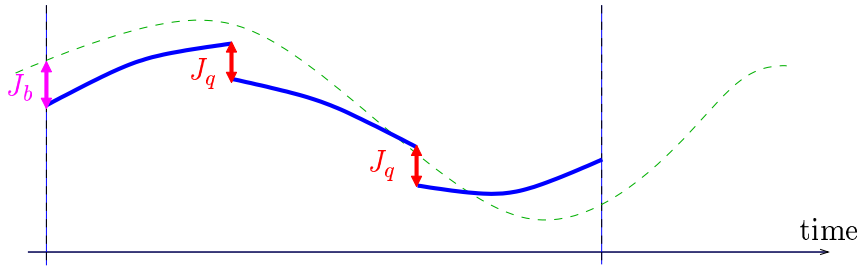
Data Assimilation with Overlapping Windows

Yannick Trémolet, Elias Hólm

ECMWF

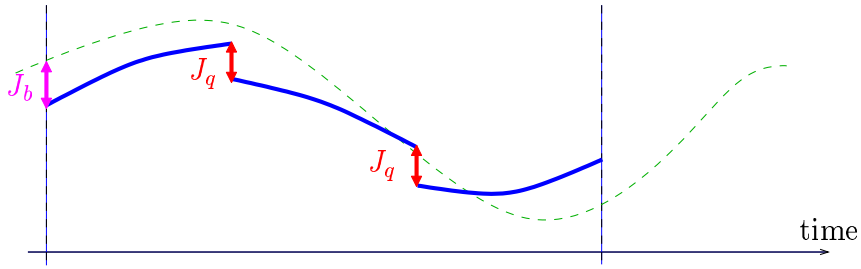
ISDA, Reading, July 2016

Long Window Weak Constraint 4D-Var

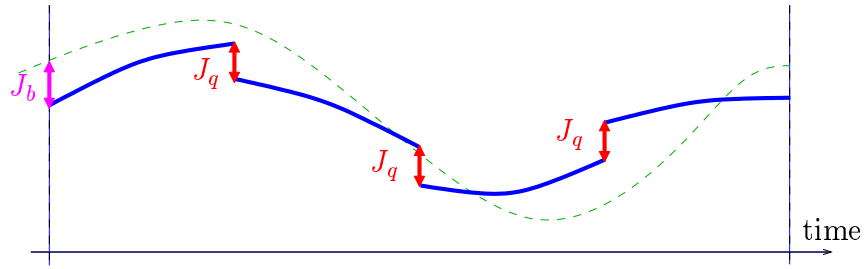


(1) Weak constraint 4D-Var

Long Window Weak Constraint 4D-Var

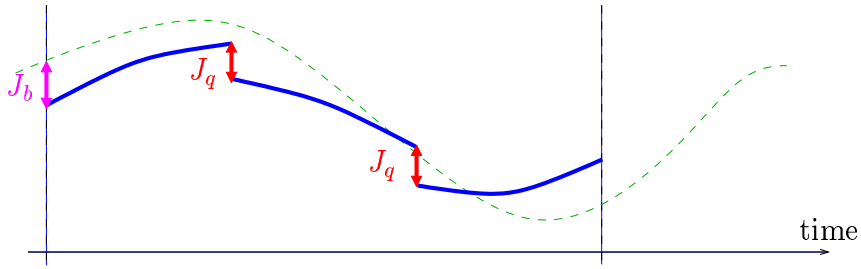


(1) Weak constraint 4D-Var

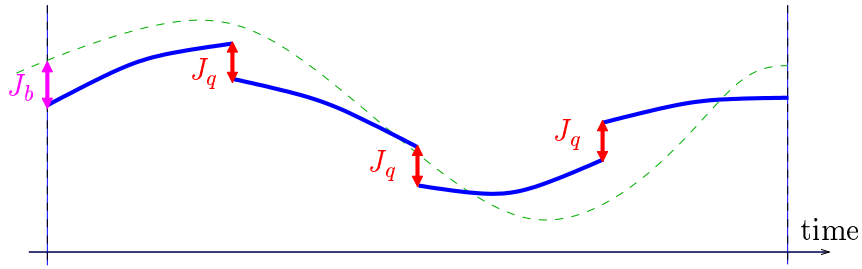


(2) Extended window

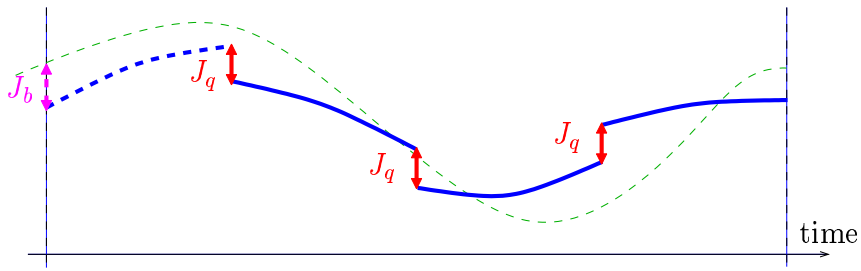
Long Window Weak Constraint 4D-Var



(1) Weak constraint 4D-Var

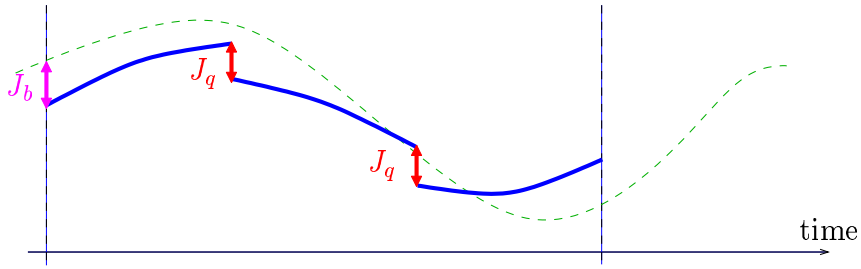


(2) Extended window

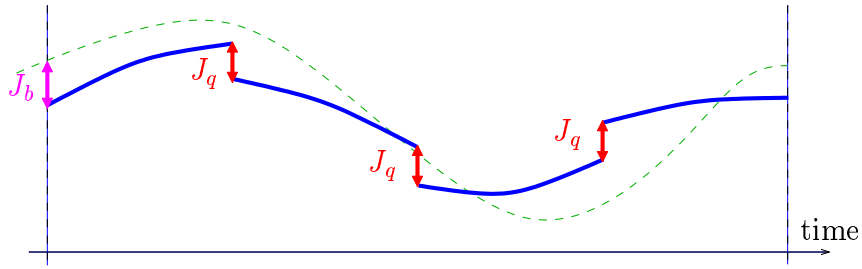


(3) Initial term has converged

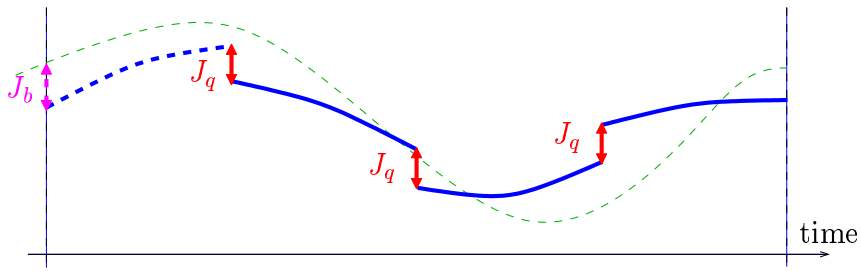
Long Window Weak Constraint 4D-Var



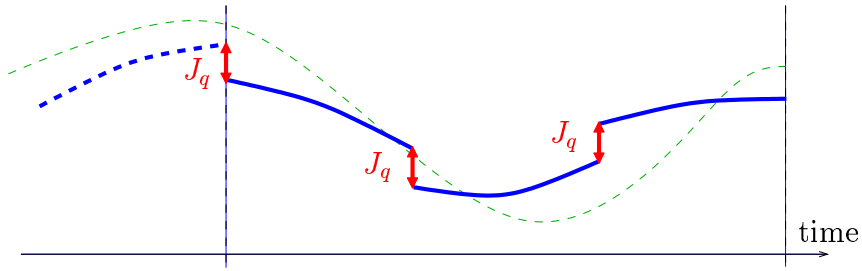
(1) Weak constraint 4D-Var



(2) Extended window

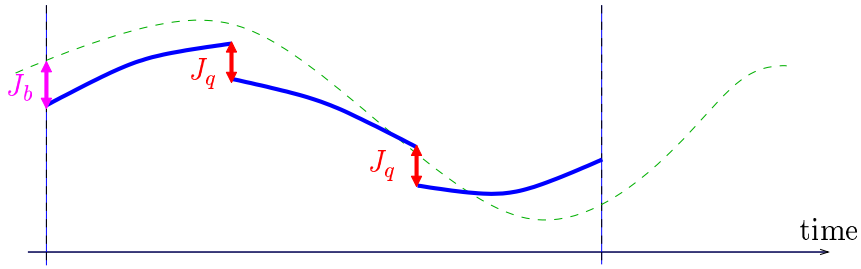


(3) Initial term has converged

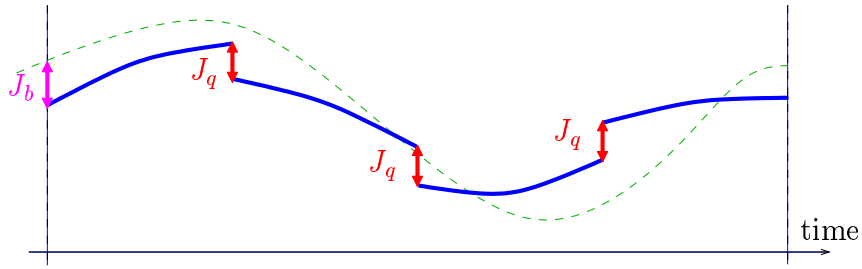


(4) Assimilation window is moved forward

