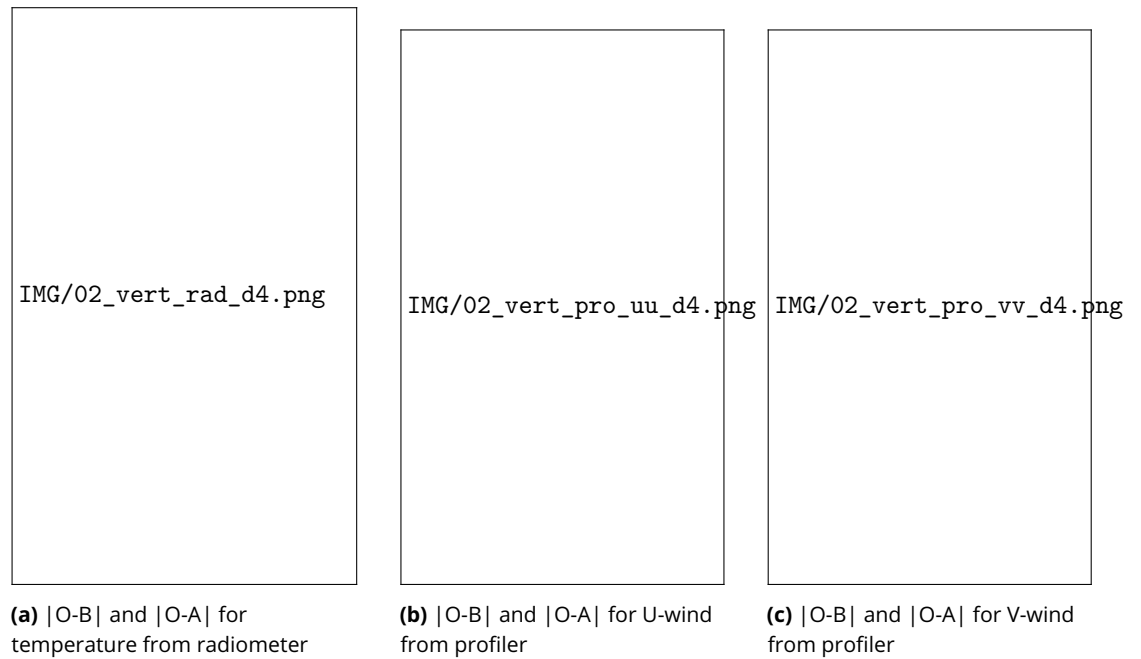**(b)** |O-B| and |O-A| for humidity

Figure 10. |Observation-Background| and |Observation-Analysis| for temperature and humidity (D04)

(a) |O-B| and |O-A| for pressure


(b) |O-B| and |O-A| for wind velocity

Figure 11. Similar to 10 but for pressure and wind velocity (D04)

(a) |O-B| and |O-A| for temperature from radiometer

(b) |O-B| and |O-A| for U-wind from profiler

(c) |O-B| and |O-A| for V-wind from profiler

Figure 12. |Observation-Background| and |Observation-Analysis| for temperature and wind components in the vertical (D04)

3.2 Spatiotemporal skill

For FSS in domain D03, the radar reflectivity data (500m) is interpolated to match the model output resolution at domain D03 (900 m). In order to compute the FSS in domain D04, the model outputs at a resolution of 300 m were interpolated to match the radar resolution (500 m). As described above, reflectivity estimation from model data is not suitable for direct comparison with radar data, the real interest here is to evaluate the shape of the storm, so the threshold used for FSS is 10 dBZ.

(a) Scales 7.2, 14.4, 28.8, 57.6 km, domain D03 (900m)

