International Workshop on
Hydrometeorological and Hydrologic Ensemble Prediction
19-22 July 2005
Boulder, Colorado

Non-Gaussian error statistics and ensemble data assimilation

Milija Zupanski

Cooperative Institute for Research in the Atmosphere
Colorado State University
Fort Collins, CO 80523-1375
E-mail: ZupanskiM@CIRA.colostate.edu

In collaboration with:

Colorado State University	S. Fletcher, D. Zupanski, D. Randall, R. Heikes
Florida State University	I.M. Navon, B. Uzunoglu
Computational Support	NCAR SCD (bluesky)

Framework for Assimilation/Prediction

- **Uncertainties (errors) are considered random variables**
  - Observations, Parameters, Model bias, Boundary conditions
  - Input forcing from atmospheric or climate models
  (e.g., temperature, precipitation, top soil moisture)

- **Normal (Gaussian) PDF generally used**
  - Variational methods (3D-Var, 4D-Var)
  - Kalman filters (extended, reduced-rank)
  - Ensemble Kalman filters
  - Least-square methods
  - Well understood
  - Relatively simple to use
  - Satisfactory results

$$PDF(\boldsymbol{x} = \boldsymbol{x}^f) \sim \exp\left[-\frac{1}{2} \left(\boldsymbol{x} - \boldsymbol{x}^f \right)^T \boldsymbol{P}_f^{-1} \left(\boldsymbol{x} - \boldsymbol{x}^f \right) \right]$$

$$J_b(\boldsymbol{x}) = -\ln\left[PDF(\boldsymbol{x} = \boldsymbol{x}^f) \right] = \frac{1}{2} \left(\boldsymbol{x} - \boldsymbol{x}^f \right)^T \boldsymbol{P}_f^{-1} \left(\boldsymbol{x} - \boldsymbol{x}^f \right)$$

Limitations of Normal (Gaussian) framework

- **Gaussian errors imply linear assumption**
  - Prediction models are nonlinear
  - Observation transformation operators are nonlinear
    - Remote sensing (radar, satellite)
    - Precipitation is a nonlinear function of T and q

- **Gaussian errors imply symmetric error structure**
  - Equally likely $\boldsymbol{x}=\boldsymbol{x}^f+\varepsilon$ and $\boldsymbol{x}=\boldsymbol{x}^f-\varepsilon$
  - Not all variables (observations) have symmetric errors
    - Moisture
    - Precipitation
    - Concentration (pollutant, aerosol)

- **Extreme events (floods, droughts)**
  - Belong to a non-Gaussian PDF (Tails of a Gaussian PDF)
  - Extremely large (positive/negative) system perturbations

- **Quality Control**
  - Large departures of observations from a forecast guess are typically rejected
  - Can belong to a non-Gaussian PDF, observations should be used

How to relax Gaussian PDF assumption ?

- **Develop non-Gaussian framework for assimilation/prediction**

  - Control variable errors can be mixed Gaussian/non-Gaussian

  - Observation errors can be mixed Gaussian/non-Gaussian

  - Improve understanding of nonlinear processes, prediction of extreme events

Further motivation for inclusion of non-Gaussian errors

- **Generality**

  - *non-Gaussian assimilation/prediction framework allows greater flexibility of the system*

- **Chaos**

  - Results from nonlinear dynamics suggest a *lognormal* error distribution of the initial ensemble perturbations (e.g., bred vectors, singular vectors)

  - Implication: the posterior PDF should be non-Gaussian (lognormal)

Ensemble Methodology

- **Maximum Likelihood Ensemble Filter (MLEF)**

- Originally developed to work with Normal (Gaussian) error assumption, extended for use with mixed Normal-Lognormal observation errors
- Estimate of the *conditional mode* of the posterior PDF
- Ensembles used to estimate the *uncertainty* of the conditional mode
- *Non-differentiable* cost-function minimization
- Implicit Hessian preconditioning
- Augmented control variable: *initial conditions*, *model bias*, *empirical parameters*, *boundary conditions*
- Related to: (*i*) Variational data assimilation, (*ii*) Iterative Kalman filters, and (*iii*) Ensemble Transform Kalman Filter – ETKF

References:

Zupanski 2005, *Mon. Wea. Rev.*

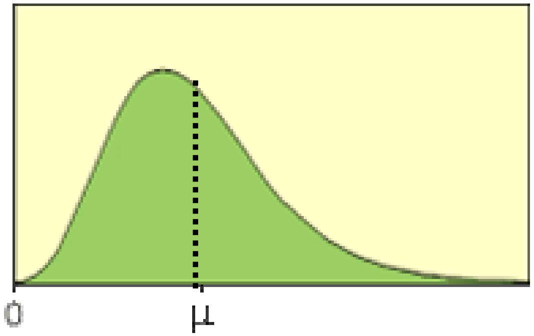
Zupanski and Zupanski 2005, *Mon. Wea. Rev.*
[ftp://ftp.cira.colostate.edu/milija/papers/MLEF_model_err.pdf]

Fletcher and Zupanski 2005, *Proc. Roy. Soc. London A*

[ftp://ftp.cira.colostate.edu/milija/papers/Steven_nongauss.pdf]

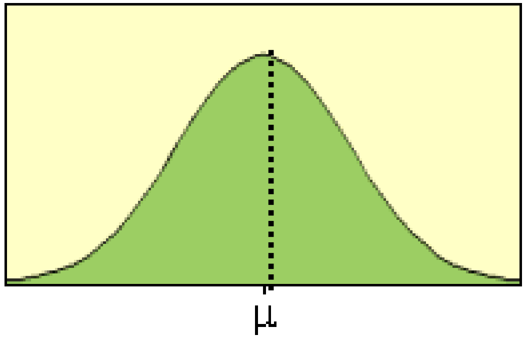
Non-Gaussian MLEF framework

Log-Normal PDF



$$\phi(x)=\begin{cases}\dfrac{exp\left(-\dfrac{1}{2}\left(\dfrac{log(x)-m}{s}\right)^2\right)}{xs\sqrt{2\pi}} & x>0\\ 0 & \text{otherwise}\end{cases}$$

Gaussian PDF



$$\phi(x)=\frac{exp\left(-\dfrac{1}{2}\left(\dfrac{x-\mu}{\sigma}\right)^2\right)}{\sigma\sqrt{2\pi}}$$

Posterior conditional PDF:

$$\phi(X\mid Y)=\frac{\phi(X)\phi(Y\mid X)}{\phi(Y)}$$

***Log-likelihood cost-function*:**

$$J(x)=-\ln[\phi(x)]$$

Maximize posterior PDF ⇔ minimize cost function

Non-Gaussian MLEF framework
Lognormal height observation errors

- **Gaussian prior probability distribution (state vector)**

$$\phi(X) \sim Gaussian$$

- **Observations**

$$\phi(Y_{HEIGHT} \mid X) \sim Lognormal \Rightarrow \qquad \boxed{\frac{\boldsymbol{y}_{HEIGHT}}{\mathcal{H}(\boldsymbol{x})} = \boldsymbol{\varepsilon}_L}$$

$$\phi(Y_{WIND} \mid X) \sim Gaussian \Rightarrow \qquad \boxed{\boldsymbol{y}_{WIND} - \mathcal{H}(\boldsymbol{x}) = \boldsymbol{\varepsilon}_G}$$

