WMO DATA ASSIMILATION SYMPOSIUM

11-15 September 2017

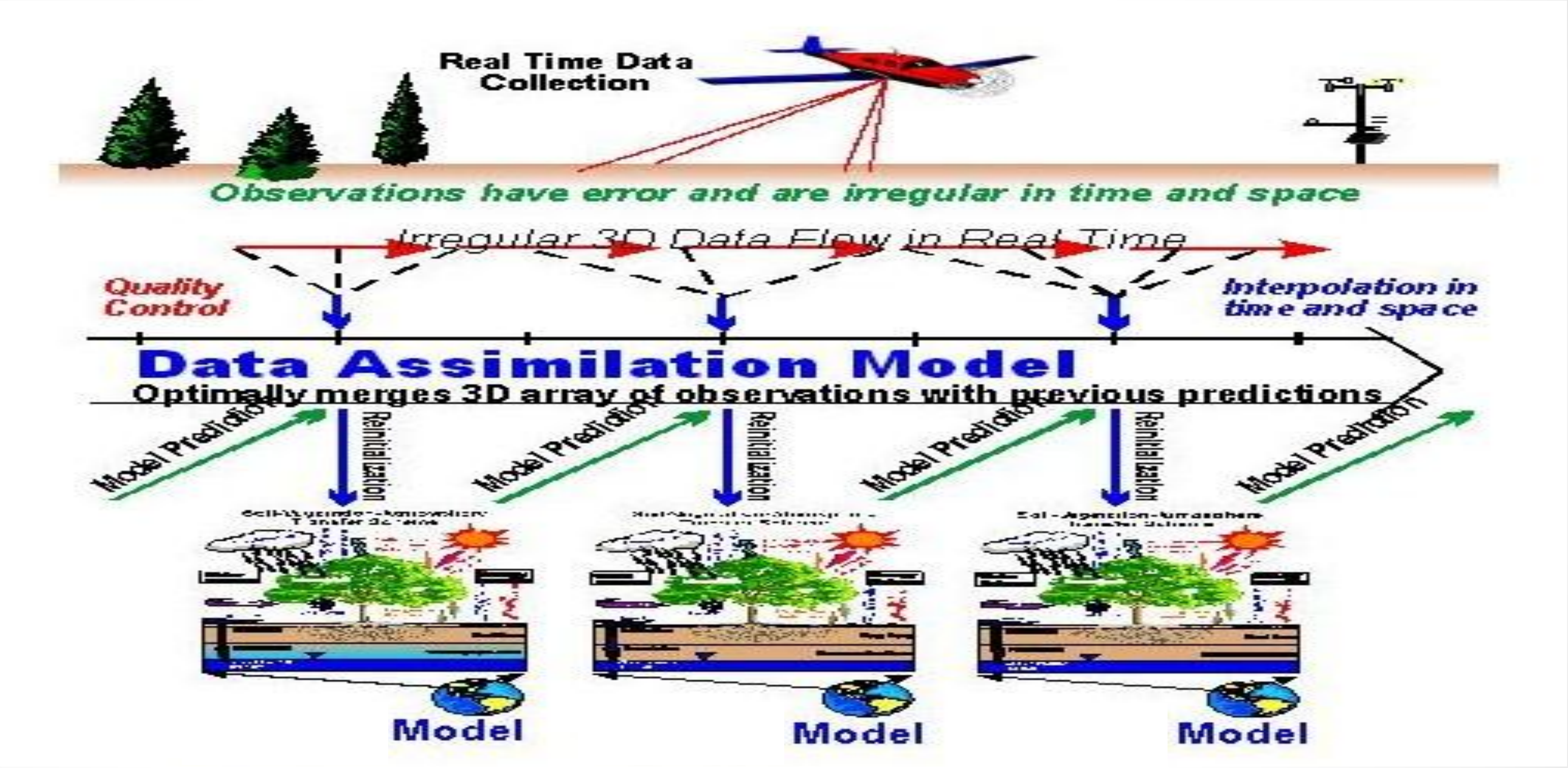
Supervised neural network for data assimilation on atmospheric general circulation model

Rosangela S. Cintra[1], Haroldo F. de Campos Velho[1], Steven Cocke[2]

[1] National Institute for Space Research, (INPE), S. J. Campos, SP, Brazil
[2] Florida State University (FSU), Tallahassee, FL, USA

Data Assimilation

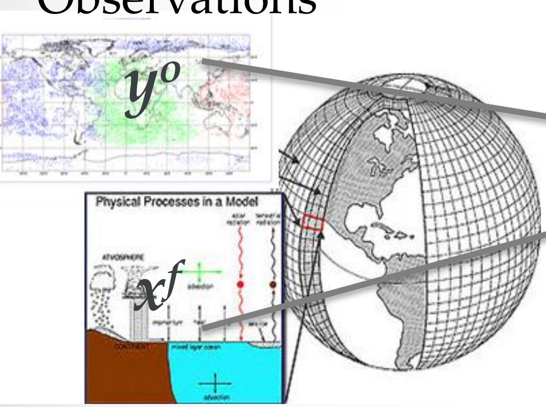
DA cycle



Nowadays: challenge for DA with use the exponential growth of the number of observations available + high resolution models. The weather predictions need to be ready to deal with this scenario.

