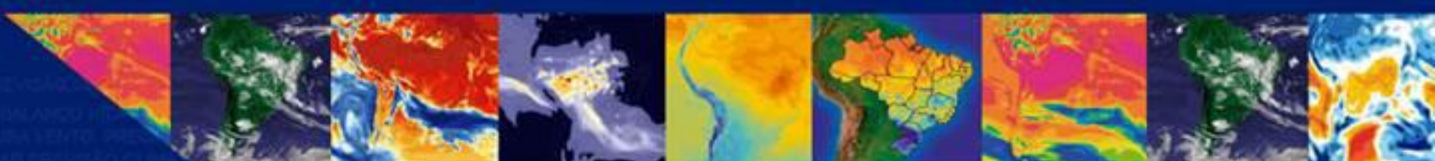
Dynamic machine learning applied to merged GEO-IR and LEO-microwave data to improve the analysis and nowcasting of severe convection

Haddad Z. (PI), Tapiador F., Turk J., Seo E-K, Goyal S., Hristova-Veleva S., Sawaya R. and many others

Daniel Vila (daniel.vila@inpe.br - presenter)

CWG Meeting Online 6-8 April 2021



Motivation

- **Current algorithms to forecast the evolution of convection rely exclusively on geostationary IR data (geoIR)**. Several approaches (including Vila et al., 2008; Autonès, 2012; Fiolleau and Roca, 2013; Liu et al., 2015) have been developed to detect, delineate, track and forecast the evolution of convective systems based on geoIR brightness temperatures ⇒ ***very limited sensitivity of the measurements to the storm under the cloud tops***

- **Satellite measurements that are more directly sensitive to the convection below the cloud tops are available today from low-Earth-orbit microwave radiometers**, but the constellation of microwave radiometers is heterogeneous ⇒ different sensors have different channel combinations, different viewing geometries, different spatial resolutions, and different measurement uncertainties – and does not provide data with a regular temporal sampling.

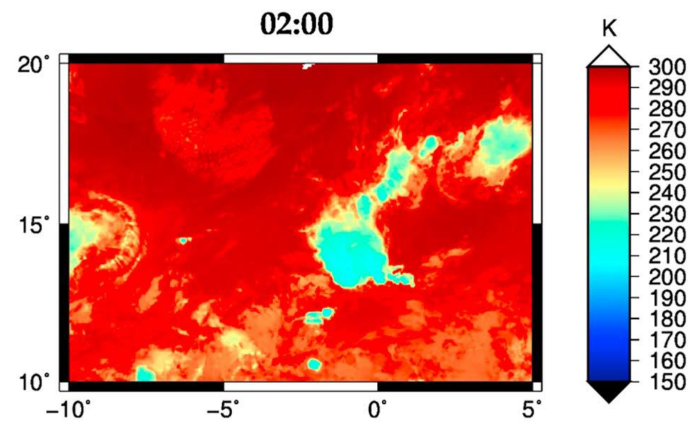
⇓

Develop a novel approach which can ingest all the available storm-sensitive satellite data, namely the regularly available geostationary IR as well as intermittent low-Earth-orbit passive microwave, for a specific storm at a specific time, and produce an analysis of the storm structure at neighboring instants in time, to analyze the evolution of the storm in time and nowcast its evolution in the near future.

Motivation

(a)

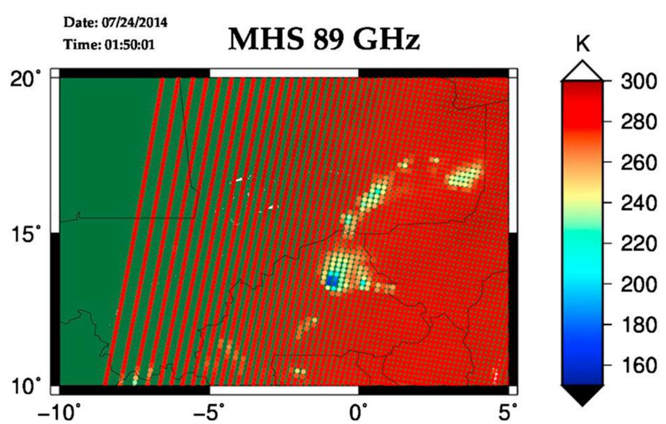
Global – 30 minutes

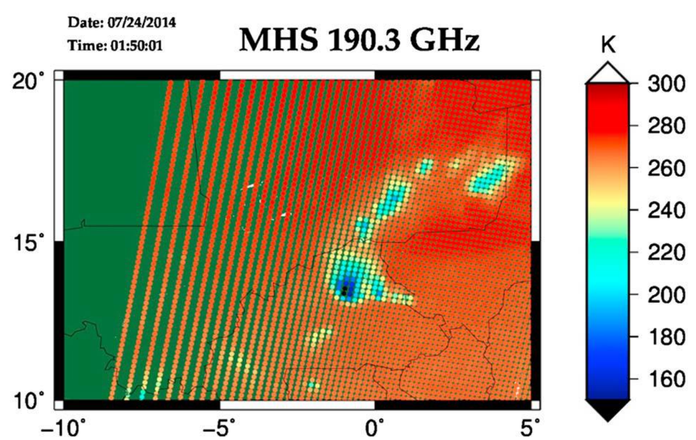


(b)