- ***Minimize mixed Normal-Lognormal cost function:***

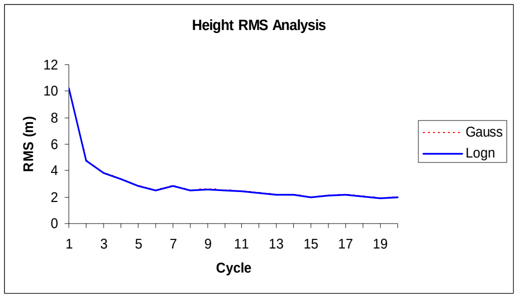
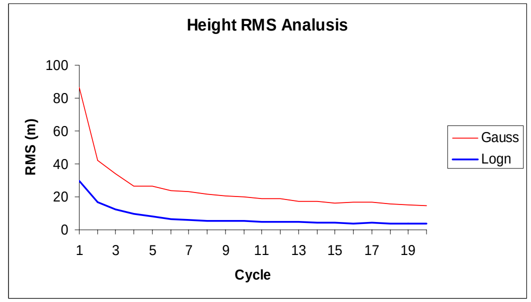
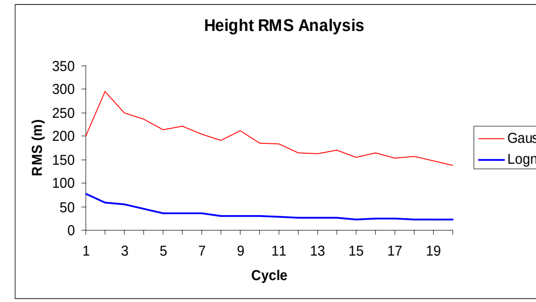
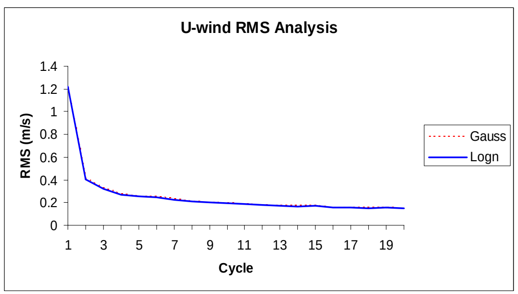
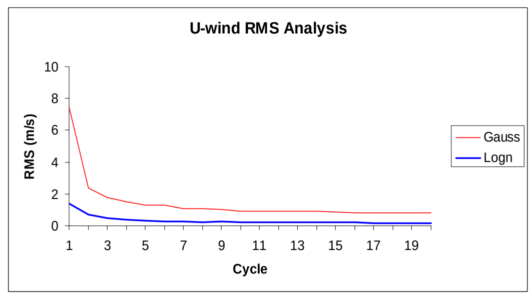
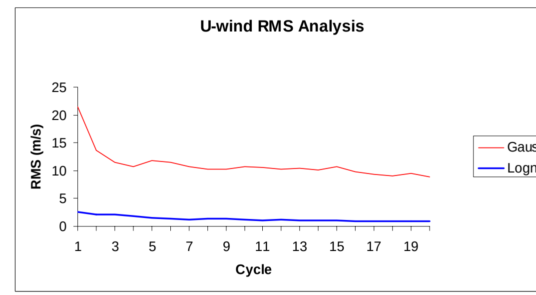
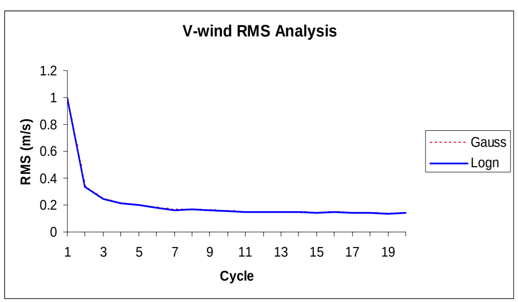
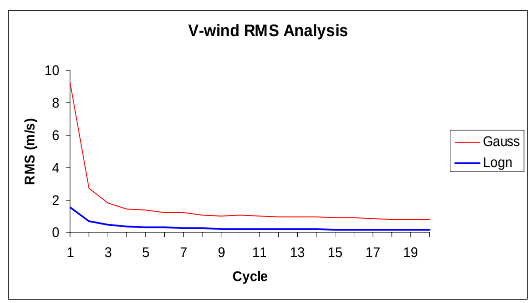
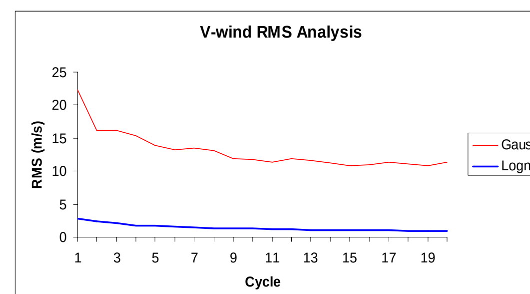
$$J(\boldsymbol{x}) = \frac{1}{2}\left(\boldsymbol{x} - \boldsymbol{x}^f\right)^T \boldsymbol{P}_f^{-1}\left(\boldsymbol{x} - \boldsymbol{x}^f\right) + \frac{1}{2}\left(y_{WIND} - \mathcal{H}(\boldsymbol{x})\right)^T \boldsymbol{R}^{-1}\left(y_{WIND} - \mathcal{H}(\boldsymbol{x})\right)$$