Data Assimilation

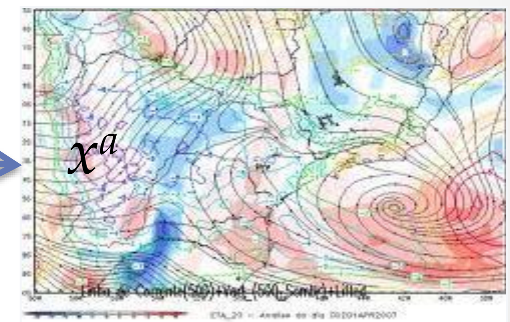
Observations



General Circulation
Model forecast

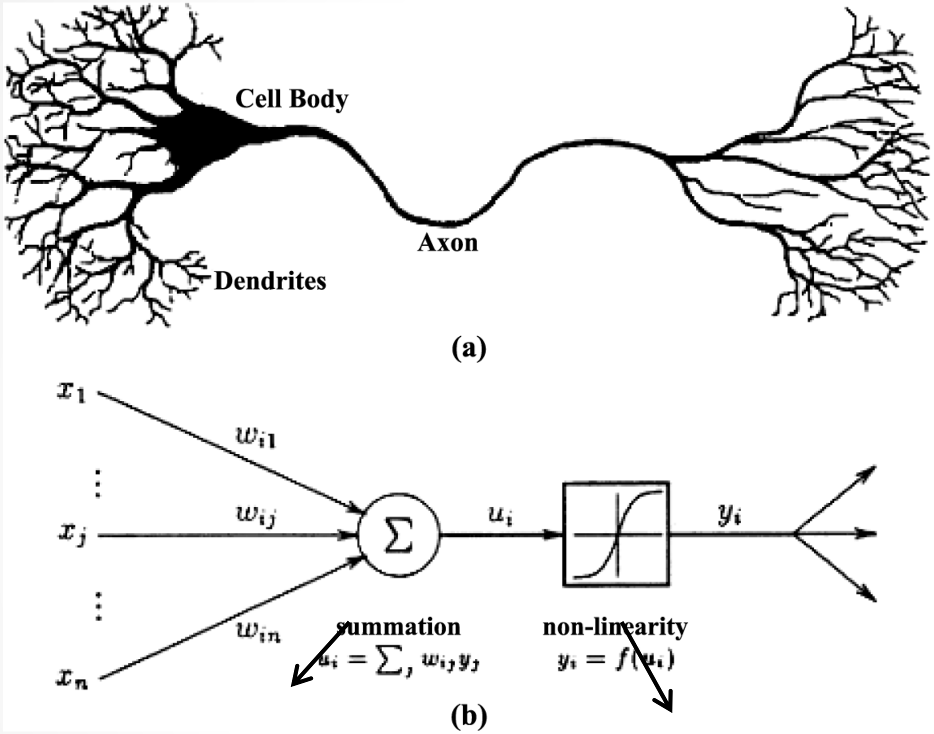
DATA ASSIILATION METHODS

global analysis



$$x^a = f_{NN}[y^o, x^f]$$

Artificial Neural Network (ANN)



Biological neuron

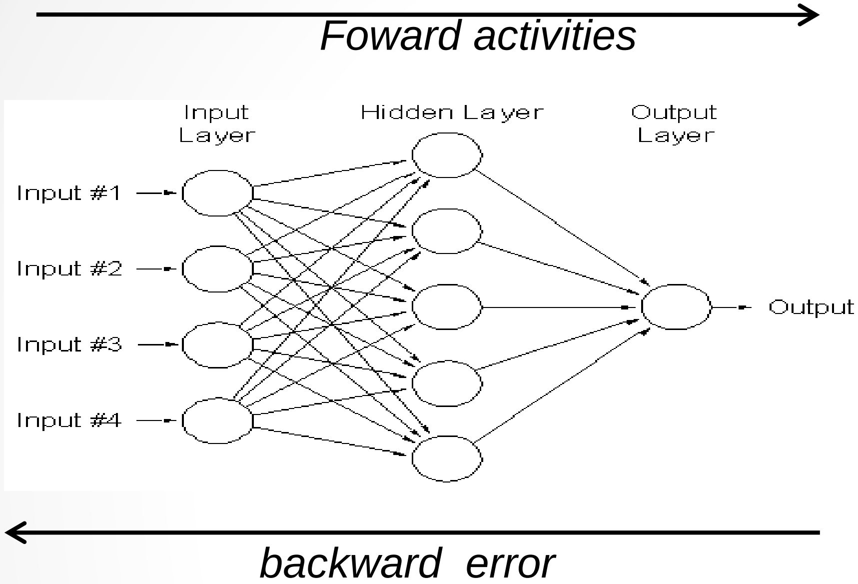
Artificial neuron

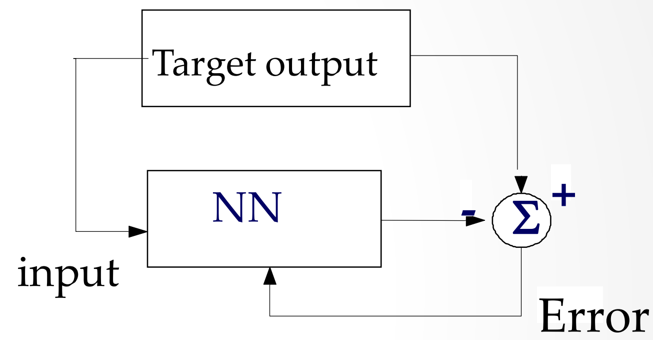
$$u_i = f\left(\sum w_{ij} x_j + b_j\right) \qquad y_i = \varphi\left(u_i\right)$$

Some artificial neurons interconnected based on a topology, make an Artificial Neural Network

Topology of NN

Multilayer Perceptron (MLP)





Supervised training

Backpropagation is algorithm used for NN training, e.g. the adjustments to the weights are conducted by back propagating of the error.

Training: an iterative process for adjusting the weights establishing the mapping of input and target vector pairs (supervisor)

Activation: for which NN receives new inputs and calculates the corresponding outputs, once it is trained.

The ANN Data assimilation Research:

KF	KF	KF 3d-Var Particle Filter	LETKF	KF Representer	LETKF
MLP	MLP Recurrent NN	MLP	MLP	MLP	MLP
(Nowosad) 2001	(Harter) 2004	(Furtado) 2008	(Cintra) 2010	(Furtado) 2012	(Cintra et al.) 2014
Lorenz63	SW 1D Solar Dynamics	Lorenz63	SPEEDY AGCM	Wave 1D SW 2D	FSUGSM AGCM
Full domain	Grid point domain		Pseudo-obs Sub-domain		Pseudo-obs Sub-domain

The ANN Data assimilation Research:

Development with the supervised ANN for Data Assimilation:

Important: full domain to grid point: implying into a dramatic reduction of algorithm complexity (Harter, 2004)