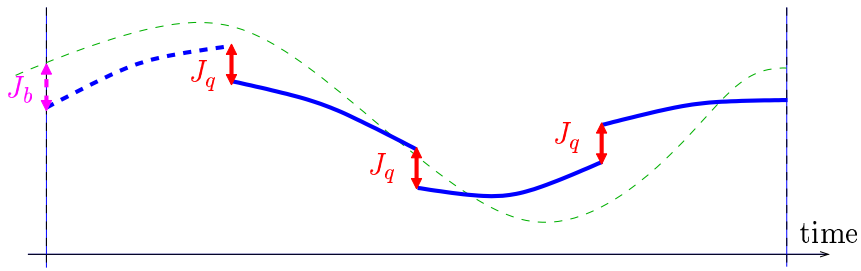
Long Window Weak Constraint 4D-Var



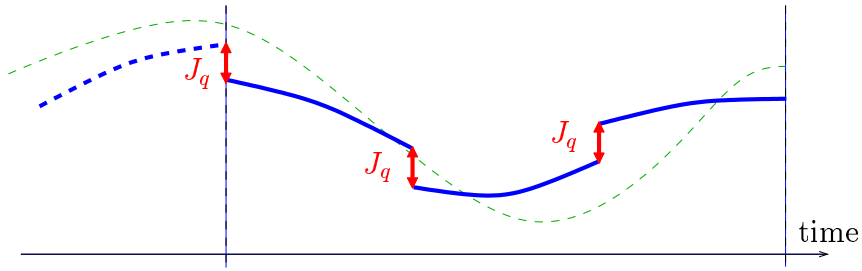
(1) Weak constraint 4D-Var



(2) Extended window



(3) Initial term has converged



(4) Assimilation window is moved forward

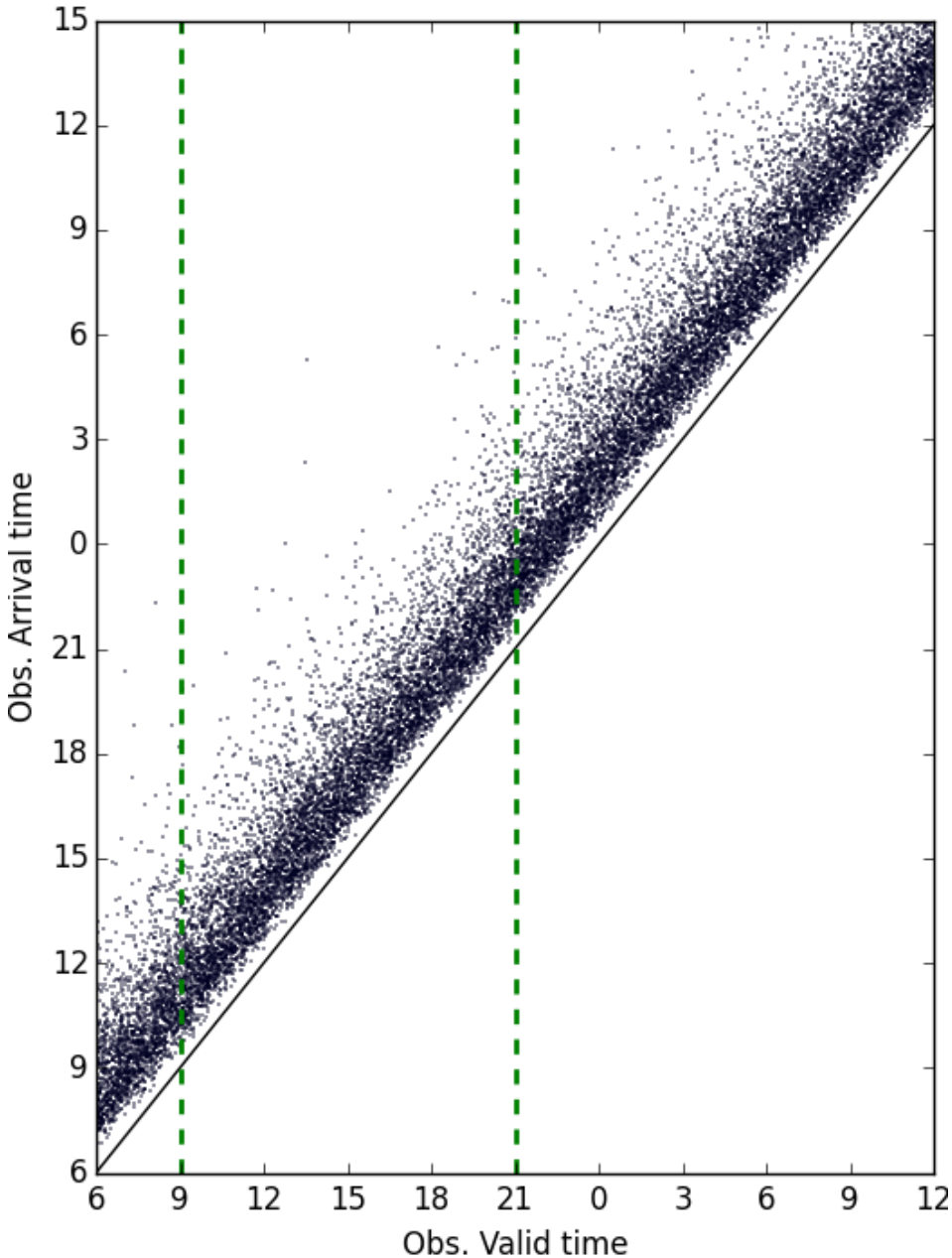
- This implementation is an approximation of weak constraint 4D-Var with an assimilation window that extends (almost) indefinitely in the past...
- ...which is equivalent to a full rank Kalman smoother (Fisher *et al.*, 2005, Ménard and Daley, 1996) that has been running indefinitely.
- In principle **B** is a problem of the past, only the error characteristics of the fundamental ingredients of the DA problem remain.

We are not there yet...

Weak constraint 4D-Var is difficult to implement in the IFS (both for scientific and technical reasons).

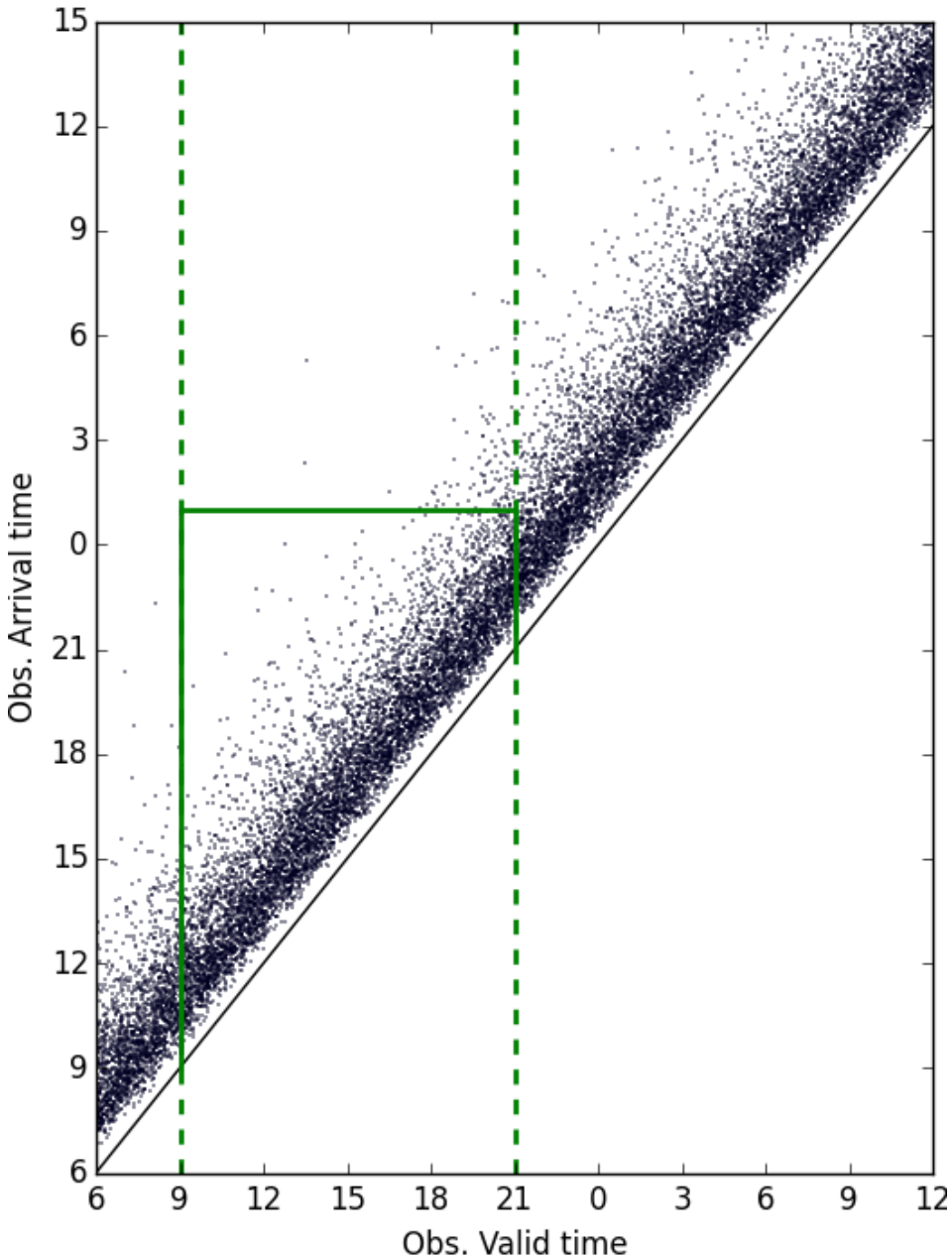
1. Overlapping windows with strong constraint 4D-Var
2. Results in reanalysis context
3. Results in NWP context
4. Comments about re-use of observations

