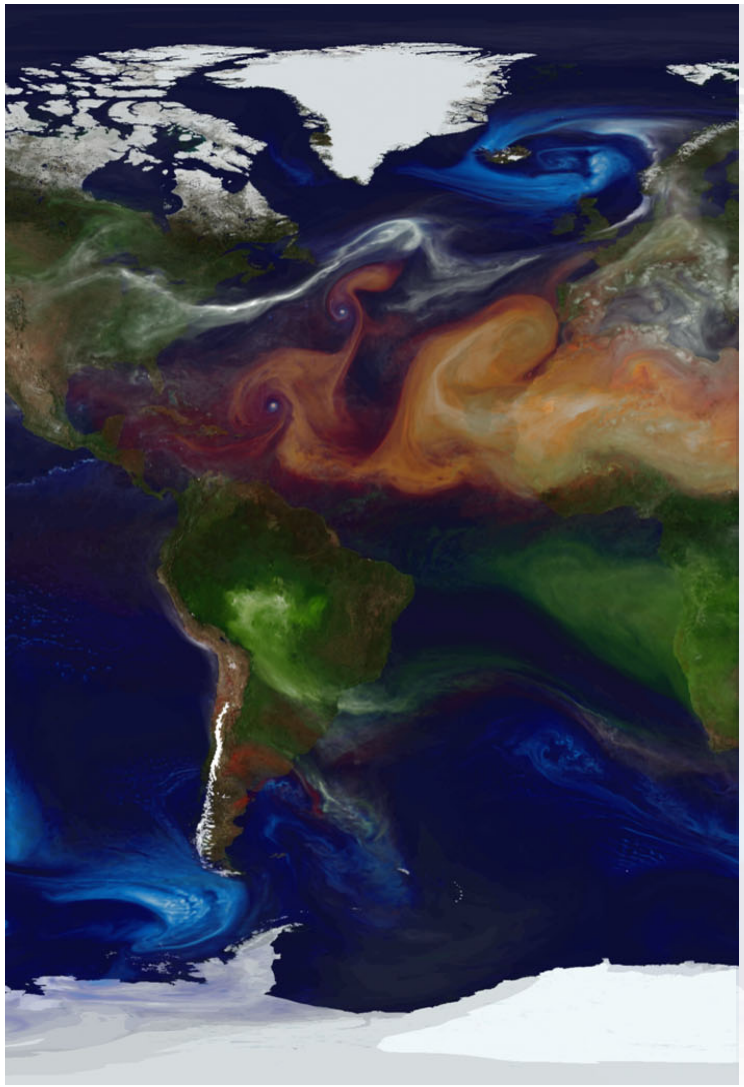


Understanding the role of analysis error in the convergence of reanalysis production streams. A MERRA-2 case study

Amal El Akkraoui[a,b]

R. Errico[a,c], N. Prive[a,c], D. Carvalho, M. Bosilovich[a]

[a]GMAO, Earth Sciences Division, Sciences and Exploration Directorate, NASA GSFC.

[b]Science Systems and Applications, Inc.

[c]Goddard Earth Sciences Technology and Research Center, Universities Space Research Association (GESTAR)

ECMWF/EUMETSAT NWP SAF Workshop on the treatment of random and systematic errors in satellite data assimilation for NWP. November 2020

Reanalysis production streams

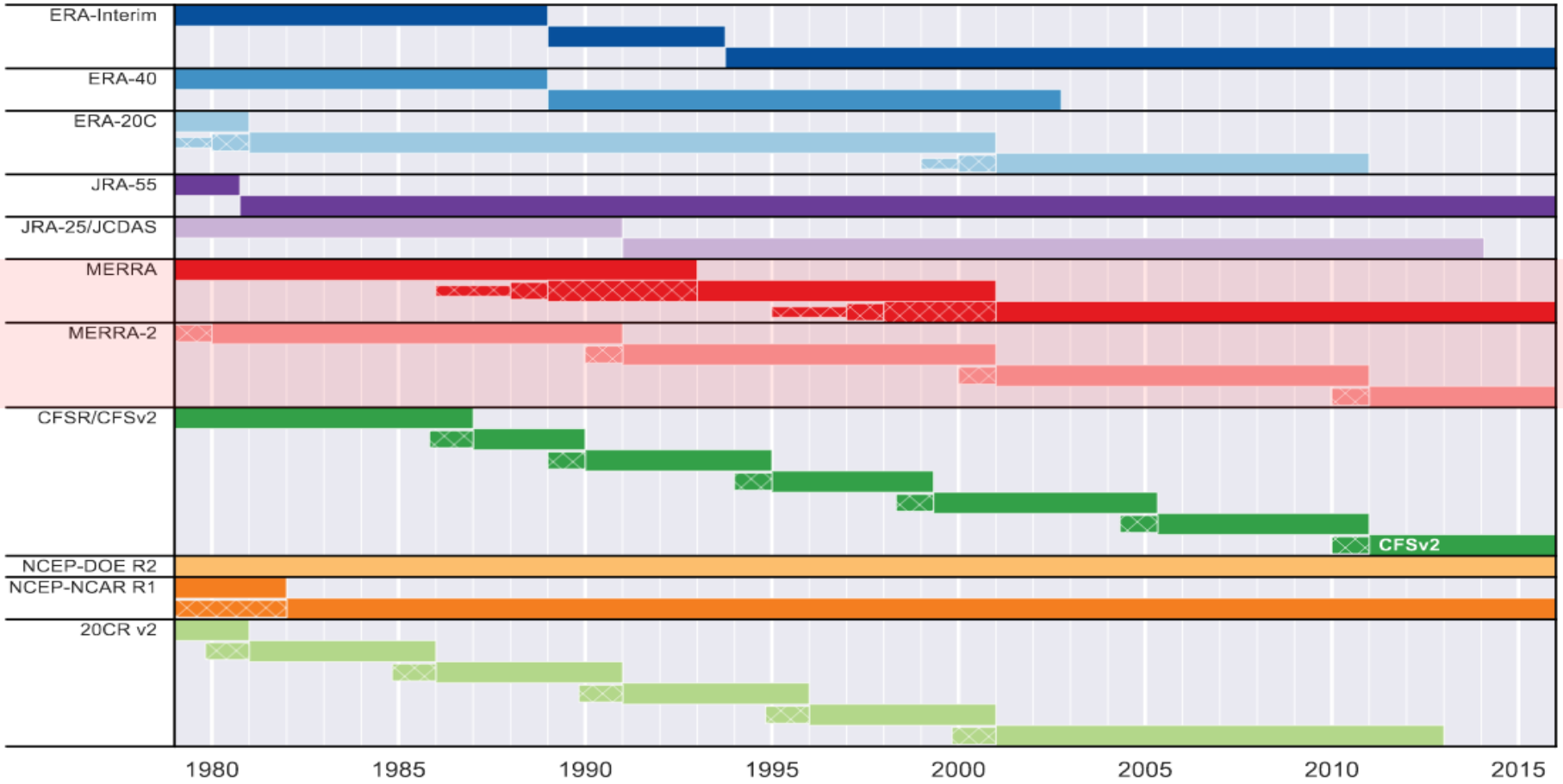
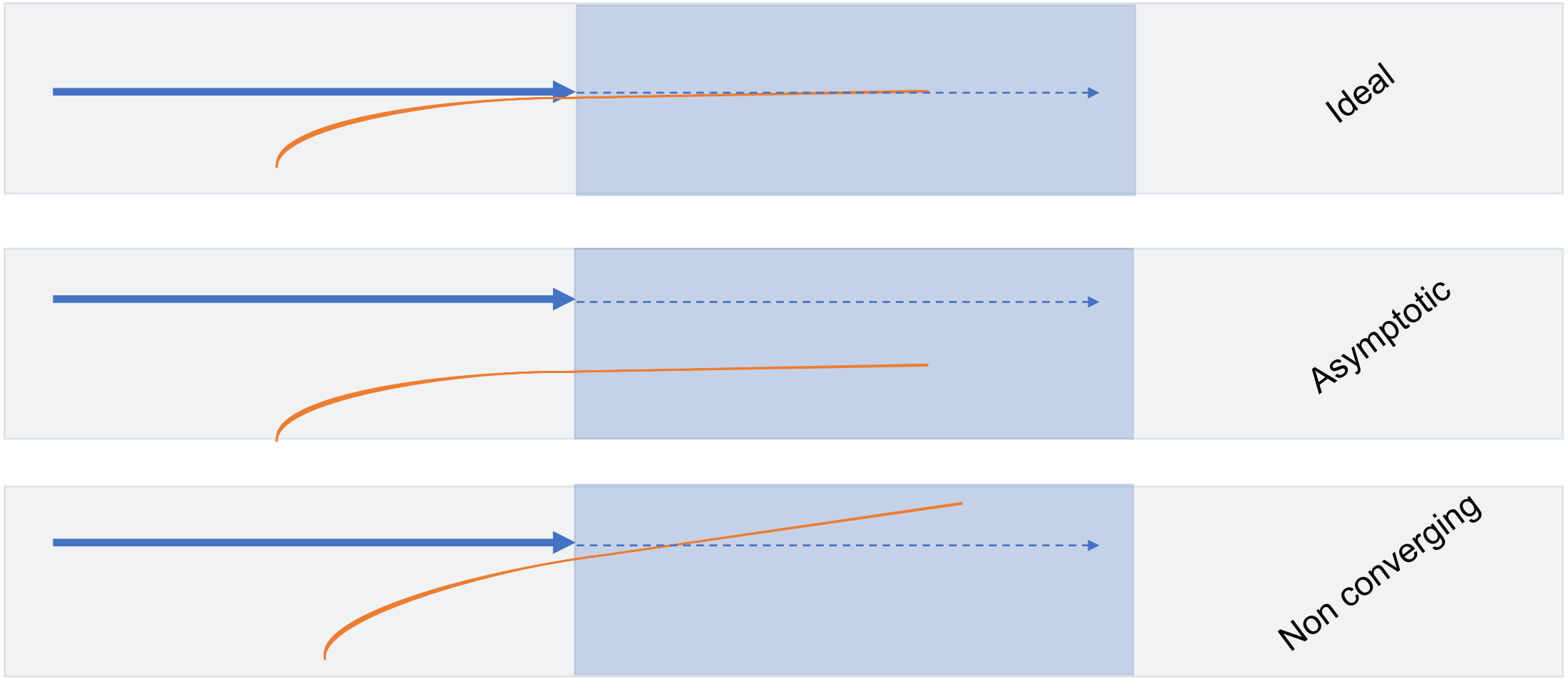