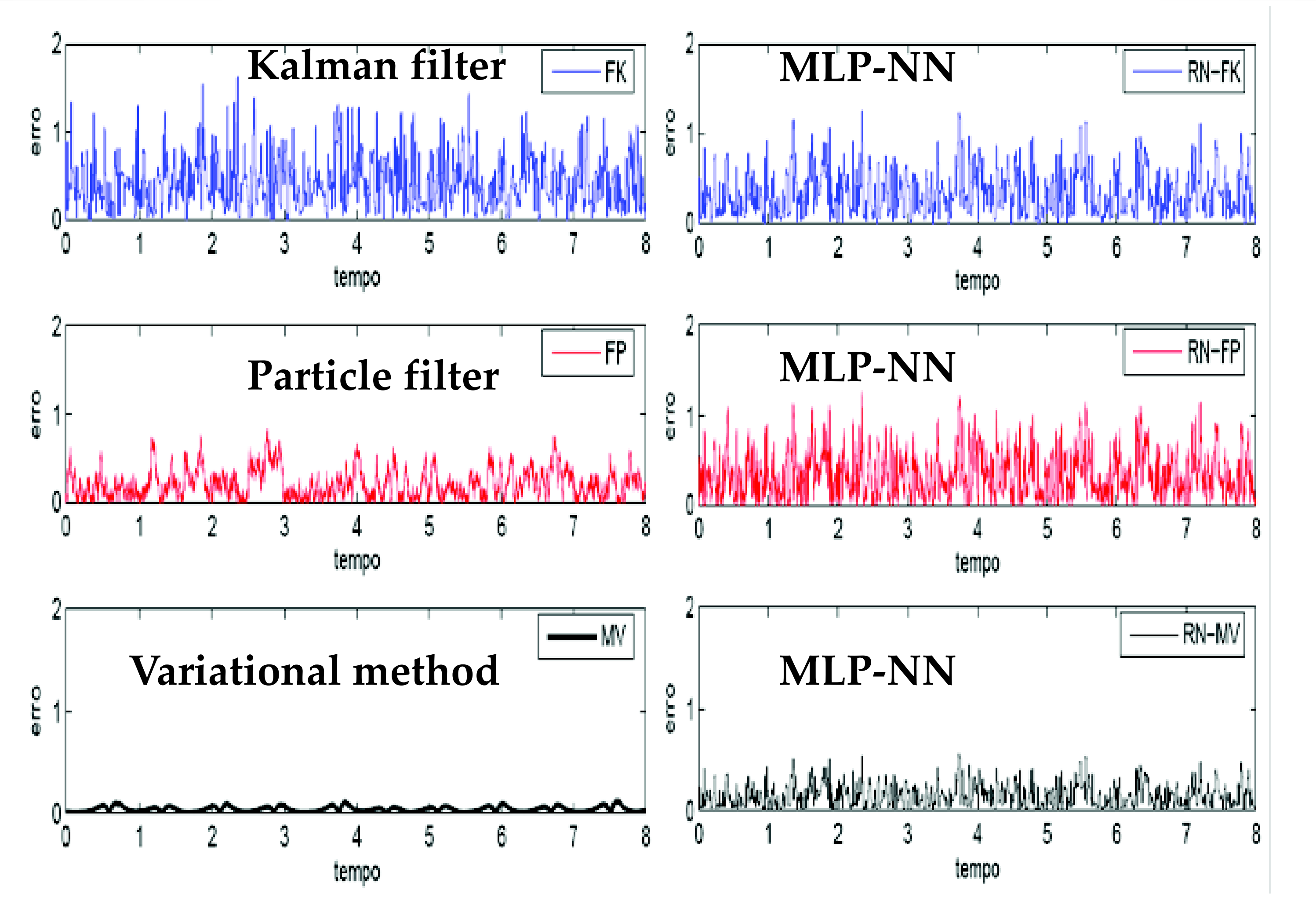
Pseudo-observation concept vs. correlation matrix (Cintra, 2010)
Introduce the "influence of observation" = A radius of influence around a grid point (without obs) is considered: weighted average projected on the grid point

Sub-domains (regions for global) = different ANN

Configure supervised ANN automatically with a optimization tool (Anochi et al, 2015)

The ANN Data assimilation Research:

Error evolution (Lorenz system under chaotic regime):

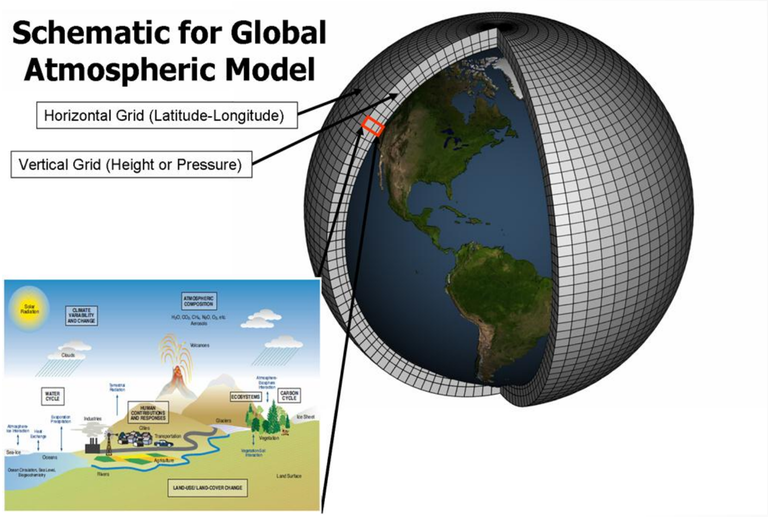


FSUGSM

Forward model (x^f):

SPEED model

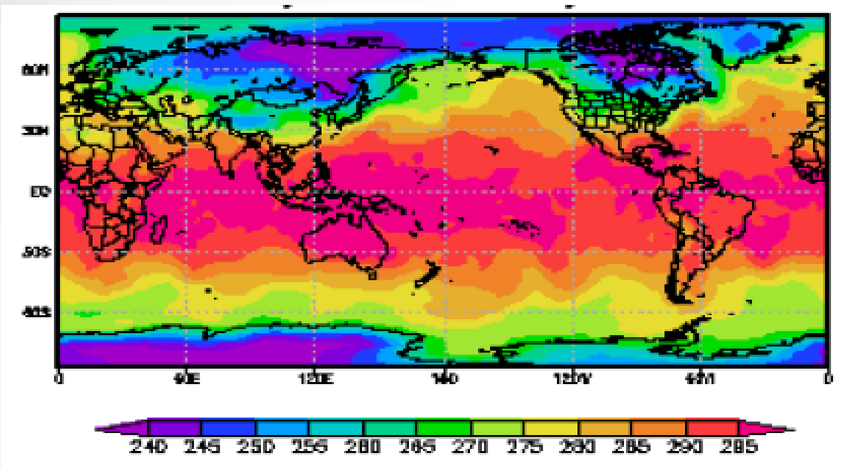
Atmospheric general
circulation model
3D spectral model T30L7
simplified parameterization
LETKF DA (UMD)