Operational System: Observations Delay



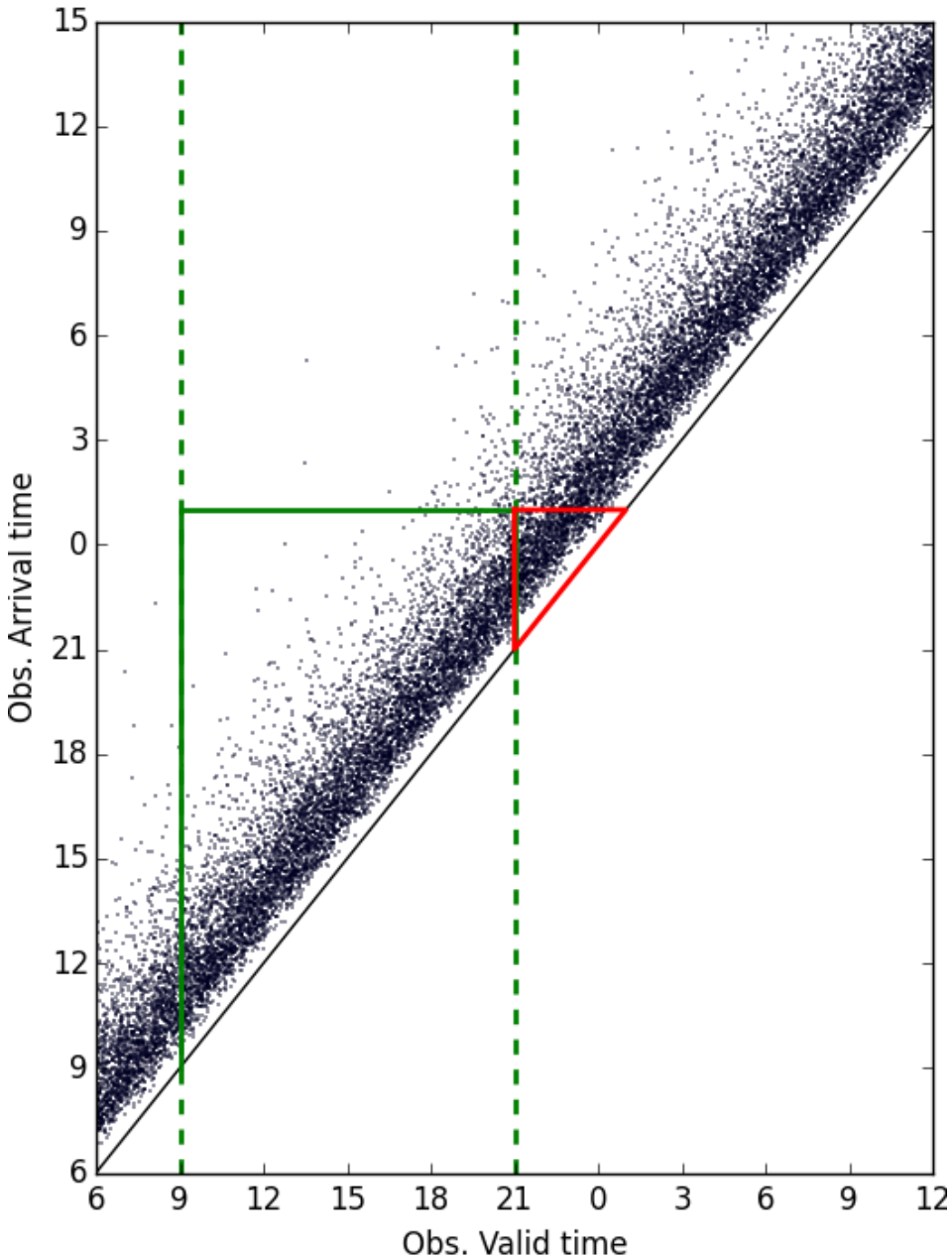
- Observations are not available instantly

Operational System: Observations Delay



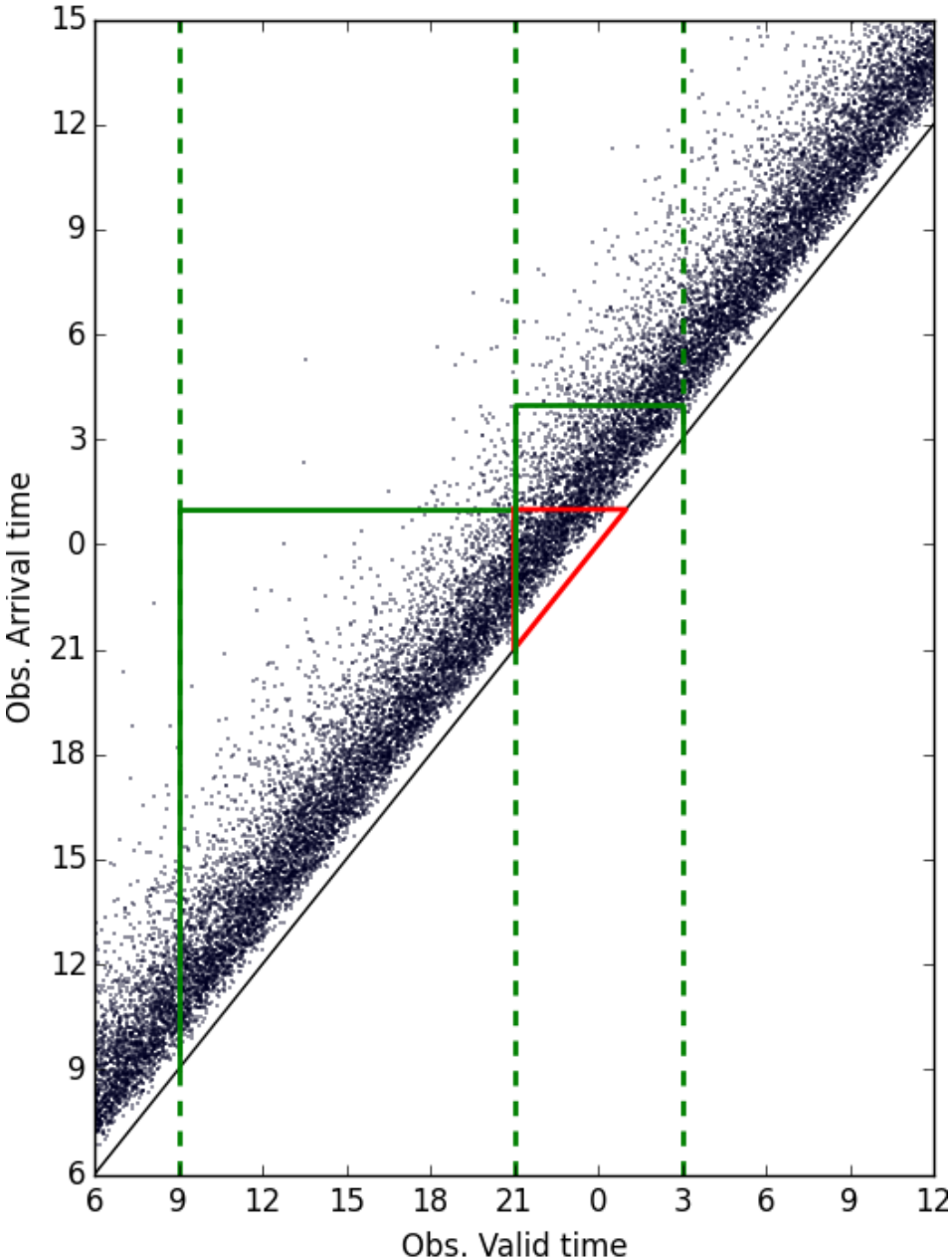
- Observations are not available instantly
- Operational systems work with a long-ish cut-off

Operational System: Observations Delay



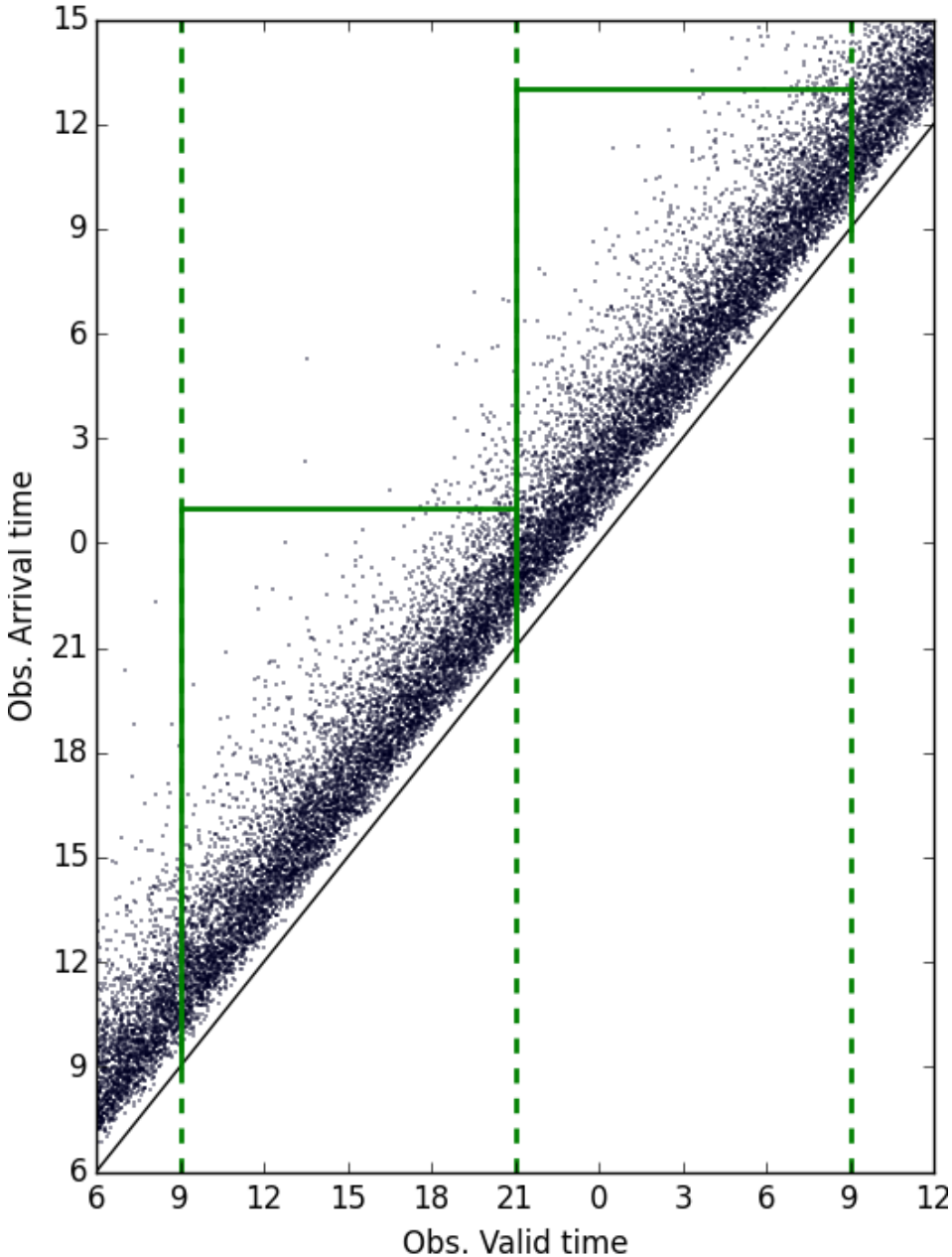
- Observations are not available instantly
- Operational systems work with a long-ish cut-off
- Issue: not all observations that have arrived are used!

