The Abdus Salam
International Centre for Theoretical Physics

Spring Colloquium on
'Regional Weather Predictability and Modeling'
April 11 - 22, 2005

1) Workshop **on Design and Use of Regional Weather Prediction Models,** *April 11 - 19*

2) Conference on **Current Efforts Toward Advancing the Skill of Regional Weather Prediction. Challenges and Outlook,** *April 20 - 22*

301/1652-5

Information content in ensemble data assimilation

D. Zupanski
Cooperative Institute for Research in the Atmosphere, CSU, Ft. Collins
USA

Information content in ensemble data assimilation

Dusanka Zupanski
CIRA/Colorado State University, Fort Collins, CO

Conference on
Current Efforts Toward Advancing the Skill of Regional Weather Prediction Challenges and Outlook
April 11-22, 2005
ICTP,Trieste, Italy

Collaborators
- M. Zupanski, L. Grasso, Scott Denning Group (Colorado State University)
- A. Y. Hou, S. Zhang (NASA/GMAO)
- M. DeMaria (NOAA/NESDIS)

Research partially supported by NOAA Grant NA17RJ1228 and NASA Grants 621-15-45-78, NAG5-12105, and NNG04GI25G

OUTLINE

- Challenges of advancing skill of regional weather prediction
- Link between information theory and ensemble data assimilation
- Experimental results
- Conclusions and future work

Dusanka Zupanski, CIRA/CSU
Zupanski@CIRA.colostate.edu

Challenges of advancing skill of regional weather prediction

- Employ state-of-the-art non-linear atmospheric models (without neglecting model errors)
- Assimilate observations with high special and temporal resolution
- Calculate analysis and forecast uncertainty

Ensemble data assimilation

- Determine amount of new information given by the observations
- Quantify predictability (calculate entropy reduction)

Information theory

Can we define a link between ensemble data assimilation and information theory?

Dusanka Zupanski, CIRA/CSU
Zupanski@CIRA.colostate.edu

METHODOLOGY

Maximum Likelihood Ensemble Filter (MLEF)
(Zupanski 2005; Zupanski and Zupanski 2005)

Developed using ideas from

- Variational data assimilation (3DVAR, 4DVAR)
- Iterated Kalman Filters
- Ensemble Transform Kalman Filter (ETKF, Bishop et al. 2001)

MLEF is designed to provide optimal estimates of

- model state variables
- empirical parameters
- model error (bias)

MLEF also calculates uncertainties of all estimates (in terms of P_a and P_f)

Dusanka Zupanski, CIRA/CSU
Zupanski@CIRA.colostate.edu

MLEF APPROACH

Minimize cost function J

$$J = \frac{1}{2}[\boldsymbol{x} - \boldsymbol{x}_b]^T \boldsymbol{P}_f^{-1}[\boldsymbol{x} - \boldsymbol{x}_b] + \frac{1}{2}[H(\boldsymbol{x}) - \boldsymbol{y}_{obs}]^T \boldsymbol{R}^{-1}[H(\boldsymbol{x}) - \boldsymbol{y}_{obs}] = \min$$

Analysis error covariance

$$P_a^{1/2} = P_f^{1/2}(I + C)^{-1/2}$$

Link between information theory and ensemble data assimilation

$$C = P_f^{T/2} H^T R^{-1} H P_f^{1/2} = (R^{-1/2} H P_f^{1/2})^T (R^{-1/2} H P_f^{1/2})$$

Forecast error covariance

$$P_f^{1/2} = [p_1^f \quad p_2^f \quad \ldots \quad p_{Nens}^f]$$

$$p_i^f = M(x + p_i^a) - M(x)$$

x - model state vector of dim $Nstate >> Nens$

M - non-linear forecast model

C - information matrix of dim $Nens \times Nens$

Dusanka Zupanski, CIRA/CSU
Zupanski@CIRA.colostate.edu

EXPERIMENTAL DESIGN

- NASA GEOS-5 column precipitation model
- Tropical Western Pacific site (130E,15N)
- 50 6-h DA cycles: 00UTC 7 May 1998- 00 UTC 17 May 1998
- 40 vertical layers
- Control variable: T, q (dim=80)
- Model simulated observations with random noise (80 obs per DA cycle)
- Nens=40
- Iterative minimization of J (1 iteration only)