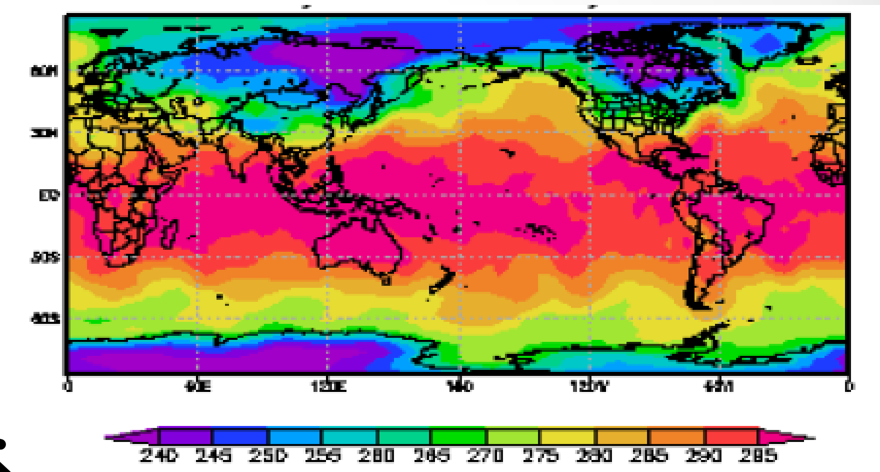


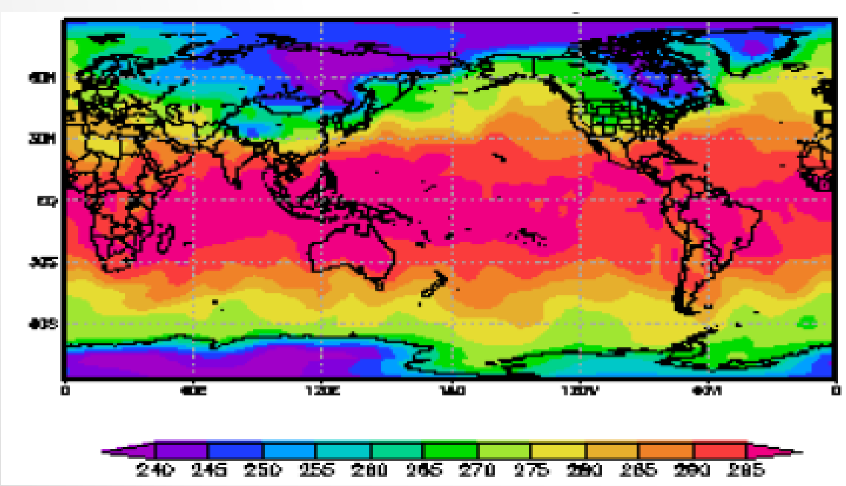
Vertical coordinates: $\sigma = p_s/p$.
Horizontal coordinates: (lat , lon) on a regular grid

The spectral model: T30 horizontal resolution and 7 vertical levels
Observations: 12035 (00 and 12 UTC) = 415 x 4 x 7 + 415
Observations: 2075 (06 and 18 UTC) = 415 x 5 (only surface)

FSUGSM



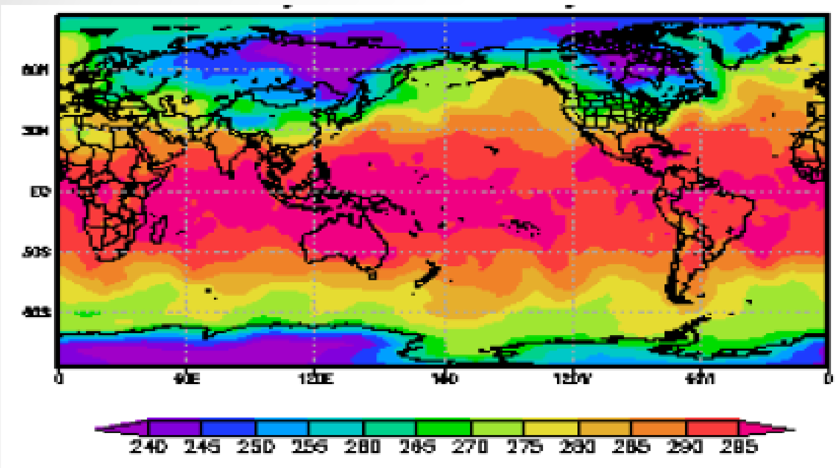


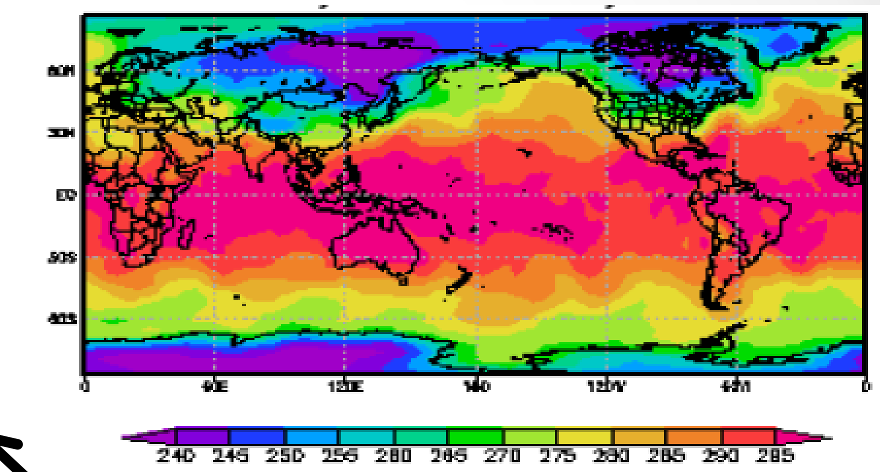


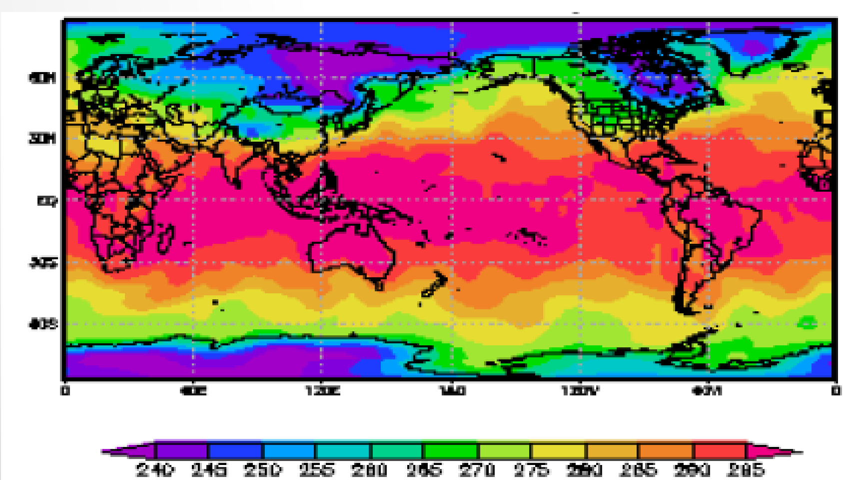
LETKF

MLP-NN
(5x6 = 30 NNs)

