Overview of WRF Data Assimilation

Tom Auligné

National Center for Atmospheric Research, Boulder, CO USA

WRFDA Tutorial - August 3-5, 2010
Acknowledgments and References

- WRF Tutorial Lectures (H. Huang and D. Barker)
- Data Assimilation concepts and methods (ECMWF Training Course, F. Bouttier and P. Courtier)
- Data Assimilation Research Testbed (DART) Tutorial (J. Anderson et al.)
# Table of contents

1. **Introduction**
2. **Simple Scalar Example**
   - Kalman Filter equations
3. **Modern Implementations**
   - Sequential Algorithms
     - Ensemble Kalman Filter
     - 3D Variational (3DVar)
   - Smoothers
     - 4D Variational (4DVar)
4. **WRFDA Overview**
Motivation

- A sufficiently accurate knowledge of the state of the atmosphere at the initial time. 
  (Today’s weather)
- A sufficiently accurate knowledge of the laws according to which one state of the atmosphere develops from another. 
  (Tomorrow’s weather)

Vilhelm Bjerknes (1904) 
(Peter Lynch)
Motivation

- A sufficiently accurate knowledge of the state of the atmosphere at the initial time. *(Today’s weather)*
- A sufficiently accurate knowledge of the laws according to which one state of the atmosphere develops from another. *(Tomorrow’s weather)*

Vilhelm Bjerknes (1904) *(Peter Lynch)*
Motivation

- Initial conditions for Numerical Weather Prediction (NWP)
- Calibration and validation
- Observing system design, monitoring and assessment
- Reanalysis
- Better understanding (Model errors, Data errors, Physical process interactions, etc)
From Empirical to Statistical methods

- Successive Correction Method (SCM, Cressman 1959)
  Each observation within a radius of influence $L$ is given a weight $w$ varying with the distance $r$ to the model grid point:
  \[ w(r) = \frac{L^2 - r^2}{L^2 + r^2} \quad (r \leq L) \]
- Nudging
- Physical Initialization (PI), Latent Heat Nudging (LHN)

However...

- Relaxation functions are somewhat arbitrary
- Good forecast can be replaced by bad observations
- Noisy observations can create unphysical analysis

So...

Modern DA techniques are usually statistical
From Empirical to Statistical methods

- Successive Correction Method (SCM, Cressman 1959)
  Each observation within a radius of influence $L$ is given a weight $w$ varying with the distance $r$ to the model grid point:
  \[ w(r) = \frac{L^2 - r^2}{L^2 + r^2} \quad (r \leq L) \]

- Nudging
- Physical Initialization (PI), Latent Heat Nudging (LHN)

However...

- Relaxation functions are somewhat arbitrary
- **Good** forecast can be replaced by **bad** observations
- Noisy observations can create unphysical analysis

So...

Modern DA techniques are usually statistical
Notations

- $x_t$: ”True” state
- $x_o$: Observation
- $x_b$: Background information
- $d = x_o - x_b$: Innovation or Departure

Hypotheses

- Observation and Background errors are uncorrelated, unbiased, normally distributed, with variance $R$ and $B$ resp.
- Analysis $x_a$ is ”optimal” in RMSE sense
- Linear Analysis: $x_a = \alpha x_o + \beta x_b = x_b + \alpha(x_o - x_b)$
What is the temperature in this room?

Notations
- $x_t$: ”True“ state
- $x_o$: Observation
- $x_b$: Background information
- $d = x_o - x_b$: Innovation or Departure

Hypotheses
- Observation and Background errors are uncorrelated, unbiased, normally distributed, with variance $R$ and $B$ resp.
- Analysis $x_a$ is ”optimal” in RMSE sense
- Linear Analysis: $x_a = \alpha x_o + \beta x_b = x_b + \alpha(x_o - x_b)$
Best Linear Unbiased Estimate

The analysis value is \( x_a = x_b + \alpha (x_o - x_b) \) and its error variance:
\[
A = (x_a - x_t)(x_a - x_t) = (1 - \alpha)^2 B + \alpha^2 R
\]

\[
\frac{\partial A}{\partial \alpha} = 2\alpha(B + R) - 2B \\
\frac{\partial A}{\partial \alpha} = 0 \Rightarrow \alpha = \frac{B}{B + R}
\]
Best Linear Unbiased Estimate