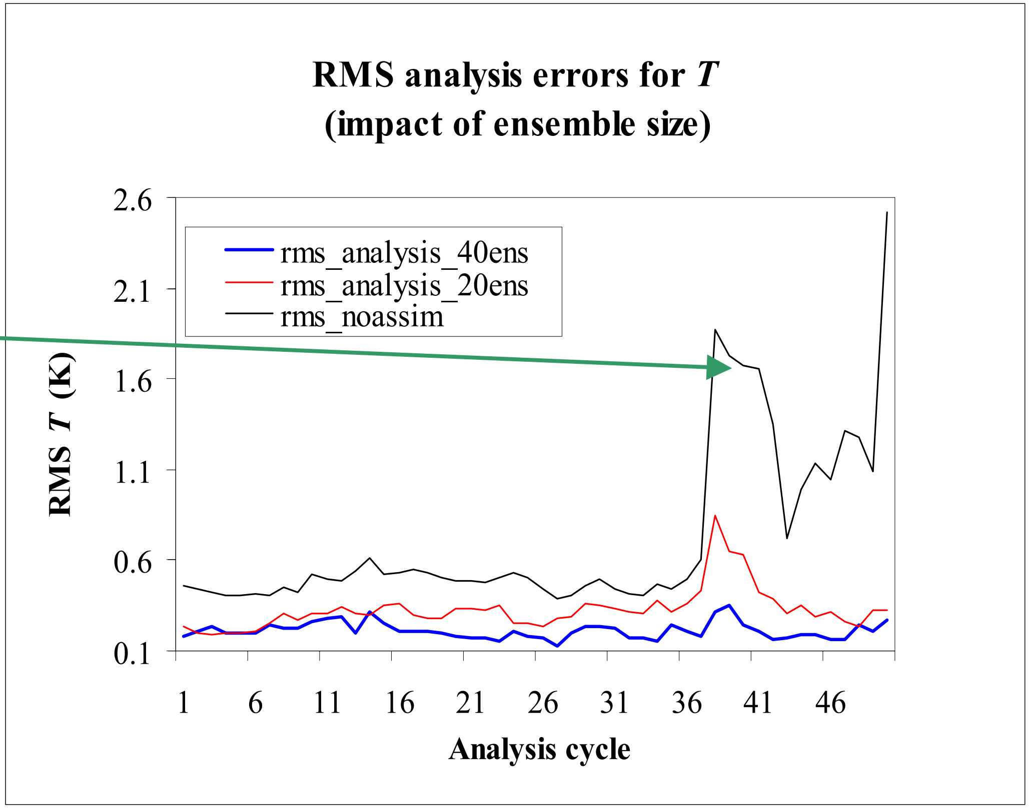
Dusanka Zupanski, CIRA/CSU
Zupanski@CIRA.colostate.edu

RESULTS using GEOS-5 column model

RMS errors of temperature with respect to the "truth"

Largest errors in the experiment without assimilation in last 13 cycles.

Data assimilation reduces RMS errors, by using information from observations.



Dusanka Zupanski, CIRA/CSU
Zupanski@CIRA.colostate.edu

RESULTS using GEOS-5 column model

Information matrix

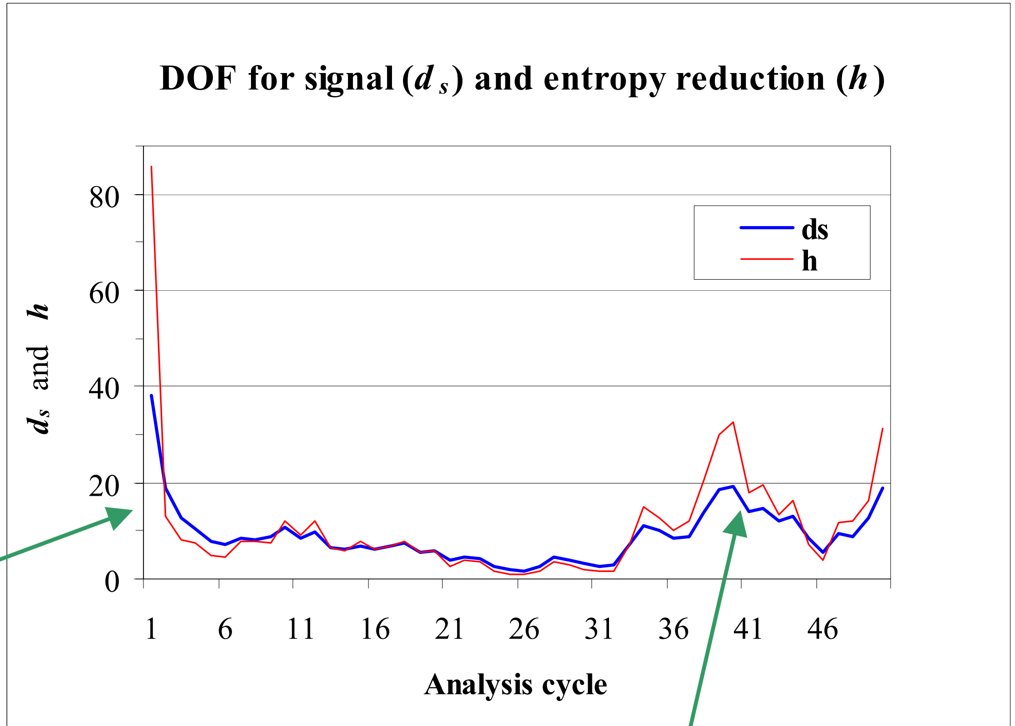
$$C = P_f^{T/2} H^T R^{-1} H P_f^{1/2} = (R^{-1/2} H P_f^{1/2})^T (R^{-1/2} H P_f^{1/2})$$

Degrees of freedom (DOF) for signal

$$d_s = tr[(\boldsymbol{I} + \boldsymbol{C})^{-1} \boldsymbol{C}] = \sum_i \frac{\lambda_i^2}{(1+\lambda_i^2)}$$