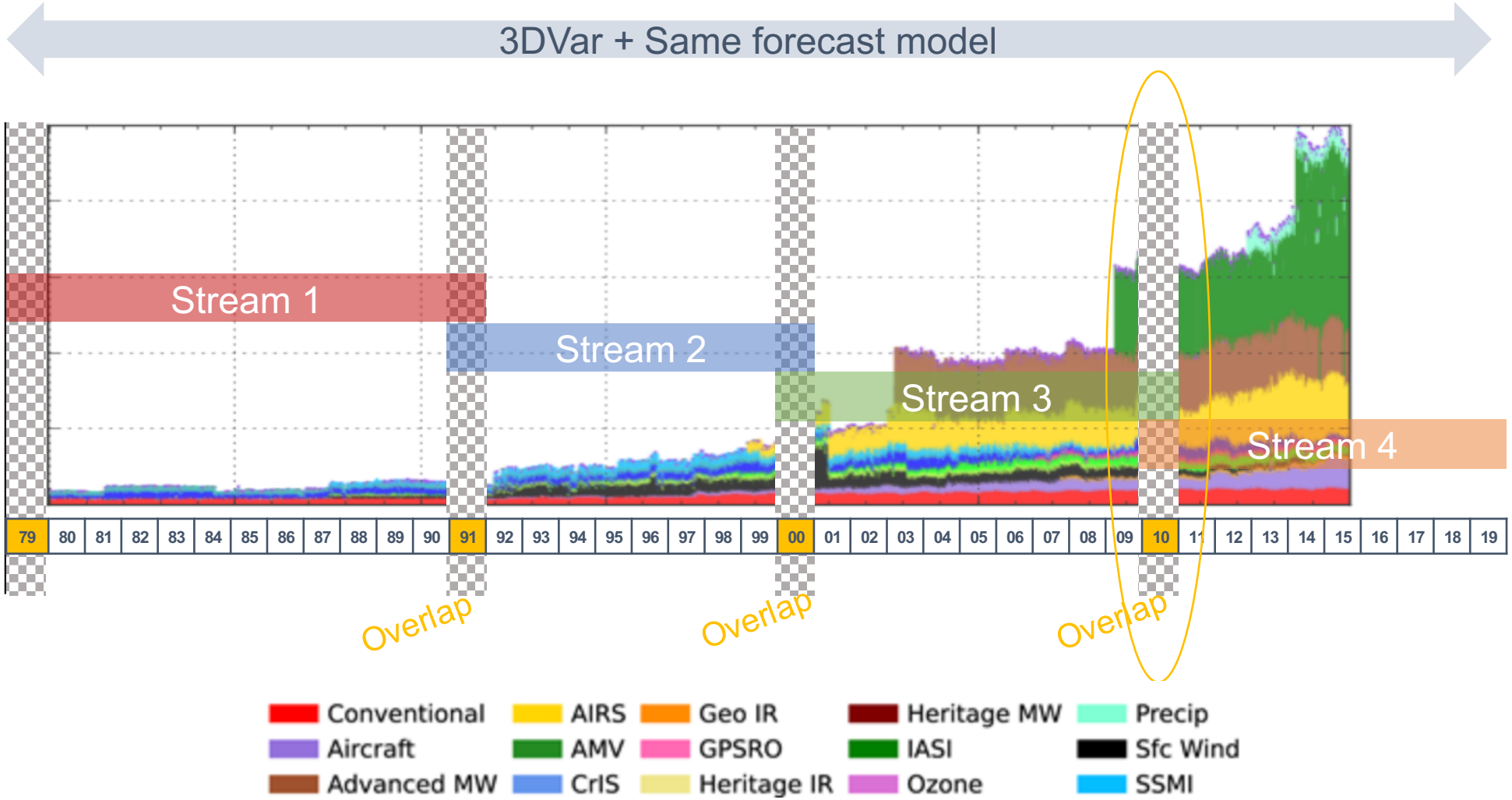
Figure 2. Summary of the execution streams of the reanalysis systems from January 1979 through December 2015. The narrowest cross-hatched sections indicate known spin-up periods, while the wider cross-hatched sections indicate overlap periods.