The analysis value is $x_a = x_b + \alpha(x_o - x_b)$ and its error variance:

$$A = (x_a - x_t)(x_a - x_t) = (1 - \alpha)^2B + \alpha^2R$$

$$\frac{\partial A}{\partial \alpha} = 2\alpha(B + R) - 2B$$

$$\frac{\partial A}{\partial \alpha} = 0 \implies \alpha = \frac{B}{B + R}$$

Best Linear Unbiased Estimate (BLUE)

$x_a = x_b + K(x_o - x_b)$ with the definition of the Kalman Gain:

$$K = B(B + R)^{-1}$$

and the analysis error variance: $A^{-1} = B^{-1} + R^{-1}$

Statistically, the analysis is better than the observation ($A < R$) and the background ($A < B$)
Best Linear Unbiased Estimate

The analysis value is $x_a = x_b + \alpha(x_o - x_b)$ and its error variance:

$$A = (x_a - x_t)(x_a - x_t) = (1 - \alpha)^2 B + \alpha^2 R$$

$$\frac{\partial A}{\partial \alpha} = 2\alpha(B + R) - 2B$$
$$\frac{\partial A}{\partial \alpha} = 0 \Rightarrow \alpha = \frac{B}{B + R}$$

**Best Linear Unbiased Estimate (BLUE)**

$x_a = x_b + K(x_o - x_b)$ with the definition of the *Kalman Gain*:

$$K = B(B + R)^{-1}$$

and the analysis error variance: $A^{-1} = B^{-1} + R^{-1}$

Statistically, the analysis is better than the observation ($A < R$) and the background ($A < B$)
This solution is equivalent to minimizing the cost function:

\[ J(x) = \frac{1}{2} (x - x_b)^T B^{-1} (x - x_b) + \frac{1}{2} (x - x_o)^T R^{-1} (x - x_o) = J_b + J_o \]

Proof:

\[ \nabla J = B^{-1} (x - x_b) + R^{-1} (x - x_o) = 0 \]

\[ \Rightarrow x_a = x_b + \frac{B}{B + R} (x_o - x_b) \]

\[ = x_b + K (x_o - x_b) \]
This solution is equivalent to minimizing the cost function:

\[ J(x) = \frac{1}{2} (x - x_b)^T B^{-1} (x - x_b) + \frac{1}{2} (x - x_o)^T R^{-1} (x - x_o) = J_b + J_o \]

Proof:

\[ \nabla J = B^{-1} (x - x_b) + R^{-1} (x - x_o) = 0 \]

\[ \Rightarrow x_a = x_b + \frac{B}{B + R} (x_o - x_b) \]

\[ = x_b + K (x_o - x_b) \]
Analysis Accuracy

from Bouttier and Courtier 1999

Quality of the Analysis

The precision is defined by the convexity or Hessian $A = J''^{-1}$
Conditional Probabilities

According to Bayes Theorem, the joint pdf of \( x \) and \( x_0 \) is:

\[
P(x \land x_0) = P(x|x_0)P(x_0) = P(x_0|x)P(x)
\]

Since \( P(x_0) = 1 \), \( P(x|x_0) = P(x_0|x)P(x) \)

We assumed the background and observation errors were Gaussian:
\[
P(x) = \lambda_b e^{\left[\frac{1}{2B}(x_b-x)^2\right]}
\]
and
\[
P(x_0|x) = \lambda_0 e^{\left[\frac{1}{2R}(x_0-x)^2\right]}
\]

\[
\Rightarrow P(x|x_0) = \lambda_a e^{\left[\frac{1}{2R}(x_0-x)^2 + \frac{1}{2B}(x_b-x)^2\right]} = \lambda_a e^{-J(x)}
\]

Maximum Likelihood

The minimum of the cost function \( J \) is also the estimator of \( x_t \) with the maximum likelihood.
Conditional Probabilities