Shannon information content, or entropy reduction (used for quantifying predictability)

$$h = \frac{1}{2} \sum_i \ln(1+\lambda_i^2)$$



Inadequate P_f at the beginning of data assimilation

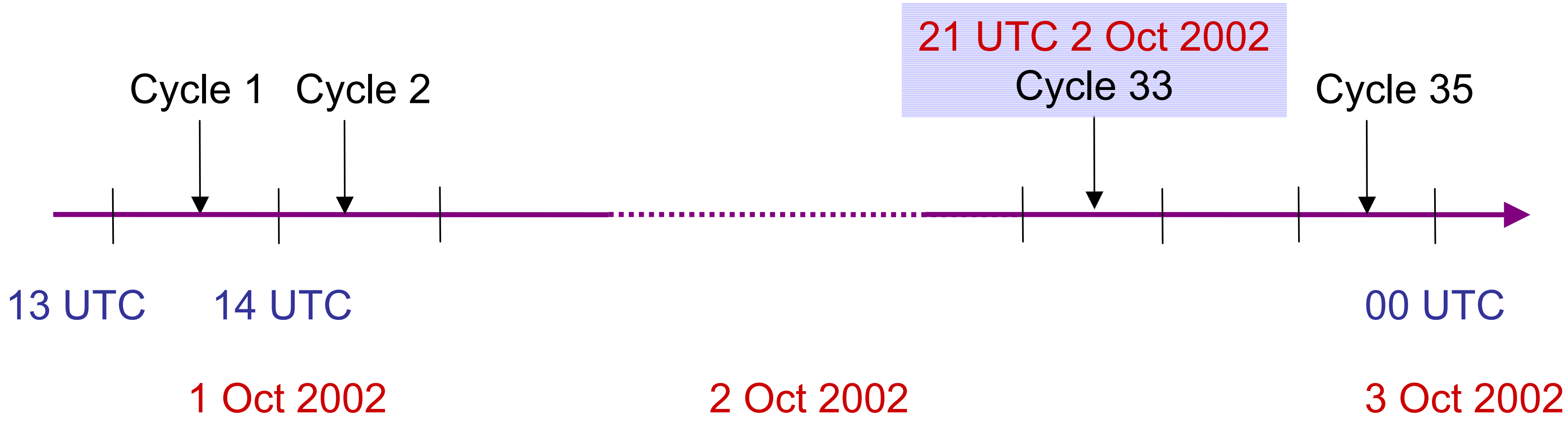
New observed information

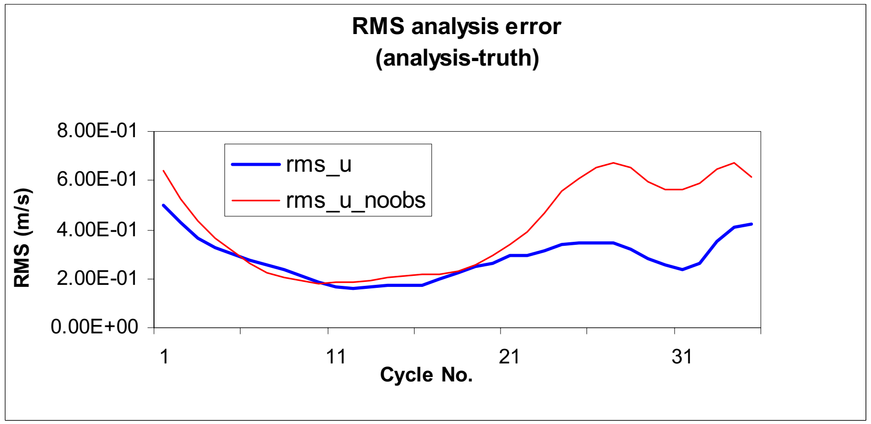
EXPERIMENTAL DESIGN

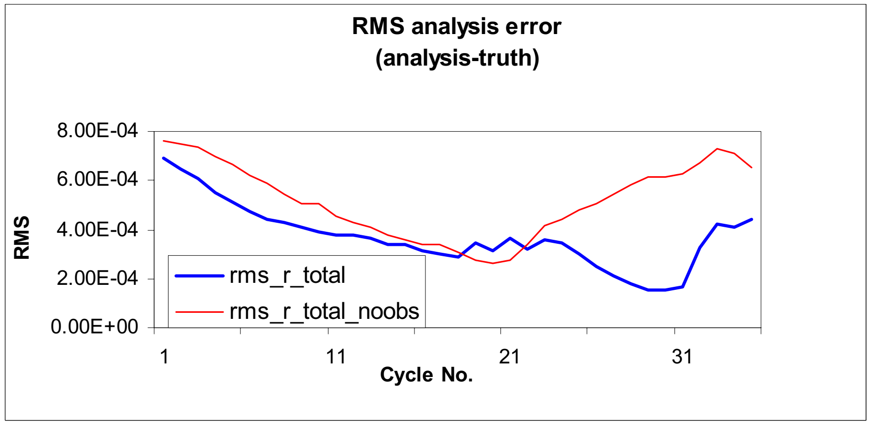
- **CSU-RAMS non-hydrostatic model**
- **Hurricane Lili case**
- **35 1-h DA cycles: 13UTC 1 Oct 2002 – 00 UTC 3 Oct**
- **30x20x21 grid points, 15 km grid distance (in the Gulf of Mexico)**
- **Control variable: u,v,w,theta,Exner, r_total (dim=54000)**
- **Model simulated observations with random noise (7200 obs per DA cycle)**
- **Nens=50**
- **Iterative minimization of J (1 iteration only)**

Dusanka Zupanski, CIRA/CSU
Zupanski@CIRA.colostate.edu

Experimental design (continued)