True







LETKF

MLP-NN

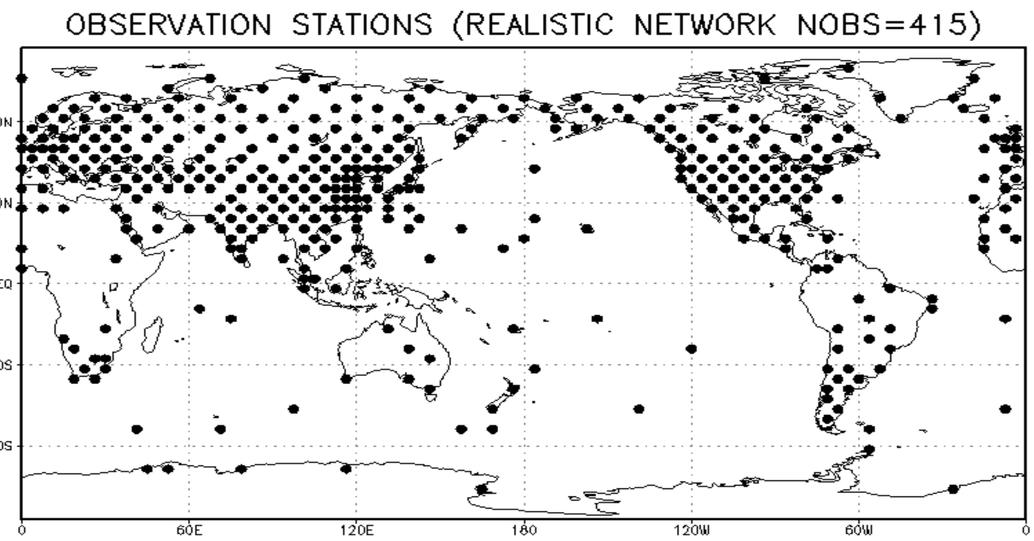
(5x6 = 30 NN)

(many NNs to be desinged)

True

FSUGSM

Forward model (x^f): **SPEED** model



Execution time	
LETKF	**MLP-NN**
04:20:39	00:02:53
hours : minutes : seconds	

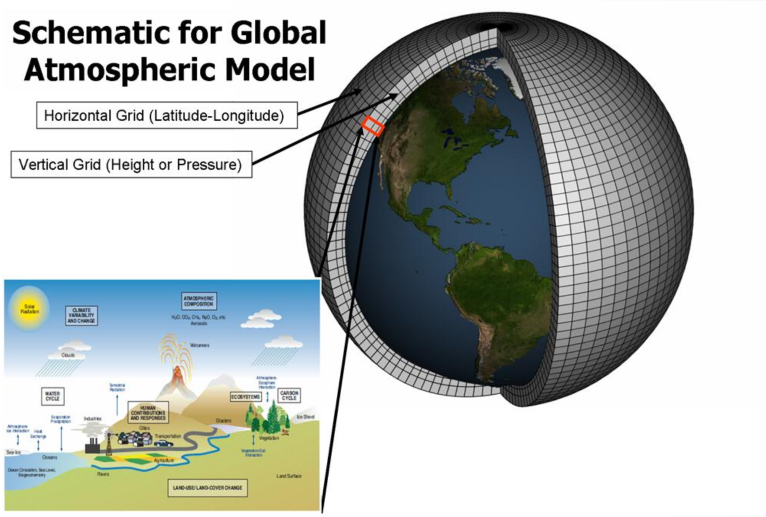
FSUGSM

Forward model (x^f):

FSUGSM

(Florida State University global spectral model) Atmospheric general circulation model

LETKF DA(UMD)



Vertical coordinates: $\sigma = p_s/p$.
Horizontal coordinates: (lat , lon) on a Gaussian grid

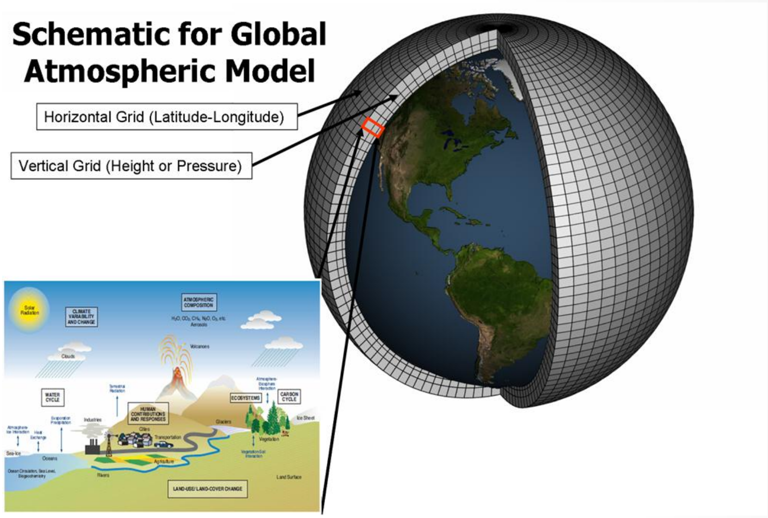
The spectral model: T63 horizontal resolution (approximately 1.875°) and 27 unevenly spaced vertical levels.
(~ regular grid 96 x 192 x 27)

FSUGSM

Forward model (x^f):

FSUGSM
(Florida State University global spectral model – T63L27)
Complete physical Parametrizations

~ regular grid 96 x 192 x 27



Globe divided into 4 regions (horizontal sub-domains) with 9 sets of three layers for upper-air variable (vertical sub-domains). One NN for each meteorological variable: 4 upper-air (T, u, v, q) and surface pressure (ps) at each sub-domain

neural networks: 4 x 9 x 4 + 4 = 148 networks.

How can we do that?

NN topology

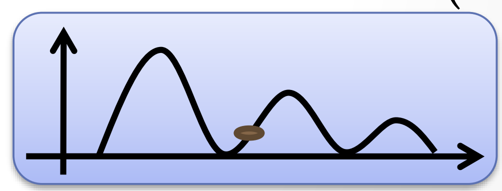
How does someone configure an artificial neural network?

1. Empirically - dependent of time to test

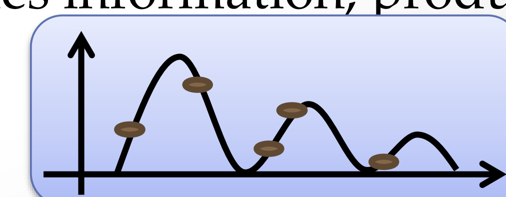
1. Automatic – dependent of one tool

Automatic configuration formulates as an optimization problem, solved by the **MPCA (**Multi-Particle Collision Algorithm) based

PCA (Particle Collision Algorithm): meta-heuristic mimic a neutron (or "particle") traveling inside of nuclear reactor



MPCA algorithm uses *n* particles sharing particles information, producing better solution than standard PCA.



MPCA Self-configuration tool

The objective function to be minimized has two terms:

$$F_{obj} = \text{penality} * \frac{(r_1 * E_{train} + r_2 * E_{activ})}{r_1 + r_2}$$

$$\text{penality} = \underbrace{\left(c_1 * \left(e^{\#\text{neuron}}\right)^2\right)}_{\text{complexity factor-1}} \times \underbrace{\left(c_2 * (\#\text{epoch})\right)}_{\text{complexity factor-2}} + 1$$

a) Penalty: measuring neural network complexity

b) The square error (differences between NN output and the target data) for learning process (E_{train}) and square error for generalization (E_{activ})

MPCA-NN for MLP

Parameters to be evaluated for NN configuration:

Parameters	Value
# hidden layer	|1| |2| |3|
# neurons	|1| ... |32|
Learning rate	|0| ... |1|
Momentum	|0| ... |0.9|
Activation function	|hiperbolic tangente| |Logistic| |Gaussian|

MPCA-NN algorithm uses a training data set for a determined problem.