Normal

$$+\frac{1}{2}\left(\ln\left[\frac{y_{HEIGHT}}{\mathcal{H}(\boldsymbol{x})}\right] - m\right)^T \boldsymbol{R}_S^{-1}\left(\ln\left[\frac{y_{HEIGHT}}{\mathcal{H}(\boldsymbol{x})}\right] - m\right) + \sum_{i=1}^{N_{obs}} \ln\left[\frac{y_{HEIGHT}}{\mathcal{H}(\boldsymbol{x})}\right]_i$$

Lognormal
(additional nonlinear terms)

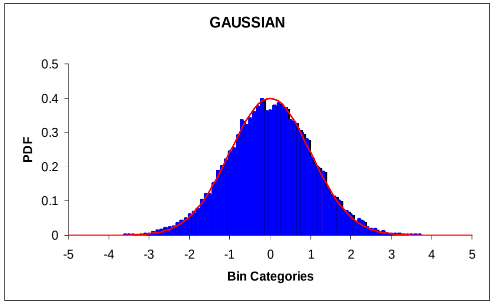
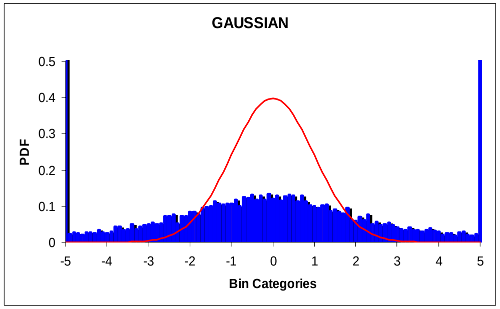
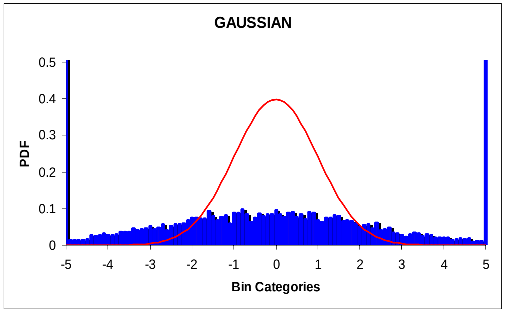
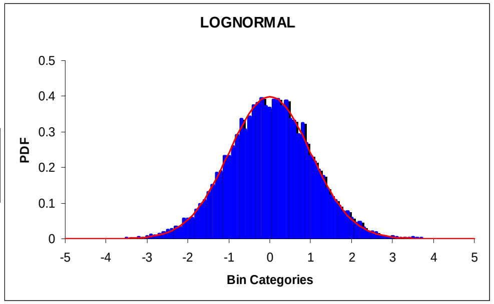
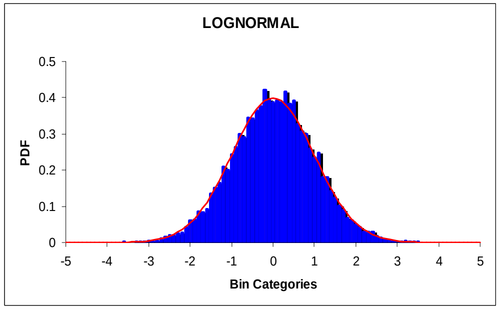
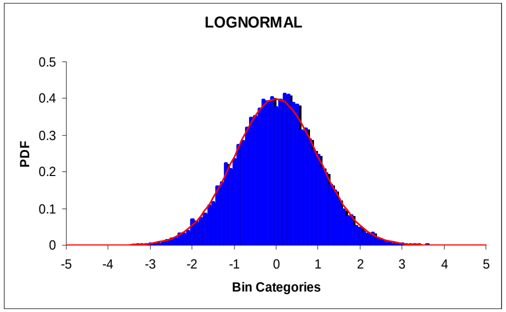
Results with CSU global shallow-water model:
Analysis RMS errors

	σ_{Logn}= 1.e-3 ~ σ_{Gauss}= 5 m	σ_{Logn}= 1.e-2 ~ σ_{Gauss}= 60 m	σ_{Logn}= 5.e-2 ~ σ_{Gauss}= 300 m
height			
U-wind			
V-wind			

Gaussian framework works only for *small* observation errors

Lognormal framework works for *all* error magnitudes

Results with CSU global shallow-water model:
Innovation Histogram (obs-fcst)