Fujiwara et al. 2017

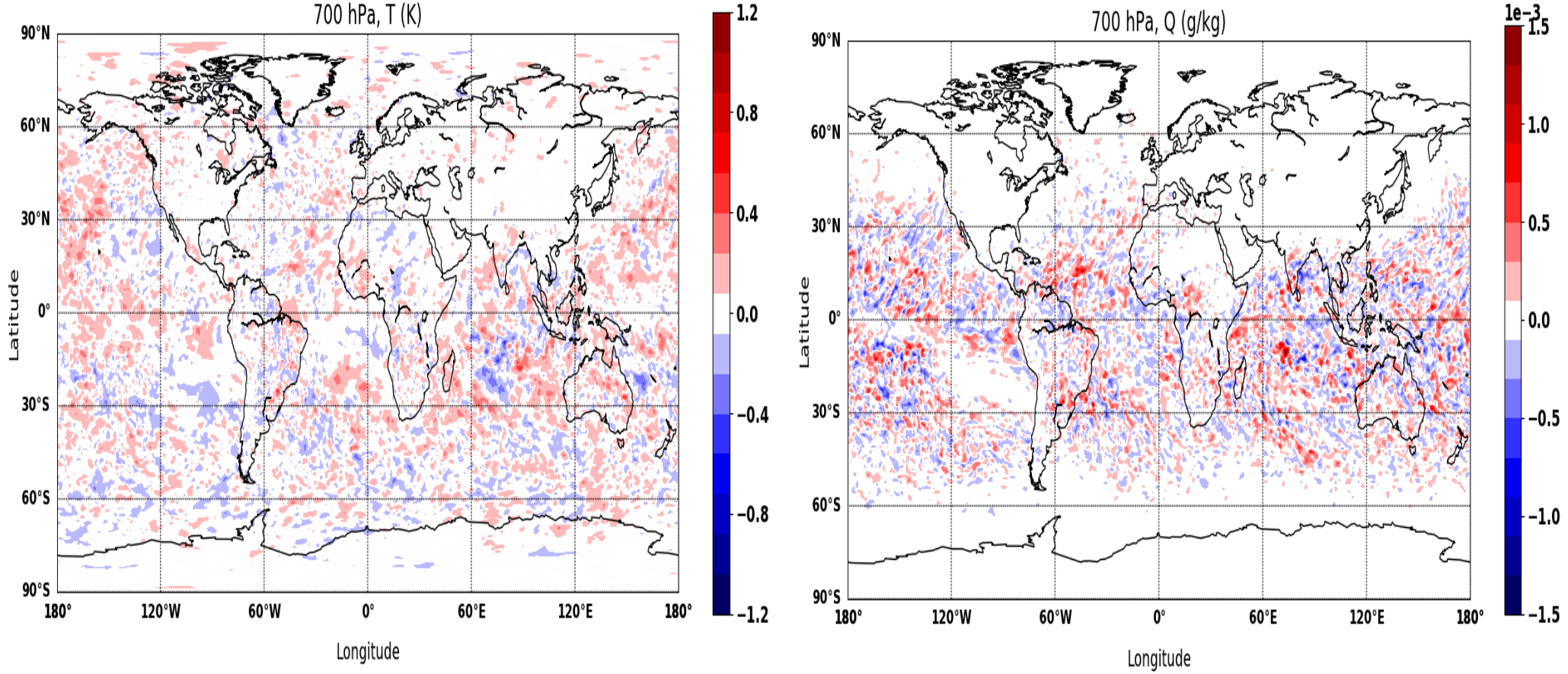
Convergence of overlapping streams



MERRA-2 Production streams

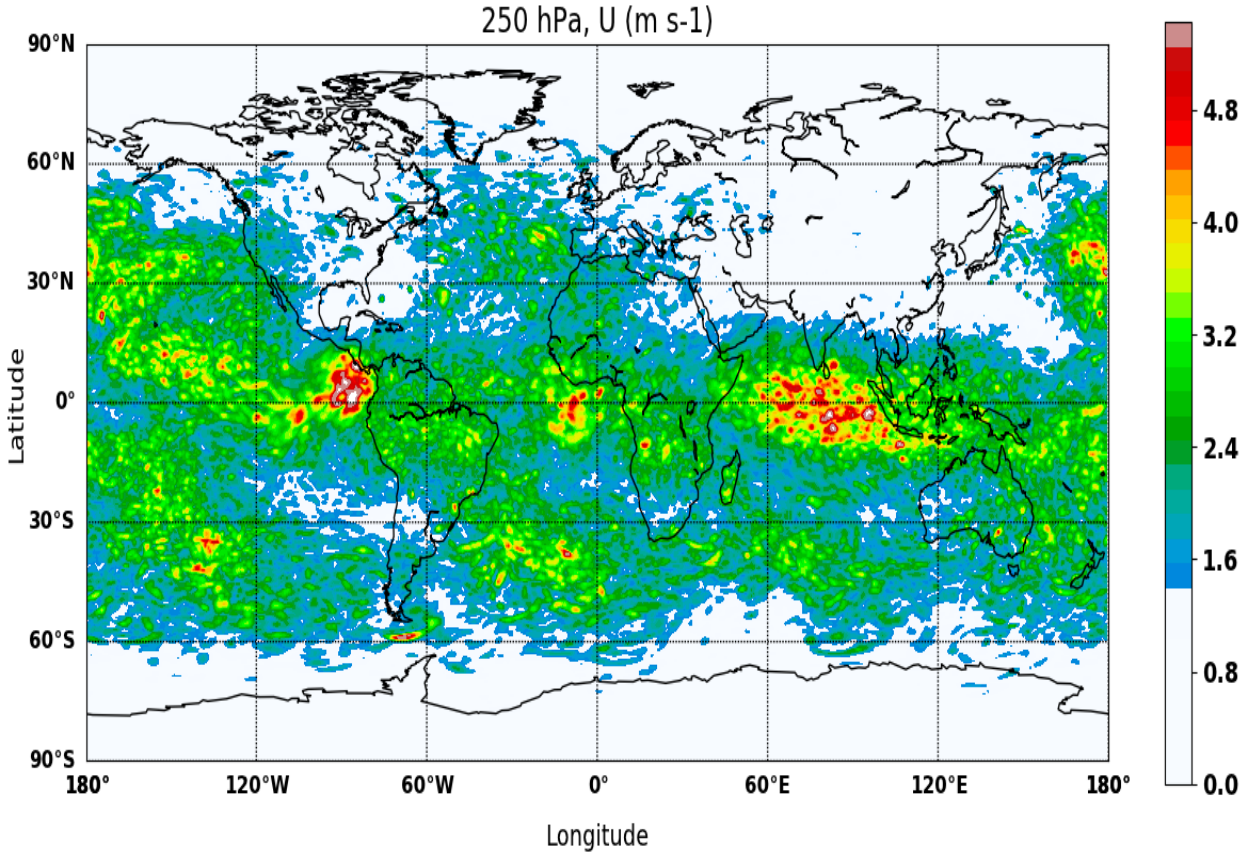


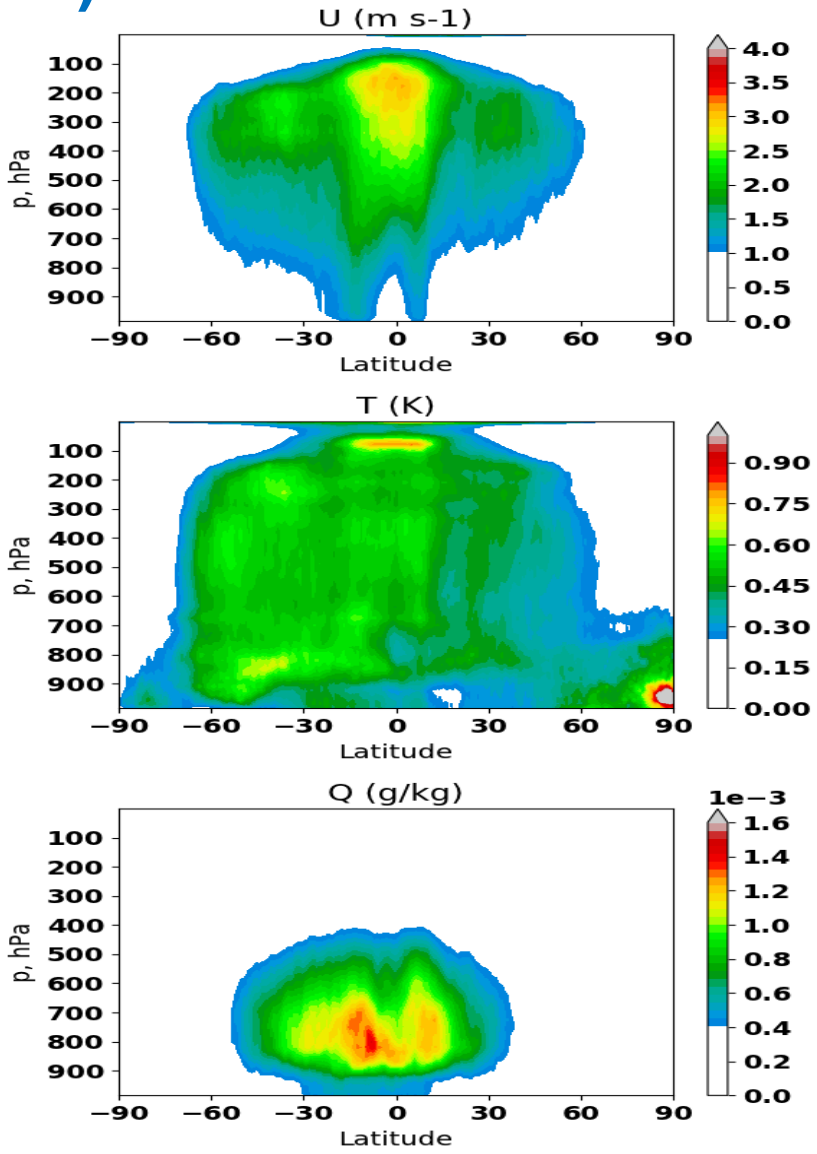
Stream differences (Dec 2010)



Monthly Mean of the differences between streams 3 and 4 for December 2010, on model level 700 hPa for Temperature (left) and specific humidity (right)

Stream differences (Dec 2010)





Standard deviation of the differences between streams 3 and 4 for December 2010. Zonal wind at 250 hPa (left). Square roots of the zonal means of the temporal variance of the differences for zonal wind (top right), Temperature (middle right), and specific humidity (bottom right).

- Two overlapping streams ≈ two solutions to the same data assimilation problem (**same cost function**) since the observations and error statistics are the same, only the background state is evolving differently.
- The error with respect to the truth in each case is **one realization of analysis error** (e).
- The realizations can be further considered **equally probable**, yet correlated events.
- The squared mean of their respective errors is then indistinguishable from the squared mean of the analysis error($S(\mathbf{e}_1)$=$S(\mathbf{e}_2)$=$S(\mathbf{e})$).