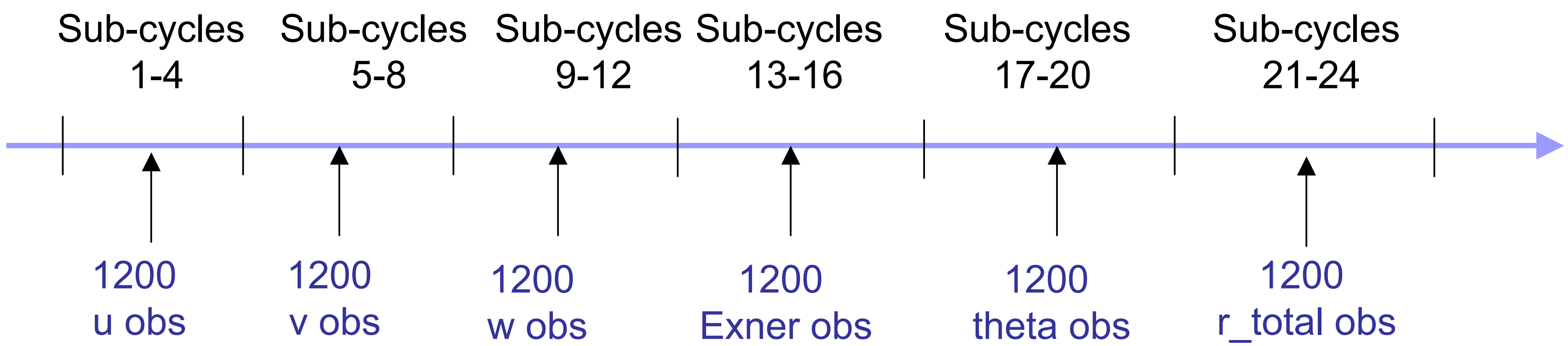


Dusanka Zupanski, CIRA/CSU
Zupanski@CIRA.colostate.edu

Experimental design (continued)

- Split cycle 33 into 24 sub-cycles
- Calculate eigenvalues of $(I+C)^{-1/2}$ in each sub-cycle (information content)

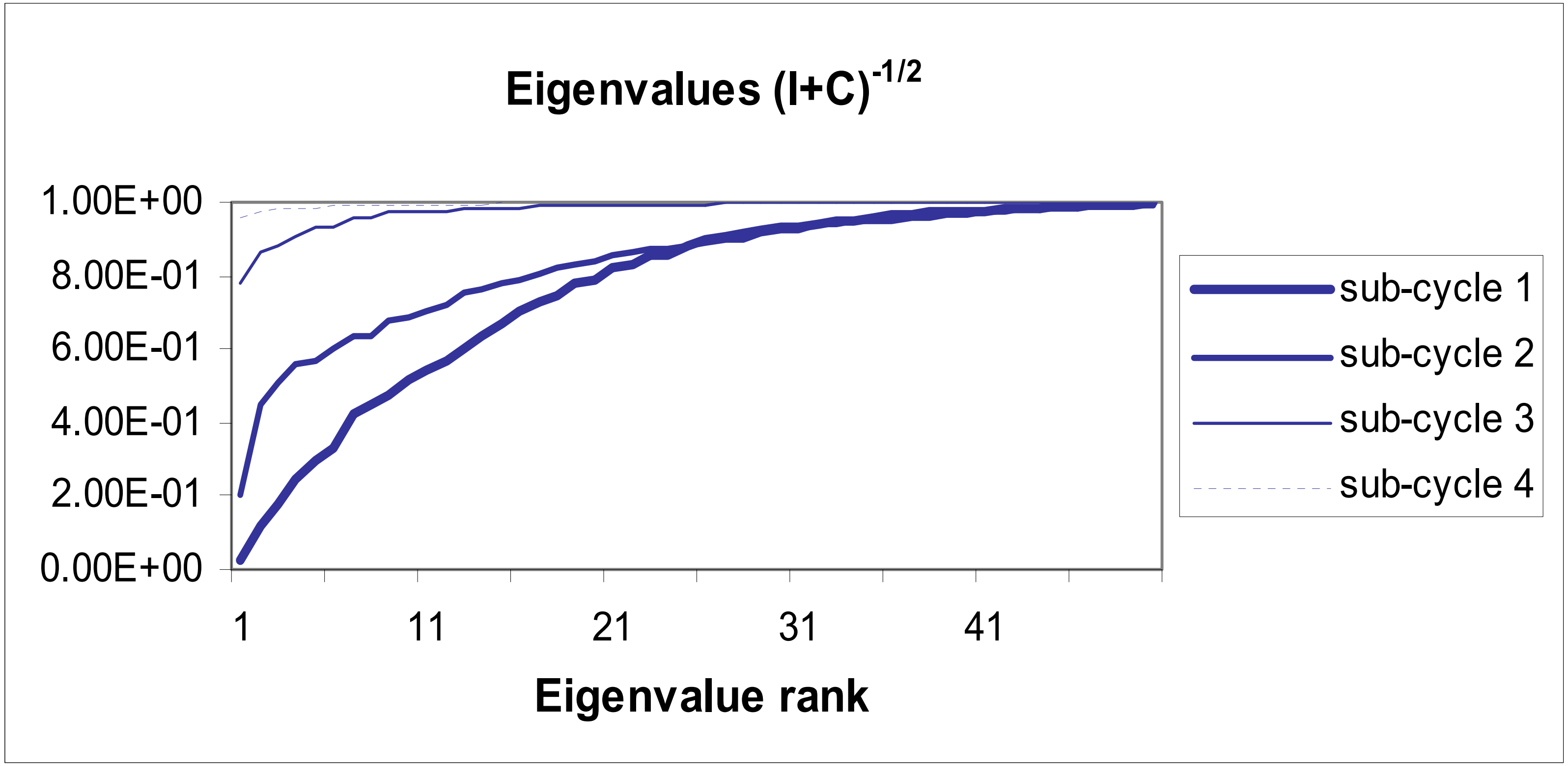
Information content of each group of observations