	σ_{Logn}= 1.e-3 ~ σ_{Gauss}= 5 m	σ_{Logn}= 1.e-2 ~ σ_{Gauss}= 60 m	σ_{Logn}= 5.e-2 ~ σ_{Gauss}= 300 m
Gaussian			
Lognormal			

$$\textit{Normalized} \quad \textit{Innovation} = \left[\boldsymbol{H}\boldsymbol{P}_f \boldsymbol{H}^{\mathrm{T}} + \boldsymbol{R}\right]^{-1/2} [\boldsymbol{y} - H(\boldsymbol{x}^a)]$$

Generalized non-Gaussian MLEF framework can handle Gaussian, Lognormal, or mixed PDF errors !

Conclusions and Future Research Directions

Mixed Normal/Lognormal Observation errors

- Improved performance of a non-Gaussian algorithm
- Similar results for small errors, significant deficiencies of Gaussian error assumption for large errors

Development of a fully non-Gaussian algorithm

- Allow for non-Gaussian *state variable* errors (initial conditions, parameters)
- Generalized algorithm with a list of PDFs

Assume lognormal errors of initial ensemble perturbations

- Evaluate theoretical results from space-time chaos
- Improve robustness

THORPEX Research: MLEF with GFS and real observations

- NCEP observations with Global Forecasting System
- Double-resolution MLEF: high-resolution control, low-resolution ensembles

Thank you !

Literature

Anderson, J. L. 2003: A local least squares framework for ensemble filtering. *Mon. Wea. Rev.*, **131**, 634-642.

Bishop, B., Etherton, J. and Majmudar, S. J. 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev*., **129**, 420-436.

Review paper

Cohn, S.E., 1997: An introduction to estimation theory. *J. Meteor. Soc. Japan*, **75**, 257-288.

Review paper

Evensen, G. 2003: The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynamics*, **53**, 343-367.

Fletcher, S. J., and M. Zupanski, 2005: A framework for data assimilation which allows for non-Gaussian errors. Submitted to *Proc. Royal Soc. of London A*. [ftp://ftp.cira.colostate.edu/milija/papers/Steven_nongauss.pdf]

Hamill, T.M., and C. Snyder, 2000: A hybrid ensemble Kalman filter-3D variational analysis scheme. *Mon. Wea. Rev.*, **128**, 2905-2919.

Houtekamer, P.L., H.L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen, 2005: Atmospheric data assimilation with an Ensemble Kalman Filter: Results with real observations. *Mon. Wea. Rev.*, **133**, 604-620.

General

Jazwinski, A.H., 1970: *Stochastic processes and filtering theory*. Academic Press, New York, 376 pp.

Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J., Corazza, M., Kalnay, E., Patil, D. J. and Yorke, J. A. 2004: A Local Ensemble Kalman Filter for Atmospheric Data Assimilation. *Tellus*, **56A**, No. 4, 273-277.

Reichle, R. H., McLaughlin, D. B. and Entekhabi, D. 2002a: Hydrologic data assimilation with the Ensemble Kalman Filter. *Mon. Wea. Rev*., **130**, 103-114.

Snyder, C., and Zhang, F. 2003: Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev*., **131**, 1663-1677.

Review paper

Tippett, M., J.L. Anderson, C.H. Bishop, T.M. Hamill, and J.S. Whitaker, 2003: Ensemble square-root filters. Mon. Wea. Rev., 131, 1485-1490.

Whitaker, J. S., and Hamill, T. M. 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev*., **130**,1913-1924.

Zupanski, M. 2005: Maximum Likelihood Ensemble Filter: Theoretical Aspects. *Mon.Wea.Rev.*, **133**, 1710-1726.

Zupanski, D. and M. Zupanski, 2005: Model error estimation employing ensemble data assimilation approach. Submitted to *Mon. Wea. Rev*. [ftp://ftp.cira.colostate.edu/milija/papers/MLEF_model_err.pdf]

Zupanski, M., S.J. Fletcher, I.M. Navon, B. Uzunoglu, R.P. Heikes, D.A. Randall, T.D. Ringler, and D. Daescu, 2005: A method for initialization of ensemble data assimilation. Submitted to *Tellus*. [ftp://ftp.cira.colostate.edu/milija/papers/Tellus_feb2005.pdf]