$$S(\boldsymbol{x}_1 - \boldsymbol{x}_2) = 2S(\boldsymbol{e}) - 2\{\boldsymbol{e}_1\boldsymbol{e}_2\} = \begin{cases} 2S(e) & Uncorrelated \\ 0 & Perfectly\ correlated \\ 4S(e) & Anti-correlated \end{cases}$$

$$0 \leq S(\boldsymbol{x}_1 - \boldsymbol{x}_2) \leq 2\,S(\boldsymbol{e})$$

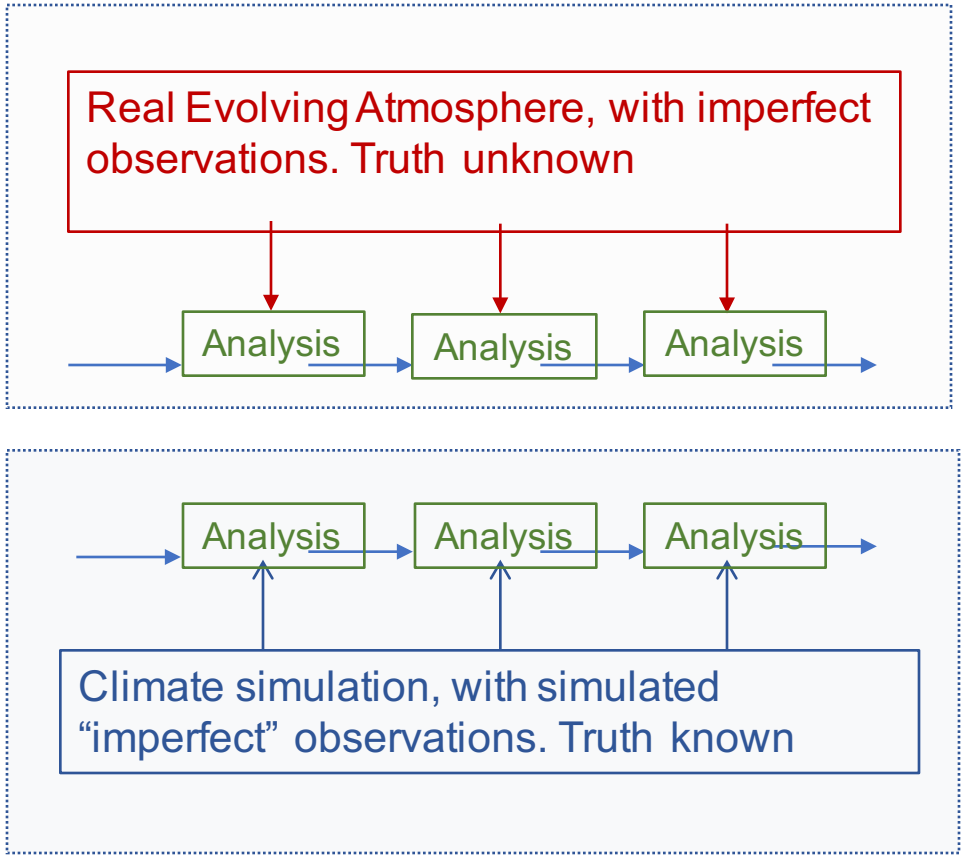
Squared mean of stream difference

Squared mean of analysis error

Observing System Simulation Experiment

An OSSE is a modeling experiment used to evaluate the impact of new observing systems on operational forecasts when actual observational data is not available.

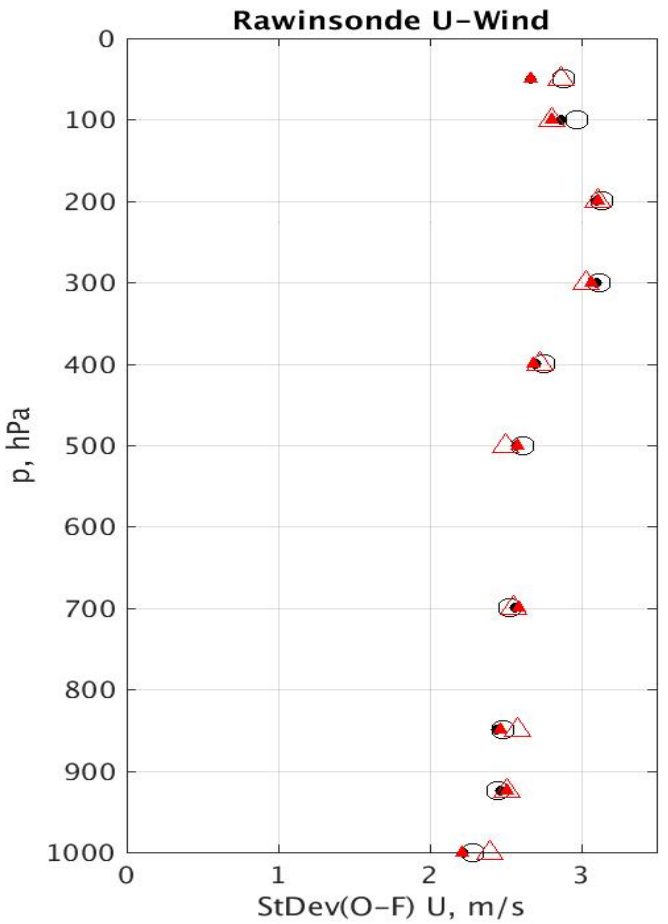
- A long free model high-resolution run is used as the "truth" - the Nature Run.
- Synthetic observations are sampled from the Nature Run fields. Suitable errors are added to mimic the behavior of the real system.
- The availability of a complete true state of the atmosphere allows the explicit calculation of some quantities not possible in the real world:
  - analysis error
  - short-term forecast error
  - efficacy of the data assimilation system

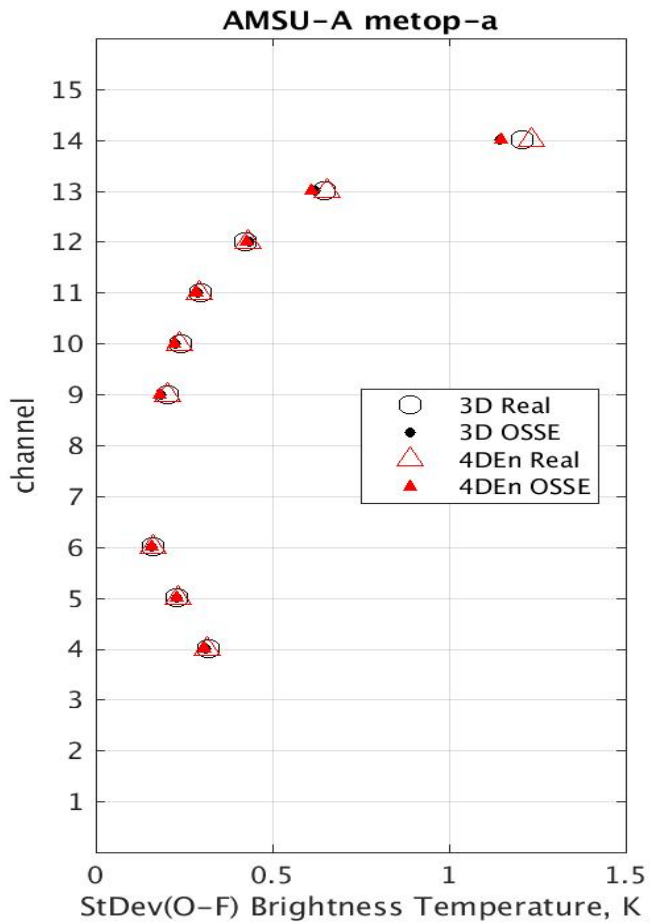


Courtesy of Ron Errico and Nikki Prive

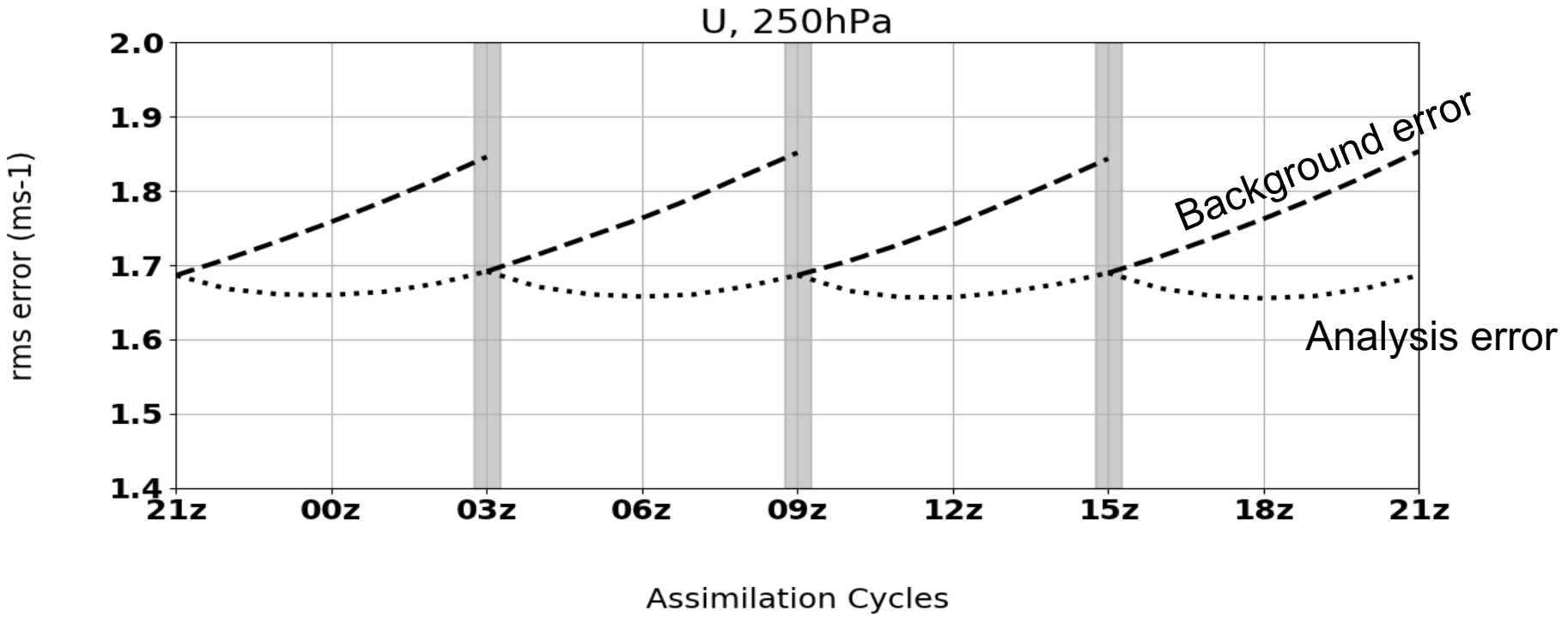
Observing System Simulation Experiment

One caveat:
specific findings are applicable within the OSSE, but extensive validation is needed before extrapolating results to the real world.