(c)

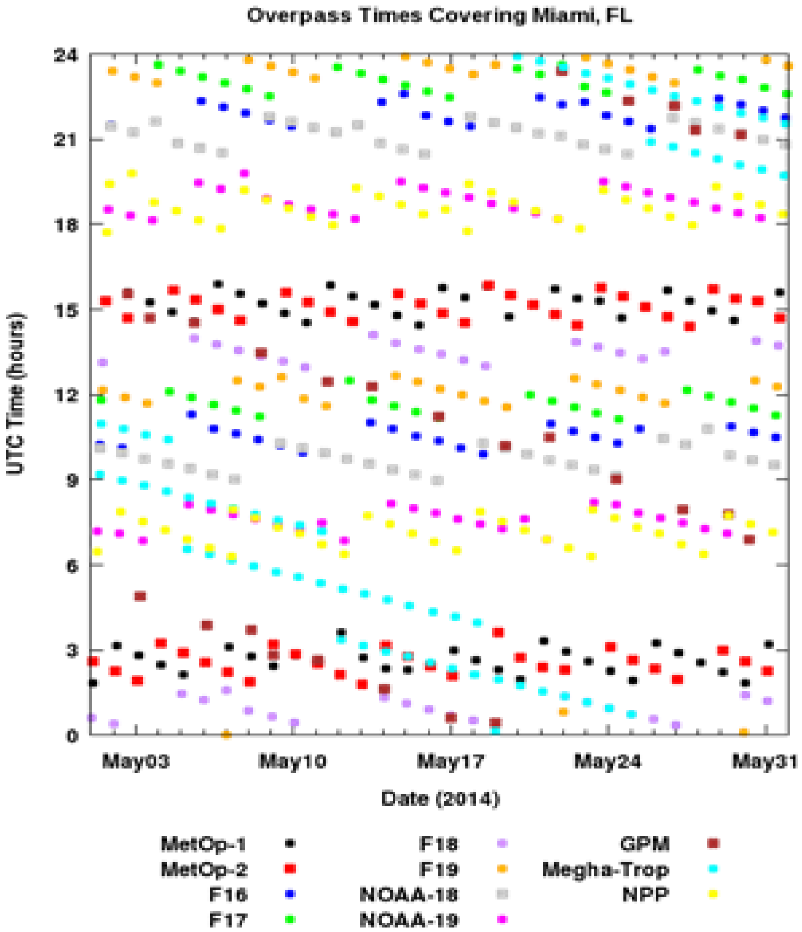
Global – Intermitent



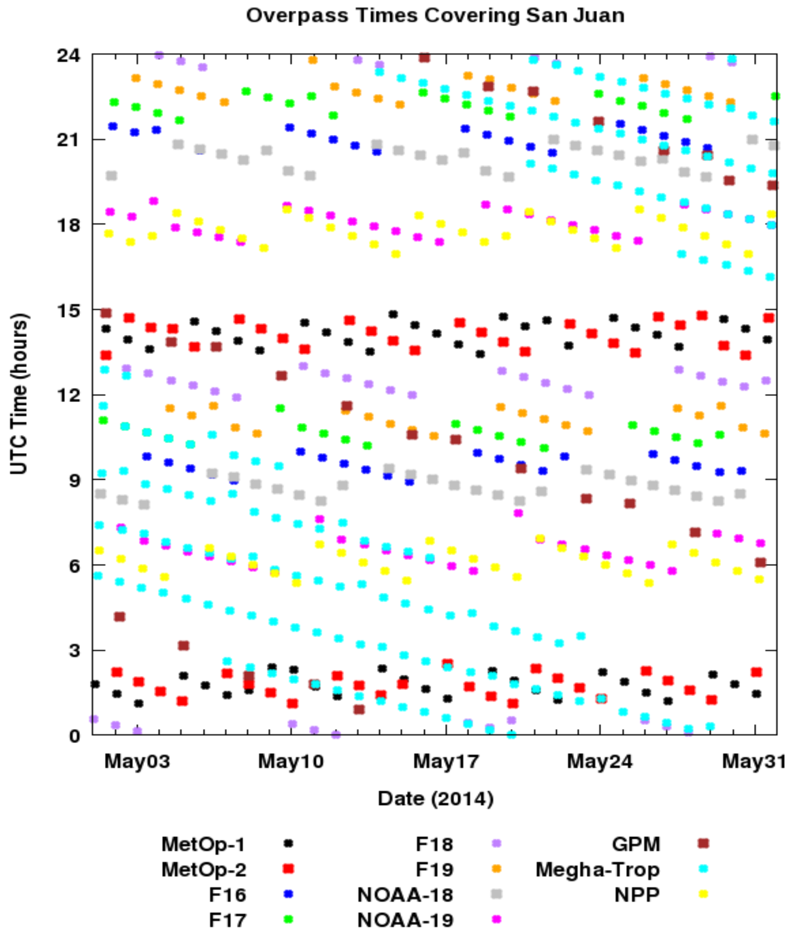


Mesoscale convective system over Burkina Faso and Mali as seen by the geostationary IR (Meteosat 12μm brightness temperatures, upper panel) and by a microwave sounder (NOAA-19 MHS's 89 and 190.3 GHz channel, bottom panel). Note how the sounder clearly detects the cores of deep convection within the cloud deck delineated by the geoIR. The ill-fated Air Algérie flight AH 5017 took off from Ouagadougou (1.57ºW 12.31ºN) at 01:15 