Operational System: Observations Delay



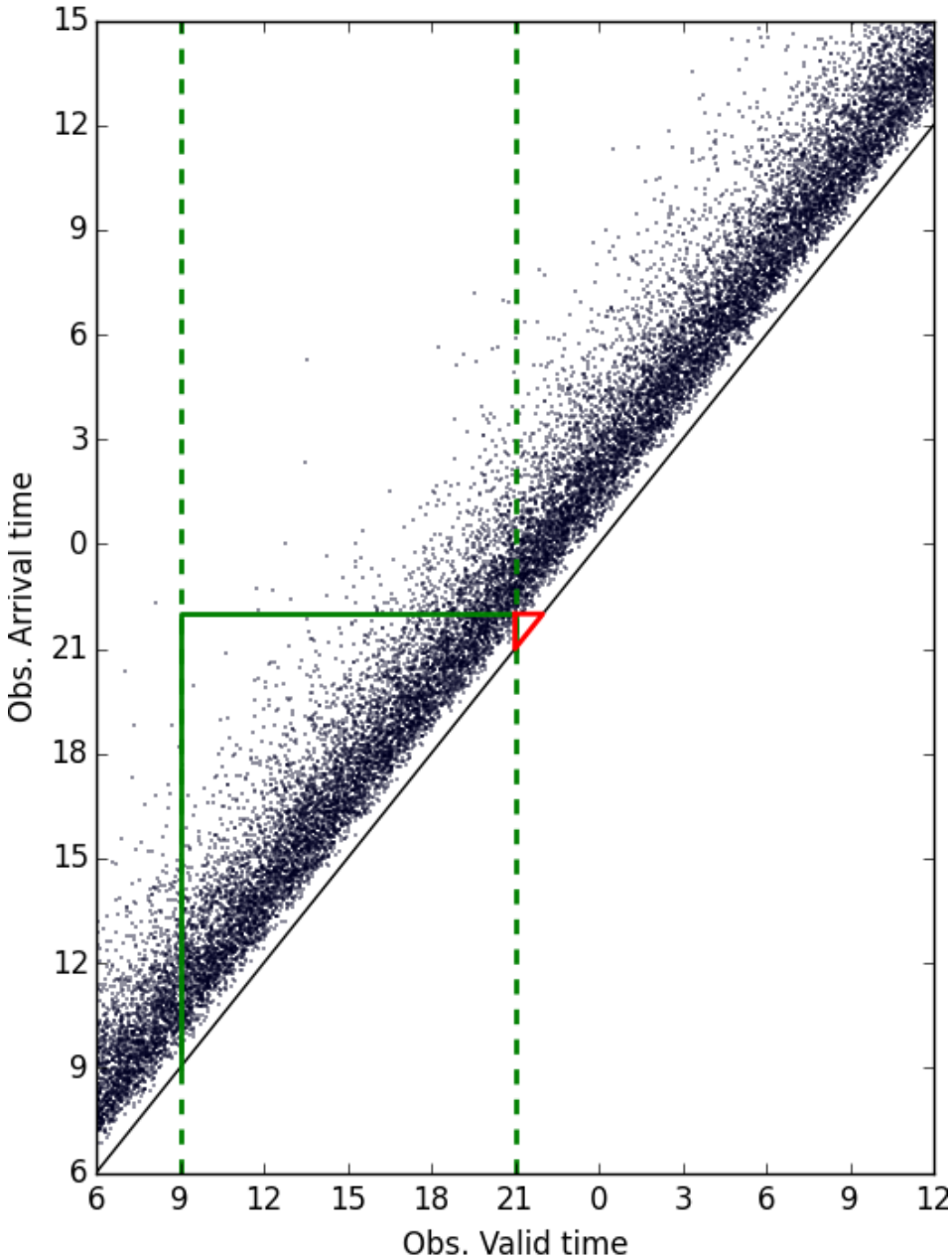
- Observations are not available instantly
- Operational systems work with a long-ish cut-off
- Issue: not all observations that have arrived are used!
- An *early-delivery* assimilation is run

Operational System: Observations Delay



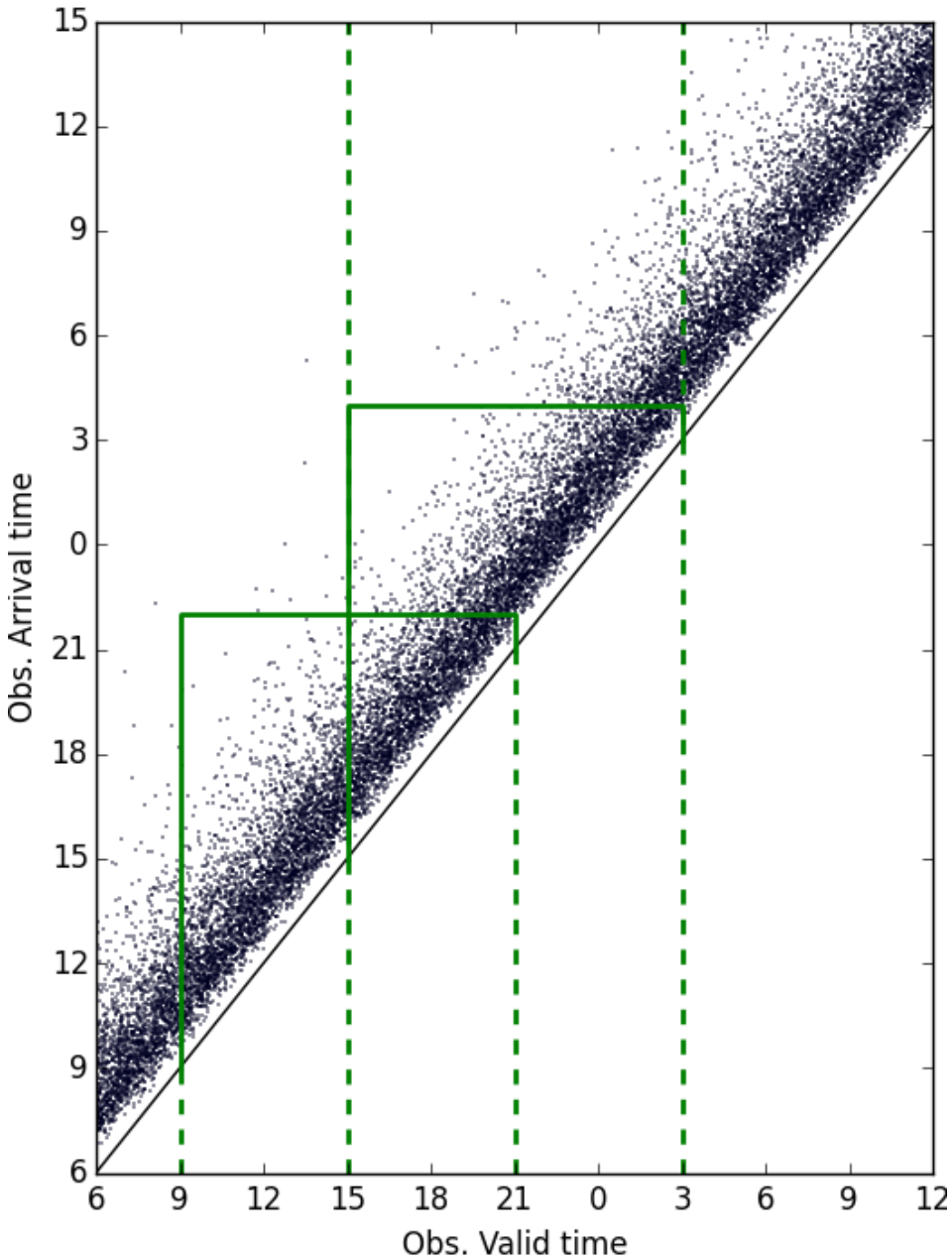
- Observations are not available instantly
- Operational systems work with a long-ish cut-off
- Issue: not all observations that have arrived are used!
- An *early-delivery* assimilation is run
- When cycling, information from the early delivery run is lost