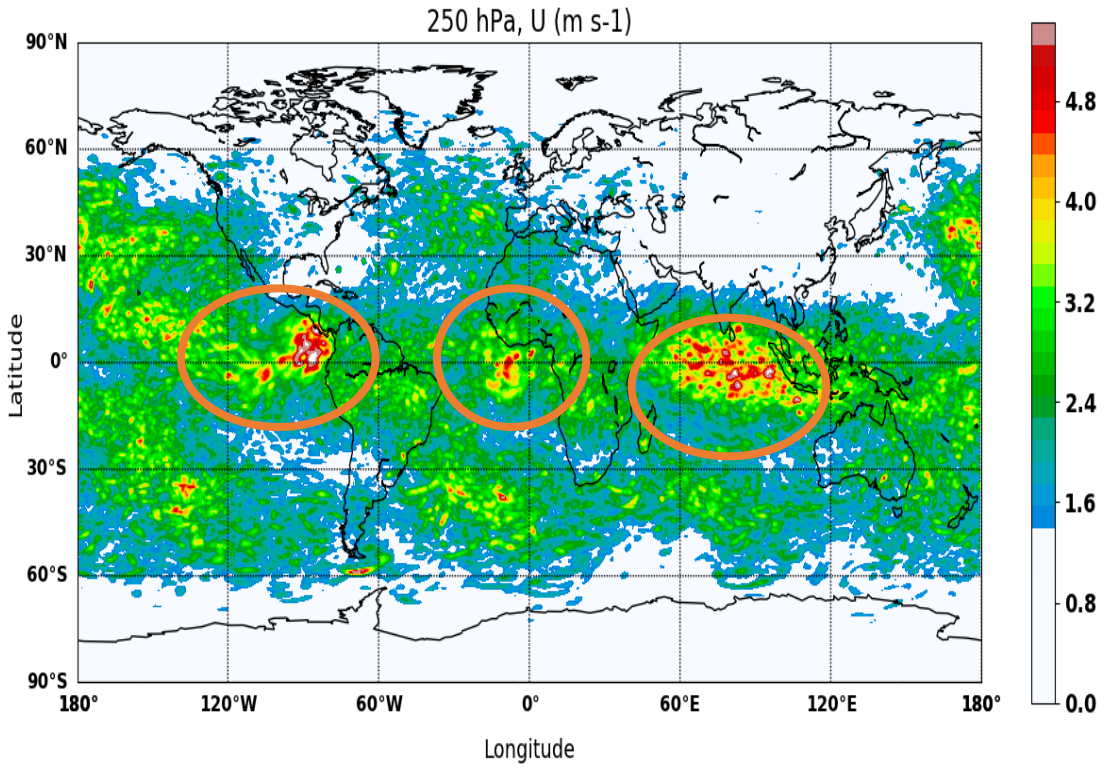
Analysis Error and Data assimilation Cycle



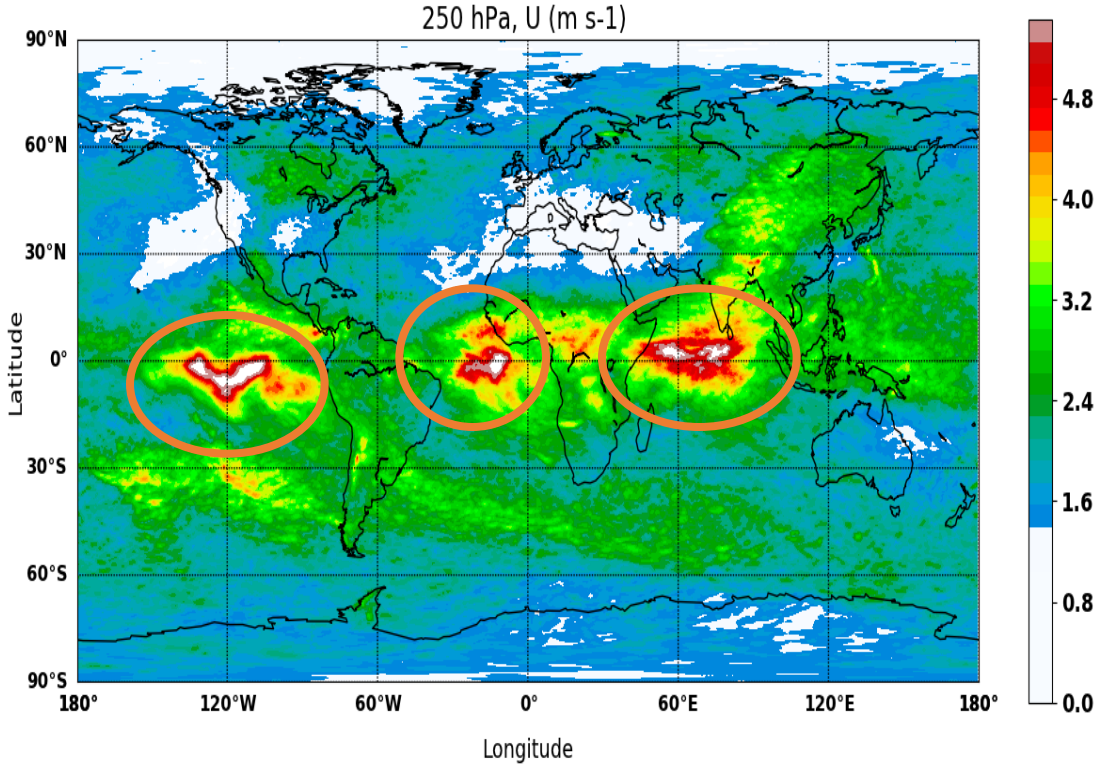
Hourly estimates of analysis and background error during a cycling OSSE
(statistics computed over a two-week cycling period)

Stream differences (Dec 2010)