Z, ingested frozen hydrometeors in its engines when it flew over the microwave-coldest portion of this system (the darkest navy pixel in the panel to the right — see Haddad et al., 2017) and crashed at 01:47 Z near 1.08ºW 15.13ºN. The panels illustrate the two data types that we propose to merge.

Motivation

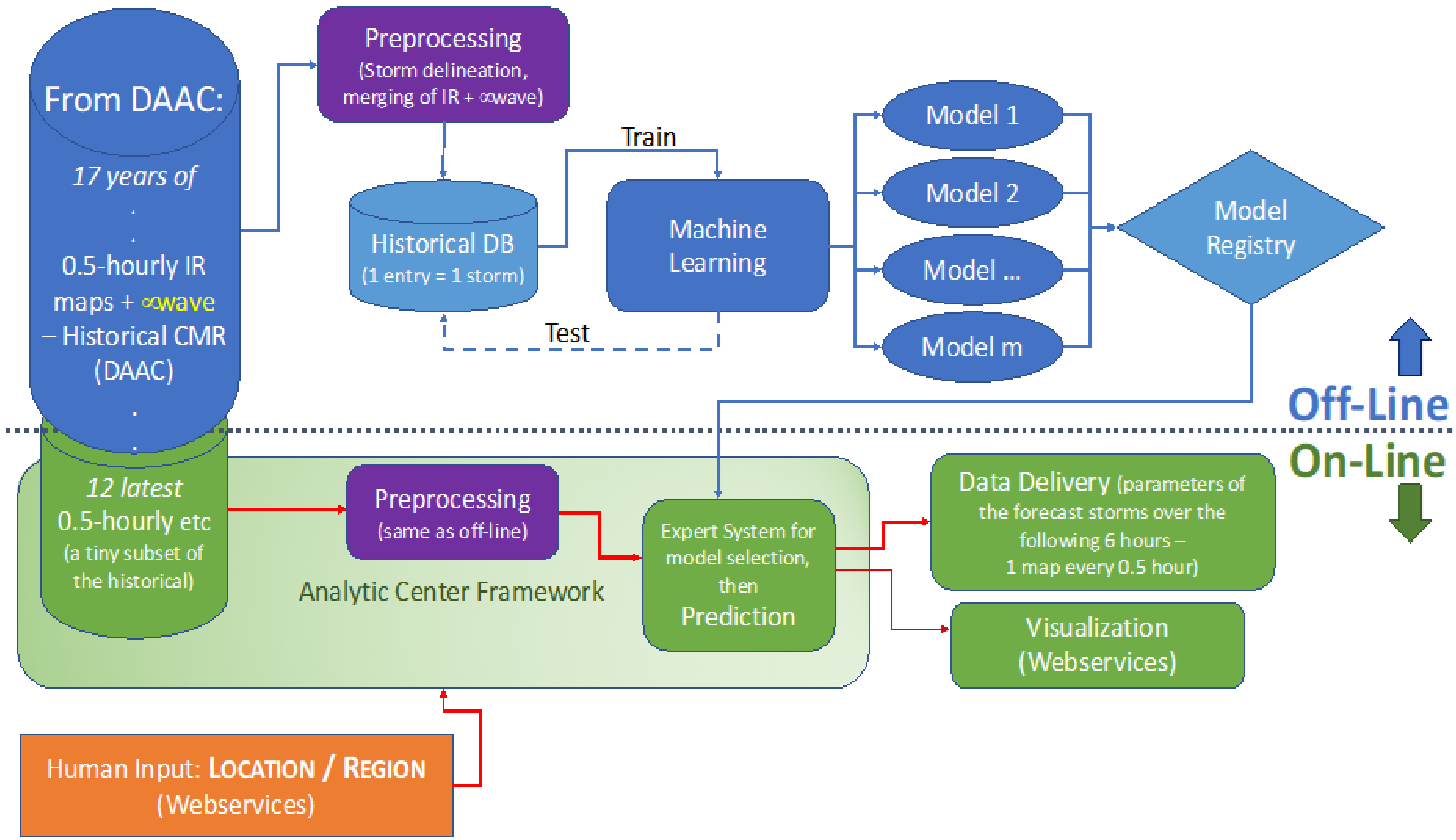


Overpass Times Covering Miami, FL
UTC Time (hours)
0
3
6
9
12
15
18
21
24
May03
May10
May17
May24
May31
Date (2014)
MetOp-1
MetOp-2
F16
F17
F18
F19
NOAA-18
NOAA-19
GPM
Megha-Trop
NPP