According to Bayes Theorem, the joint pdf of \( x \) and \( x_o \) is:

\[
P(x \land x_o) = P(x| x_o)P(x_o) = P(x_o|x)P(x)
\]

Since \( P(x_o) = 1 \), \( P(x| x_o) = P(x_o|x)P(x) \)

We assumed the background and observation errors were Gaussian:

\[
P(x) = \lambda_b e^{\frac{1}{2B}(x_b-x)^2} \quad \text{and} \quad P(x_o|x) = \lambda_o e^{\frac{1}{2R}(x_o-x)^2}
\]

\[
\Rightarrow P(x| x_o) = \lambda_a e^{\frac{1}{2R}(x_o-x)^2 + \frac{1}{2B}(x_b-x)^2} = \lambda_a e^{-J(x)}
\]

Maximum Likelihood

The minimum of the cost function \( J \) is also the estimator of \( x_t \) with the maximum likelihood.
Partial Conclusions

Under the aforementioned hypotheses, the BLUE:

- can be determined analytically through the Kalman gain $K$
- is also the minimum of a cost function $J = J_b + J_o$
- is optimal for minimum variance and maximum likelihood
Forecast model $M_{i \rightarrow i+1} = M$ from step $i$ to $i+1$

$$x_{i+1}^t = M(x_i^t) + q_i$$

where $q_i$ is the model error. As $q_i$ is unknown and $x_i^a$ is the best estimate of $x_i^t$, usually: $x_{i+1}^f = M(x_i^a)$

Forecast error

$$x_{i+1}^f - x_{i+1}^t = M(x_i^a) - M(x_i^t) - q_i \approx M_i(x_i^a - x_i^t) - q_i$$

$M$ is called the **Tangent-Linear** code of the non-linear model $M$
Forecast model $M_{i\rightarrow i+1} = M$ from step $i$ to $i+1$

$$x_{i+1}^t = M(x_i^t) + q_i$$

where $q_i$ is the model error. As $q_i$ is unknown and $x_i^a$ is the best estimate of $x_i^t$, usually: $x_{i+1}^f = M(x_i^a)$

Forecast error

$$x_{i+1}^f - x_{i+1}^t = M(x_i^a) - M(x_i^t) - q_i \approx M_i(x_i^a - x_i^t) - q_i$$

$M$ is called the **Tangent-Linear** code of the non-linear model $M$

Forecast error covariance matrix

$$P_{i+1}^f \approx M_i(x_i^a - x_i^t)(x_i^a - x_i^t)^T M_i + q_i q_i^T = M_i P_i^a M_i^T + Q_i$$
Sequential Data Assimilation

Forecast model $M_{i \rightarrow i+1} = M$ from step $i$ to $i+1$

$$x_{i+1}^t = M(x_i^t) + q_i$$

where $q_i$ is the model error. As $q_i$ is unknown and $x_i^a$ is the best estimate of $x_i^t$, usually: $x_{i+1}^f = M(x_i^a)$

Forecast error

$$x_{i+1}^f - x_{i+1}^t = M(x_i^a) - M(x_i^t) - q_i \approx M_i(x_i^a - x_i^t) - q_i$$

$M$ is called the **Tangent-Linear** code of the non-linear model $M$

Forecast error covariance matrix

$$P_{i+1}^f \approx M_i(x_i^a - x_i^t)(x_i^a - x_i^t)^T M_i + q_i q_i^T = M_i P_i^a M_i^T + Q_i$$
Sequential Data Assimilation

We can use the forecast as background for the **BLUE** calculation

\[
K_i = P_i^f (P_i^f + R)^{-1}
\]

\[
x_i^a = x_i^f + K(x_i^o - x_i^f)
\]

\[
(P_i^a)^{-1} = (P_i^f)^{-1} + R^{-1} \Rightarrow P_i^a = (I - K_i)P_i^f
\]