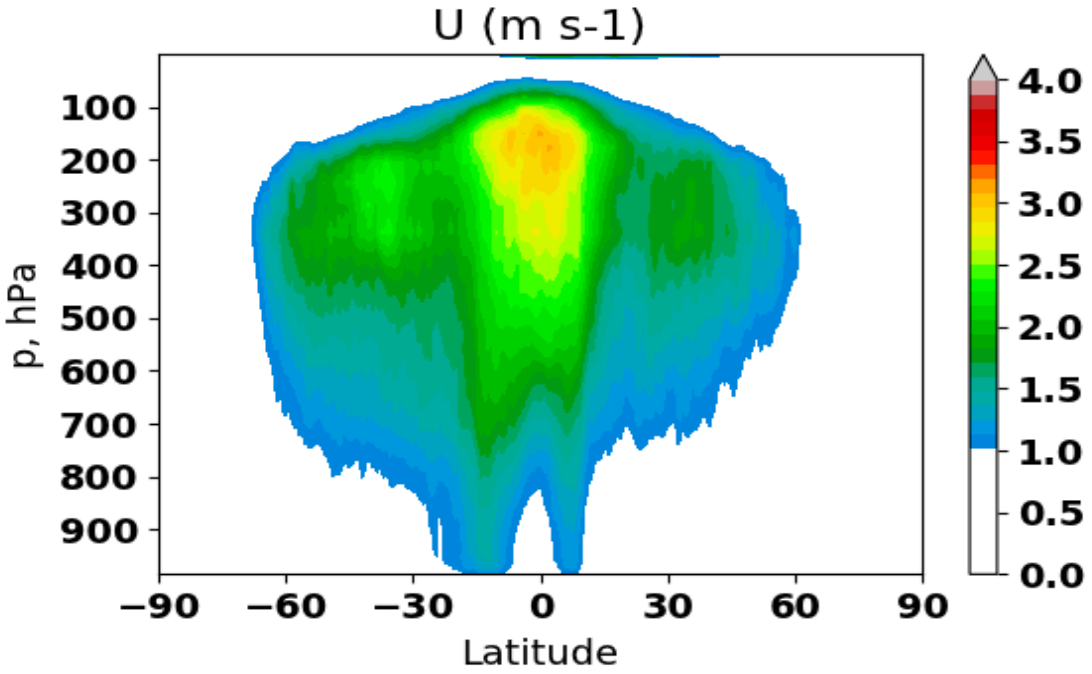
Std Stream 3&4 difference
Dec, 2010



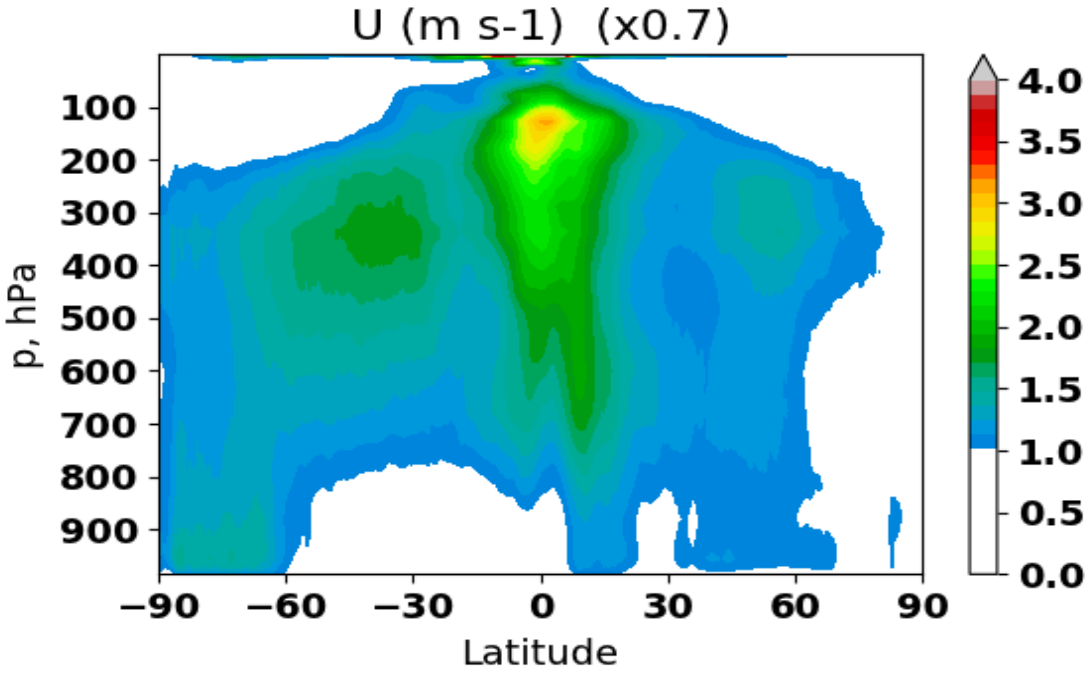
Std Analysis error
estimate from GMAO
3DVar OSSE



Stream differences (Dec 2010)

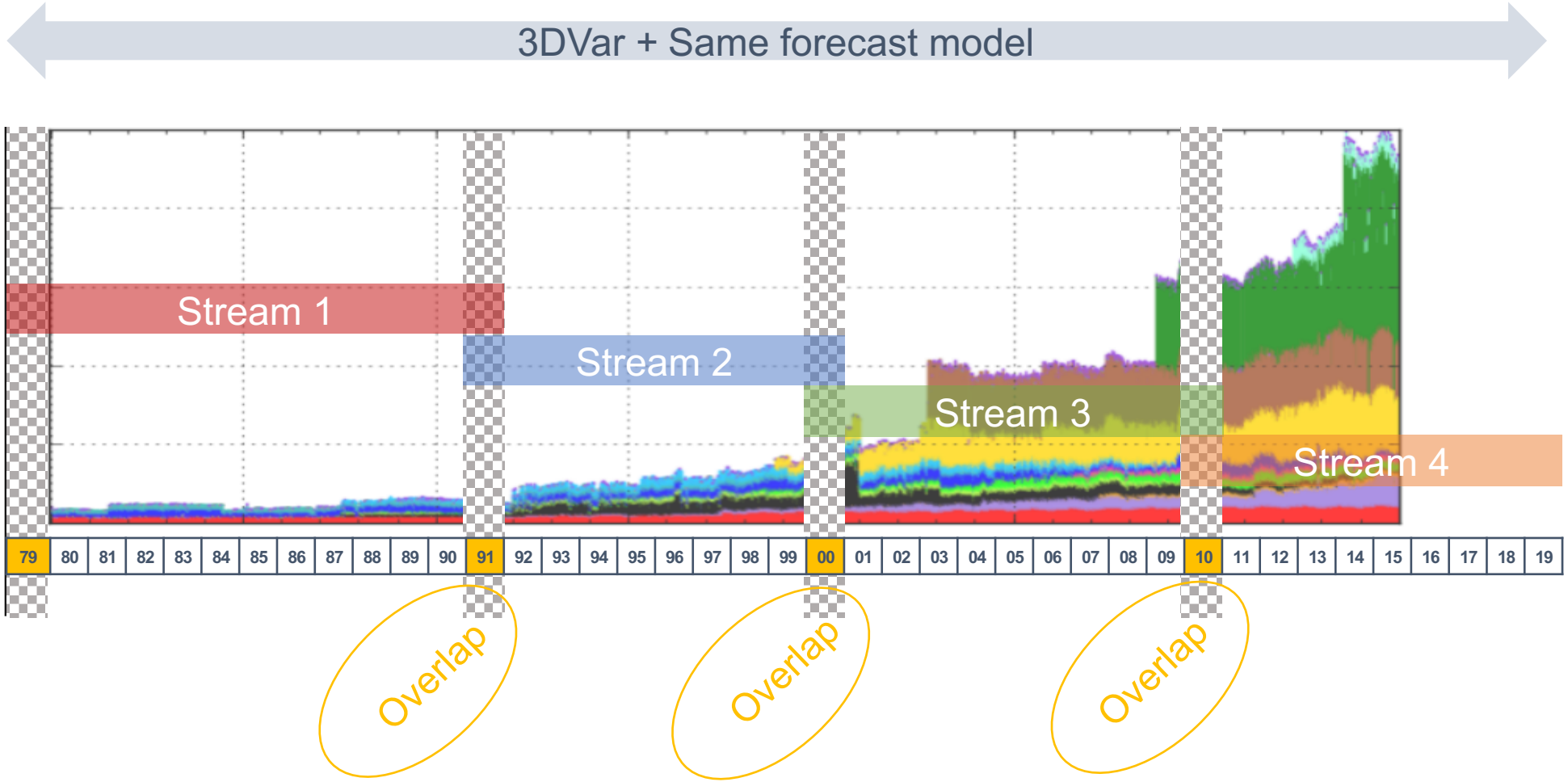


Std stream 3&4 differences
Dec, 2010

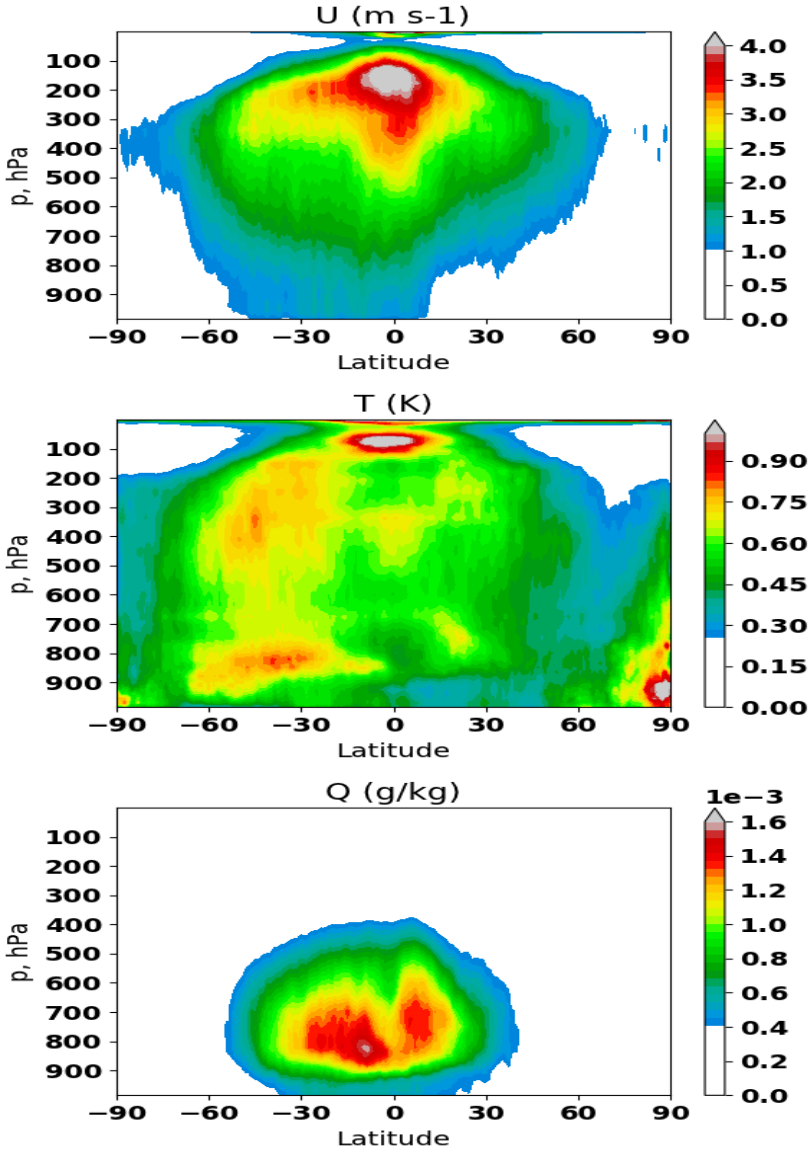


Std analysis error
(3DVar-OSSE)