Dusanka Zupanski, CIRA/CSU
Zupanski@CIRA.colostate.edu

RESULTS using RAMS model

Sub-cycles 1-4
u- obs groups

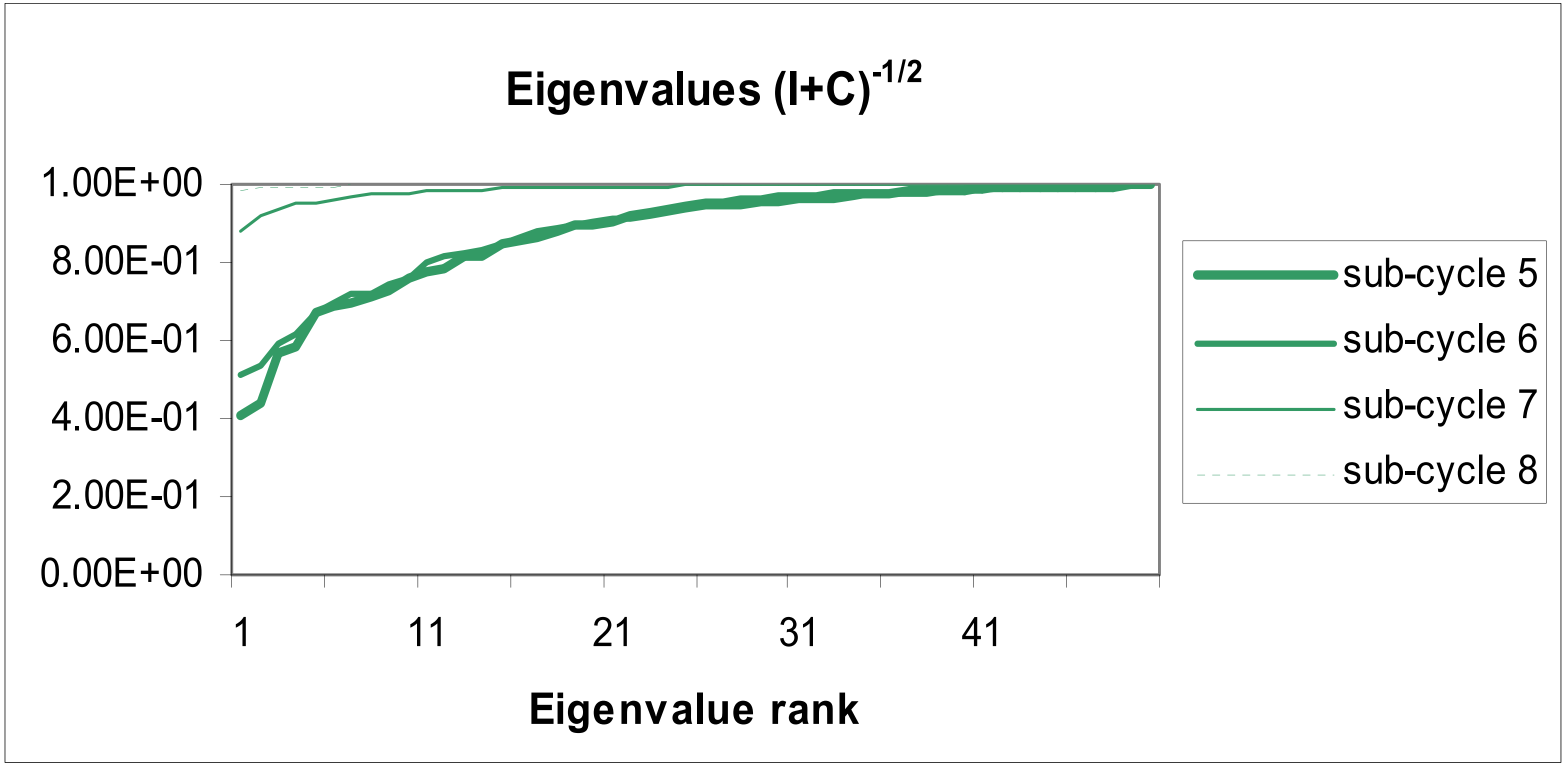


System is "learning" about the truth via updating analysis error covariance.