For data assimilation, the MPCA topology is done by trained some multilayer perceptrons with input data (observation and forecast) and target (x^{tg}) data (analysis to mimic). $x^a = f_{NN}[y^o, x^f, (x^{tg})]$

MLP for Moisture Data Assimilation (MLP-DA)

The MPCA executes 25 realizations (stochastic alorithm) to find the best fitness to 148 Multilayer Perceptrons:

The MPCA topology to each MLP:

- Four input nodes,
    - the synthetic observation vector (y^o)
    - the 6-hours forecast model vector (x^f),
    - the grid point horizontal coordinate (*i*)
    - the grid point vertical coordinate (*j*)

  One target node for the analysis vector results (x^a) (only for training)

- One hidden layer
- The hyperbolic tangent as the activation function:

.

$$\varphi(u_i) = \tanh(u_i) = \frac{1-e^{u_i}}{1-e^{u_i}}$$

MLP for Moisture Data Assimilation (MLP-DA)

- Each MLP has its own learning rate, momentum rate, and the number of neuron in the hidden layer configured by MPCA:

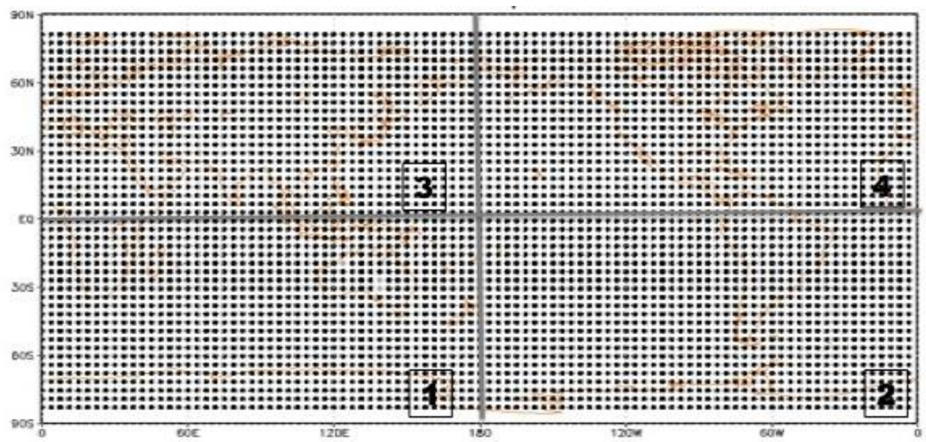
NETWORK (var/reg/layer)	NEURONS (Hidden Layer)	LEARNING RATE η	MOMENTUM RATE α
qq0101	09	0.424676	0.735560
qq0102	07	0.695070	0.836189
qq0103	09	0.128201	0.987913
qq0201	09	0.091828	0.621134
qq0202	10	0.247087	0.997031
qq0203	09	0.068212	0.994036
qq0301	05	0.601685	0.447649
qq0302	06	0.543795	0.980525
qq0303	10	0.852829	0.909061
qq0401	06	0.517619	0.949778
qq0402	10	0.355837	0.975882
qq0403	10	0.438510	0.995963

- Training steps when the cost function is less than 10^{-5}.

Observation Simulation System Experiment (y^o)

Synthetic observation is obtained by adding Gaussian random noise at the grid point values on the true fields.

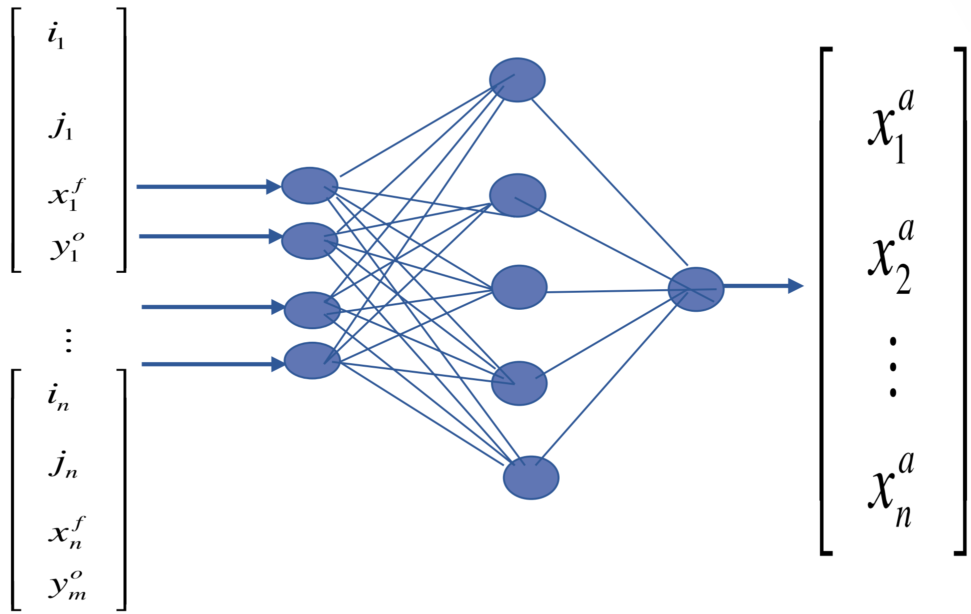
The observational grid is a regularly distributed in a dense network:
This grid localization is every other latitude/longitude grid point of the FSUGSM native grid.



Observations localizations of observations divides in four regions of global area each is (90º x 180º) size. The dot points represent stations.

