UCLA JIFRESSE

Joint Institute for Regional Earth System Science & Engineering

NASA

Assimilating precipitation observations into HWRF while avoiding the pitfalls of microphysical representations

Alex Fore, Svetla Hristova-Veleva, Hsiao-Chieh Tseng, Jeff Steward, Shu-Hua Chen, Tomi Vukicevic and Ziad Haddad

zhaddad@jifresse.ucla.edu

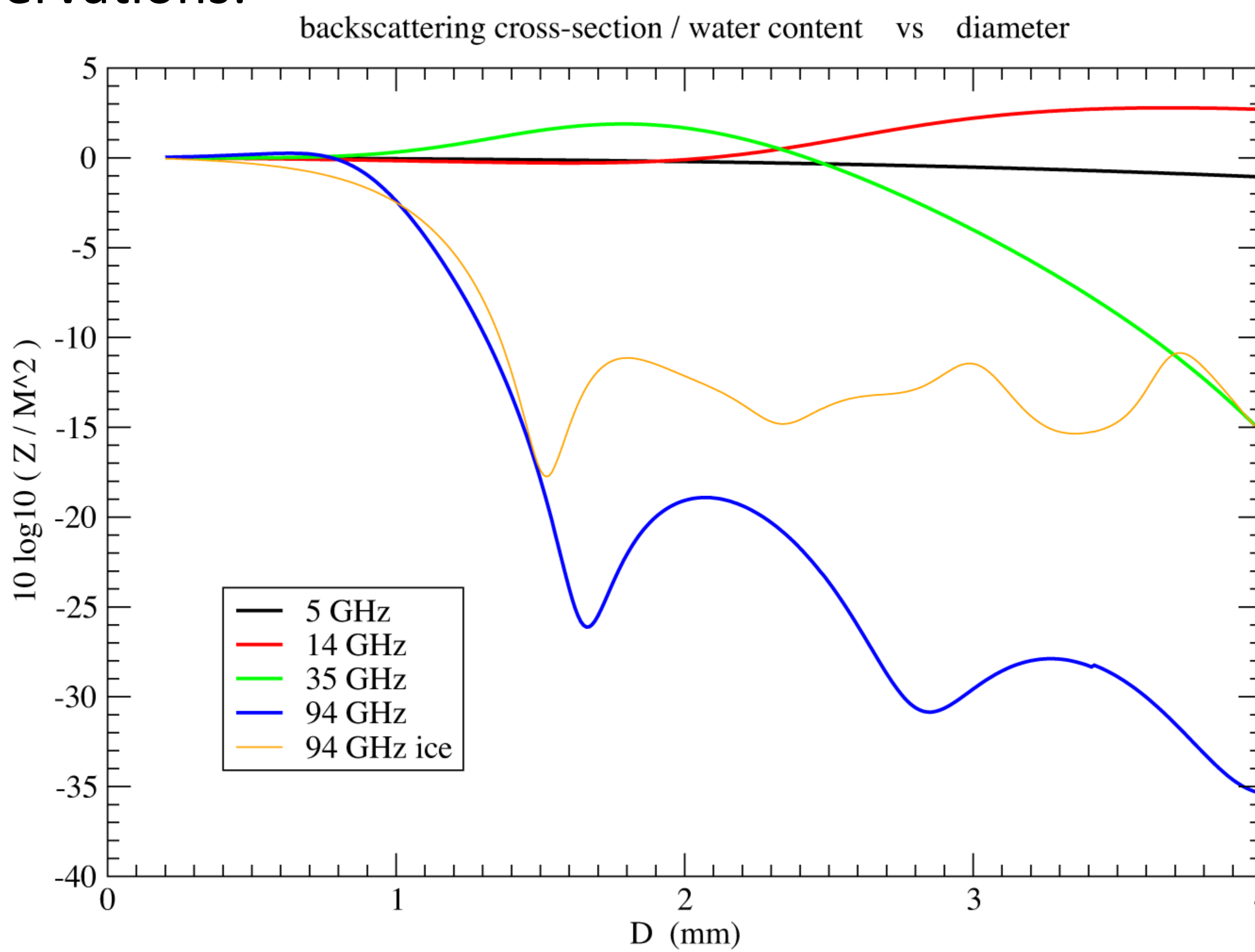
Assimilating rain observations requires addressing 4 or 5 issues

1) Actual observations are highly non-linear functions of the underlying rain variables **as well as other state variables** – and the latter are not retrieved when rain estimates are made from the observations, they are assumed to have certain values or a given joint behavior
2) Joint behavior of condensation variables with other state variables has never been characterized or quantified
3) Condensation does not always exist – so the "joint behavior" statistics (in particular background covariances) may have to be conditioned on existence, and model position errors must be addressed to have meaningful assimilation
4) There are no realistic dynamics for the detail precip variables – worse, the models do incorporate "fake realism" (gamma-distributed drop diameters, e.g.)
5) It would be highly desirable to conserve water through the assimilation

“fake realism” in drop diameter distributions:

The size of the hydrometeors is very important to correctly interpret the microwave observations:

Even with
2 or 3 frequencies,
don’t necessarily
end up with
independent
measurements



$\Rightarrow$ Assume closed-form diameter distributions (e.g. exponential or Γ)

“fake realism” in drop diameter distributions:

$\Rightarrow$ Assume closed-form diameter distributions (e.g. exponential or Γ):

$$N(D) = N_0 \; D^{\mu} \; e^{-\Lambda D}$$

Let’s try to interpret the parameters in terms of physically meaningful quantities:

$$D_m = \frac{\int D \; D^3 N(D) \; dD}{\int D^3 N(D) \; dD} = \frac{\mu + 4}{\Lambda}$$

$$q = \int \frac{4}{3}\pi \left(\frac{D}{2}\right)^3 \rho \; N(D) \; dD = \frac{\pi}{6} \frac{\rho \; \Gamma(\mu + 1) \; D_m^{\mu+1}}{(\mu + 4)^{\mu+1}} \; N_0$$

$$\sigma_m = \left(\frac{\int (D - D_m)^2 \; D^3 N(D) \; dD}{\int D^3 N(D) \; dD} \right)^{1/2} = \frac{D_m}{\sqrt{\mu + 4}}$$

“fake realism” in drop diameter distributions:

$\Rightarrow$ Assume closed-form diameter distributions (e.g. exponential or Γ):

$$N(D) = N_0 \; D^{\mu} \; e^{-\Lambda D}$$

Let’s try to interpret the parameters in terms of physically meaningful quantities:

$$D_m = \frac{\int D \; D^3 N(D) \; dD}{\int D^3 N(D) \; dD} = \frac{\mu + 4}{\Lambda}$$

$$q = \int \frac{4}{3}\pi \left(\frac{D}{2}\right)^3 \rho \; N(D) \; dD = \frac{\pi}{6} \frac{\rho \; \Gamma(\mu+1) \; D_m^{\mu+1}}{(\mu+4)^{\mu+1}} N_0$$

$$\sigma_m = \left(\frac{\int (D - D_m)^2 \; D^3 N(D) \; dD}{\int D^3 N(D) \; dD} \right)^{1/2} = \frac{D_m}{\sqrt{\mu + 4}}$$

“fake realism” in drop diameter distributions:

$\Rightarrow$ Assume closed-form diameter distributions (e.g. exponential or Γ):

In the models, typically assume N_0 constant and $\mu = 0$.

What that implies is:

$$q = \int \frac{4}{3}\pi\left(\frac{D}{2}\right)^3 \rho\ N(D)\ dD = \frac{\pi}{6}\frac{\rho\ \Gamma(\mu+1)\ D_m^{\mu+1}}{(\mu+4)^{\mu+1}} N_0$$

$$= \frac{\pi\rho}{24} N_0\ D_m$$

“fake realism” in drop diameter distributions:

$\Rightarrow$ Assume closed-form diameter distributions (e.g. exponential or Γ):

In the models, typically assume N_0 constant and μ = 0.

What that implies is: $$D_m = \frac{24}{\pi \rho N_0} q$$

In particular,

- D_m/q = constant, and
- max(D_m)/min(D_m) = max(q)/min(q)

“fake realism” in drop diameter distributions:

$\Rightarrow$ Assume closed-form diameter distributions (e.g. exponential or Γ):

In the models, typically assume N_0 constant and $\mu = 0$.

What that implies is: $$D_m = \frac{24}{\pi \rho N_0} q$$

In particular,

- D_m/q = constant, and
- max(D_m)/min(D_m) = max(q)/min(q)

But 3.5 mm / 0.5 mm $\neq$ max($R^{0.9}$)/min($R^{0.9}$) $\approx$ 100 mm/hr / 0.1 mm/hr

“fake realism” in drop diameter distributions:

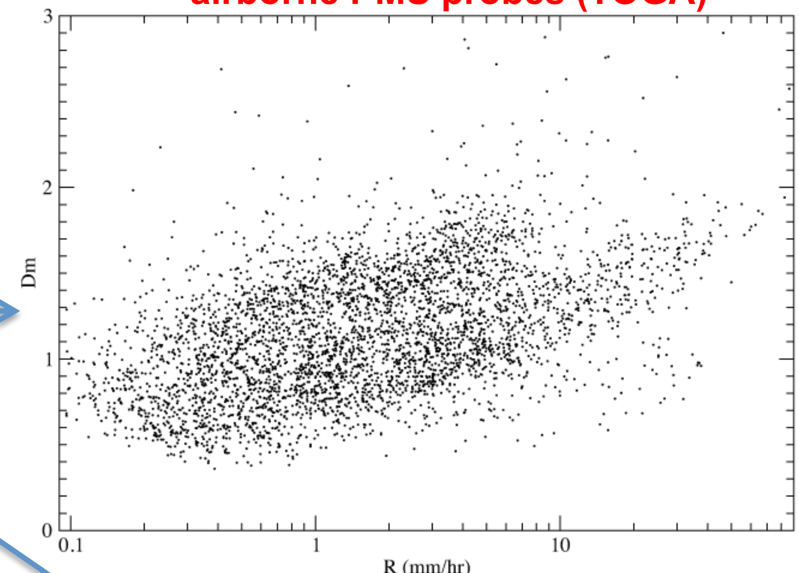
⇒ In fact, hydrometeor data suggests $D_m \sim q^{0.2}$ ± white noise