Dusanka Zupanski, CIRA/CSU
Zupanski@CIRA.colostate.edu

RESULTS using RAMS model

Sub-cycles 5-8
v- obs groups

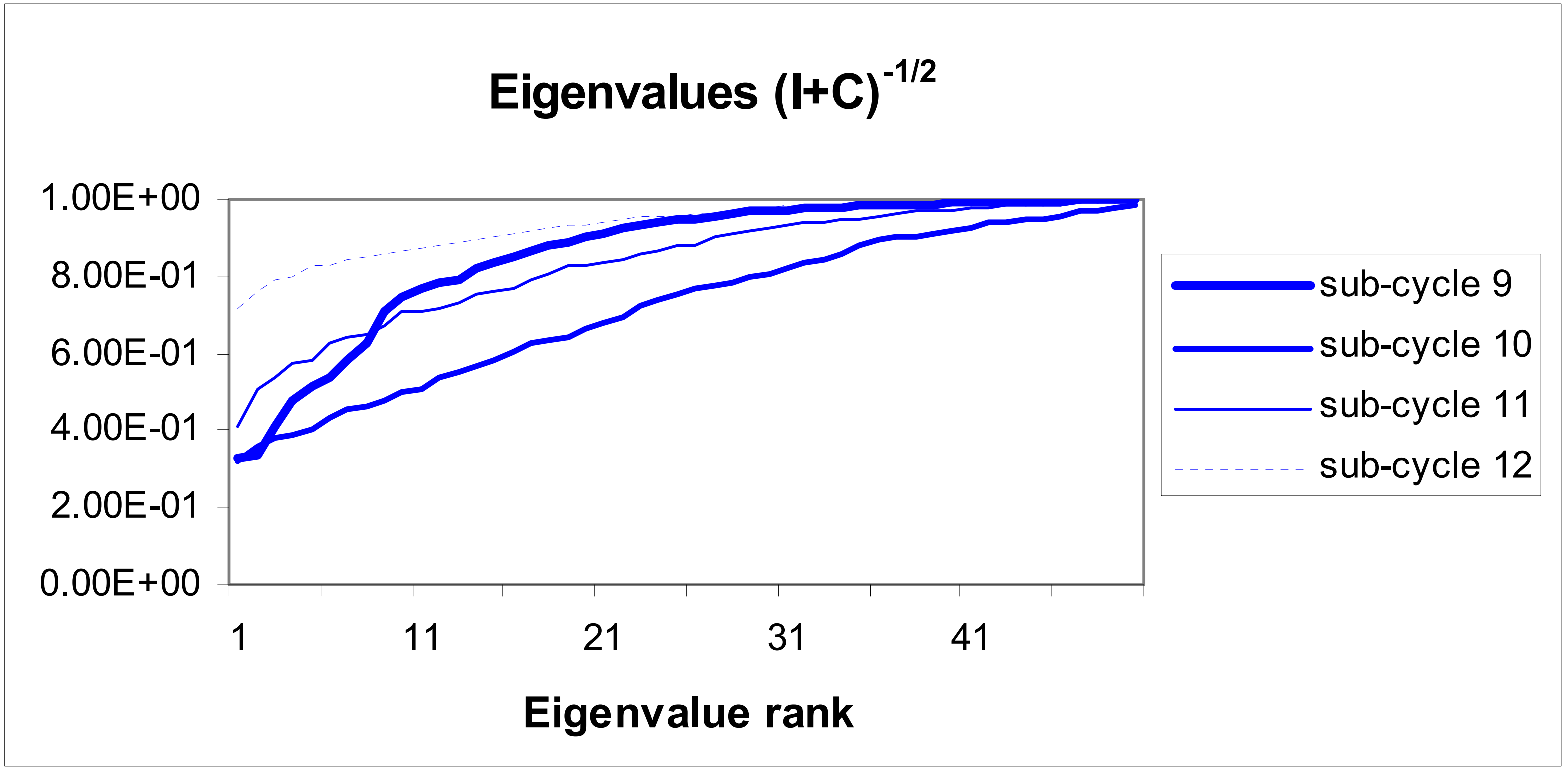


Most information in sub-cycles 5 and 6.