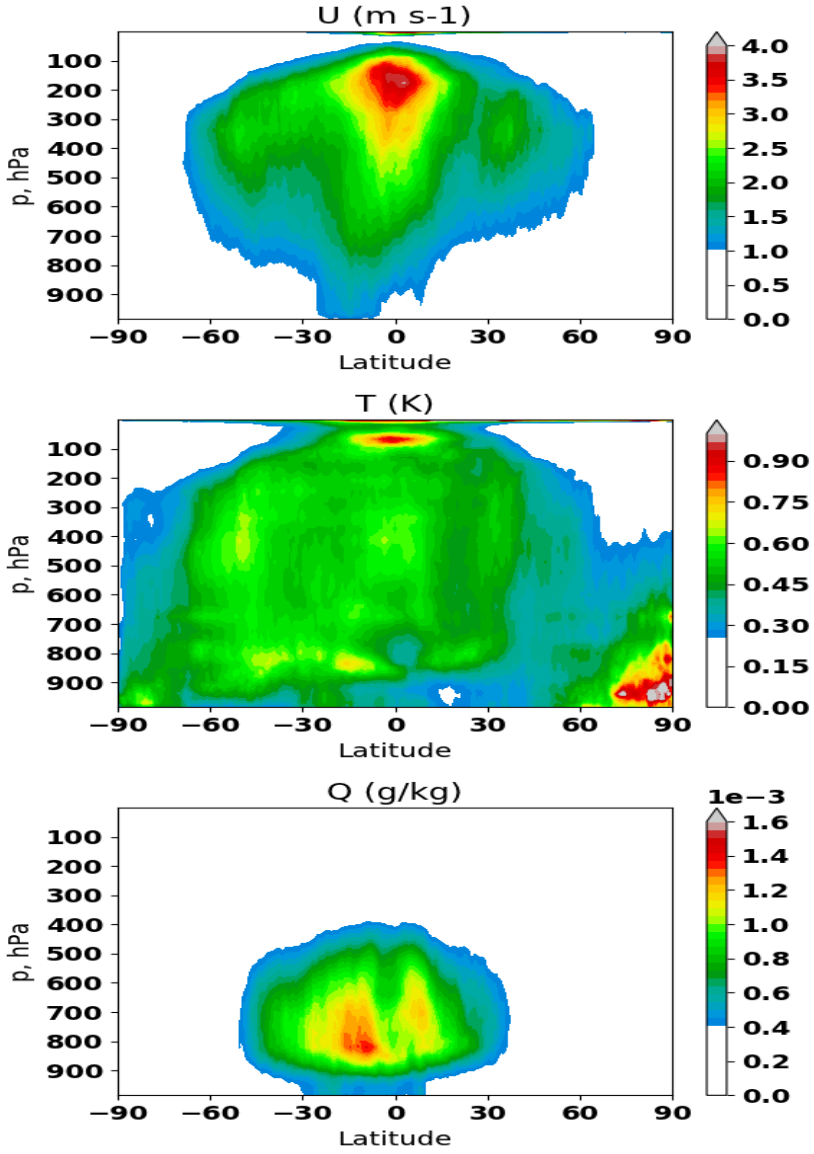
MERRA-2 Production streams



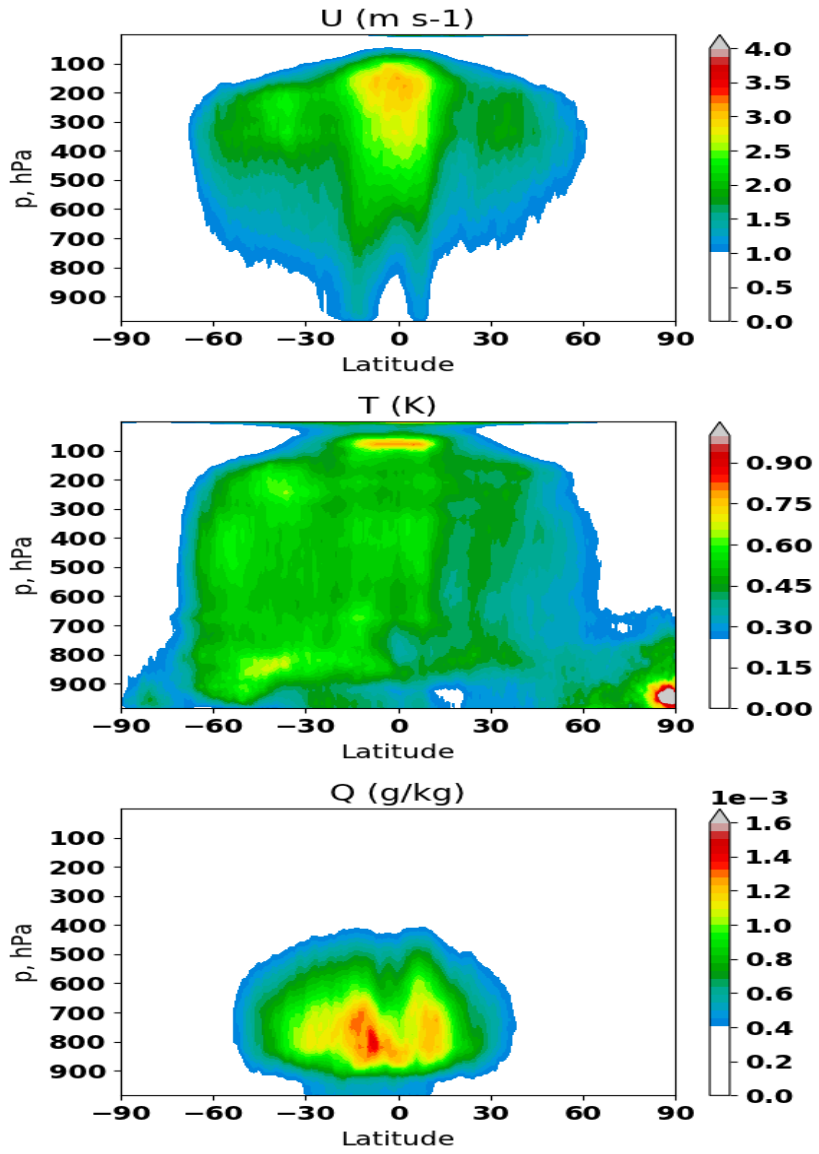
Dec 1991



Dec 2000



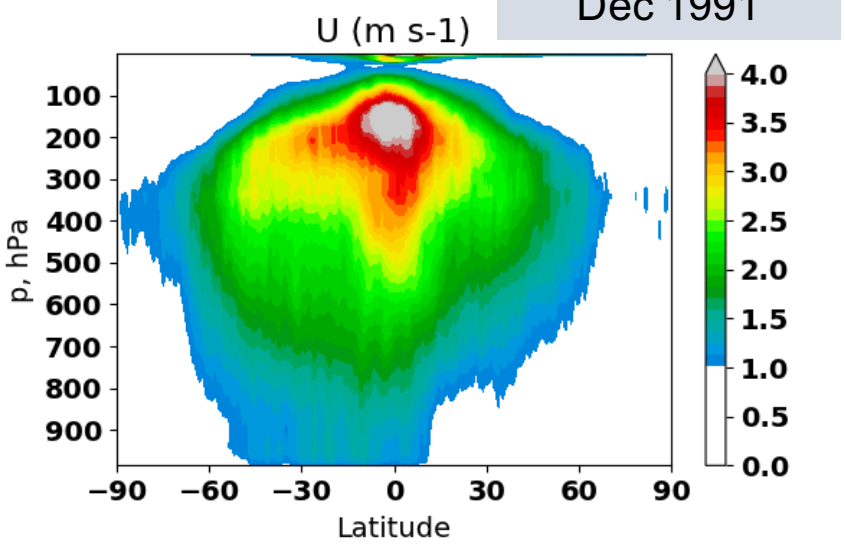
Dec 2010



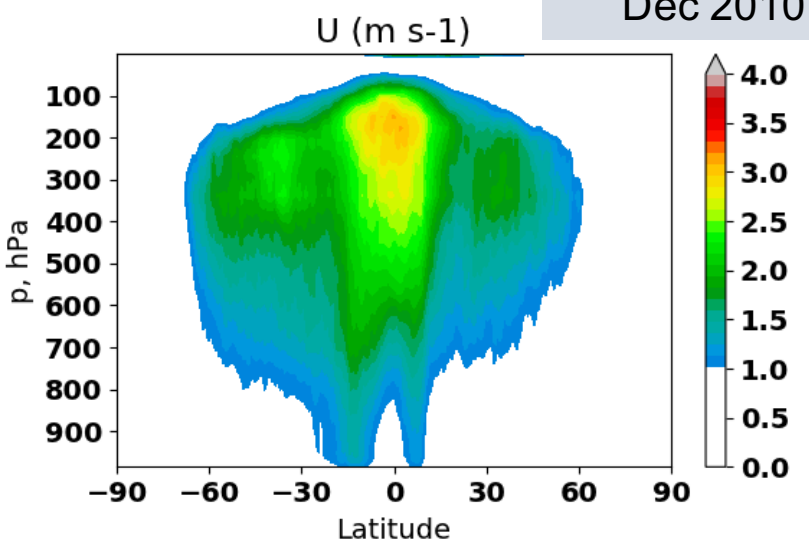
Std stream differences

Std analysis error (3DVar-OSSE)

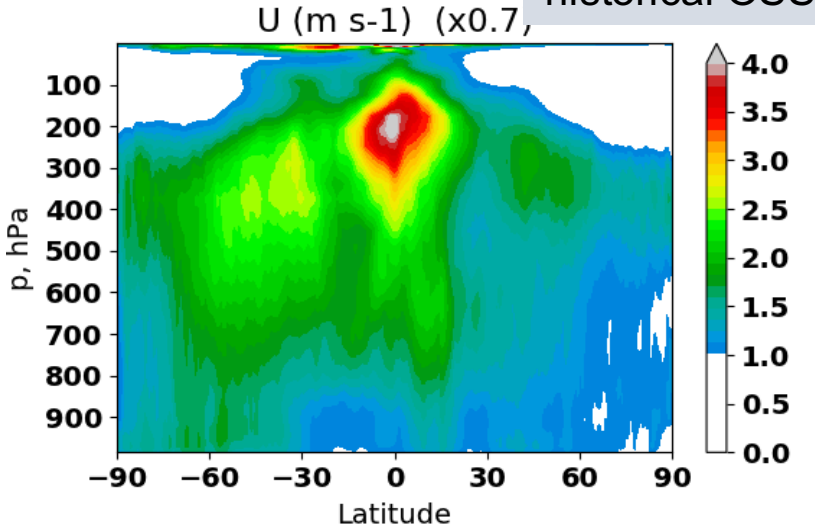