$D_m \sim q^{0.2}$ behavior in profiler data

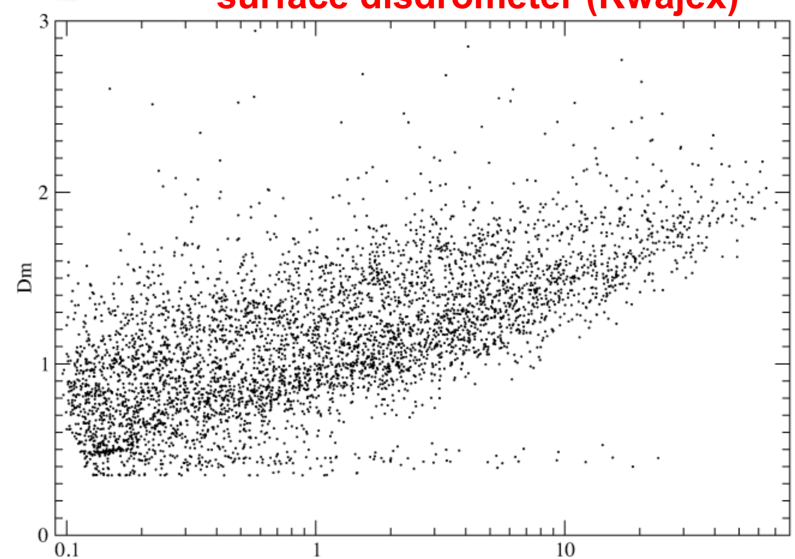
is consistent with TOGA-COARE

and Kwajex data …

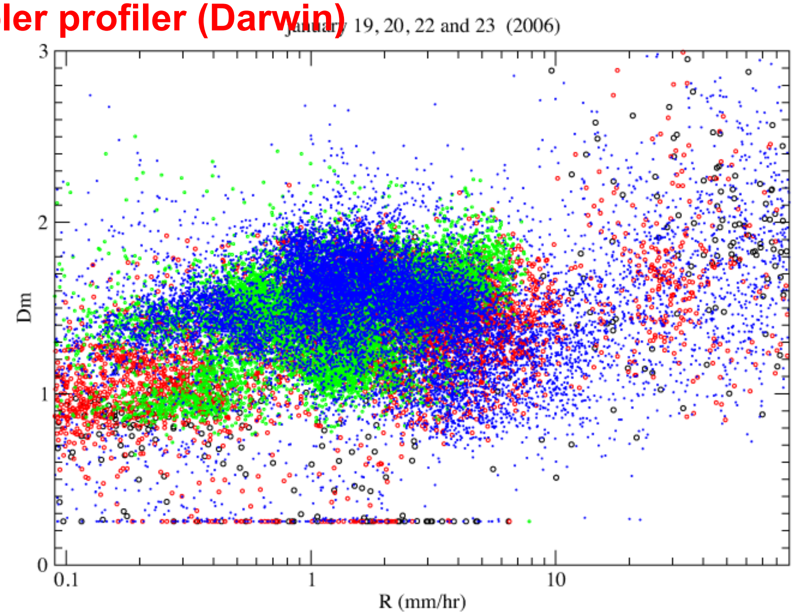
airborne PMS probes (TOGA)



surface disdrometer (Kwajex)



Doppler profiler (Darwin)



“fake realism” in drop diameter distributions:

$\Longrightarrow$ Assume closed-form diameter distributions (e.g. exponential or Γ):

$$N(D) = N_0 \; D^{\mu} \; e^{-\Lambda D}$$

Fixing Λ is at least as problematic:

$$D_m = \frac{\int D \; D^3 N(D) \; dD}{\int D^3 N(D) \; dD} = \frac{\mu + 4}{\Lambda}$$

$$\sigma_m = \left(\frac{\int (D - D_m)^2 \; D^3 N(D) \; dD}{\int D^3 N(D) \; dD} \right)^{1/2} = \frac{D_m}{\sqrt{\mu + 4}}$$

The above imply:

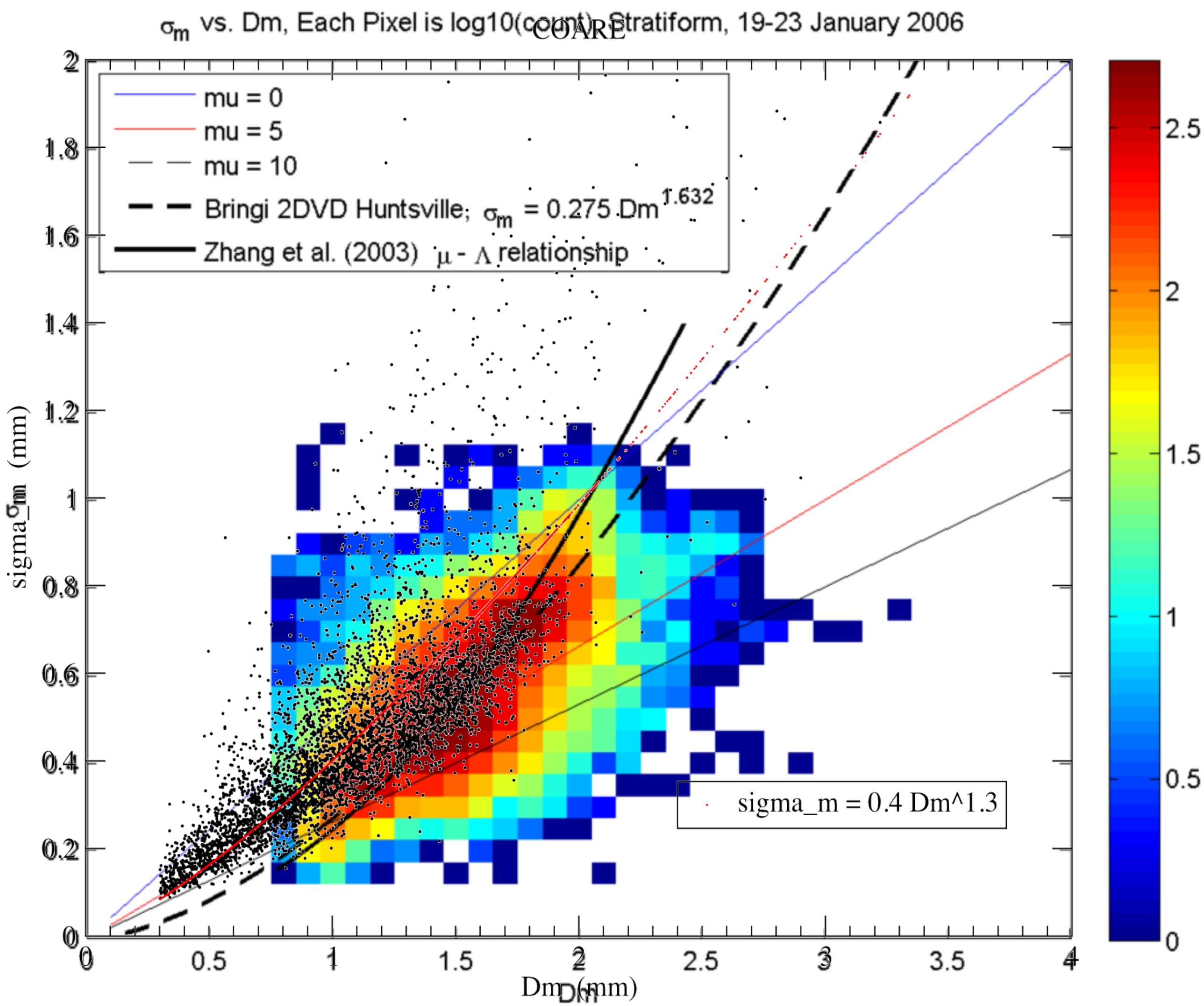
$$D_m = \frac{D_m^2 / \sigma_m^2}{\Lambda} \quad \Longrightarrow \quad \frac{D_m}{\sigma_m^2} = const$$

Observations indicate:

$$\sigma_m = const \cdot D_m^{1.6} \pm noise$$

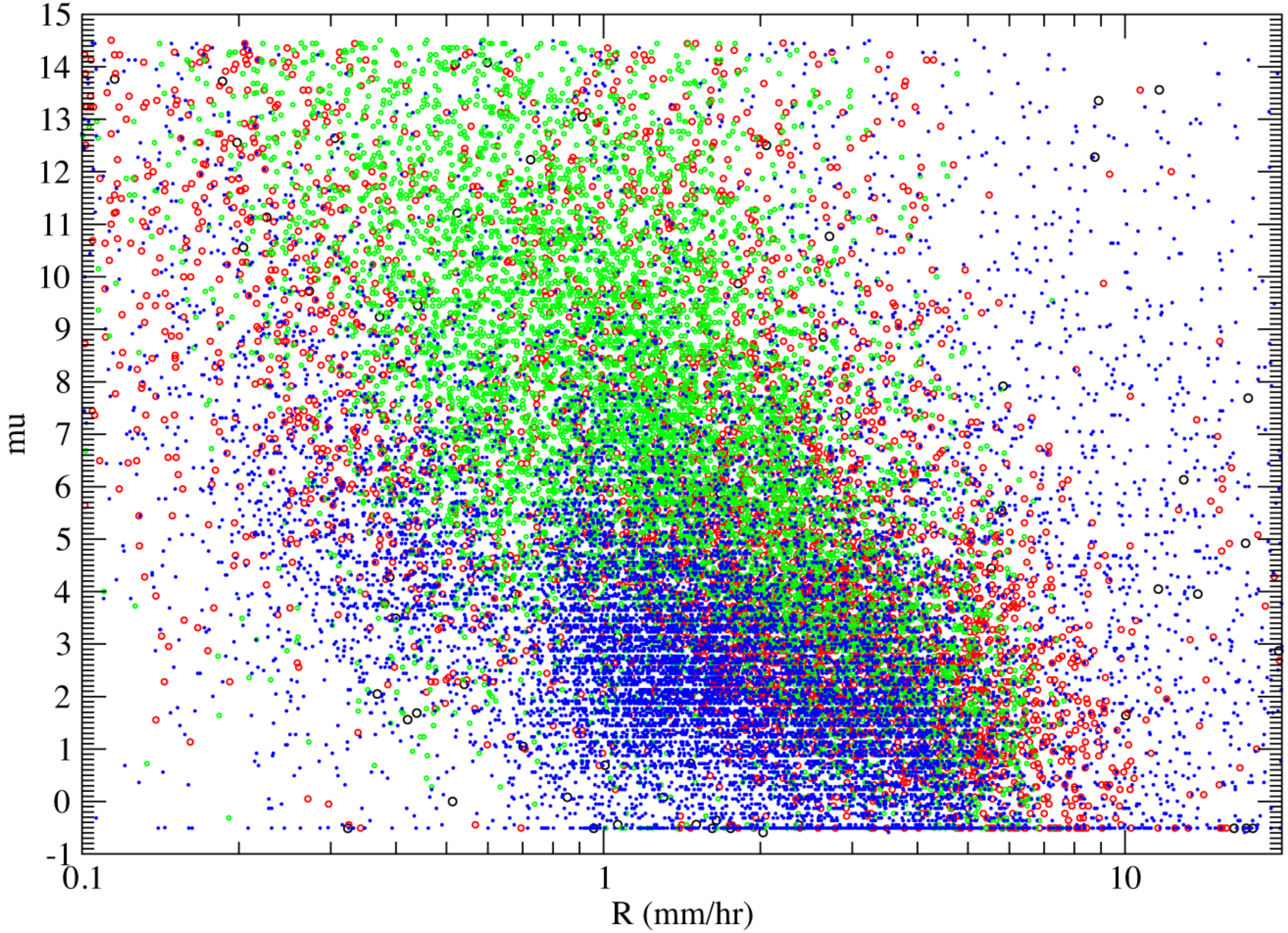
“fake realism” in drop diameter distributions:

$\Rightarrow$ Observations indicate: $$\sigma_m = const \cdot D_m^{1.5} \pm noise$$



Furthermore, μ is neither 0 nor constant (& neither are N_0, Λ):

Darwin profiler, January 19+20 (blue), 22 (red) and 23 (green), 2006



Coup de grâce:

$$\frac{\partial \mu}{\partial t} + V \cdot \nabla \mu = ?$$

$$\frac{\partial \sigma_m}{\partial t} + V \cdot \nabla \sigma_m = ?$$

Alternative:

$$T_b = \underbrace{\varepsilon(w)\ T_S\ e^{-\int_0^\infty k_{ext}}}_{\text{surface } \wedge} + \underbrace{\int k_{abs}(h)\ T(h)\ e^{-\int_h^\infty k_{ext}}\, dh}_{\text{condensation } \wedge} + \underbrace{\left(\int k_{abs}(h)\ T(h)\ e^{-\int_0^h k_{ext}}\, dh\right)\big(1-\varepsilon(w)\big)e^{-\int_0^\infty k_{ext}}}_{\text{condensation } \vee\text{, reflected } \wedge}$$

$$= \varepsilon(w)\ T_S\ A(atmosphere) + B(atmosphere)\ + C(atmosphere)\big(1-\varepsilon(w)\big)$$

$$= \varepsilon(w)\ \Big(\ T_S\ A(atmosphere)\ +\ F(atmosphere)\Big) +\ G(atmosphere)$$

$\Rightarrow$ assimilate the first two T_b principal components $\mathbf{T}_1$ and $\mathbf{T}_2$, expressed in terms of the first two vertical principal components q_1' and q_2' of the water variables + the first four vertical components of the remaining dynamical variables

Alternative:

$$T_b = \underbrace{\varepsilon(w)\; T_S\; e^{-\int_0^\infty k_{ext}}}_{\text{surface up}} + \underbrace{\int k_{abs}(h)\; T(h)\; e^{-\int_h^\infty k_{ext}}\, dh}_{\text{condensation up}} + \underbrace{\left(\int k_{abs}(h)\; T(h)\; e^{-\int_0^h k_{ext}}\, dh \right) \big(1-\varepsilon(w)\big) e^{-\int_0^\infty k_{ext}}}_{\text{condensation down, reflected up}}$$

$$= \varepsilon(w)\; T_S\; A(atmosphere) + B(atmosphere) \;\; + C(atmosphere)\big(1-\varepsilon(w)\big)$$

$$= \varepsilon(w) \left(\; T_S\; A(q_1', q_2', x_1, \ldots, x_4) \;\; + \;\; F(q_1', q_2', x_1, \ldots, x_4) \right) + \;\; G(q_1', q_2', x_1, \ldots, x_4)$$

$\Rightarrow$ assimilate the first two T_b principal components $\mathbf{T}_1$ and $\mathbf{T}_2$, expressed in terms of the first two vertical principal components q_1’ and q_2’ of the water variables + the first four vertical components of the remaining dynamical variables

The main reason to separate out the condensed-water-content variables is that we want to force their inclusion in our truncated background-covariances representation …

Alternative:

The (x_1, x_2) that minimizes

$$(x_1-x_{b1})^2/\sigma_{b1}^2 + (x_2-x_{b2})^2/\sigma_{b2}^2 \qquad + \qquad (f(x_1)-O)^2/r^2$$

will necessarily have $x_2 = x_{b2}$

The main reason to separate out the condensed-water-content variables is that we want to force their inclusion in our truncated background-covariances representation …

Alternative:

The (x_1, x_2) that minimizes

$$(x_1 - x_{b1})^2/\sigma_{b1}^2 + (x_2 - x_{b2})^2/\sigma_{b2}^2 \qquad + \qquad (f(x_1) - O)^2/r^2$$

will necessarily have $x_2 = x_{b2}$

$\Rightarrow$ May not need to worry, as 1st exercise with all 300 variables together had PC_1 and PC_2 dominated by w and rain (& snow in PC_2) … and P_{surf} (?)

The main reason to separate out the condensed-water-content variables is that we want to force their inclusion in our truncated background-covariances representation …

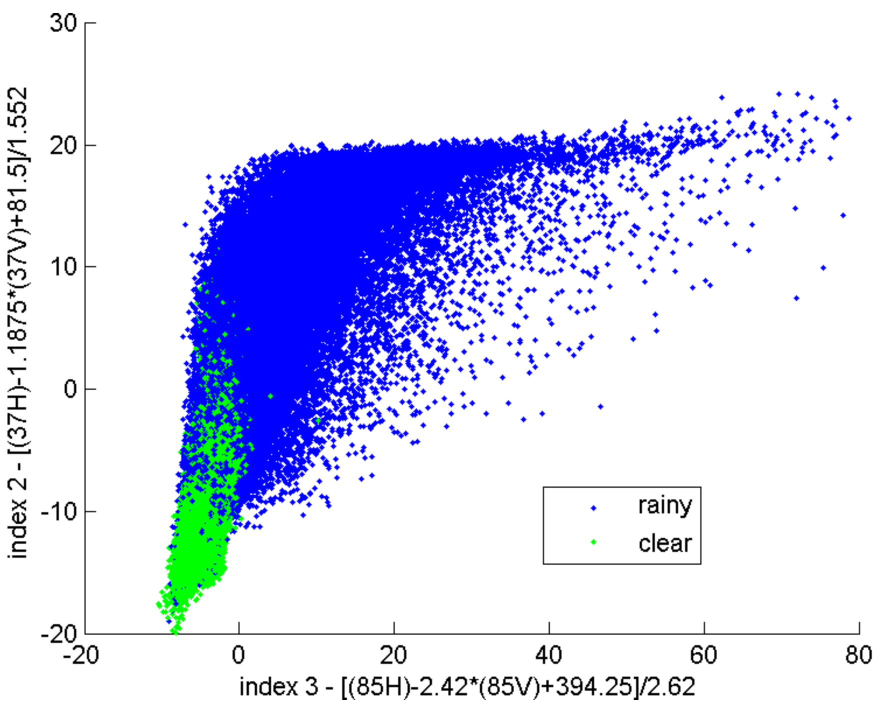
Alternative:

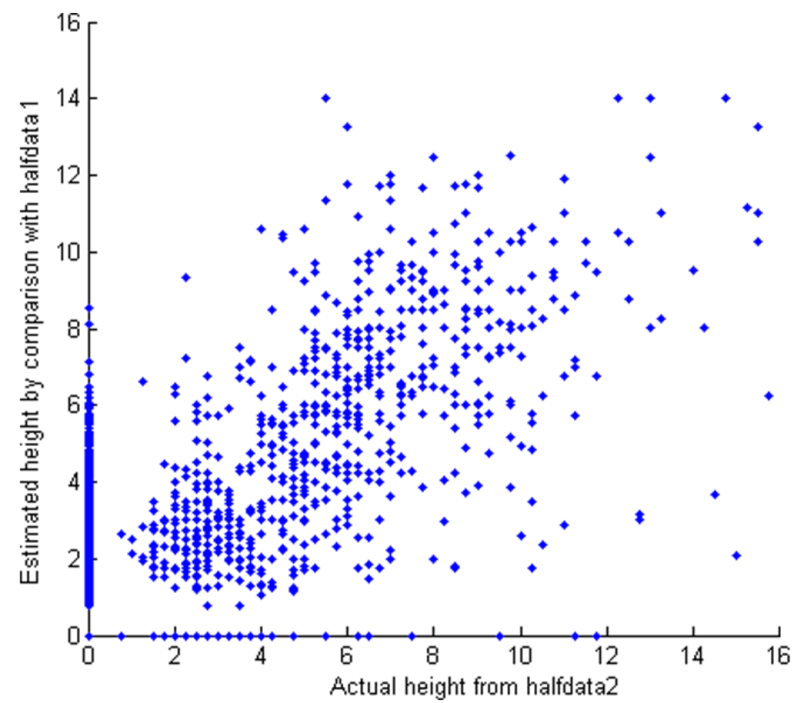
In addition to assimilating T_b principal components 1&2 in the form

$$T_i' = \varepsilon(w)\left(T_S \; A_i(q_1', q_2', x_1, \ldots, x_4) \; + \; F_i(q_1', q_2', x_1, \ldots, x_4)\right) + \; G_i(q_1', q_2', x_1, \ldots, x_4) + noise$$

assimilate rain detection too, i.e. observe q’=0+noise when our detector says there is no rain:

Example from analysis of all 2005 AMSR-E and TMI measurements over Atlantic hurricanes





2) covariances of state variables (including condensation) can be represented: change variables to use vertical principal components, so the new variables are vertically uncorrelated, then verify that they are also horizontally uncorrelated:

Variables 1-30 are the horizontal wind speeds V

Variables 31-60 are the vertical wind speeds W

Variables 61-90 are the temperatures T

Variables 91-120 are the pressures P

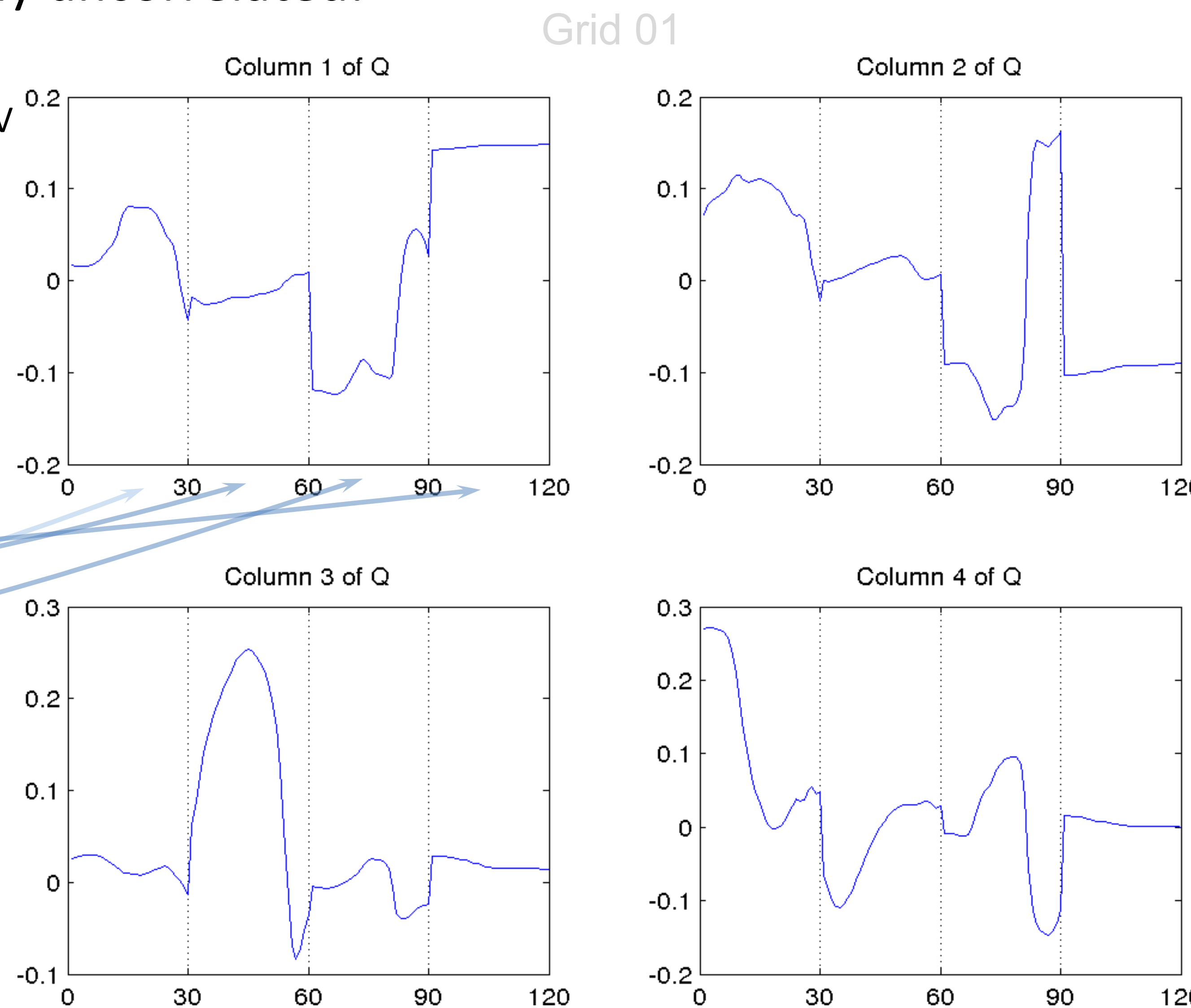
=>

PC_1 is essentially P/T

PC_2 is $V - T_{lo} + T_{hi} - P$

PC_3 is the average W

PC_4 is $V_{surf} - W_{surf} - T_{hi}$



2) covariances of state variables (including condensation) can be represented: change variables to use vertical principal components, so the new variables are vertically uncorrelated, then verify that they are also horizontally uncorrelated:

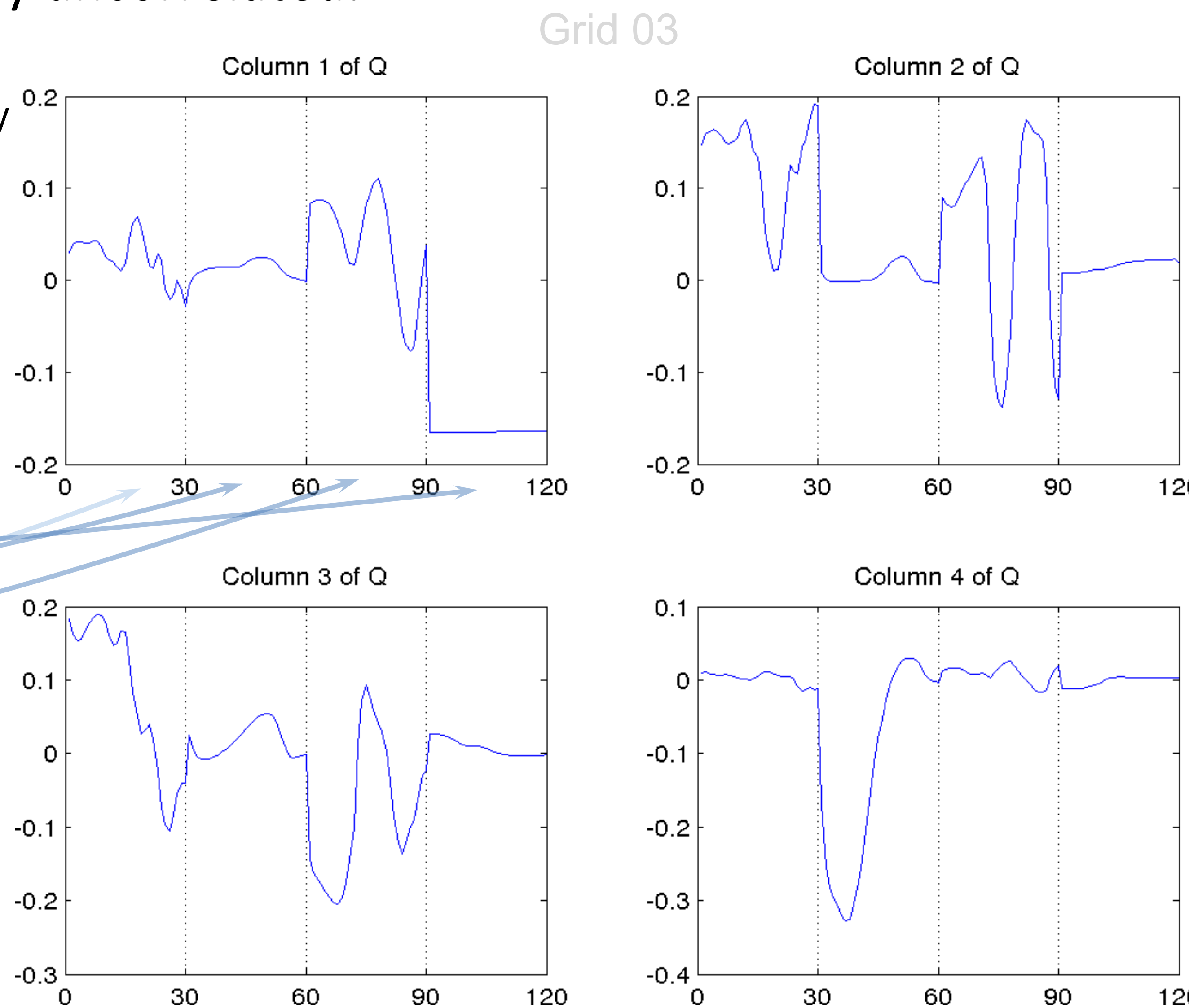
Variables 1-30 are
the horizontal wind speeds V

Variables 31-60 are
the vertical wind speeds W

Variables 61-90 are
the temperatures T

Variables 91-120 are
the pressures P

=>



PC_1 is essentially T/P

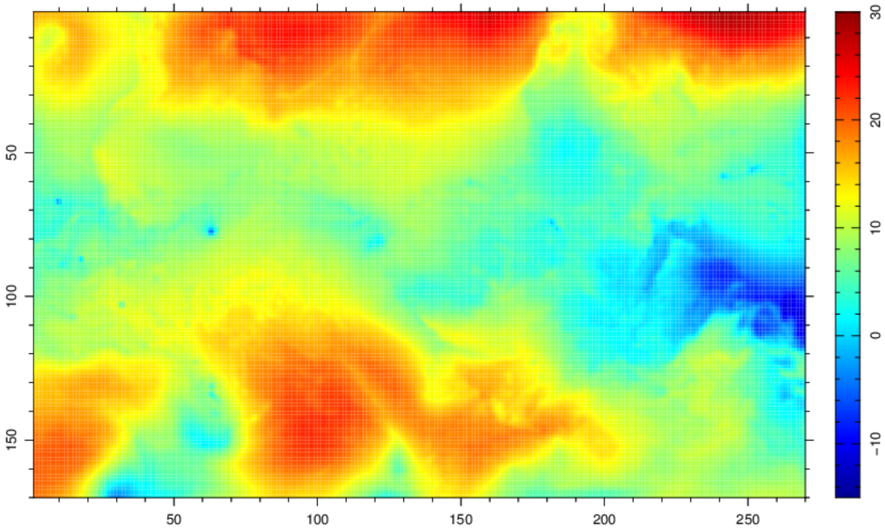
PC_2 is $V + T_{lo} - T_{mid} + T_{hi}$

PC_4 is the average - W

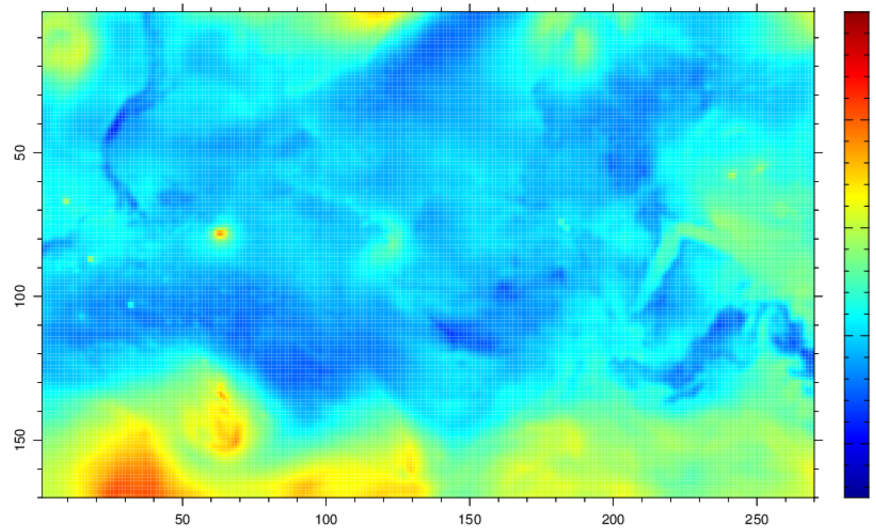
PC_3 is $V_{surf} - T_{lo} + T_{mid} - T_{hi}$