Dusanka Zupanski, CIRA/CSU
Zupanski@CIRA.colostate.edu

RESULTS using RAMS model

Sub-cycles 9-12
w- obs groups

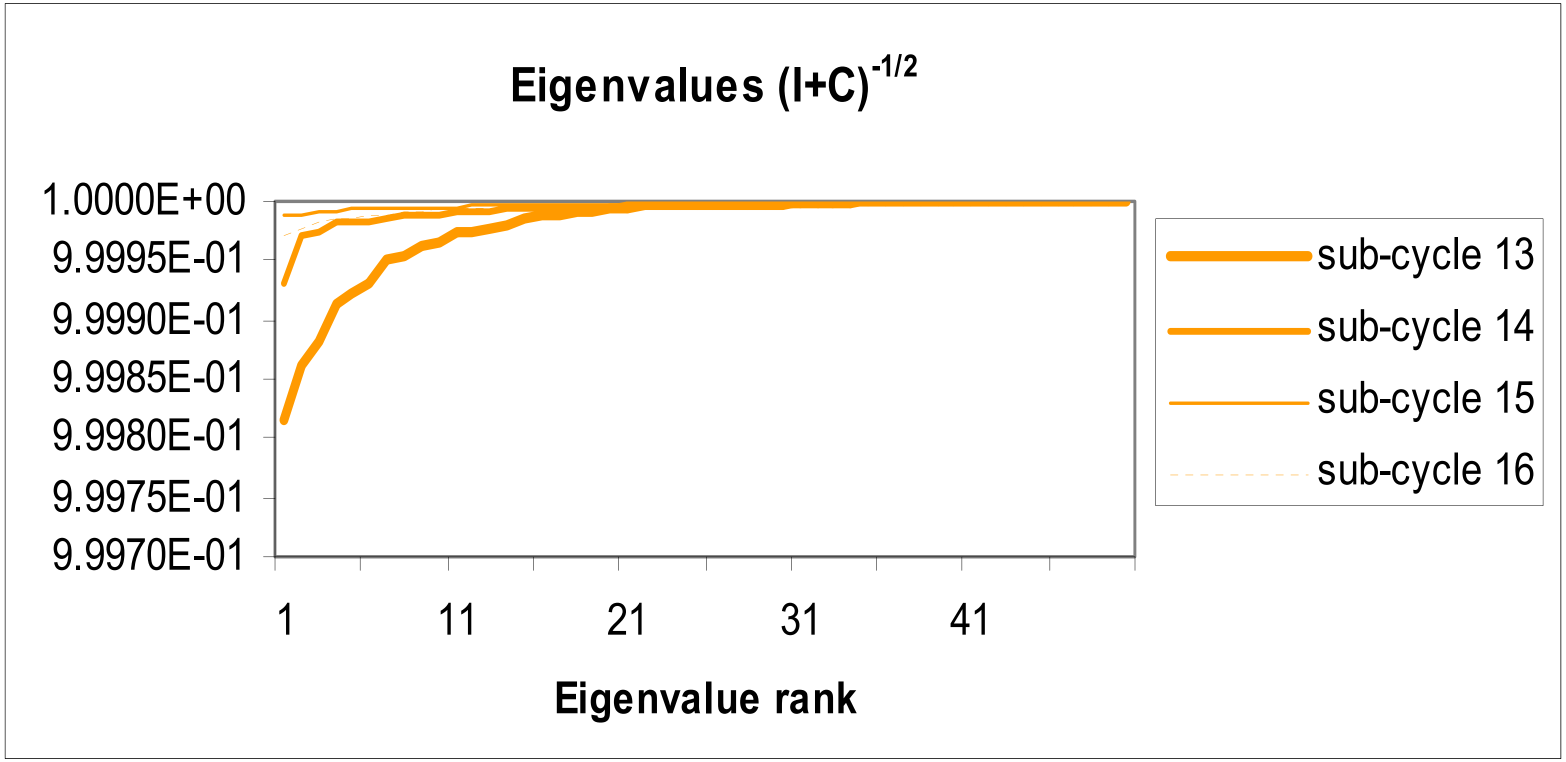


Most information in sub-cycle 10.