Finally, we can distinguish the model space \(x\) from the observation space \(y\) and introduce an Observation Operator \(H : x \mapsto y\), which is linearized: \(H(x_i^a) - H(x_i^f) \approx H(x_i^a - x_i^f)\)

\[
K_i = P_i^f H_i^T (H_i P_i^f H_i^T + R)^{-1}
\]

\[
x_i^a = x_i^f + K(y_i^o - x_i^f)
\]

\[
P_i^a = (I - K_i H_i)P_i^f
\]
Sequential Data Assimilation

We can use the forecast as background for the **BLUE** calculation

\[ K_i = P_i^f (P_i^f + R)^{-1} \]
\[ x_i^a = x_i^f + K(x_i^o - x_i^f) \]
\[ (P_i^a)^{-1} = (P_i^f)^{-1} + R^{-1} \Rightarrow P_i^a = (I - K_i)P_i^f \]

Finally, we can distinguish the model space \( x \) from the observation space \( y \) and introduce an Observation Operator \( H : x \mapsto y \), which is linearized: \( H(x_i^a) - H(x_i^f) \approx H(x_i^a - x_i^f) \)

\[ K_i = P_i^f H_i^T (H_i P_i^f H_i^T + R)^{-1} \]
\[ x_i^a = x_i^f + K(y_i^o - x_i^f) \]
\[ P_i^a = (I - K_i H_i)P_i^f \]
**The Extended Kalman Filter Algorithm**

Analysis step $i$:

\[
K_i = P_i^f H_i^\top [H_i P_i^f H_i^\top + R]^{-1}
\]  
(1)

\[
x_i^a = x_i^f + K_i [y^o - H x_i^f]
\]  
(2)

\[
P_i^a = [I - K_i H_i] P_i^f
\]  
(3)

Forecast step from $i$ to $i + 1$:

\[
x_{i+1}^f = M(x_i^a)
\]  
(4)

\[
P_{i+1}^f = M_i P_i^a M_i^\top + Q_i
\]  
(5)

**Hypotheses**

- Gaussian distributions of errors
- $M$: Linearization around non-linear Model $M$
- $H$: Linearization around non-linear Observation Operator $H$
The Extended Kalman Filter Algorithm

Analysis step $i$:

$$K_i = P_i^f H_i^T [H_i P_i^f H_i^T + R]^{-1}$$

$$x_i^a = x_i^f + K_i [y^o - H x_i^f]$$

$$P_i^a = [I - K_i H_i] P_i^f$$

Forecast step from $i$ to $i + 1$:

$$x_{i+1}^f = M(x_i^a)$$

$$P_{i+1}^f = M_i P_i^a M_i^T + Q_i$$

Hypotheses

- Gaussian distributions of errors
- $M$: Linearization around non-linear Model $M$
- $H$: Linearization around non-linear Observation Operator $H$
From scalar to vector: dimensions

\( x \rightarrow x \)

Number of grid points \( \approx 10^7 \)
Dimension of \( P^f, P^a \approx 10^7 \times 10^7 \)

\( y^o \rightarrow y^o \)
Number of observations \( \approx 10^6 \)
Dimension of \( R \approx 10^6 \times 10^6 \)
Hypotheses

- Monte Carlo approximation to pdfs
- Gaussian distributions used for computing update
- Localization in space: for each model grid point, only a few observations are used to compute the analysis increment.
Ensemble Kalman Filter (EnKF)

Hypotheses

- Monte Carlo approximation to pdfs
- Gaussian distributions used for computing update
- Localization in space: for each model grid point, only a few observations are used to compute the analysis increment.
Ensemble Kalman Filter (EnKF)

Hypotheses

- Monte Carlo approximation to pdfs
- Gaussian distributions used for computing update
- Localization in space: for each model grid point, only a few observations are used to compute the analysis increment.

Compare with observation and observational error distribution.
Hypotheses

- Monte Carlo approximation to pdfs
- Gaussian distributions used for computing update
- Localization in space: for each model grid point, only a few observations are used to compute the analysis increment.
Hypotheses

- Monte Carlo approximation to pdfs
- Gaussian distributions used for computing update
- Localization in space: for each model grid point, only a few observations are used to compute the analysis increment.
Ensemble Kalman Filter (EnKF)

Hypotheses

- Monte Carlo approximation to pdfs
- Gaussian distributions used for computing update
- Localization in space: for each model grid point, only a few observations are used to compute the analysis increment.
Ensemble Kalman Filter (EnKF)

Hypotheses
- Monte Carlo approximation to pdfs
- Gaussian distributions used for computing update
- Localization in space: for each model grid point, only a few observations are used to compute the analysis increment.