Overpass Times Covering San Juan
UTC Time (hours)
0
3
6
9
12
15
18
21
24
May03
May10
May17
May24
May31
Date (2014)
MetOp-1
MetOp-2
F16
F17
F18
F19
NOAA-18
NOAA-19
GPM
Megha-Trop
NPP

Framework and Objetives



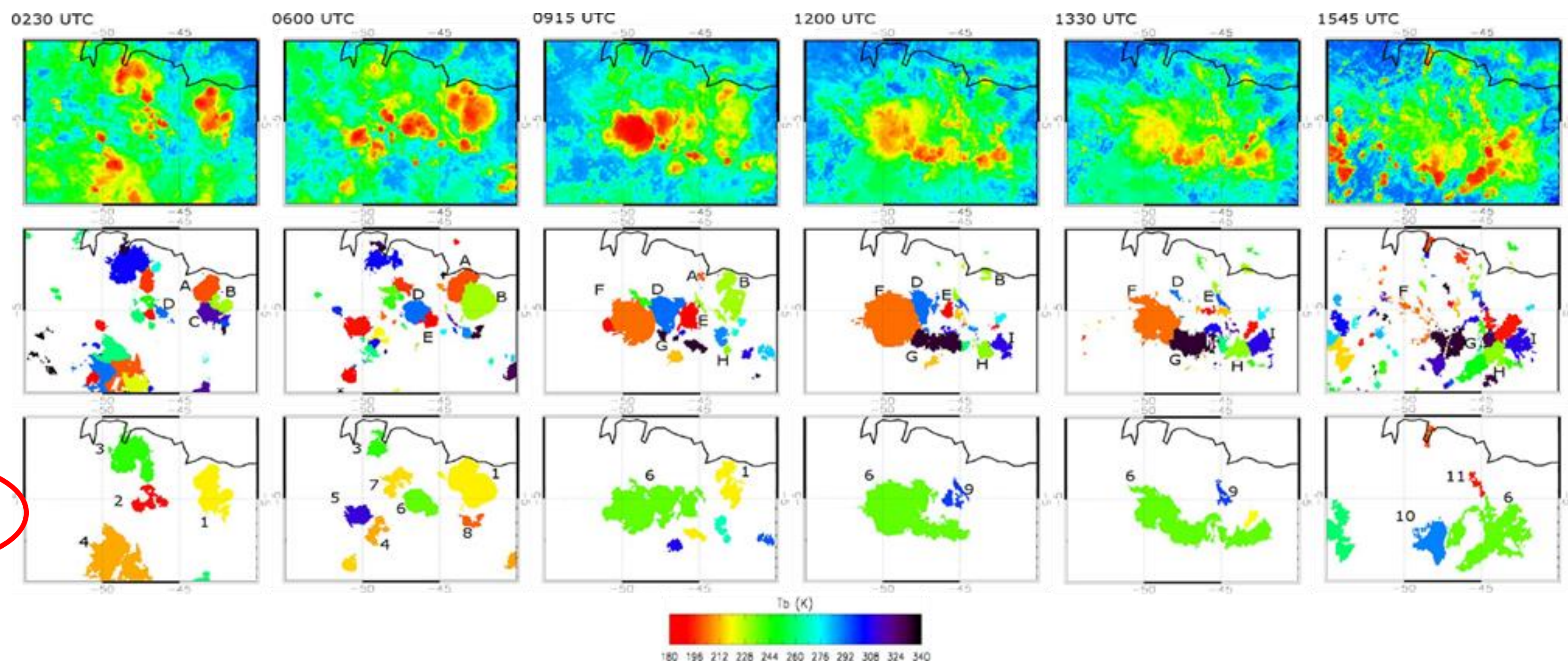
- **Automate the merging of geoIR with intermittent low-Earth-orbit passive microwave data**
- **Use ML-based algorithms to analyze storm structure and evolution from 17 years' worth of observations, and extract models that can nowcast storm evolution 3 hours into the future**
- **Implement an on-line system that can nowcast in real time (based on geoIR from the present back to 6 hours in the recent past plus the microwave data from the most recent microwave satellite pass)**

www.cptec.inpe.br

Proposed Methodology
GeoIR-based detection and tracking

Tracking Methodologies

- Image overlap based on fixed temperature threshold — **ForTrACC** (Vila et al., 2008)
- "Seed and Grow" based on local minima — **TOOCAN** (Fiolleau et al., 2013)