Overlapping System: Observations Delay



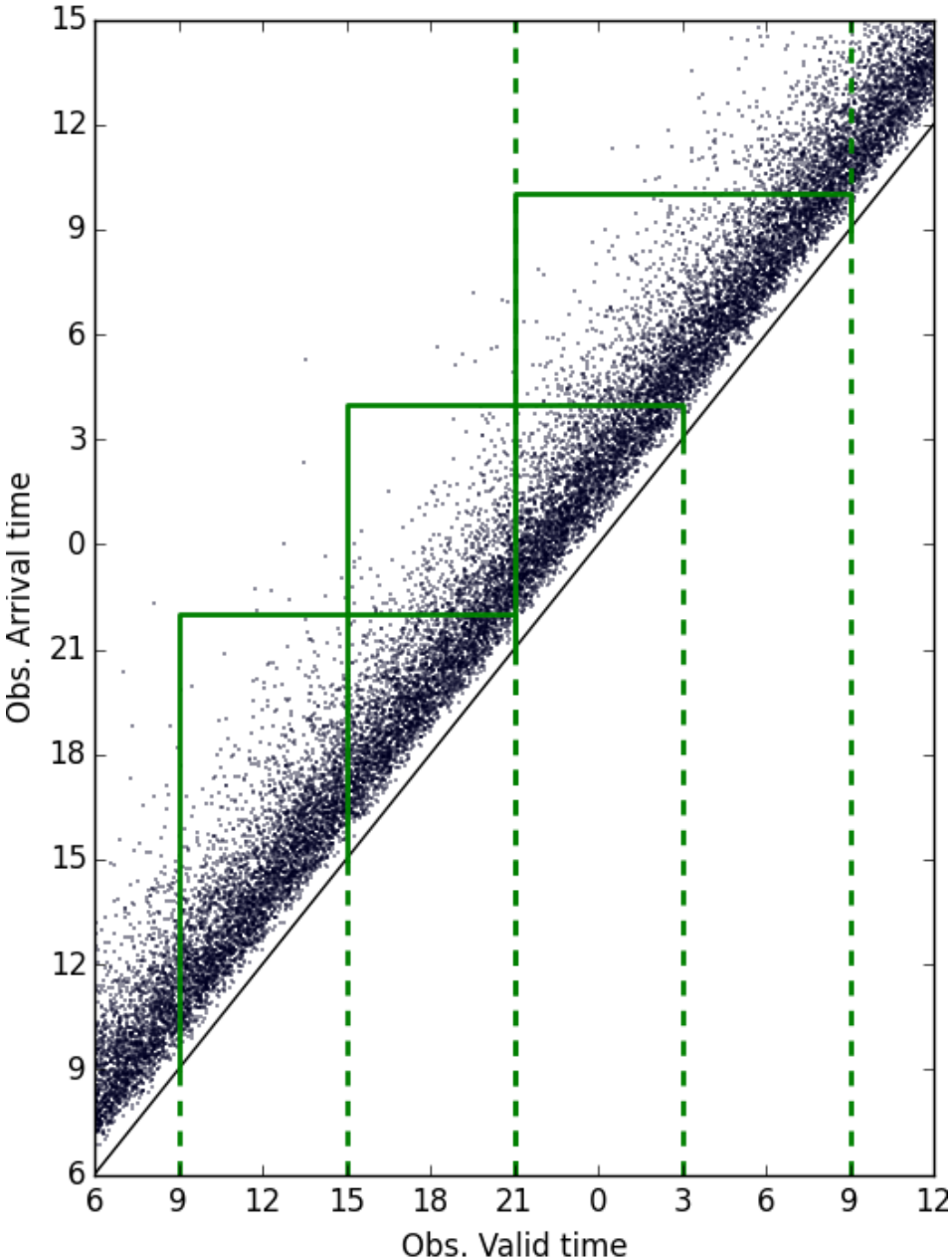
- A shorter cut-off is used

Overlapping System: Observations Delay



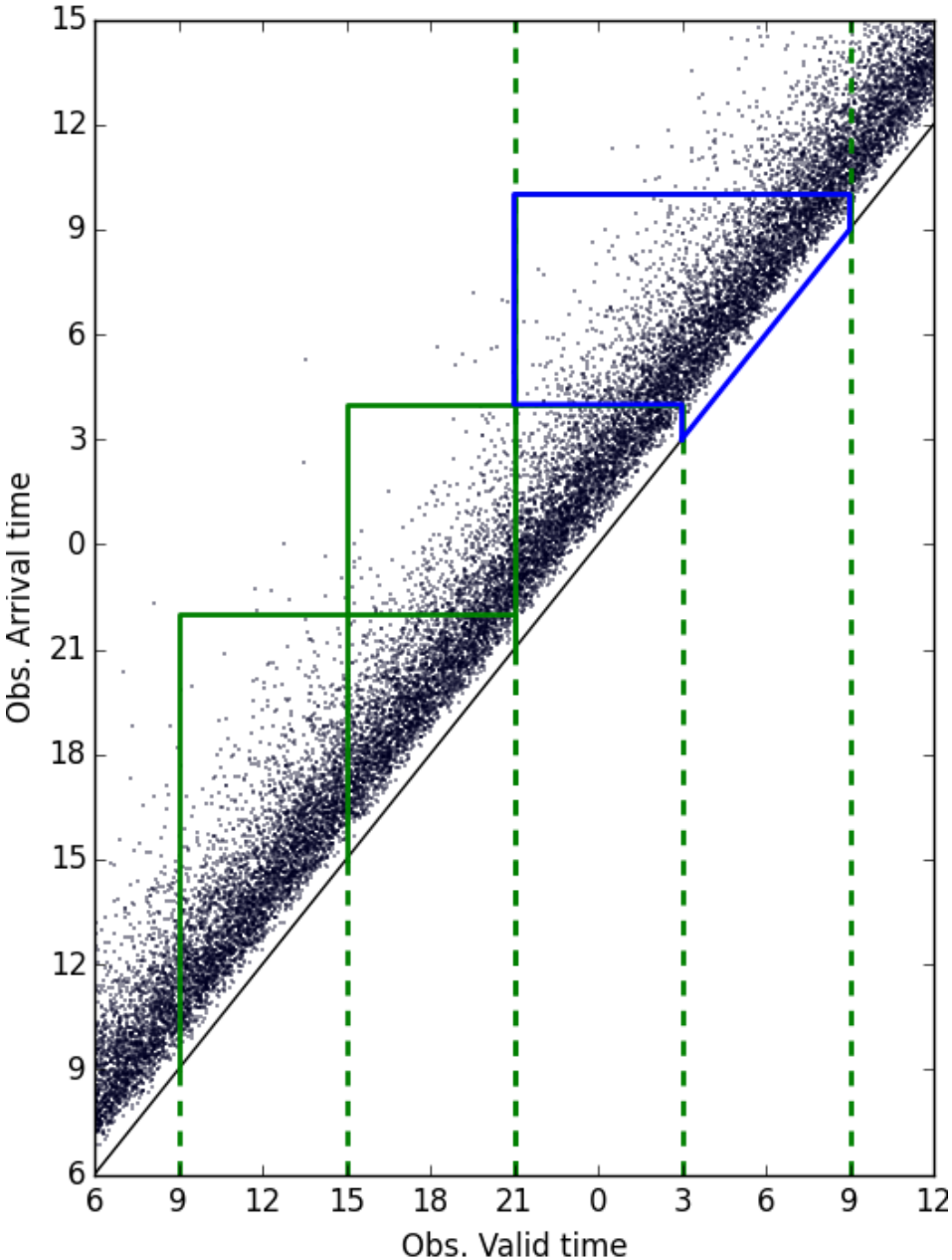
- A shorter cut-off is used
- The next overlapping cycle will use the observations that have arrived in between

Overlapping System: Observations Delay



- A shorter cut-off is used
- The next overlapping cycle will use the observations that have arrived in between
- No observations are lost
- The cost is the same: $4 \times$ 12h 4D-Var vs. $2 \times$ 12h 4D-Var $+ 4 \times$ 6h 4D-Var

Overlapping System: Observations Delay



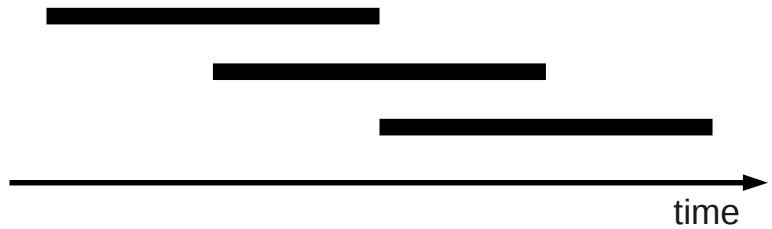
- A shorter cut-off is used
- The next overlapping cycle will use the observations that have arrived in between
- No observations are lost
- The cost is the same: $4 \times$ 12h 4D-Var vs. $2 \times$ 12h 4D-Var $+ 4 \times$ 6h 4D-Var
- It is possible to use newly arrived observations only
- This algorithm is a *quasi*-smoother

Future use of long windows

- The assimilation window can be moved in time frequently, for example by 15 minutes every 15 minutes.
- Observations that have arrived in the last 15 minutes are added.
  - Very short cut-off time, without losing observations.
- Because of the large overlap, few inner iterations are needed at each stage
  - but many effective outer iterations.
- The assimilation becomes a service that runs continuously:
  - an up-to-date analysis is always available,
  - less pressure on the time critical path,
  - can be optimised for energy consumption rather than time to solution,
  - no peak in daily computer usage (smaller machine).