Advantages
- Easy to implement and provides estimate of Analysis Accuracy
- $H$ and $M$ need not be linearized

Drawbacks
- Localization avoids degeneracy from under-sampling and reduces spurious noise, but it affects model internal balance
Hypotheses

Avoid calculating $K$ by solving the equivalent minimization problem defined by the cost function:

$$J(x) = \frac{1}{2} (x - x_b)^T B^{-1} (x - x_b) + \frac{1}{2} (y^o - H(x))^T R^{-1} (y^o - H(x))$$

$$\nabla J(x) = B^{-1} (x - x_b) - H^T R^{-1} [y - H(x)]$$

$H^T$ is called the Adjoint of the linearized observation operator
Hypotheses

Avoid calculating $K$ by solving the equivalent minimization problem defined by the cost function:

$$J(x) = \frac{1}{2}(x - x_b)^T B^{-1}(x - x_b) + \frac{1}{2}(y^o - H(x))^T R^{-1}(y^o - H(x))$$

$$\nabla J(x) = B^{-1}(x - x_b) - H^T R^{-1}[y - H(x)]$$

$H^T$ is called the Adjoint of the linearized observation operator
3D Variational Data Assimilation (3DVar)

Minimization Algorithm
- Iterative minimizer
  → several simulations
- Steepest Descent, Quasi-Newton, Conjugate Gradient, etc

Preconditioning
- Improve Condition Nb
- Faster convergence

from Bouttier and Courtier 1999
Single Observation Experiment
Hypotheses

- Avoid calculating $K$ by solving the equivalent minimization problem defined by the cost function:
  \[
  J(x) = \frac{1}{2} (x - x_b)^T B^{-1} (x - x_b) + \frac{1}{2} (y^o - H(x))^T R^{-1} (y^o - H(x))
  \]

Advantages

- Easy to use with complex observation operators
- Can add external weak or penalty constraints $J_c$

Drawbacks

- Sub-optimal for strongly non-linear observation operators
- All observations are assumed to be instantaneous
3D Variational Data Assimilation (3DVar)

Hypotheses
- Avoid calculating $K$ by solving the equivalent minimization problem defined by the cost function:
  \[ J(x) = \frac{1}{2} (x - x_b)^T B^{-1} (x - x_b) + \frac{1}{2} (y^o - H(x))^T R^{-1} (y^o - H(x)) \]

Advantages
- Easy to use with complex observation operators
- Can add external weak or penalty constraints $J_c$

Drawbacks
- Sub-optimal for strongly non-linear observation operators
- All observations are assumed to be instantaneous
Hypotheses

- Generalization of 3DVar for observations distributed in time
- Analysis variable $x$ defined at the **beginning** of time window
- Find model trajectory minimizing the distance to observations
Hypotheses

- Generalization of 3DVar for observations distributed in time
- Analysis variable $x$ defined at the *beginning* of time window
- Find model trajectory minimizing the distance to observations

The Cost Function becomes:

$$J(x) = \frac{1}{2}(x - x_b)^T B^{-1} (x - x_b) + \frac{1}{2}(y^o - HM(x))^T R^{-1} (y^o - HM(x))$$

$$\nabla J(x) = B^{-1} (x - x_b) - M^T H^T R^{-1} [y - HM(x)]$$

$M^T$ is called the **Adjoint** of the linearized forecast model
4D Variational Data Assimilation (4DVar)
4D Variational Data Assimilation (4DVar)

**Hypotheses**
- Generalization of 3DVar for observations distributed in time
- Analysis variable $x$ defined at the **beginning** of time window
- Find model trajectory minimizing the distance to observations

**Advantages**
- Model internal balance is more prone to be respected

**Drawbacks**
- The development and maintenance of the Adjoint model $M^T$ can be cumbersome
- Limitation of the ”perfect model” assumption
Hypotheses

- Generalization of 3DVar for observations distributed in time
- Analysis variable $x$ defined at the **beginning** of time window
- Find model trajectory minimizing the distance to observations