(b) Scales 4.0, 8.0, 16.0, 32.0 km, domain D04 (500m)
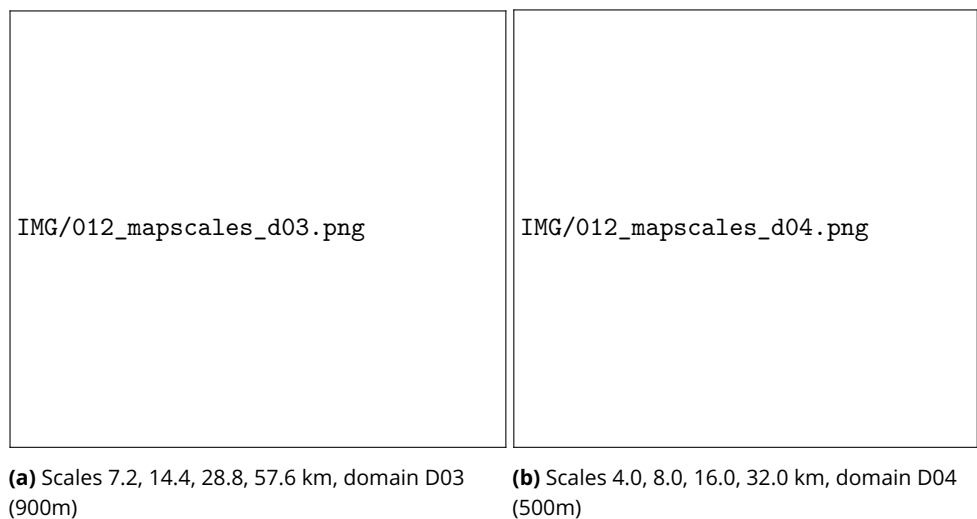
Figure 13. Domain scales for FSS. Black contour topography isoline 1800m

Figure 14. FSS (CRT and 3DV) using threshold of 10 dBZ and scale of 7.2 km for reflectivity in dom 03
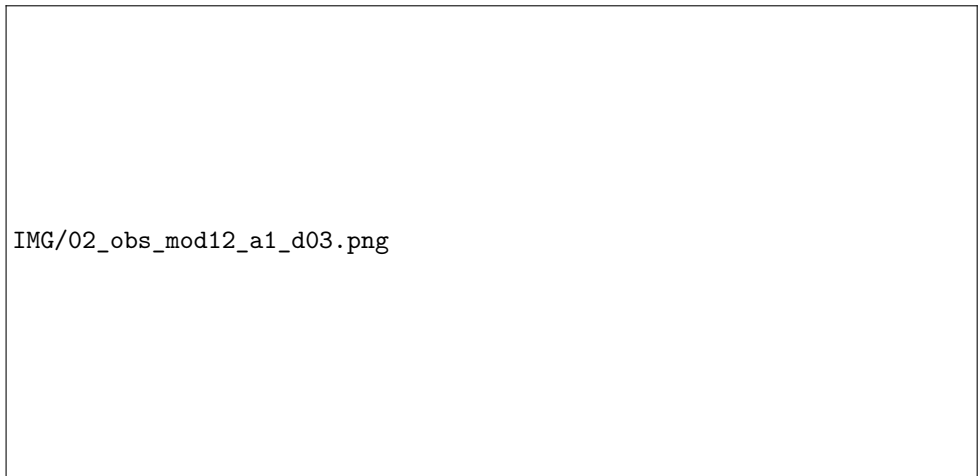

Figure 15. Similar to 14 but for d04

Scale	Average FSS CTR	Average FSS 3DV
8 - (7.2 km)	0.1810	0.2279
16 - (14.4 km)	0.2305	0.2981
32 - (28.8 km)	0.2899	0.4020
64 - (57.6 km)	0.3592	0.5462

Table 5. FSS D03

Scale	Average FSS CTR	Average FSS 3DV
8 - (4 km)	0.2137	0.2324
16 - (8 km)	0.2702	0.2921
32 - (16 km)	0.3532	0.3869
64 - (32 km)	0.4526	0.5133