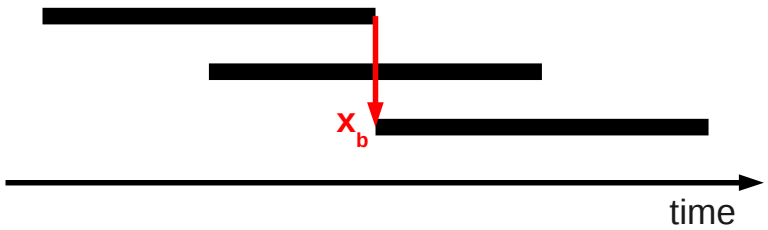
Overlapping Windows

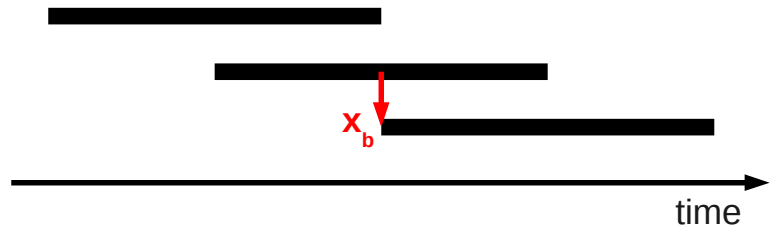
- The assimilation needs updating several times daily: long assimilation windows will overlap.



Overlapping Windows

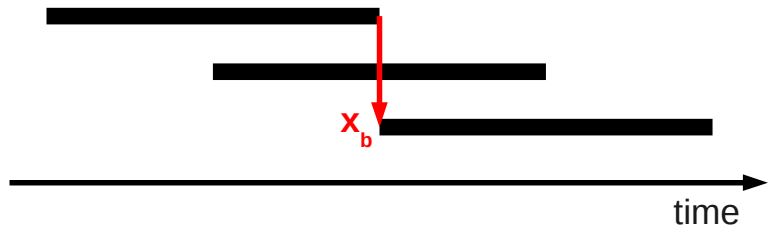
- The assimilation needs updating several times daily: long assimilation windows will overlap.



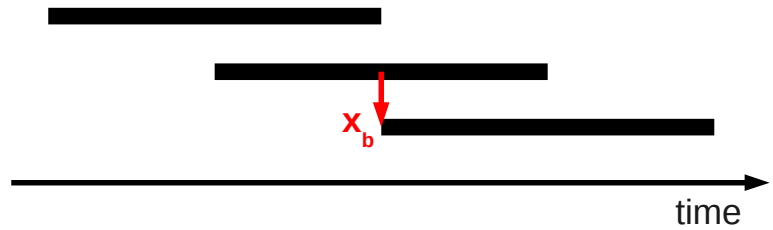


Overlapping Windows

- The assimilation needs updating several times daily: long assimilation windows will overlap.



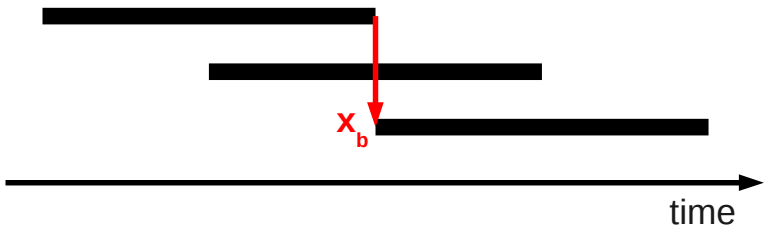
  - The background and observation errors are uncorrelated
  - The background is not the most up to date
  - Two DA streams running independently



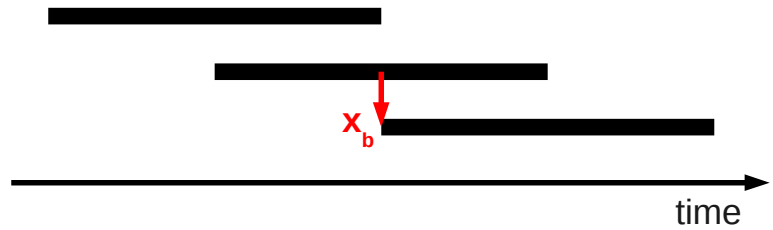
  - The background is the most accurate
  - The background and observation errors are correlated

Overlapping Windows

- The assimilation needs updating several times daily: long assimilation windows will overlap.



  - The background and observation errors are uncorrelated
  - The background is not the most up to date
  - Two DA streams running independently



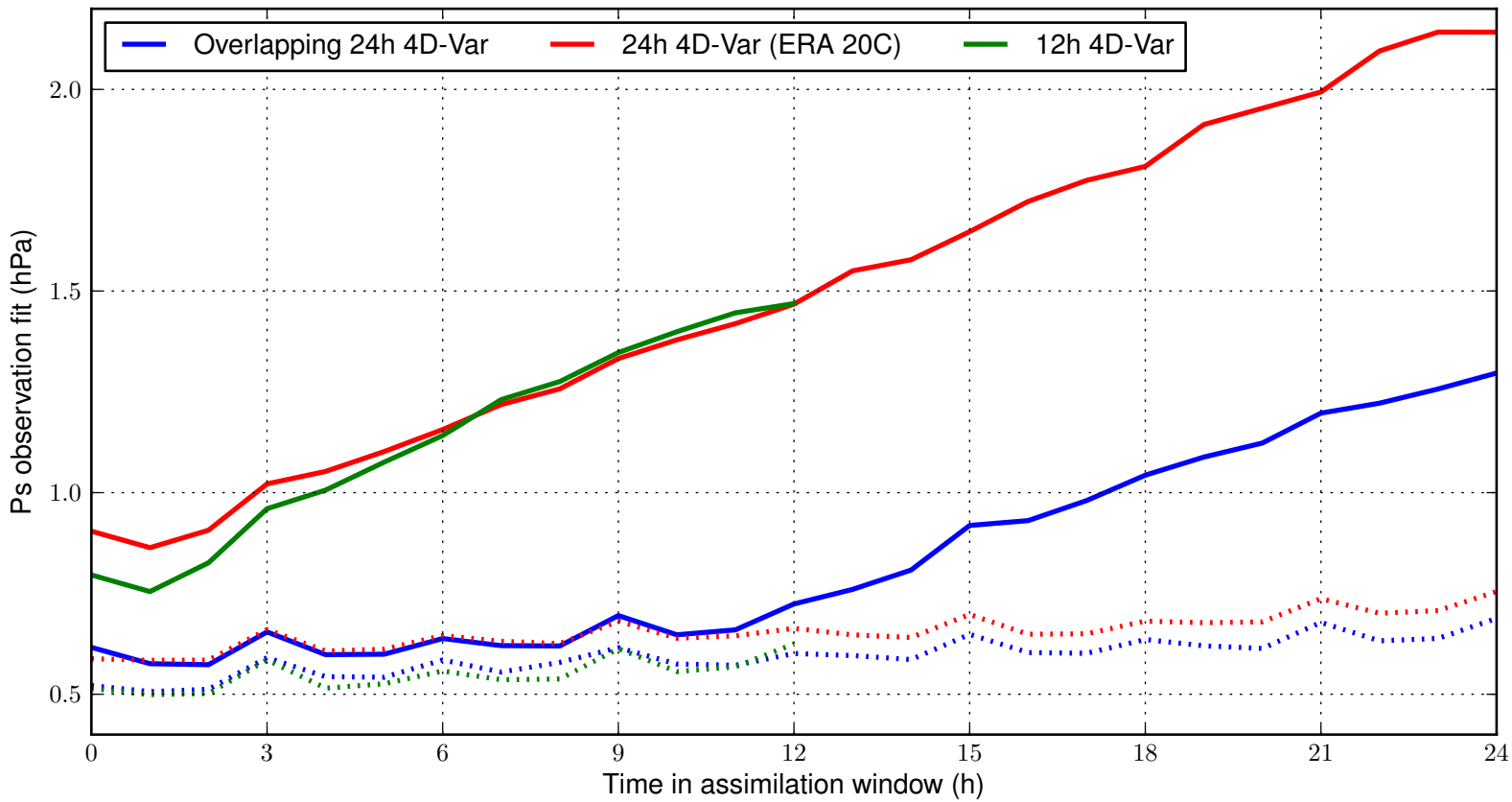
  - The background is the most accurate
  - The background and observation errors are correlated

- In practice, using the most recent background gives the best results.