MLP for Moisture Data Assimilation (MLP-DA)

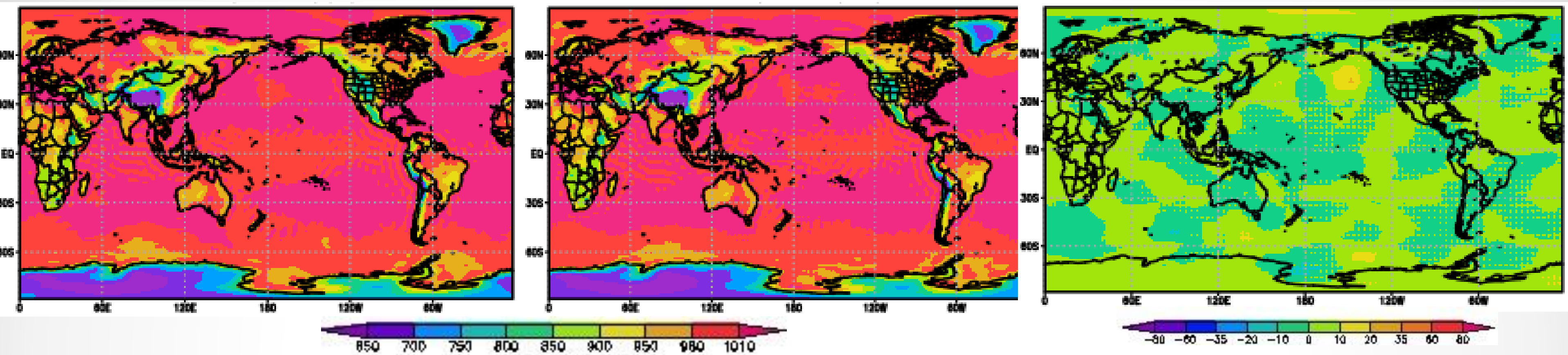
Training data sets collected observations, forecast and target analysis: Every day from Jan/2001 up to Jan/2003: 00, 06, 12, and 18 UTC, month to month.



Activation test data set collected (i, j, y^o, x^f): every day of Jan/2004 and **Generalization** phase: every day of Jan/2005: 00, 06, 12, and 18 UTC. (Generalization is the "operational" data assimilation phase)

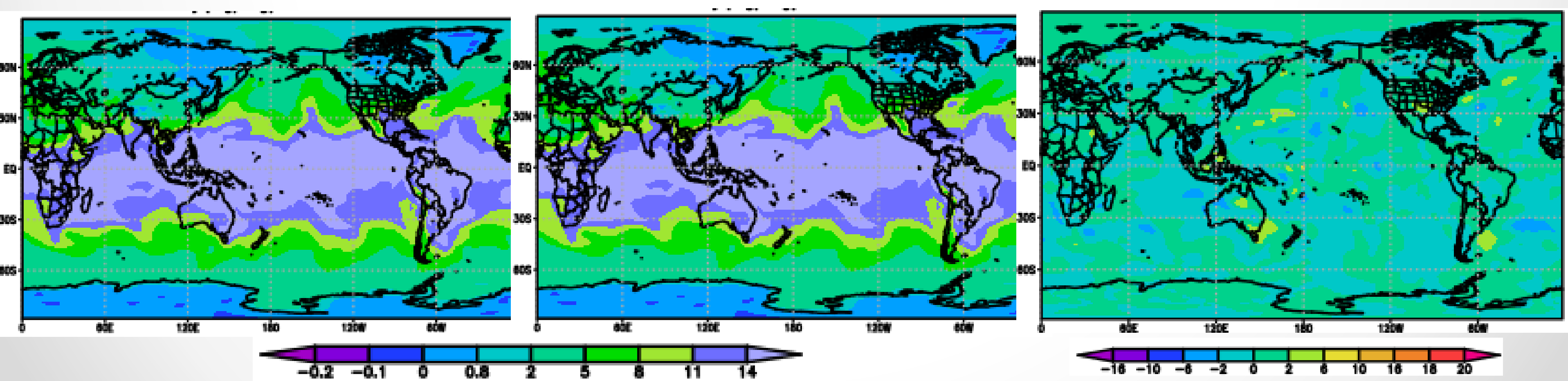
The 148 MLPs generate analysis for all prognostic variables.

Surface Pressure(Kg/Kg) generalization 04/Jan/2005 – 12 UTC



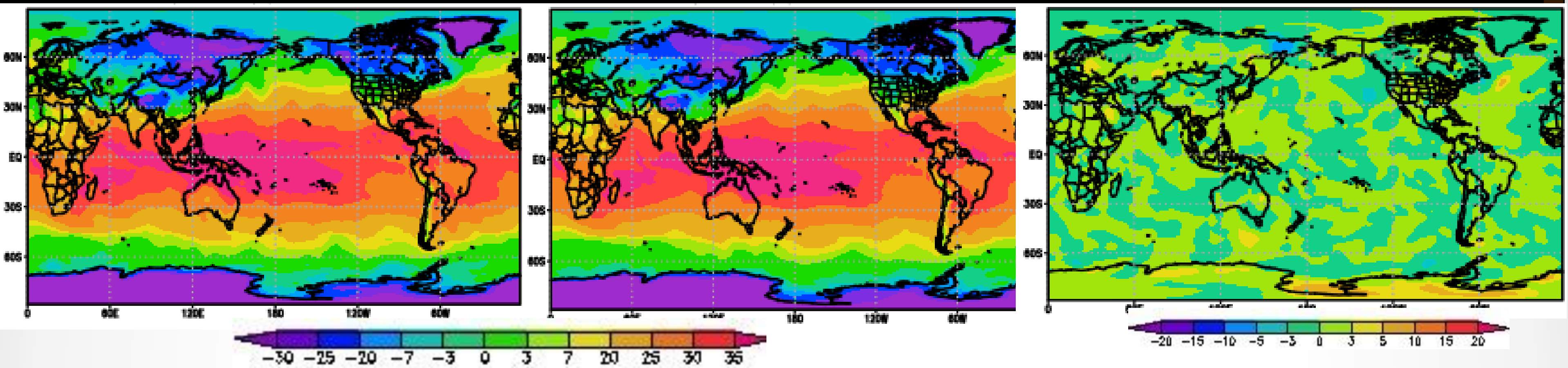
LETKF analysis MLP analysis Differences analysis