Proposed Methodology
mm-wave retrievals

Reference global datasets of near-coincidences between the observations of each radiometer type and a reference vertically-profiling radar are used to produce empirical Bayesian instantaneous estimates of a set of 6+ geophysical variables

- **FLAG** = the binary flag indicating whether condensed water in concentrations above a (low) threshold of 0.05 g/m3 is detected above the freezing level
- $\mathbf{H_{0.05}}$ = the maximum height above mean sea level (AMSL) reached by condensed water concentrations greater than 0.05 g/m3
- $\mathbf{H_{0.2}}$ = the maximum height AMSL where the condensed water concentration exceeds 0.2 g/m3
- $\mathbf{PC_1}$ (CWC) = the first vertical principal component of the Condensed Water Concentration above 5 km – essentially a vertical average of the condensed water content
- $\mathbf{PC_2}$ (CWC) = the second vertical principal component of condensed water mass above 5 km – essentially the average above 8.5 km minus the average between 5 and 8.5 km
- **RH** = the average relative humidity above 5 km
- **CLASS** = the type (convective/stratiform) of the precipitation

Half the reference (radar+radiometer) coincidence dataset is used to define the Bayesian retrieval, and the other half is used to quantify the uncertainty as a function of the observed brightness temperatures.

Proposed Methodology mm-wave retrievals

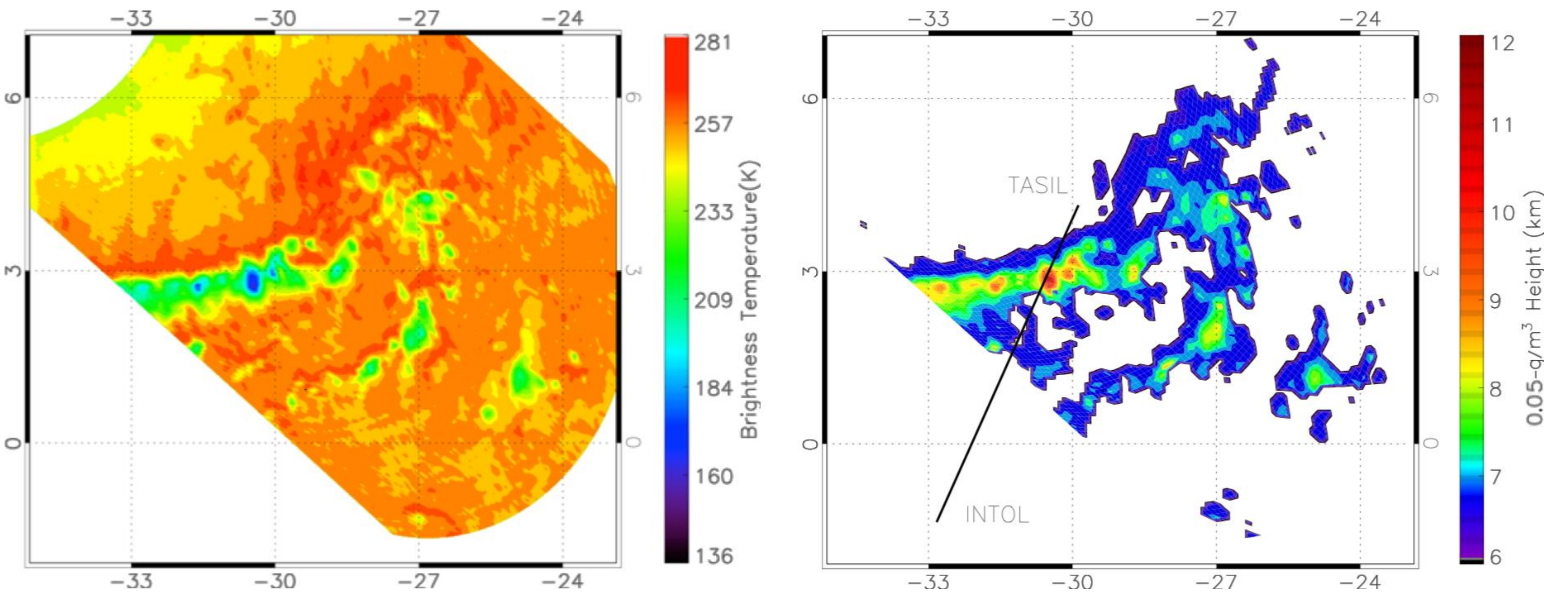


Figure 6: Example illustrating the utility of the second microwave variable $H_{0.05}$ derived from the TMI measurements made around 02:25 Z on 1 June 2009 over a portion of the Atlantic Inter-Tropical Convergence Zone between Brazil to the southwest and Senegal to the northeast. The left panel shows the brightness temperatures measured by TMI's 89 GHz V-polarized channel – the clearly visible cold (blue) regions are the columns with marked out-of-beam scattering of the upwelling radiation, indicating a deep column of ice hydrometeors. The multi-channel radiances were used to estimate the maximum height AMSL where the condensed water content exceeded 0.05 g/m^3 (as explained in Haddad and Park, 2010), and the results are shown in the panel to the right. The significance of this example is highlighted in the panel to the right, where the line segment is the trajectory of Air France flight AF447, which manifestly flew through the top of the deepest convective core less than 15 minutes before it crashed (see Haddad and Park, 2010). The r.m.s. uncertainty in the retrieved $H_{0.05}$ (estimated from the testing half of the global reference coincidence dataset) varies between 300 and 900 meters over the observed region.

www.cptec.inpe.br

Reference training/testing dataset for analysis and nowcasting

Two Modes

Analysis Mode A

IR Data

- a = area with temperature < 235K
- b = the longitude and latitude of the center of the best-fit ellipse
- c = the minimum IR temperature within the system
- d ...k = other 11 parameters

6 hours -1/2 hour interval

MW Data

- k = the area where H0.2 > 5 km
- l = the area where H0.05 > 5 km
- m= the area where PC1 > 0.2 g/m3,
- n = the area where PC2 > - 0.2 g/m3
- o = average value of RH
- p = convective/stratiform proportion