Ps-only Re-analysis

Background and Analysis fit to Observations throughout the assimilation window

2004-07-01 to 2005-04-09

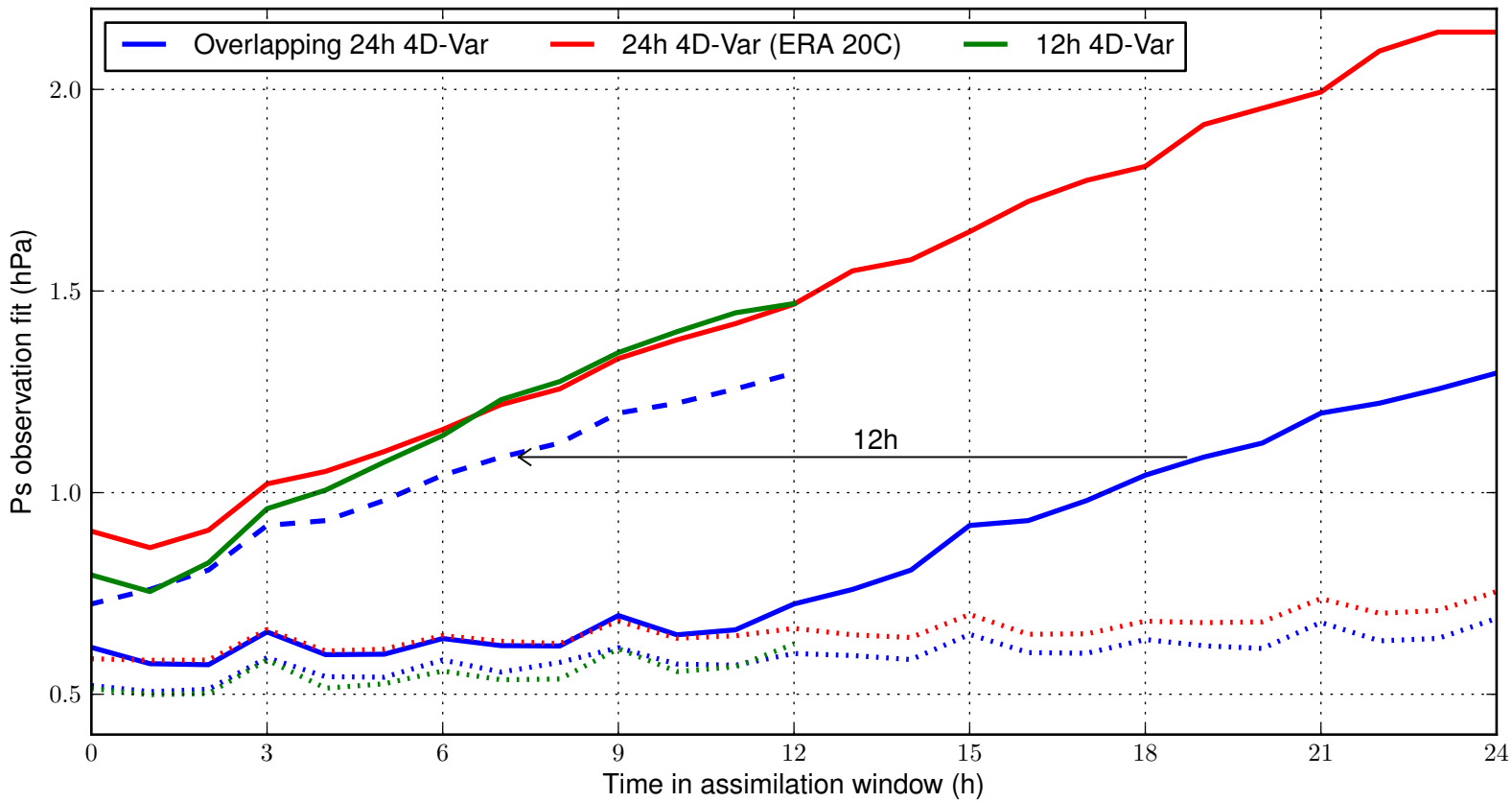


Paul Poli

Ps-only Re-analysis

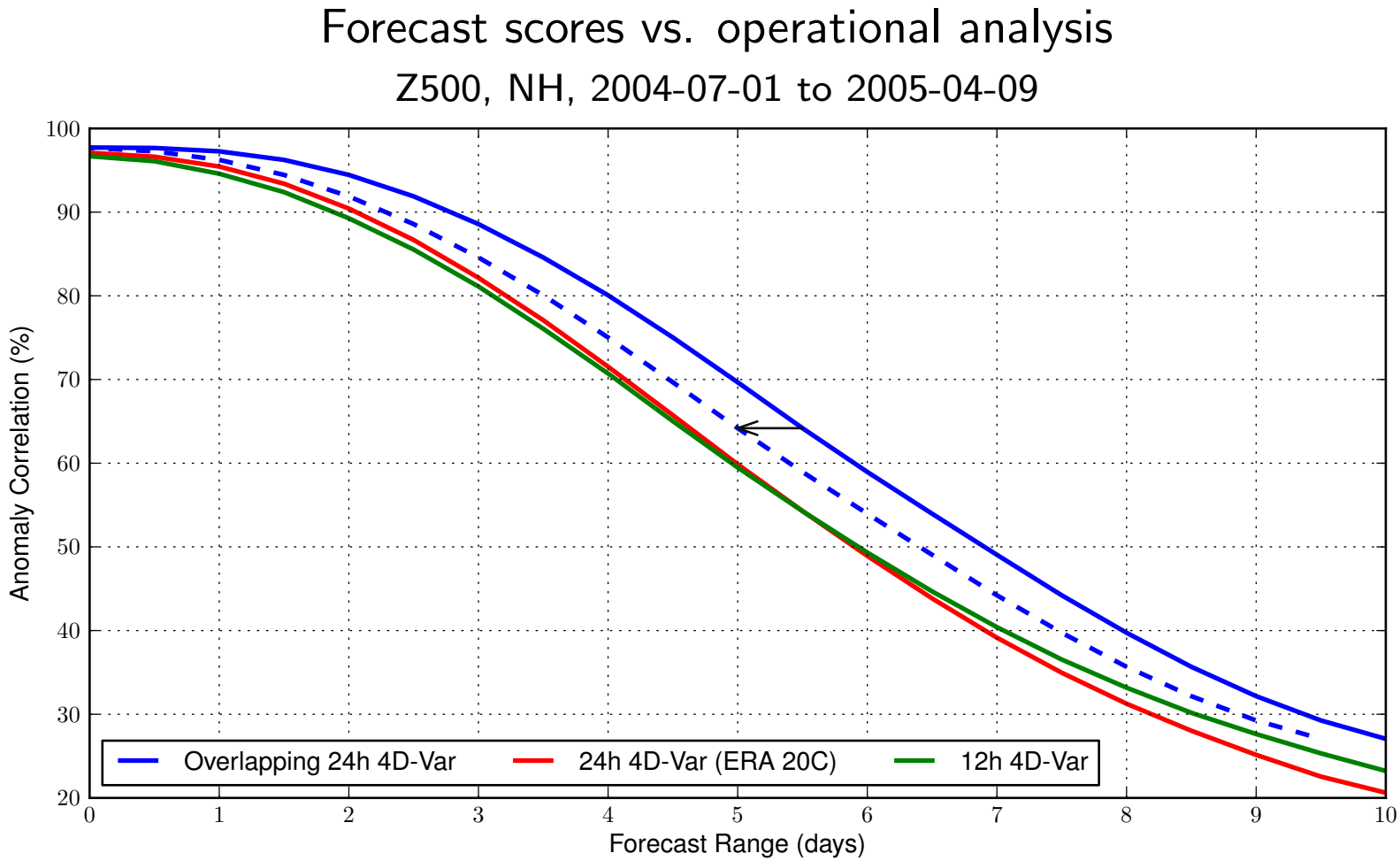
Background and Analysis fit to Observations throughout the assimilation window

2004-07-01 to 2005-04-09



Paul Poli

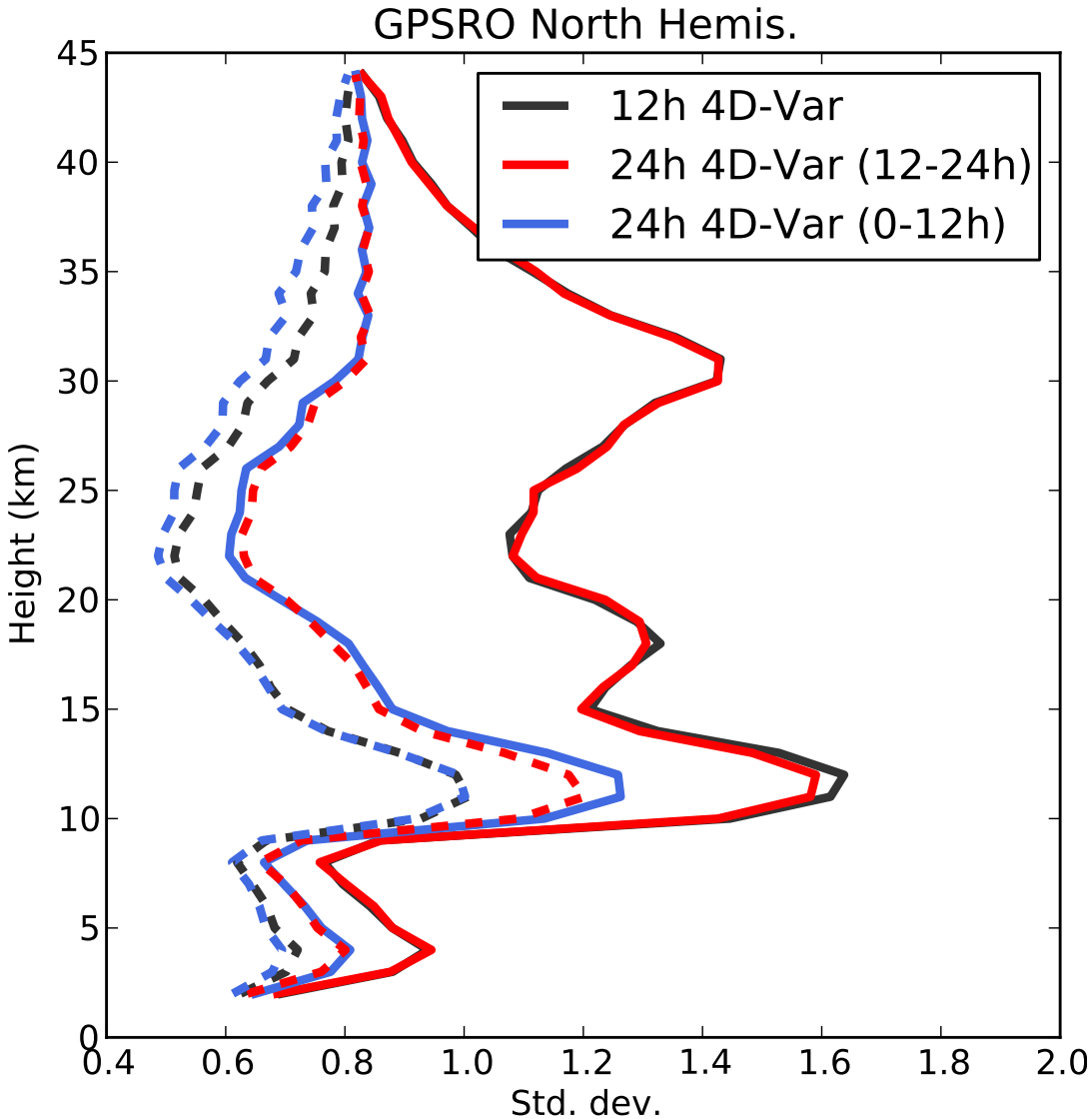
Ps-only Re-analysis



- Verification against independent (unused) observations:
  - confirms positive results with overlapping windows,
  - shows that 24h 4D-Var without overlap is slightly better than 12h 4D-Var.