Dusanka Zupanski, CIRA/CSU
Zupanski@CIRA.colostate.edu

RESULTS using RAMS model

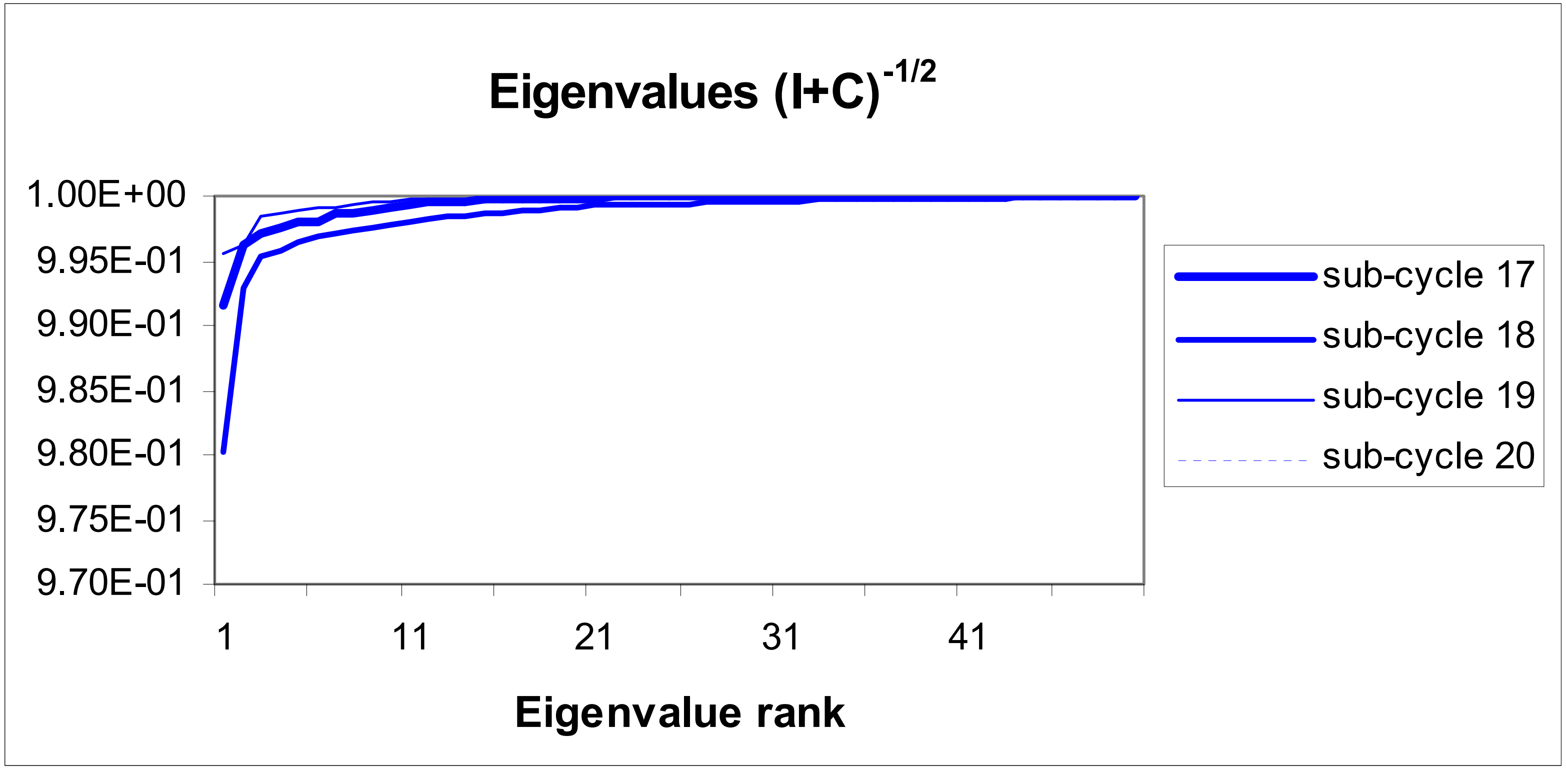
Sub-cycles 13-16
Exner- obs groups



Dusanka Zupanski, CIRA/CSU
Zupanski@CIRA.colostate.edu

RESULTS using RAMS model

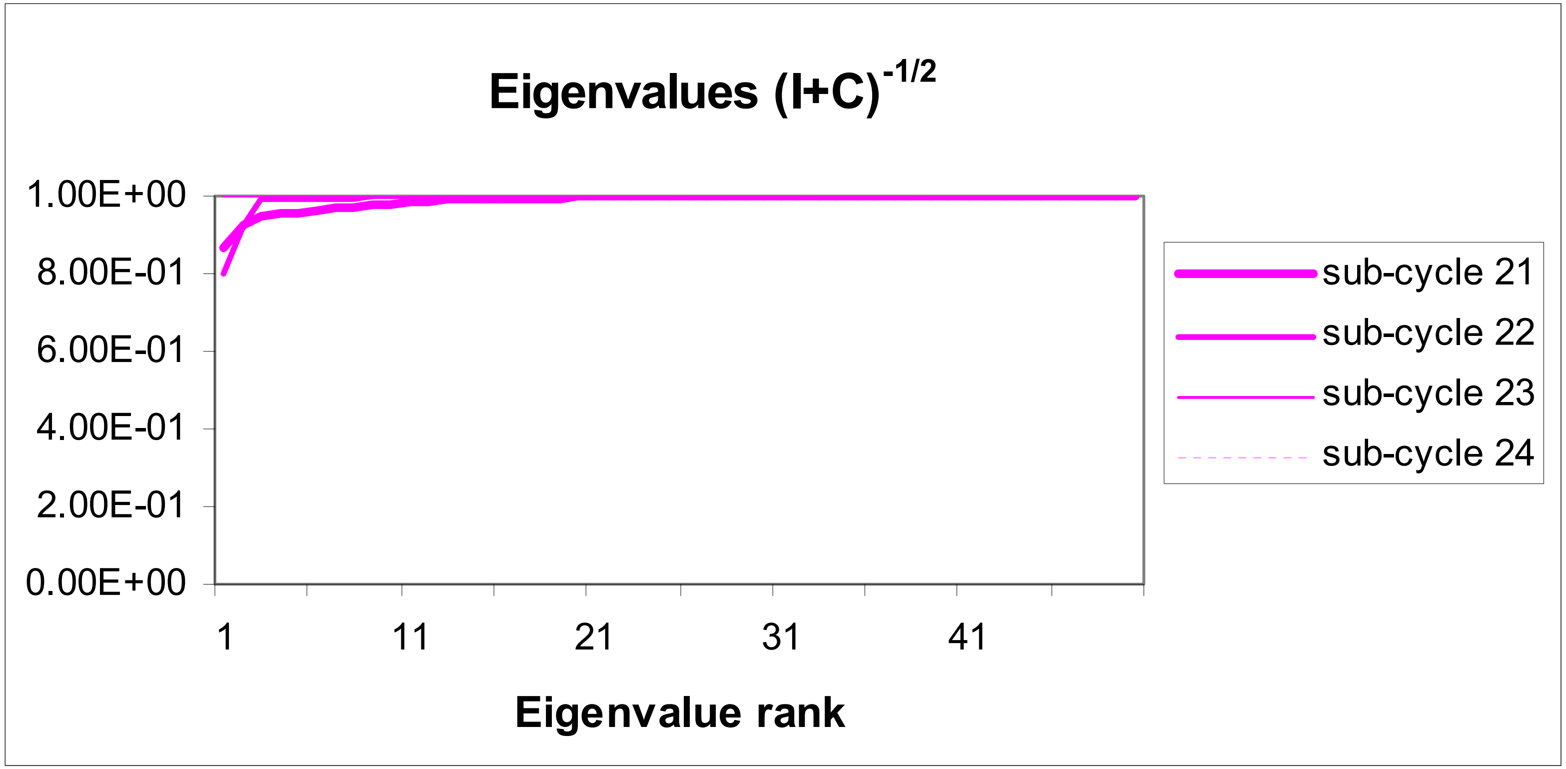
Sub-cycles 17-20
theta- obs groups



Dusanka Zupanski, CIRA/CSU
Zupanski@CIRA.colostate.edu

RESULTS using RAMS model

Sub-cycles 21-24
theta- obs groups



Sub-cycles with little information can be excluded ⇒ data selection.

Dusanka Zupanski, CIRA/CSU
Zupanski@CIRA.colostate.edu

CONCLUSIONS

- The proposed general framework can be effectively used to quantify information content of various observations.
- Measures of predictability, such as entropy reduction, can also be calculated.
- The framework is applicable to a forecast model of any complexity. Only eigenvalues of a small size matrix (*Nens* x *Nens*) need to be evaluated.
- Data assimilation system has a capability to learn form observations.

Ensemble data assimilation linked with information theory is a powerful approach for addressing challenging issues of regional weather prediction.

Dusanka Zupanski, CIRA/CSU
Zupanski@CIRA.colostate.edu