Specific Humidity (Kg/Kg) generalization 04/Jan/2005 – 12 UTC



LETKF analysis MLP analysis Differences analysis

Temperature(C°)generalization 04/Jan/2005 – 12 UTC Surface

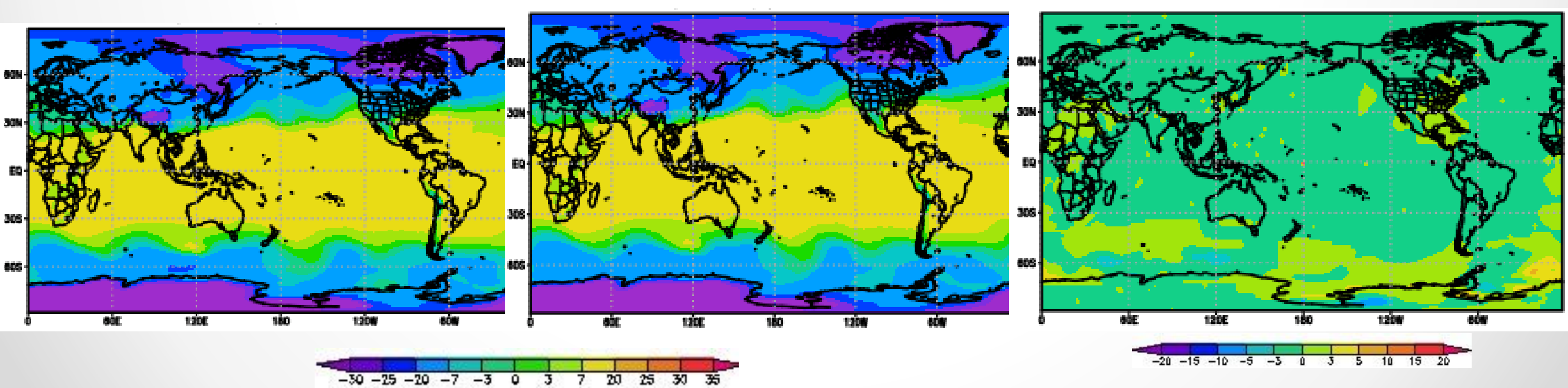


LETKF analysis

MLP analysis

Differences analysis

Temperature(C°)generalization 04/Jan/2005 – 12 UTC 500 hPa

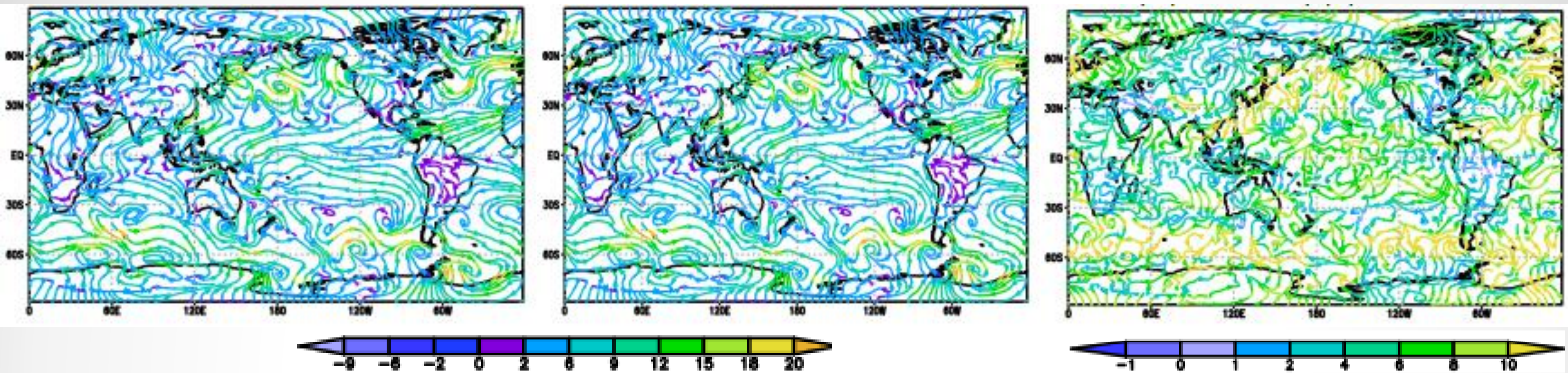


LETKF analysis

MLP analysis

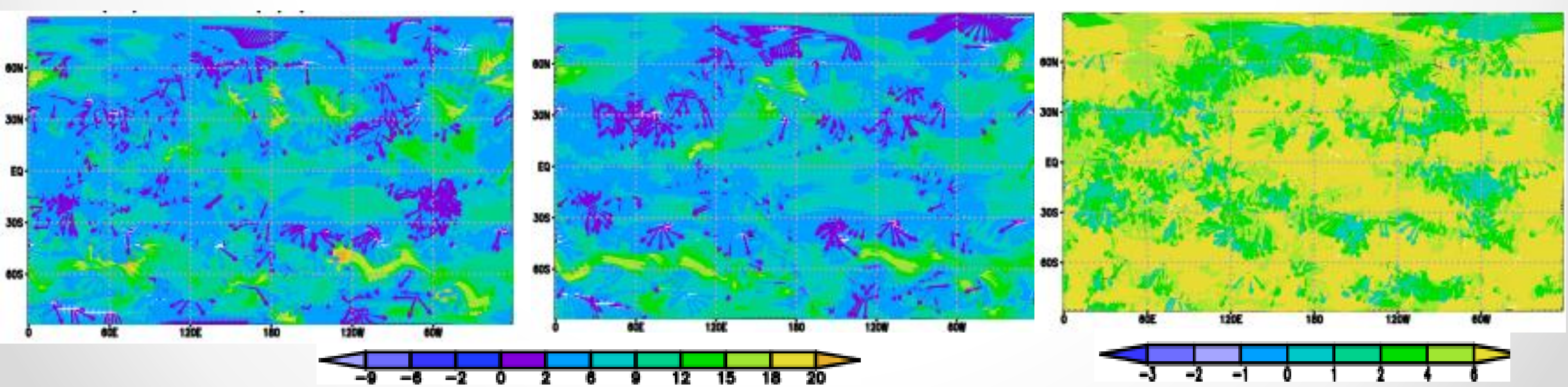
Differences analysis

Stream wind (zonal/meridional component) 08/Jan/2005 –06UTC



LETKF analysis MLP analysis Analysis difference

Vector wind (zonal/meridional component) 08/Jan/2005 –06UTC



LETKF analysis MLP analysis Analysis difference

COMPUTATIONAL PERFORMANCE

Execution of 124 cycles	MLP-DA (hour:min:sec)	LETKF (hour:min:sec)
Analysis time	00:02:29	11:01:20
Ensemble time	00:00:00	15:50:40
Parallel model time	00:27:20	00:00:00
Total Time	00:29:49	26:52:00

266 times faster

55 times faster

The LETKF analysis runs on 40 nodes at Cray XT/16 (1280 nodes, each node with 2 Opteron 12 cores, total of 30720 cores) (http://www.cptec.inpe.br/supercomputador)).

The MLPs runs with a sequential program.

MLP-DA computed analyses for the FSUGSM model:

- Analyses with similar LETKF quality
- Analysis with better computer performance.

The ANN Data assimilation Research:

New developments and challenges

1. Emulating NCEP analysis for the Global model from the CPTEC-INPE (under development)
2. Hybrid computing (CPU+FPGA or CPU+GPU or CPU+MIC) (see poster S1-51)
3. Models with adaptive mesh refining

Thank you for your attention!

Rosangela S. Cintra - INPE, S. José dos Campos, SP - rosangela.cintra@inpe.br

Haroldo F. de Campos Velho - INPE, S. José dos Campos, SP – haroldo.camposvelho@inpe.br

Steven Cocke - Florida State University, Tallahassee, FL, USA - scocke@fsu.edu

LAC Laboratório Associado de Computação e Matemática Aplicada

INPE

COAPS Center for Ocean-Atmospheric Prediction Studies

FLORIDA STATE UNIVERSITY 1851