Use of Observations (full system)

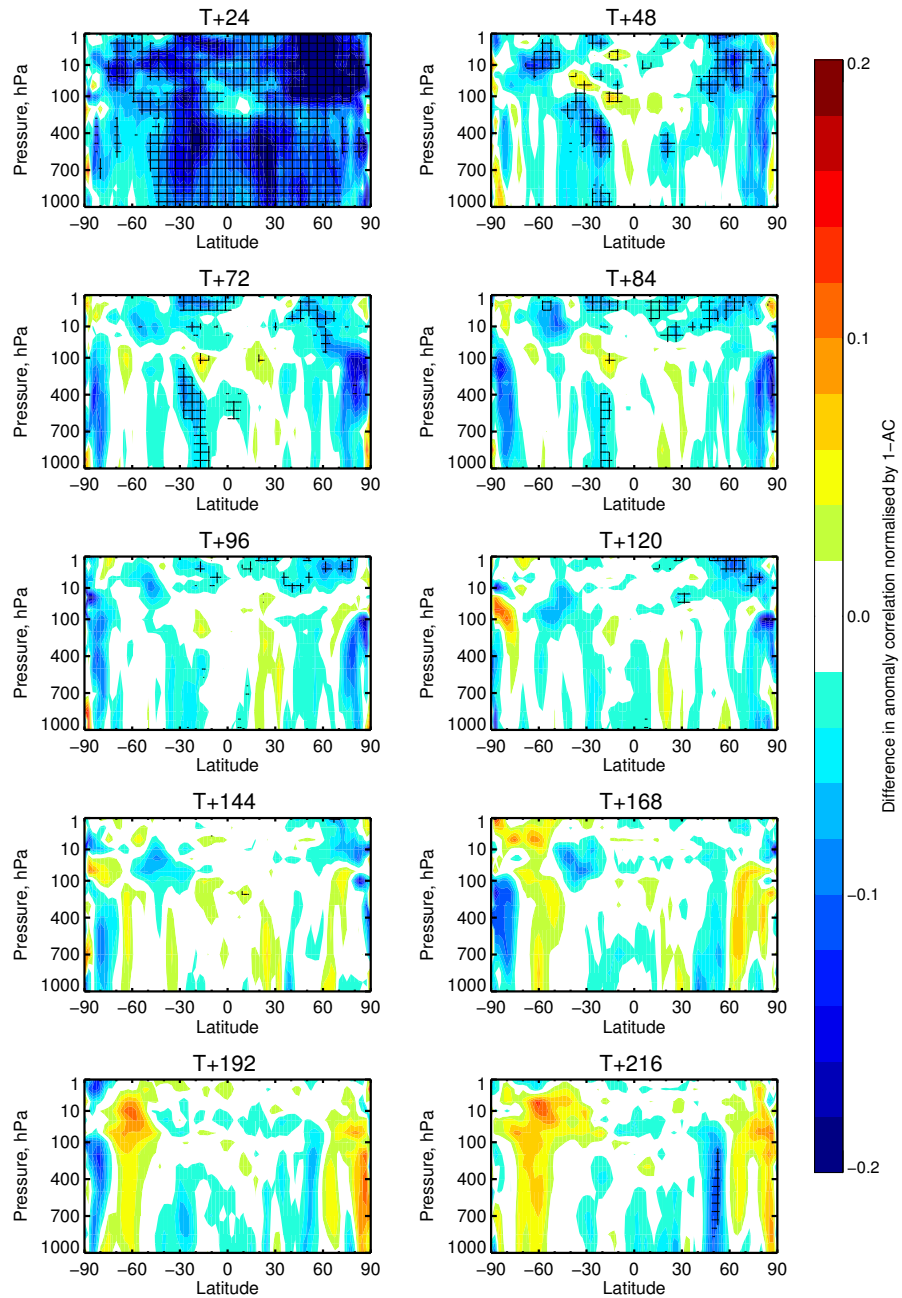
- Overlaping assimilation windows implies that observations are used twice.



- The background fit the first time observations are seen is similar to the 12h 4D-Var control
- The analysis fit is equal or better than in 12h 4D-Var the second time observations are used
- It is possible to fit observations at least as well even with more constraints

Forecast Performance

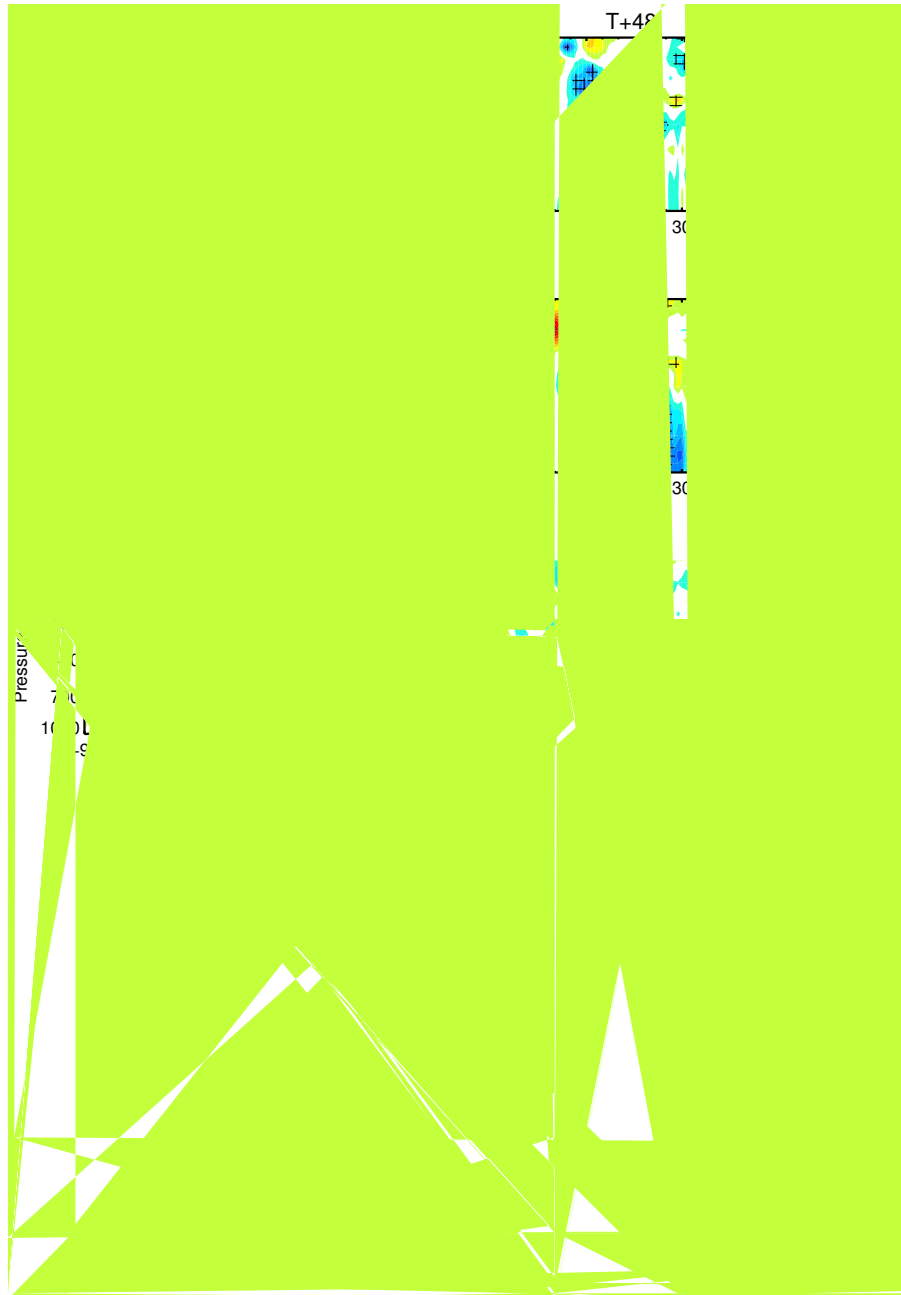
- For operational use: 12h 4D-Var every 6 hours
- Flow dependent **B** from 6-hour EDA
- In practice, re-using observations gives better performance
- Not re-using observations would require replacing **B** by an analysis error covariance matrix



New obs. only - obs. re-use, JJA 2014

Forecast Performance

- For operational use: 12h 4D-Var every 6 hours
- Flow dependent **B** from 6-hour EDA
- Without any change in **B**, comparison between overlapping windows and early delivery is mixed
- **B** will be re-tuned for the overlapping system



Overlap - Early delivery, JJA 2014

Background-Observations Errors Correlation

- Results improve when re-using observations
- Other errors in **B** and **R** might be (much) larger than the missing cross-correlation term
- Observation errors could be adjusted (Cf. talk Marc Bocquet)?
- The state used as the background is in fact already an analysis
  - **B** should be replaced by an analysis error covariance matrix (the EDA can be modified for this)
  - Keeping both **B** and the observations anchors the guess where it has already been pushed: $(J_b + J_o)_{min}$ acts as a J_a weight
  - Keeping **B** but removing the observations at the start of the window creates an information leak (not enough weight on an accurate prior)
- In the long term, with weak constraint 4D-Var, there is no **B**, only **Q**
  - No issue of cross-correlation
  - Viewing **B** as an approximation of **Q**, ignoring cross correlations is sensible