Table 6. FSS D04

3.3   Precipitation series

Figure 16. Precipitation series for CTR and 3DV in dom 03
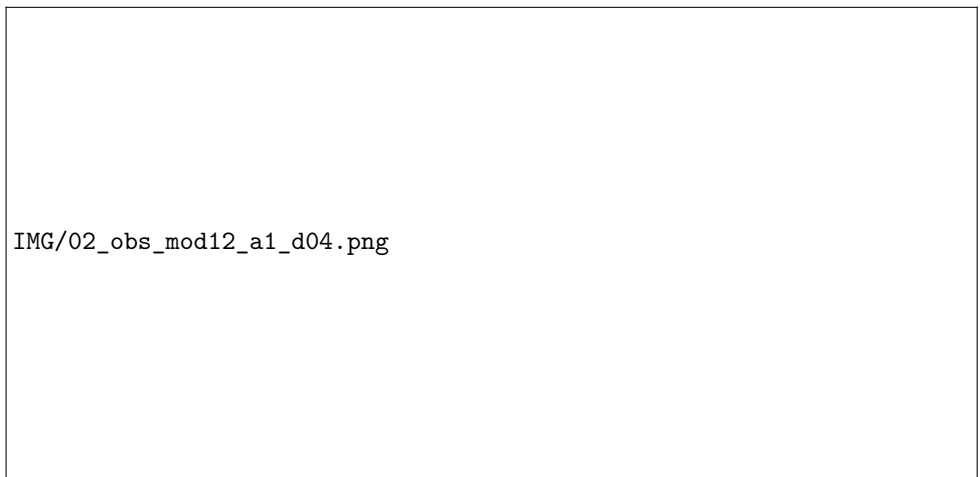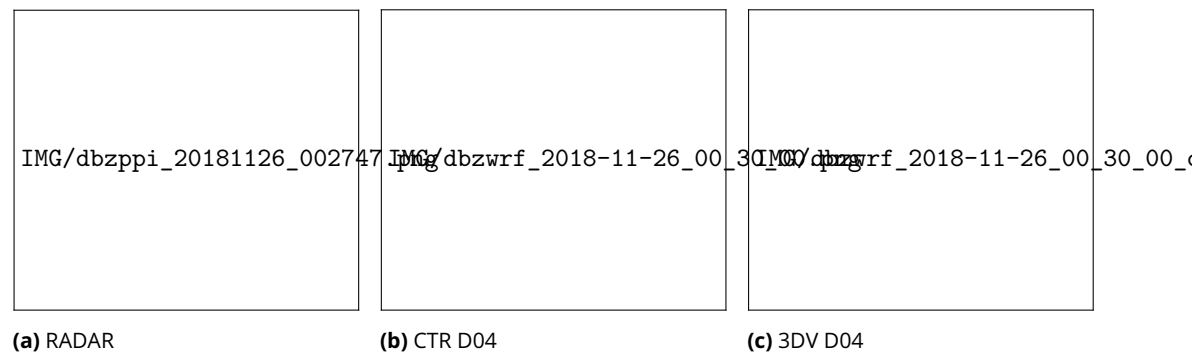

Figure 17. Similar to 16 but for dom 04

	CTR	3DV
RMSE [mm]	0.3506	0.2239

Table 7. rmse d03

	CTR	3DV
RMSE [mm]	0.2998	0.2010

Table 8. rmse d04

3.4 Reflectivity fields

(a) RADAR

(b) CTR D04

(c) 3DV D04

Figure 18. Reflectivity fields for RADAR, CTR and 3DV at 2018-11-26 00:30:00

(a) RADAR

(b) CTR D04

(c) 3DV D04
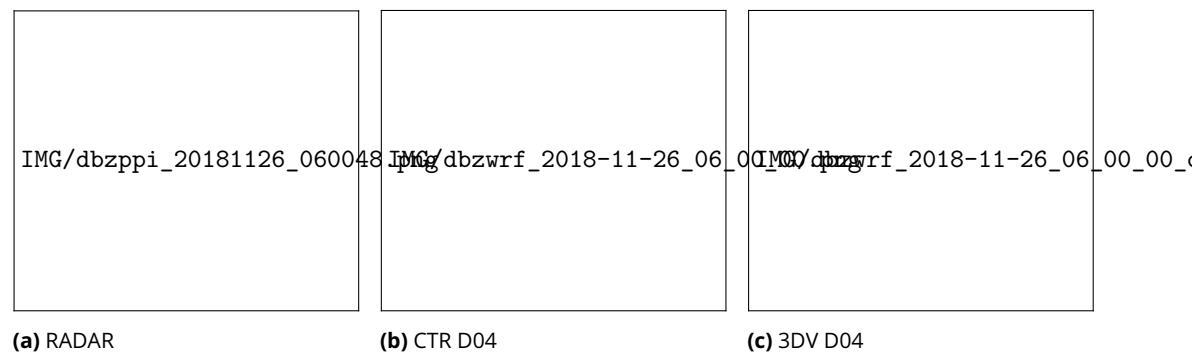
Figure 19. Similar to *Figure 18* but at 2018-11-26 06:00:00