Advantages

Model internal balance is more prone to be respected

Drawbacks

- The development and maintenance of the Adjoint model $M^T$ can be cumbersome
- Limitation of the ”perfect model” assumption
WRF Data Assimilation (WRFDA)

Overview of WRF Data Assimilation
WRF Data Assimilation (WRFDA)

**Community WRF DA System**
- Regional/Global
- Research/Operations
- Deterministic/Probabilistic

**Algorithms**
- 3DVar, 4DVar (Regional)
- Ensemble (ETKF/EnKF)
- Hybrid Var/Ens

**Model:** WRF
- ARW, NMM
WRFDA Program

- NCAR Staff: 20 FTE, 10 projects
- Ext. collaborators (AFWA, KMA, CWB, BMB): 10 FTE
- Community Users: 40
WRFDA Observations

Conventional
- Surface (SYNOP, METAR, SHIP, BUOY)
- Upper Air (TEMP, PIBAL, AIREP, ACARS, TAMDAR)

Bogus
- Tropical Cyclone Bogus
- Global Bogus
WRFDA Observations

Remotely Sensed Retrievals

- Atmospheric Motion Vectors (from GEOS and Polar)
- SATEM Thickness
- Ground-based GPS TPW/Zenith Total Delay
- SSM/I oceanic surface wind speed and TPW
- Scatterometer oceanic surface winds
- Wind Profiler
- Radar Radial Velocities and Reflectivities
- Satellite Temperature, humidity, thickness profiles
- GPS Refractivity (COSMIC)
WRFDA Observations

Satellite Radiances (RTTOV or CRTM Radiative Transfer)

- HIRS (from NOAA-16, 17, 18 and METOP-2)
- AMSU-A (from NOAA-15, 16, 18, EOS-Aqua and METOP-2)
- AMSU-B (from NOAA-15, 16, 17)
- MHS (from NOAA-18 and METOP-2)
- AIRS (from EOS-Aqua)
- SSMIS (from DMSP-16)
Welcome to the users home page for the Weather Research and Forecasting (WRF) model data assimilation system (WRFDA). The WRFDA system is in the public domain and is freely available for community use. It is designed to be a flexible, state-of-the-art atmospheric data assimilation system that is portable and efficient on available parallel computing platforms. WRFDA is suitable for use in a broad range of applications across scales ranging from kilometers of regional mesoscale to thousands of kilometers of global scales.

The Mesoscale and Microscale Meteorology Division of NCAR is currently maintaining and supporting a subset of the overall WRF code (Version 3) that includes:
Conclusions

- Observations $y^o$
- Background $x_b$
- Observation Operator $H$
- Innovations $y^o - H(x_b)$

- Observation Error $R$
- Bkg/Ana Error $P^f, P^a$
- Tangent-Linear $H, M$
- Adjoint $H^T, M^T$

(Extended) Kalman Filter (quasi-)linear statistical algorithm
Conclusions

- Observations $y^o$
- Background $x_b$
- Observation Operator $H$
- Innovations $y^o - H(x_b)$

- Observation Error $R$
- Bkg/Ana Error $P^f, P^a$
- Tangent-Linear $H, M$
- Adjoint $H^T, M^T$

(Extended) Kalman Filter (quasi-)linear statistical algorithm

Simplifications for practical implementation
- Ensemble methods: EnKF
- Variational methods: 3DVar, 4DVar
Observations $y^o$
- Background $x_b$
- Observation Operator $H$
- Innovations $y^o - H(x_b)$

Observation Error $R$
- Bkg/Ana Error $P^f$, $P^a$
- Tangent-Linear $H$, $M$
- Adjoint $H^T$, $M^T$

(Extended) Kalman Filter (quasi-)linear statistical algorithm

Simplifications for practical implementation
- Ensemble methods: EnKF
- Variational methods: 3DVar, 4DVar
Conclusions

Warning

WRFDA should NOT be used as a *black box*

- Processing of Observations (Quality Control, Bias Correction)
- Modeling of Background and Observation error covariances
- Accounting for Model errors and Non-Linearities
Thank you for your attention...