First 4 principal components in Ernesto (2006):

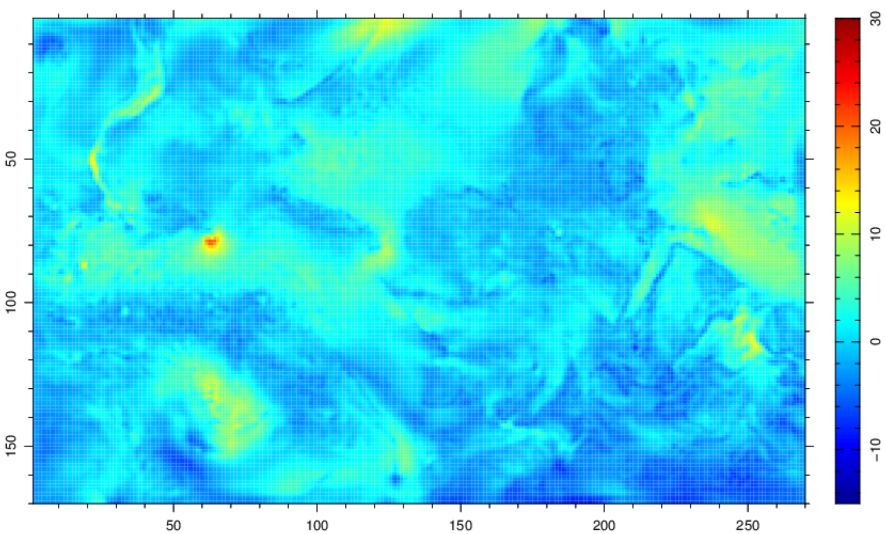
Grid 01



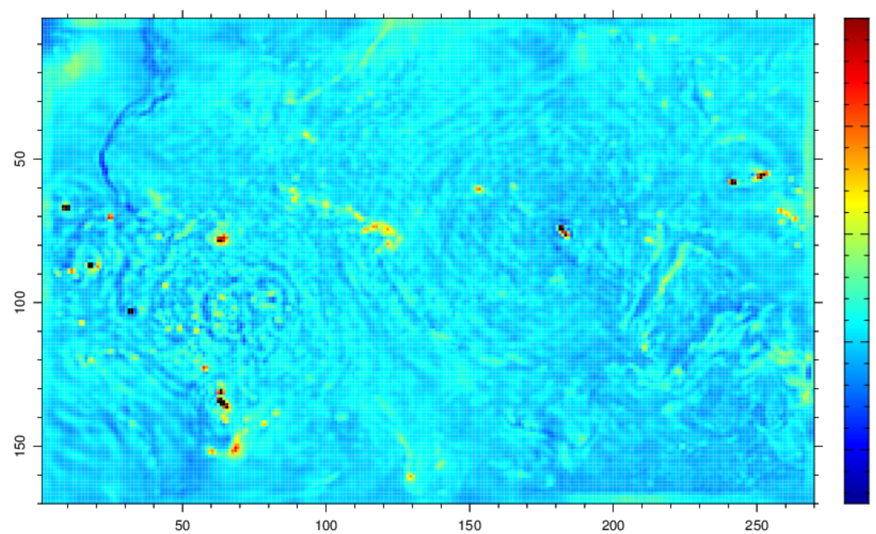
PC_1 (~ P/T)



PC_2 (~ $V - T_{lo} + T_{hi} - P$)



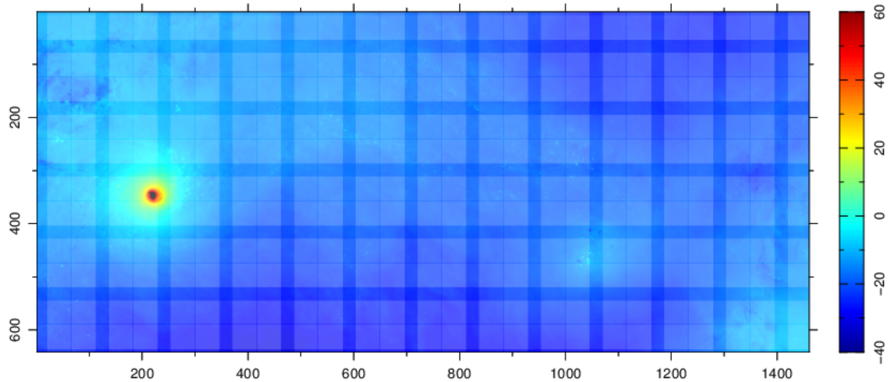
PC_4 (~ $V_{surf} - W_{surf} - T_{hi}$)

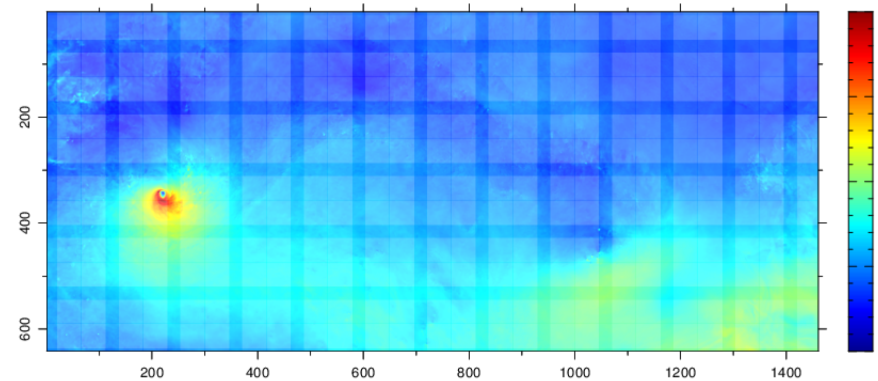


PC_3 (~ average W)

First 4 principal components in Ernesto (2006):

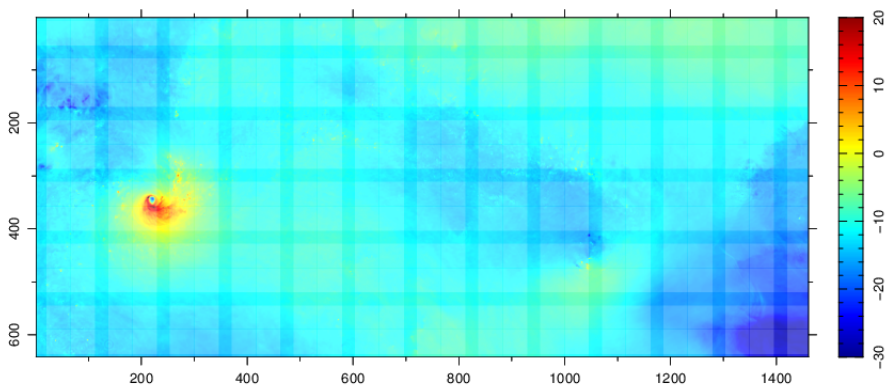
Grid 03

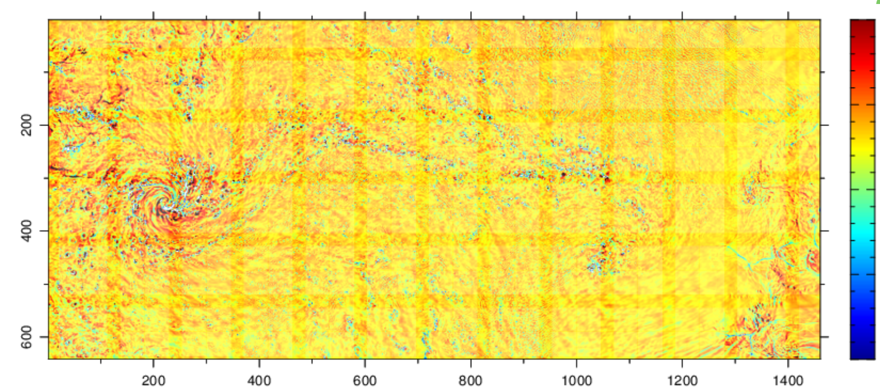




PC_1 (~ T/P)

PC_2 (~ $V + T_{lo} - T_{mid} + T_{hi}$)



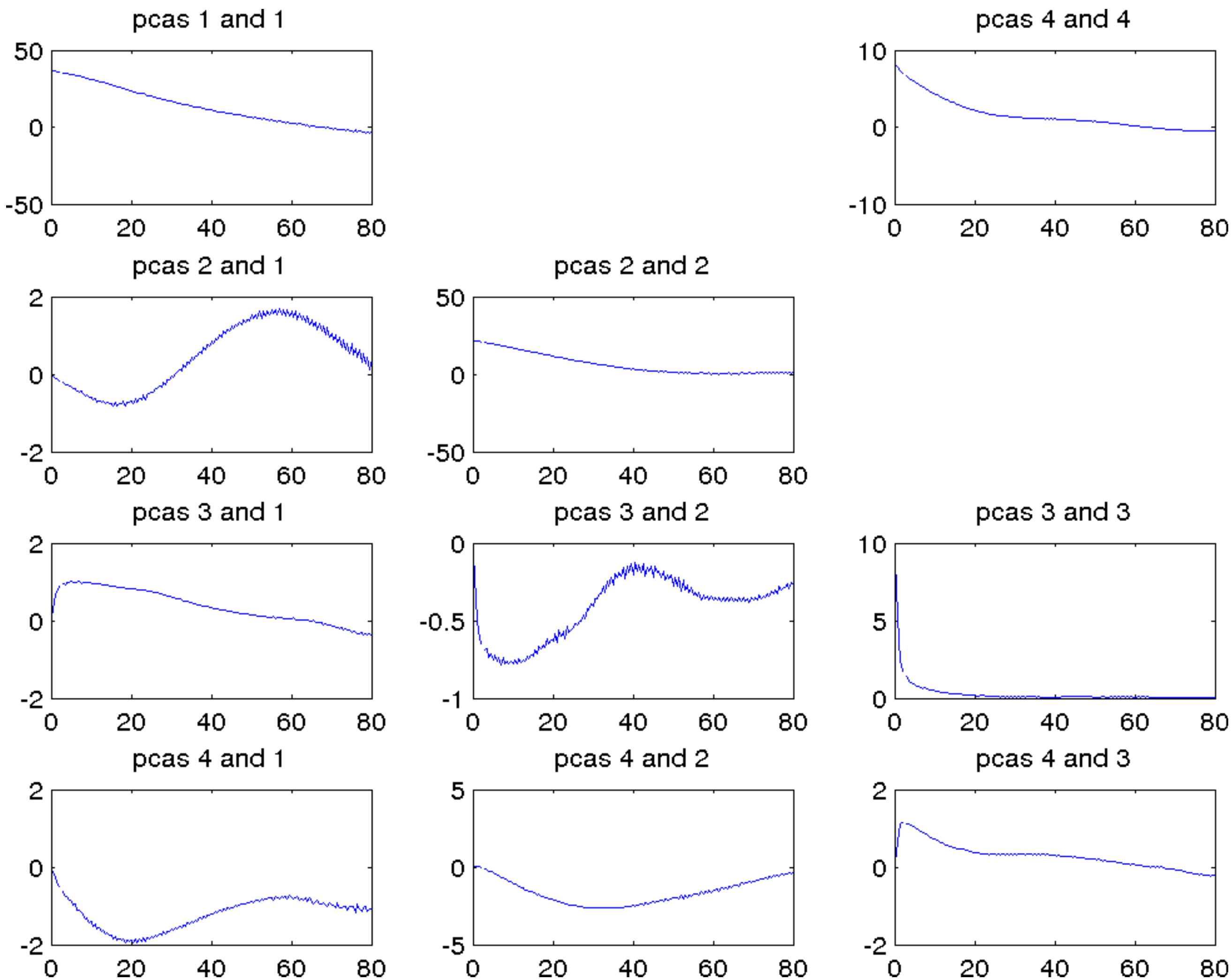


PC_3 (~ $V_{surf} - T_{lo} + T_{mid} - T_{hi}$)