3.5 FDA

(a) var1

(b) var2

Figure 20. FDA

(a) var1   **(b)** var2
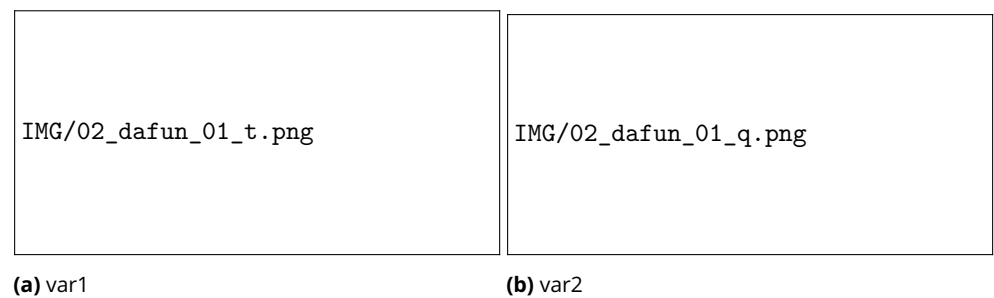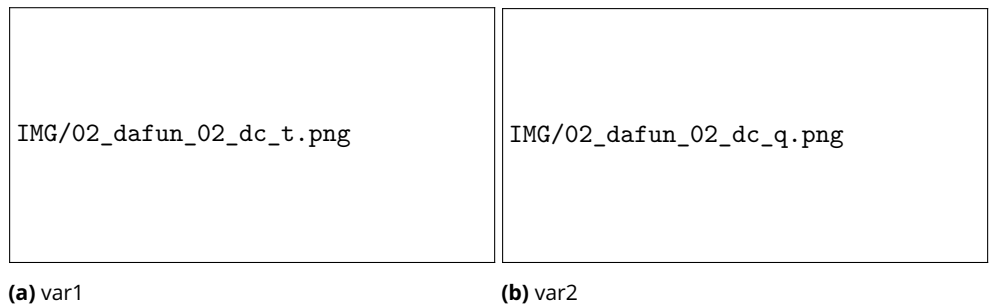
Figure 21. FDA + norm + smooth

(a) var1   **(b)** var2
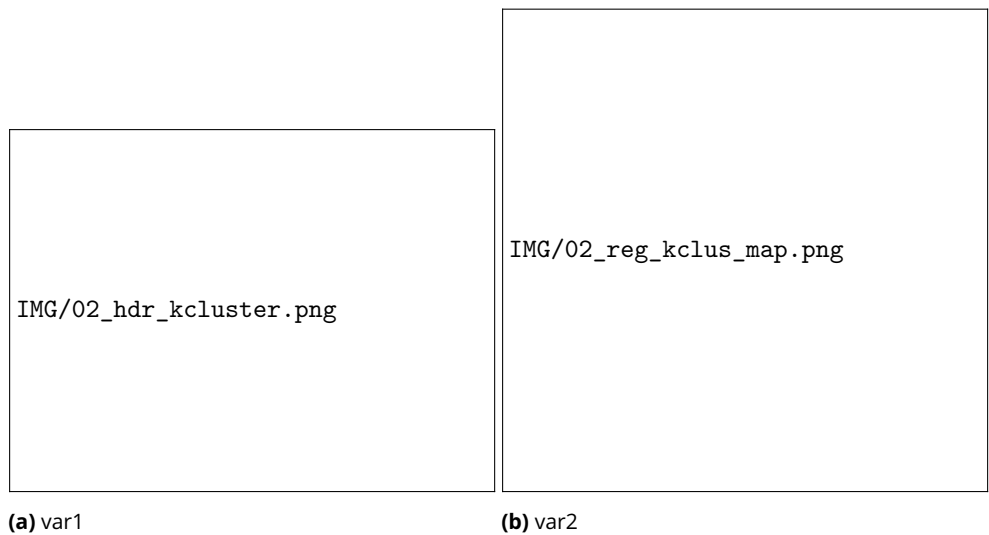
Figure 22. cluster

References

Asch M, Bocquet M, Nodet M. Data assimilation: methods, algorithms, and applications. SIAM; 2016.