MLEF Analysis Step

Use forecast (prior) error covariance square-root

$$\boxed{\boldsymbol{P}_f^{1/2} = [\boldsymbol{p}_1^f \quad \boldsymbol{p}_2^f \quad \cdots \quad \boldsymbol{p}_{N_E}^f]}$$

Minimize cost function in subspace spanned by ensemble perturbations $\boldsymbol{p}_i^f$

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

Similar to variational, however:

❑ **Non-differentiable iterative minimization with preconditioning**

- No differentiability assumption: works for all bounded operators
- Generalized gradient, generalized Hessian

❑ **Solution in ensemble subspace**

- Reduced dimensions of the analysis correction subspace
- Focus on unstable, growing perturbations in the analysis
- Search for the attractor subspace (unstable manifold span-vectors)

Hessian preconditioning

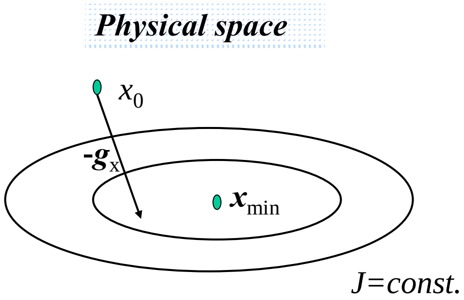
Change of variable (Hessian preconditioning)

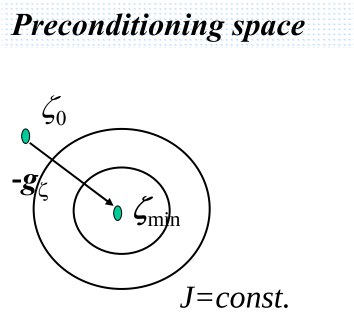
$$\boxed{\boldsymbol{x}_k = \boldsymbol{x}^f + \boldsymbol{P}_f^{1/2}\left[\boldsymbol{I} + \left(\boldsymbol{Z}^b\right)^T \boldsymbol{Z}^b\right]^{-1/2} \zeta_k}$$

$$\boxed{\boldsymbol{z}_i^b = \boldsymbol{R}^{-1/2}\left[\mathcal{H}(\boldsymbol{x}^f + \boldsymbol{p}_i^f) - \mathcal{H}(\boldsymbol{x}^f)\right]} \quad \boxed{\boldsymbol{Z}^b = \left[\boldsymbol{z}_1^b \quad \boldsymbol{z}_2^b \quad \cdot \quad \cdot \quad \boldsymbol{z}_{N_E}^b\right]}$$