Intermitent

G operator - ML approach

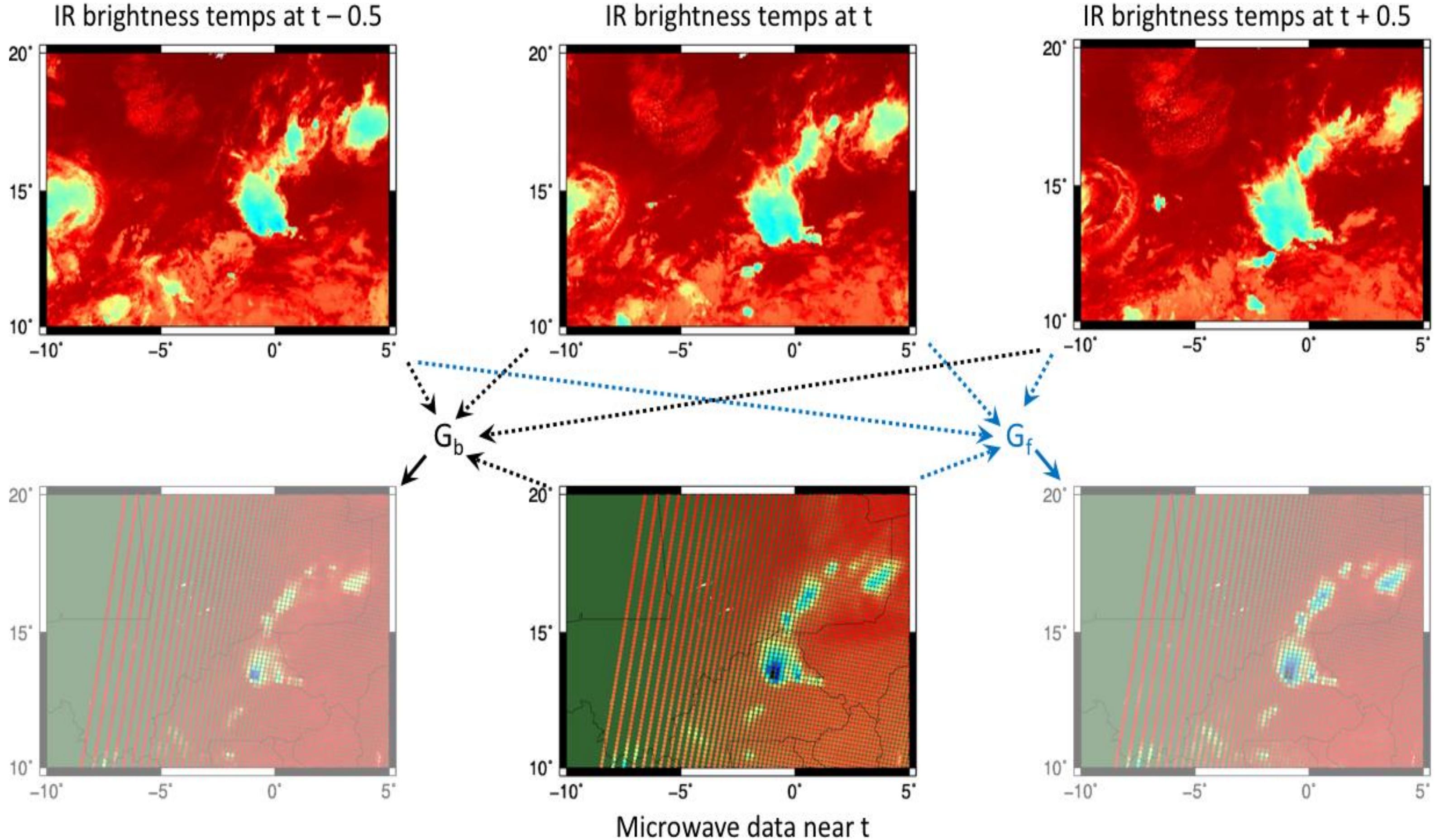
Nowcasting Mode N

F operator - ML approach

$$\mathbf{F}(a_1, b_1, c_1, d_1, \ldots, a_{12}, b_{12}, c_{12}, d_{12}; k, l, m, n, o, p, \tau) = (a_{13}, b_{13}, c_{13}, d_{13}; \ldots; a_{18}, b_{18}, c_{18}, d_{18})$$