ASF, ALOS PALSAR Radiometrically Terrain Corrected (RTC), High-Resolution (12.5 m). NASA Earthdata / ASF DAAC; 2015. https://www.earthdata.nasa.gov/data/projects/alos-palsar-rtc-project, geoTIFF, UTM; analysis-ready SAR (RTC). Processing 2014–2015. Accessed YYYY-MM-DD.

Bannister RN. A review of operational methods of variational and ensemble-variational data assimilation. Quarterly Journal of the Royal Meteorological Society. 2017; 143(703):607–633.

Carrassi A, Bocquet M, Bertino L, Evensen G. Data assimilation in the geosciences: An overview of methods, issues, and perspectives. Wiley Interdisciplinary Reviews: Climate Change. 2018; 9(5):e535.

Courtier P, Thépaut JN, Hollingsworth A. A strategy for operational implementation of 4D-Var, using an incremental approach. Quarterly Journal of the Royal Meteorological Society. 1994; 120(519):1367–1387. doi: 10.1002/qj.49712051912.

Dixon M, Javornik B, Lidar Radar Open Software Environment (LROSE) Core Software; 2016. https://doi.org/10.5065/60HZ-RY38, doi: 10.5065/60HZ-RY38.

Dudhia J. Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. Journal of Atmospheric Sciences. 1989; 46(20):3077–3107.

Friedl MA, et al. Global land cover mapping from MODIS: Algorithms and early results. Remote Sensing of Environment. 2002; 83:287–302. doi: 10.1016/S0034-4257(02)00078-0.

Gopalakrishnan D. Regional 4DVar Assimilation Studies on Weather Systems Over India Using the WRF Model. PhD thesis, Indian Institute of Space Science and Technology; 2019.

Hong SY, Noh Y, Dudhia J. A new vertical diffusion package with an explicit treatment of entrainment processes. Monthly weather review. 2006; 134(9):2318–2341.

Iacono MJ, Delamere JS, Mlawer EJ, Shephard MW, Clough SA, Collins WD. Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. Journal of Geophysical Research: Atmospheres. 2008; 113(D13).

Imhoff RO, De Cruz L, Dewettinck W, Brauer CC, Uijlenhoet R, van Heeringen KJ, Velasco-Forero C, Nerini D, Van Ginderachter M, Weerts AH. Scale-dependent blending of ensemble rainfall nowcasts and NWP in the open-source pysteps library. Quarterly Journal of the Royal Meteorological Society. 2023; n/a(n/a):1–30. https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.4461, doi: https://doi.org/10.1002/qj.4461.

Jiménez PA, Dudhia J, González-Rouco JF, Navarro J, Montávez JP, García-Bustamante E. A revised scheme for the WRF surface layer formulation. Monthly weather review. 2012; 140(3):898–918.

Kain JS. The Kain–Fritsch convective parameterization: an update. Journal of applied meteorology. 2004; 43(1):170–181.

Kalnay E. Atmospheric modeling, data assimilation and predictability. Cambridge university press; 2003.

Kotsuki S, Miyoshi T, Terasaki K, Lien GY, Kalnay E. Assimilating the global satellite mapping of precipitation data with the Nonhydrostatic Icosahedral Atmospheric Model (NICAM). Journal of Geophysical Research: Atmospheres. 2017; 122(2):631–650.

Lackmann G. Midlatitude synoptic meteorology. American Meteorological Society; 2011.

Ladwig W, wrf-python. Boulder, Colorado: UCAR/NCAR; 2017. doi: 10.5065/D6W094P1, [Software].

Lei L, Hacker JP. Nudging, ensemble, and nudging ensembles for data assimilation in the presence of model error. Monthly Weather Review. 2015; 143(7):2600–2610.

López-Pintado S, Romo J. A half-region depth for functional data. Computational Statistics & Data Analysis. 2011; 55(4):1679–1695.

Lorenz EN, Haman K. The essence of chaos. Pure and Applied Geophysics. 1996; 147(3):598–599.