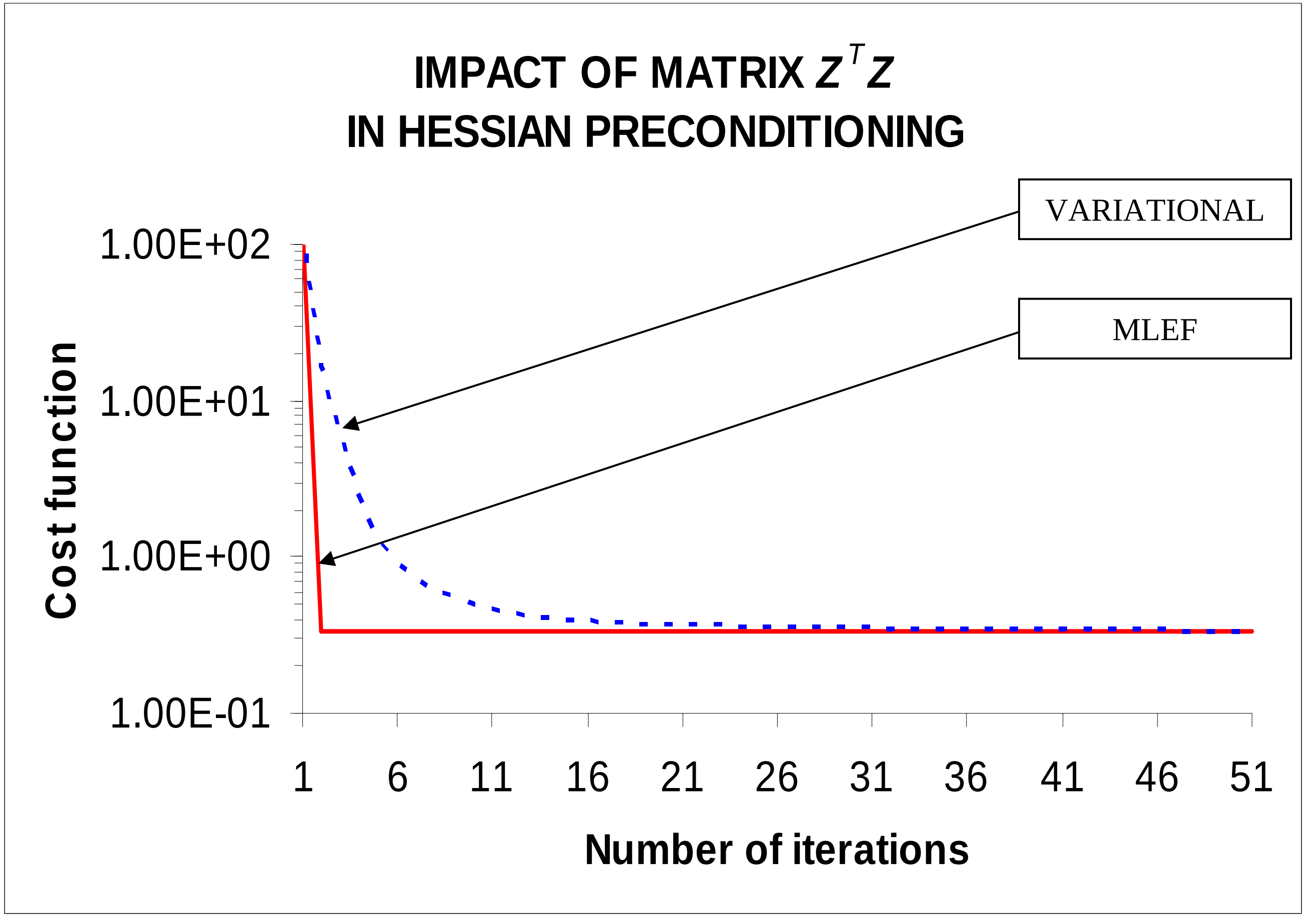
$$\boxed{\zeta_{k+1} = \zeta_k + \alpha_k \boldsymbol{d}_k}$$

k – iteration index





Hessian Preconditioning



$$\boldsymbol{P}_{VAR}^{-1} = \boldsymbol{P}_f$$

$$\mathrm{P}_{ENS}^{-1} = \boldsymbol{P}_f^{\,1/2}(\boldsymbol{I} + \boldsymbol{Z}^T \boldsymbol{Z})^{-1} \boldsymbol{P}_f^{\,T/2}$$

Analysis (posterior) error covariance

$$\boldsymbol{P}_a^{1/2} = \boldsymbol{P}_f^{1/2}(\boldsymbol{I} + \boldsymbol{Z}^T\boldsymbol{Z})^{-1/2}$$

$$\boldsymbol{z}_i = \boldsymbol{R}^{-1/2}\mathcal{H}(\boldsymbol{x} + \boldsymbol{p}_i^f) - \boldsymbol{R}^{-1/2}\mathcal{H}(\boldsymbol{x})$$

- **Analysis error covariance estimated from minimization algorithm**
- ***Assumption*: Inverse Hessian = Analysis error covariance**

How to make sure the assumption is valid?

- **$\|x^{true} - x^a\| < \varepsilon$: for small ε (compared to analysis errors)**
- **In practice, ε is only a fraction of the analysis error** (1-2 %)

Why is ε small?

- **Good Hessian preconditioning allows efficient and accurate minimization**
- **By monitoring minimization, assure the calculated solution is close to the true minimum**

Forecast (prior) error covariance

$$\boldsymbol{P}_f^{1/2} = [\boldsymbol{p}_1^f \quad \boldsymbol{p}_2^f \quad \cdots \quad \boldsymbol{p}_{N_E}^f]$$

$$\boldsymbol{p}_i^f = \mathcal{M}(\boldsymbol{x}^a + \boldsymbol{p}_i^a) - \mathcal{M}(\boldsymbol{x}^a)$$

- **Control vector is the *most likely* forecast**
- **Square-root used in the algorithm** (full covariance can be calculated)
- **No sampling of error covariance**
- **Provides dynamic continuity between the analysis and forecast**