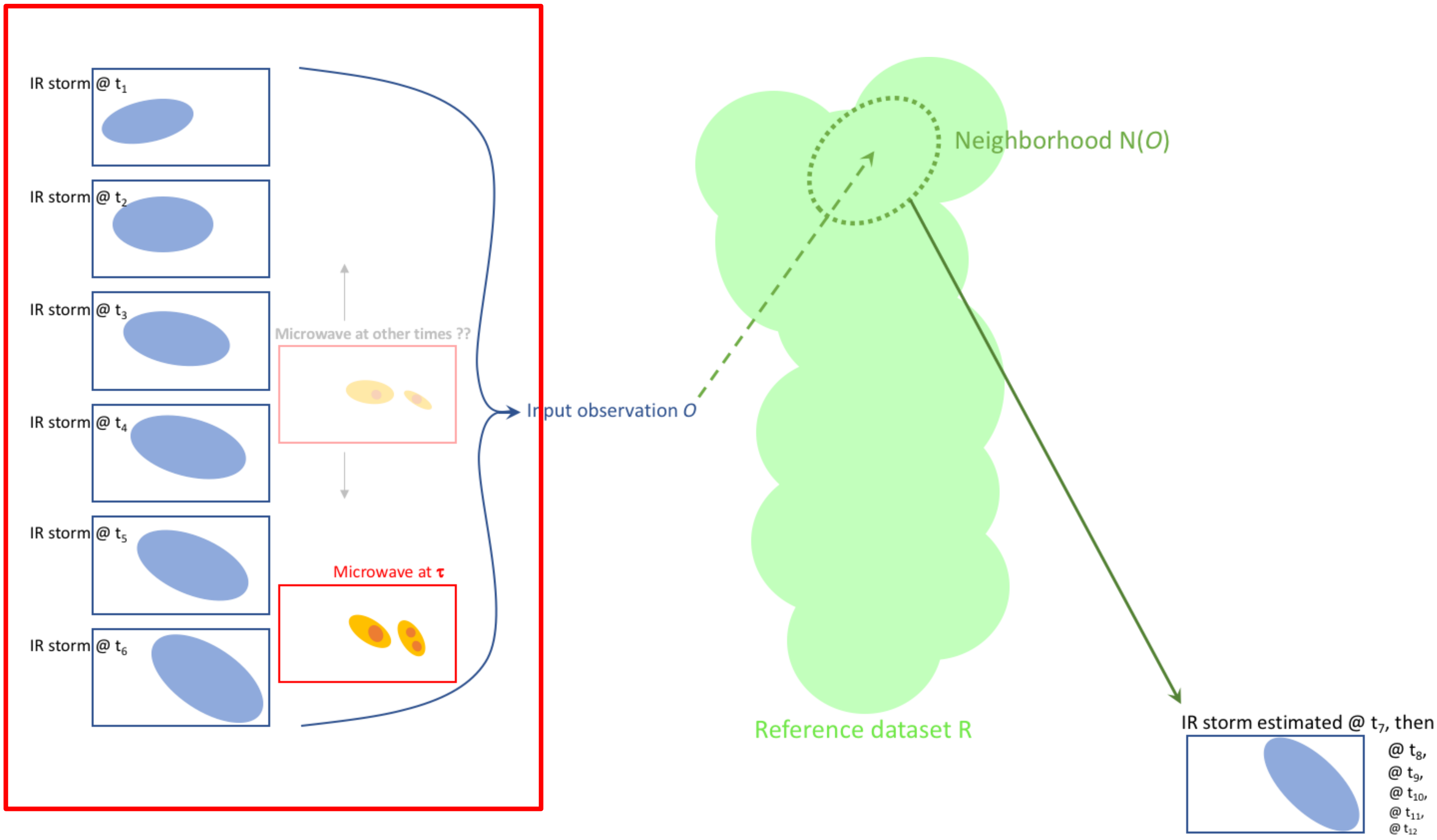
3 hours forecast -1/2 hour interval

Reference training/testing dataset for analysis and nowcasting



Schematic diagram summarizing the workflow in the microwave-pixel-level mode A

Reference training/testing dataset for analysis and nowcasting

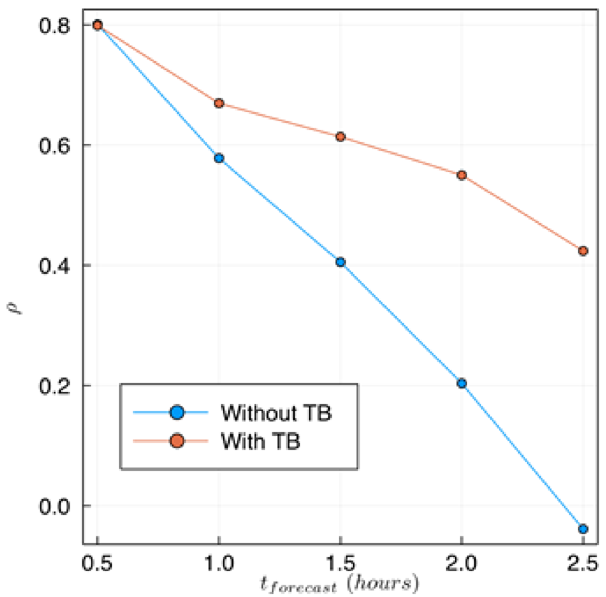


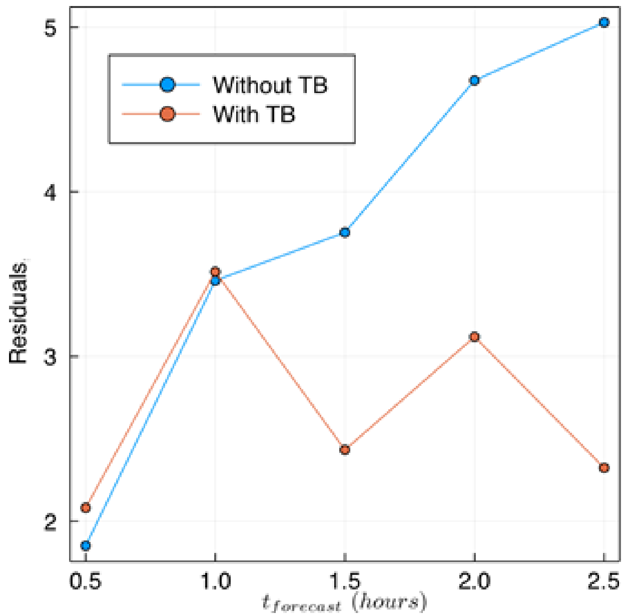
Schematic diagram summarizing the workflow in the storm-level mode N

Preliminary Results

A preliminary analysis of 18 months' worth of merged global geoIR data, which,

- Processed through the ForTrACC delineation and tracking algorithm,
- Pruned to retain only those storms that were observed by the SAPHIR sounder (the only microwave sounder with a purely tropical orbit, which guarantees relatively frequent revisits).





Correlation (left panel) and average normalized residual error (right panel) between the future storm geoIR area at different future half-hour time steps (horizontal axis) and "the past" geoIR area (in blue) or the past area augmented by the cold-differential microwave area (in red). The future area is a single scalar, "the past" is the scalar consisting of the second-order polynomial combination whose coefficients maximize the correlation with the future area. The correlation is greater and the error smaller with the microwave data.

Workplan

Construction of reference archive

- Generate estimates of storm structure from all microwave observations at pixel resolution
- Delineate storm histories from the merged IR data
- Inject 1) in 2), and divide the result into a learning reference subset R_L and a test subset R_T

Dynamic learning for nowcasting

- Implement and test advanced learning procedures to use R_L to infer the storm-level IR-evolution.
- Quantify the uncertainty using R_L

Dynamic learning for analysis

- Re-format the reference data to represent the 5x5 subgrids about a microwave grid point
- Re-segment the data according to the delay between consecutive microwave data.
- implement and test advanced learning procedures to use R_L to infer the pixel-level microwave evolution

www.cptec.inpe.br

Summary and Possible Outcomes

- The use of intermittently available passive-microwave data to analyze the three-dimensional structure and evolution of convective storms has not received specific attention by the scientific community, in spite of its clear relevance.

- At present, the main issue for the operational applicability of the proposed system is the latency between microwave data acquisition and their availability for processing. The current latency time is ~1 hour. Once the utility of the data is demonstrated by the proposed first version of the nowcasting system, the latency can be reduced to that of the operational (and new constellations) satellites.

- Another application for the quantitative storm context is to the analysis and monitoring of the Hadley circulation. While different analyses indicate a poleward expansion of the width of “the tropics” (the width of the Hadley circulation) by about two degrees over the past two decades, the changes in precipitation patterns that would accompany this expansion have not been quantified. The quantitative characterization of the storm context (instead of surface precipitation) will enable to understand this process.