Lu GY, Wong DW. An adaptive inverse-distance weighting spatial interpolation technique. Computers & geosciences. 2008; 34(9):1044–1055.

Miguez-Macho G, Stenchikov GL, Robock A. Spectral nudging to eliminate the effects of domain position and geometry in regional climate model simulations. Journal of Geophysical Research: Atmospheres. 2004; 109(D13).

NCEP, NCEP GFS 0.25 Degree Global Forecast Grids Historical Archive. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory; 2015. https://doi.org/10.5065/D65D8PWK, doi: 10.5065/D65D8PWK, dataset ID: d084001.

Oye D, Case M. REORDER: A program for gridding radar data— Installation and use manual for the UNIX version. National Center for Atmospheric Research (NCAR); 1995.

Parrish DF, Derber JC. The National Meteorological Center's spectral statistical-interpolation analysis system. Monthly Weather Review. 1992; 120(8):1747–1763. doi: 10.1175/1520-0493(1992)120<1747:TNMCSS>2.0.CO;2.

Pu Z, Kalnay E. Numerical weather prediction basics: Models, numerical methods, and data assimilation. In: *Handbook of hydrometeorological ensemble forecasting* Springer; 2018.p. 1–31.

Pulkkinen S, Nerini D, Pérez Hortal AA, Velasco-Forero C, Seed A, Germann U, Foresti L. Pysteps: an open-source Python library for probabilistic precipitation nowcasting (v1.0). Geoscientific Model Development. 2019; 12(10):4185–4219. https://gmd.copernicus.org/articles/12/4185/2019/, doi: 10.5194/gmd-12-4185-2019.

Roberts NM, Lean HW. Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. Monthly Weather Review. 2008; 136(1):78–97.

Routray A, Osuri KK, Kulkarni MA. A comparative study on performance of analysis nudging and 3DVAR in simulation of a heavy rainfall event using WRF modeling system. International Scholarly Research Notices. 2012; 2012(1):523942.

Skamarock WC, Klemp JB, Dudhia J, Gill DO, Liu Z, Berner J, Wang W, Powers JG, Duda MG, Barker DM, Huang XY. A Description of the Advanced Research WRF Version 4. National Center for Atmospheric Research; 2019.

Spero TL, Nolte CG, Mallard MS, Bowden JH. A maieutic exploration of nudging strategies for regional climate applications using the WRF model. Journal of Applied Meteorology and Climatology. 2018; 57(8):1883–1906.

Stoelinga MT. Simulated equivalent reflectivity factor as currently formulated in RIP: Description and possible improvements. White paper. 2005; 5.

Sun J, Wang H, Tong W, Zhang Y, Lin CY, Xu D. Comparison of the impacts of momentum control variables on high-resolution variational data assimilation and precipitation forecasting. Monthly Weather Review. 2016; 144(1):149–169.

Tewari N, Tewari M, Chen F, Wang W, Dudhia J, LeMone M, Mitchell K, Ek M, Gayno G, Wegiel J, et al. Implementation and verification of the unified NOAH land surface model in the WRF model (Formerly Paper Number 17.5). In: *Proceedings of the 20th conference on weather analysis and forecasting/16th conference on numerical weather prediction, Seattle, WA, USA*, vol. 14; 2004. p. 11–15.

Thompson G, Field PR, Rasmussen RM, Hall WD. Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. Monthly weather review. 2008; 136(12):5095–5115.

UNGRD. Consolidado anual de emergencias. Unidad Nacional para la Gestión del Riesgo de Desastres (UNGRD); 2018.

USGS, Global 30 Arc-Second Elevation (GTOPO30); 1996. https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-global-30-arc-second-elevation-gtopo30, accessed 2025-09-16. USGS EROS Archive.

Workneh HT, Chen X, Ma Y, Bayable E, Dash A. Comparison of IDW, Kriging and orographic based linear interpolations of rainfall in six rainfall regimes of Ethiopia. Journal of Hydrology: Regional Studies. 2024; 52:101696.

Zhang X, Bao JW, Chen B, Grell ED. A three-dimensional scale-adaptive turbulent kinetic energy scheme in the WRF-ARW model. Monthly Weather Review. 2018; 146(7):2023–2045.