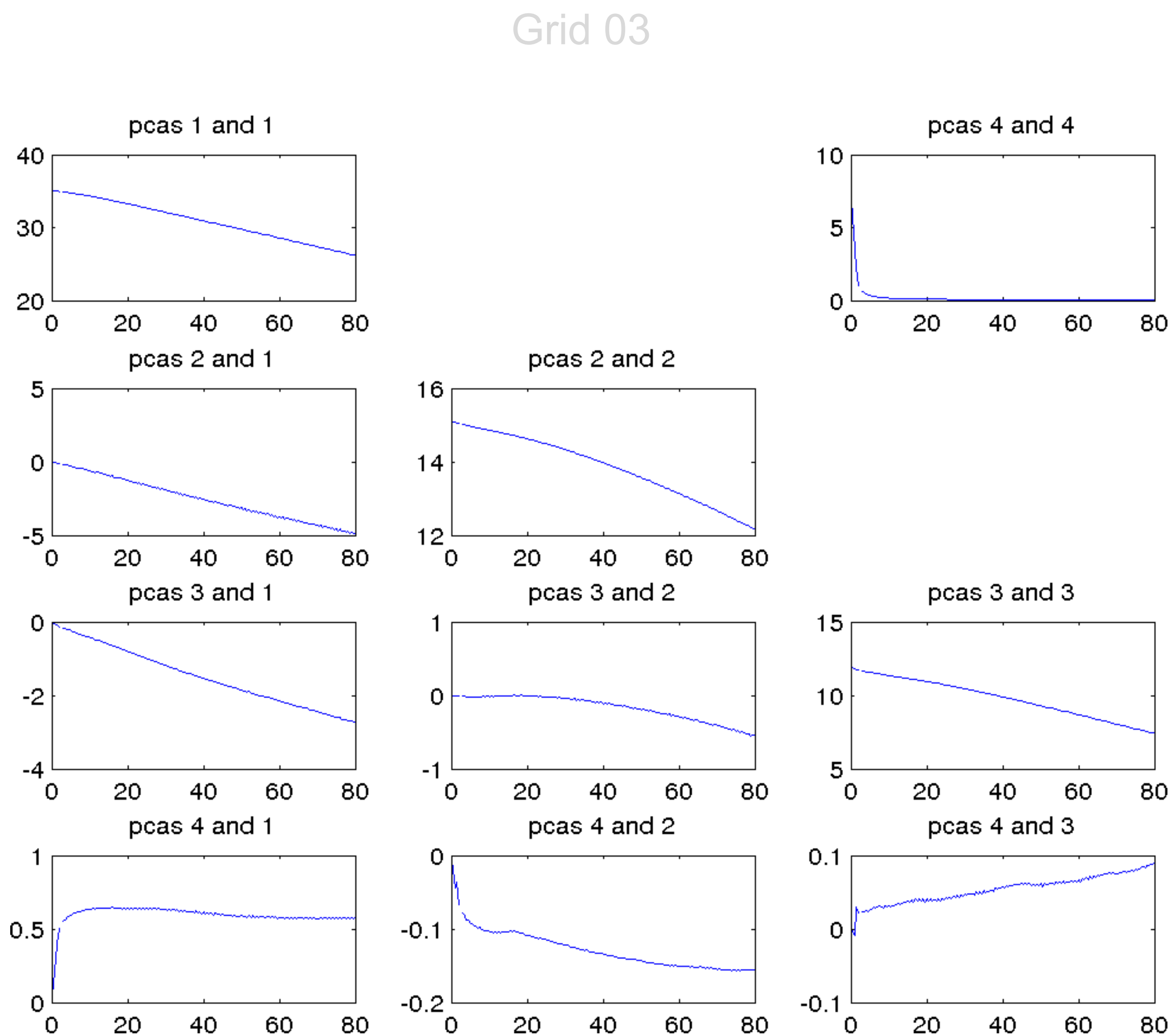
PC_4 (~ average - W)

Horizontal covariance of first 4 principal components in Ernesto:

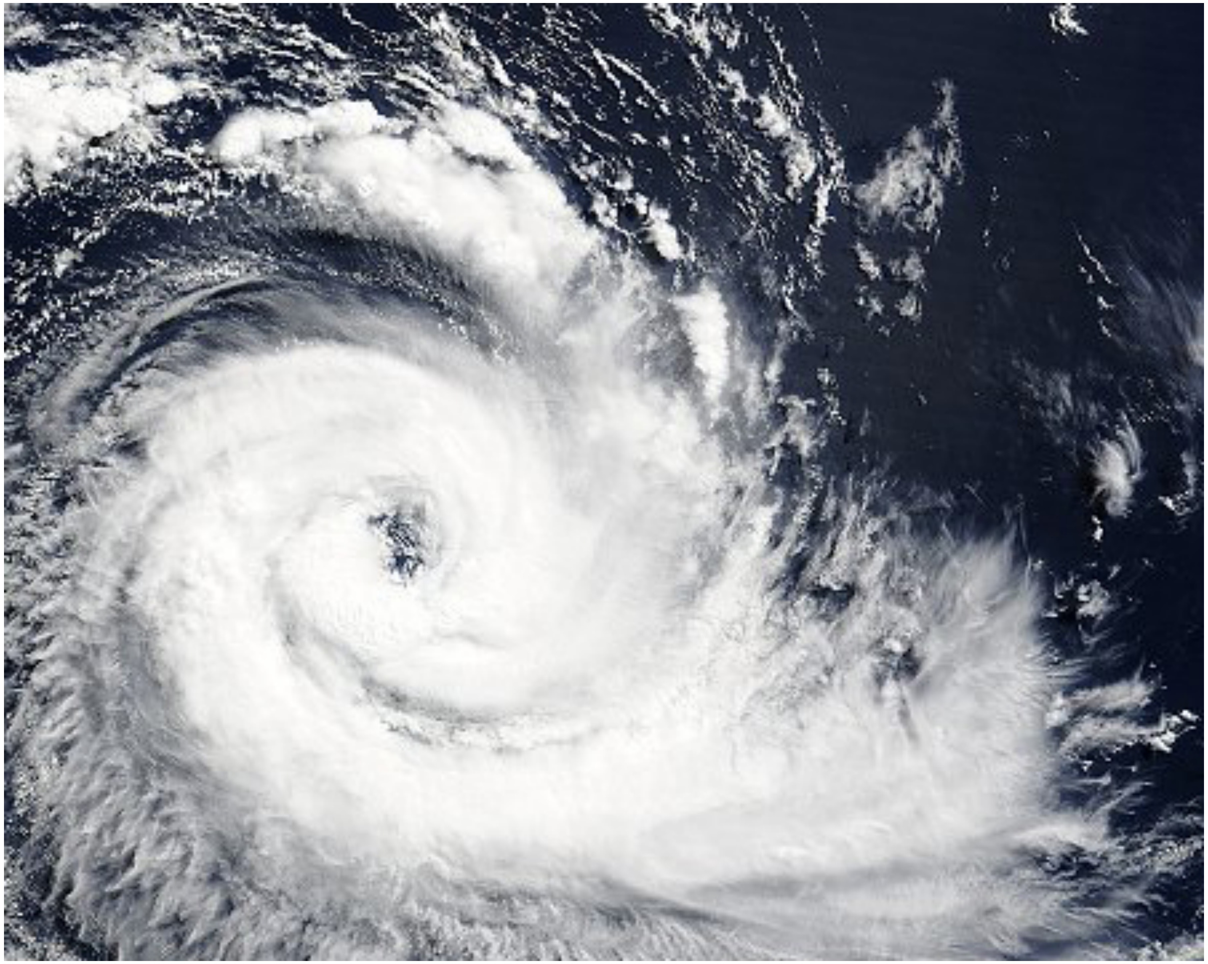
Grid 01



Horizontal covariance of first 4 principal components in Ernesto:

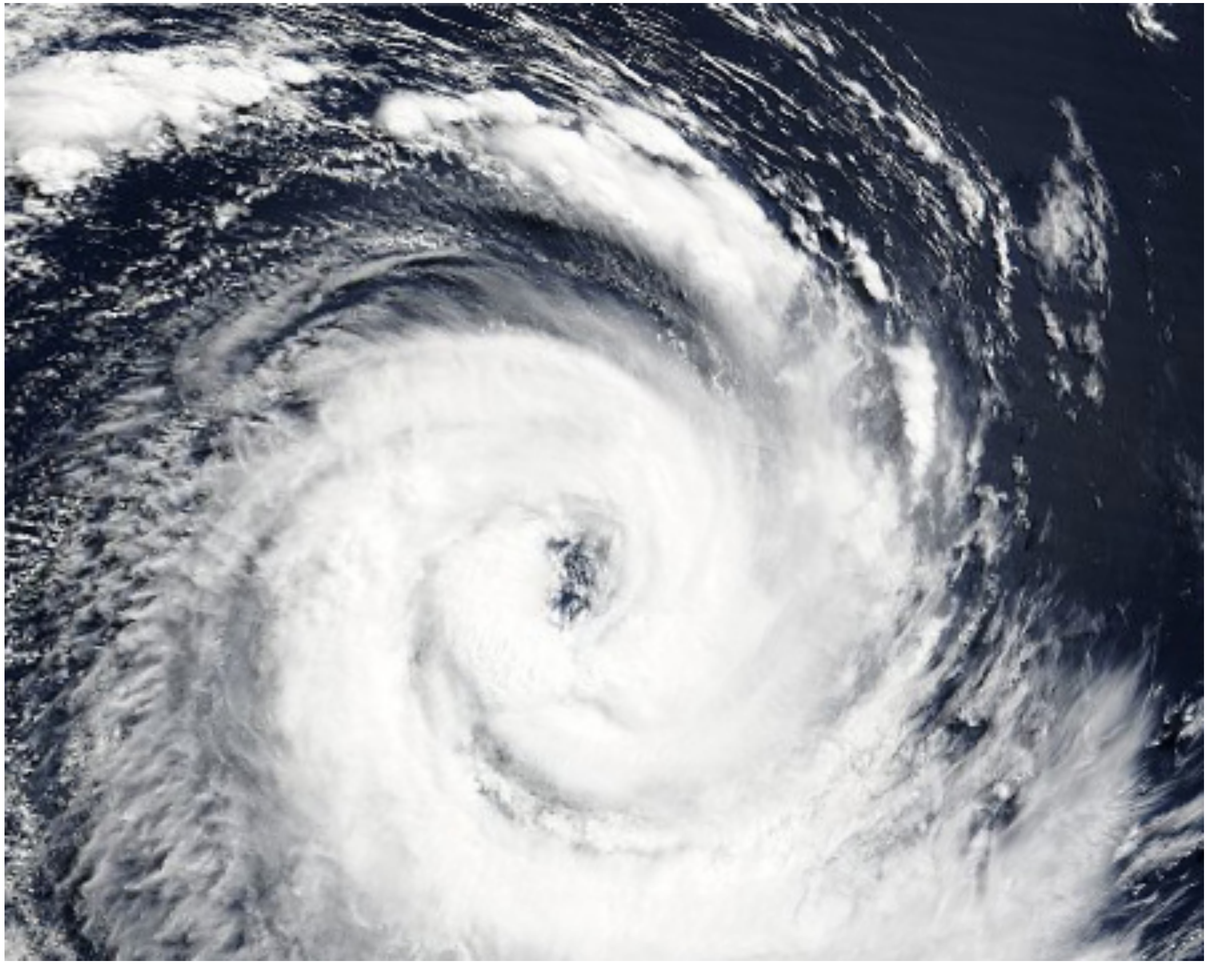


3) … model position errors (…) to have meaningful assimilation:



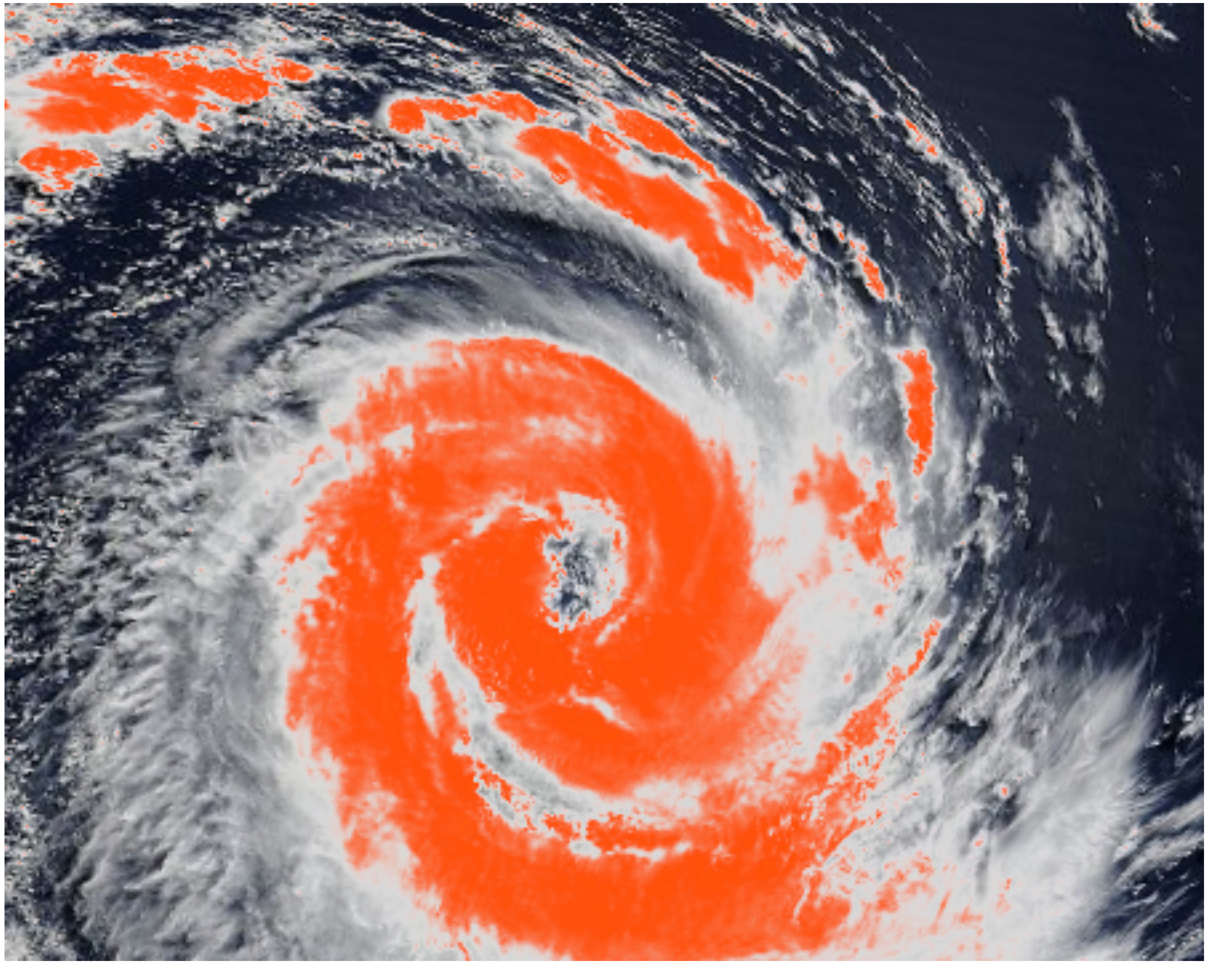
Observed condensed water
(existence independent of altitude)

3) … model position errors (…) to have meaningful assimilation:



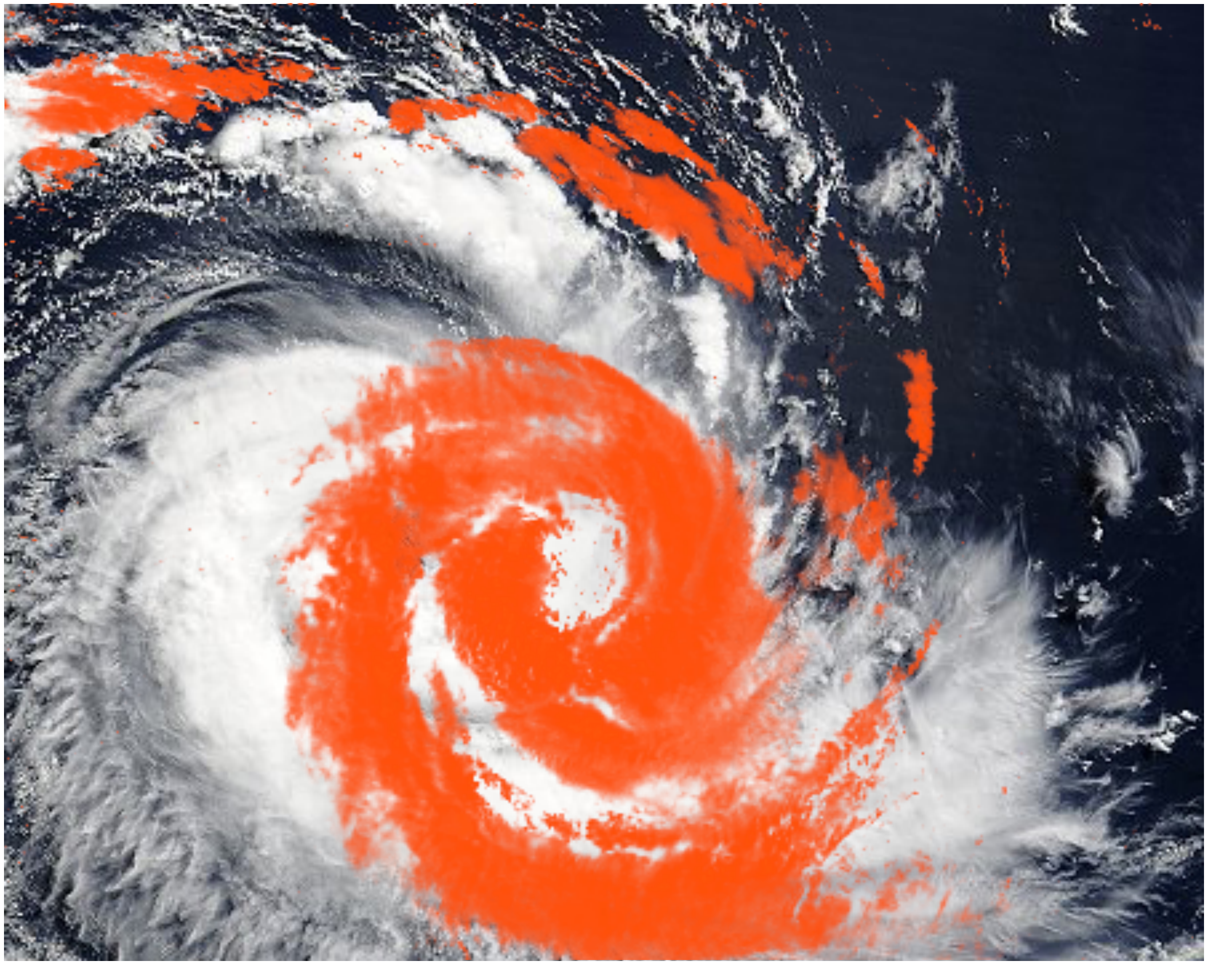
Condensed water in model forecast

3) … model position errors (…) to have meaningful assimilation:



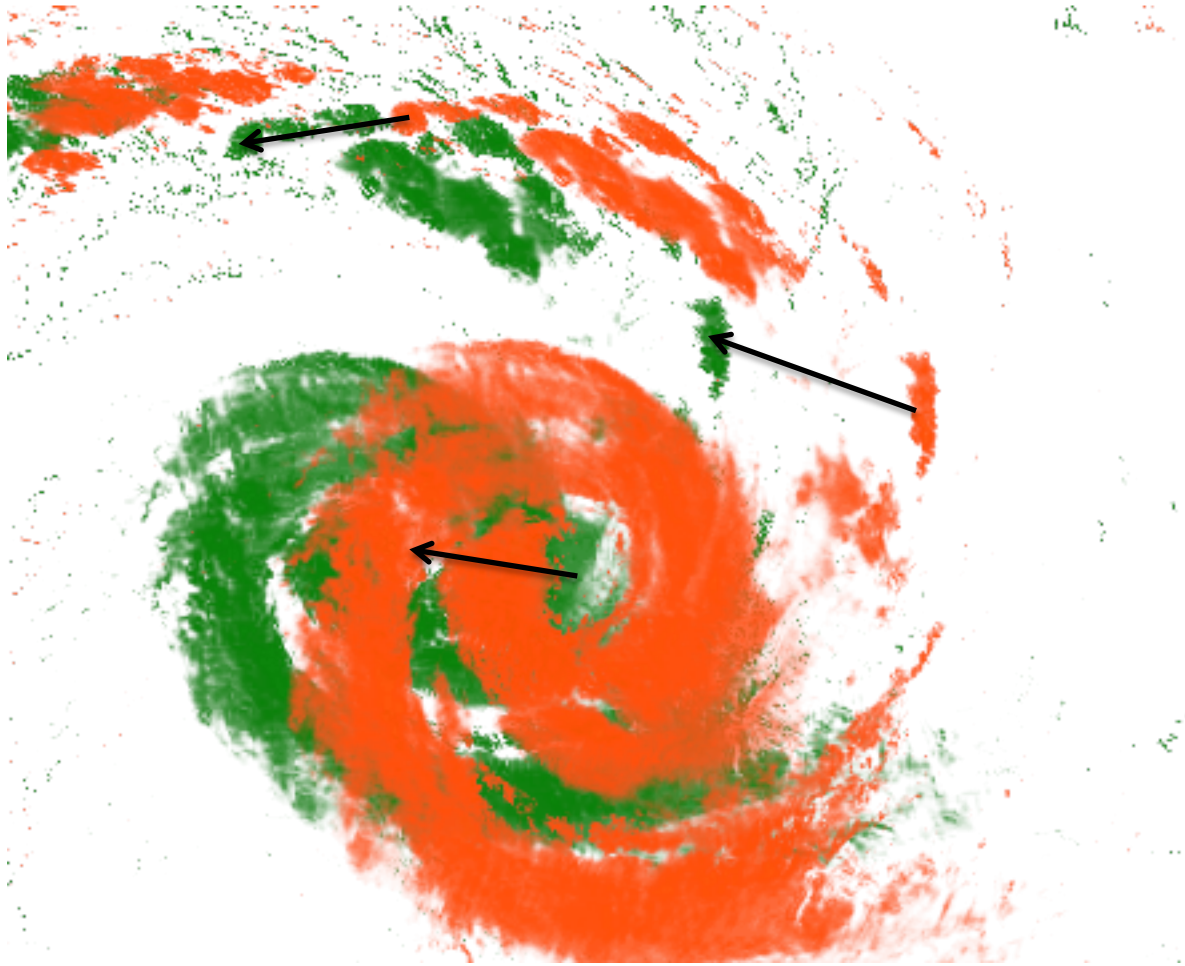
Rain (actually, thresholded 85GHz radiance, a la NHC) in model forecast

3) … model position errors (…) to have meaningful assimilation:



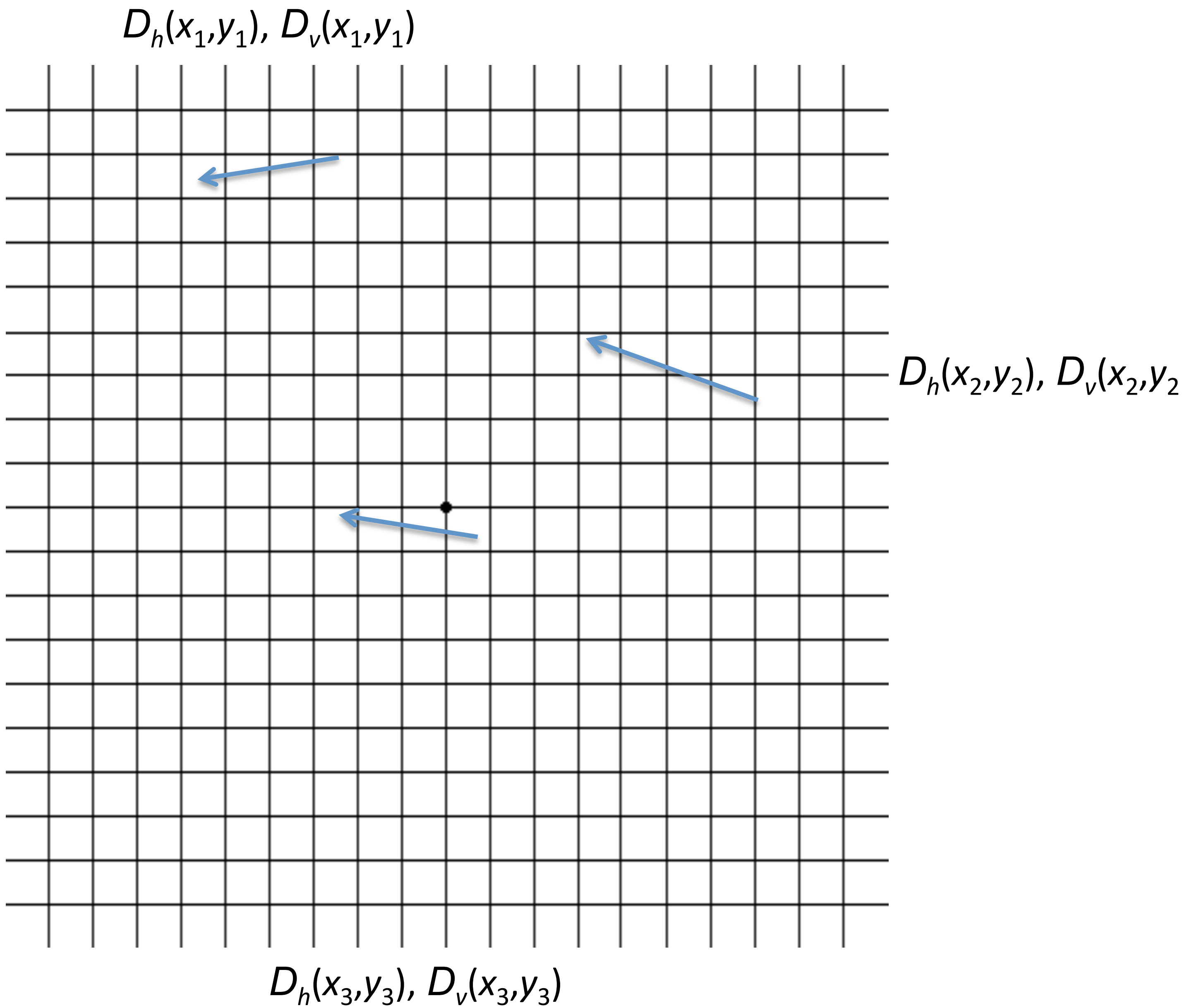
Rain (actually, thresholded 85GHz radiance, a la NHC) in model forecast is displaced relative to observed field

3) … model position errors (…) to have meaningful assimilation:



Rain (actually, thresholded 85GHz radiance, a la NHC) in model forecast is displaced relative to observed field

3) … model position errors (…) to have meaningful assimilation:



3) … model position errors (…) to have meaningful assimilation:

$\Rightarrow$ Would like to use these three (D_h, D_v) to define the position correction at any point (x, y) on the grid, so that all the state variables in the model can be re-sampled (i.e. interpolated). Natural to try:

$$\Delta_h(x,y) = \Sigma_{i=1,3}\, D_h(x_i,y_i)\, \exp[-0.5((x-x_i)^2+(y-y_i)^2)/\sigma_i^2]$$

$$\Delta_v(x,y) = \Sigma_{i=1,3}\, D_v(x_i,y_i)\, \exp[-0.5((x-x_i)^2+(y-y_i)^2)/\sigma_i^2]$$

with the σ_i inversely proportional to how unique and robust the local correction appears to be …

More rigorously …

1) Find local (64x64) translations ($D_h^{(0)}(m,n)$, $D_v^{(0)}(m,n)$) that maximize the local correlation (over the (m,n)th 64x64 subdomain) between $H(\underline{X}^f$-displaced-by-$D^{(0)})$ and T_{85} a la Cmorph – and estimate the Hessian of the correlation for each subdomain

2) Starting with ($D_h^{(0)}(m,n)$, $D_v^{(0)}(m,n)$) , look for the final optimum ($D_h(m,n)$, $D_v(m,n)$) which minimize

$$[T_{85} - H(\underline{X}^f\text{-displaced-by-}\Delta)^t\, C_T^{-1}\, (T_{85} - H(\underline{X}^f\text{-displaced-by-}\Delta)] + \text{(regularization)}$$

where

$$\Delta_h(x,y) = \Sigma_{m,n}\, D_h(m,n) \exp[-0.5((x-x_{m,n})^2+(y-y_{m,n})^2)/\sigma_{m,n}^{\;2}]$$
$$\Delta_v(x,y) = \Sigma_{m,n}\, D_v(m,n) \exp[-0.5((x-x_{m,n})^2+(y-y_{m,n})^2)/\sigma_{m,n}^{\;2}]$$

and where $\sigma_{m,n}$ is inversely proportional to the eigenvalues of the Hessian of the correlation at ($D_h^{(0)}(m,n)$, $D_v^{(0)}(m,n)$),

following Ravela's and Aonashi's approaches

5) It would be highly desirable to conserve water through the assimilation (very speculative at this stage)

A simple way to ensure that a quantity q is conserved would be to **change variables** to a new set of variables in which q would be one of the variables, and the remaining ones would be uncorrelated and therefore in its orthogonal complement, and the assimilation would be performed in this orthogonal-complement .

Making such a change of variables is not always possible. But in the case where q = total water, this quantity can be included in a linear change of variables so that

in the new set **$\{q, x_1, \ldots, x_N\}$, all (q,x_i) are uncorrelated,**
then perform the analysis on $\{x_1, \ldots, x_N\}$ only.

This is probably best done with a multi-scale approach, otherwise the correlations are insignificant …