Dec 1991



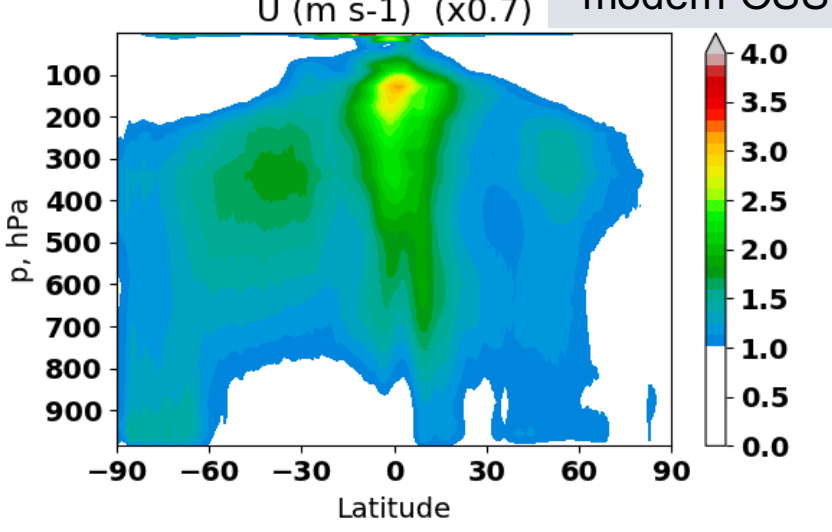
Dec 2010



historical OSSE



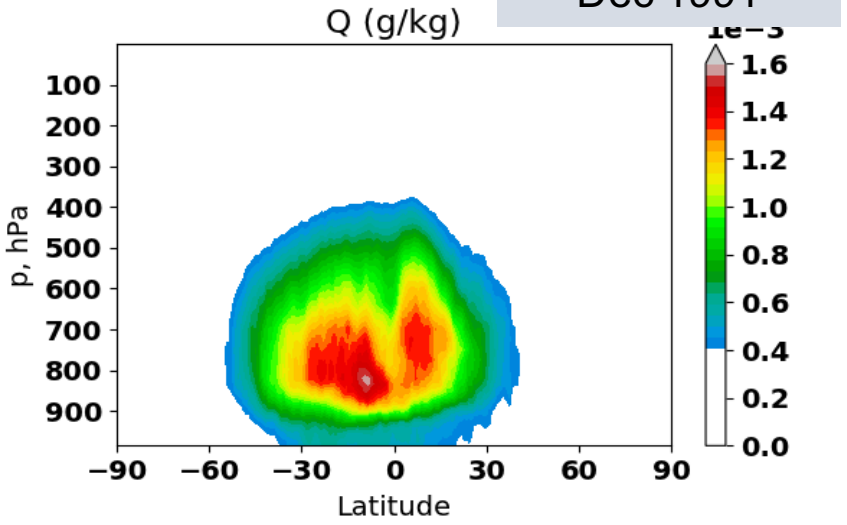
modern OSSE



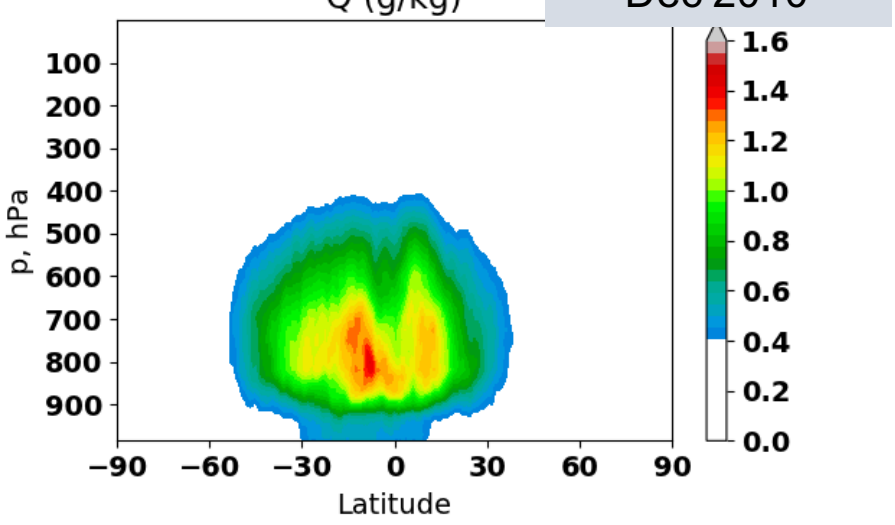
Std stream differences

Std analysis error (3DVar-OSSE)

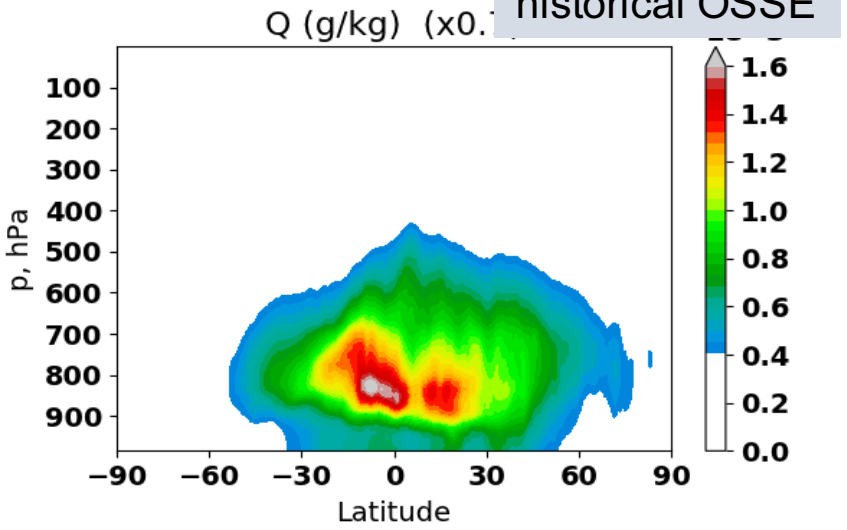
Dec 1991



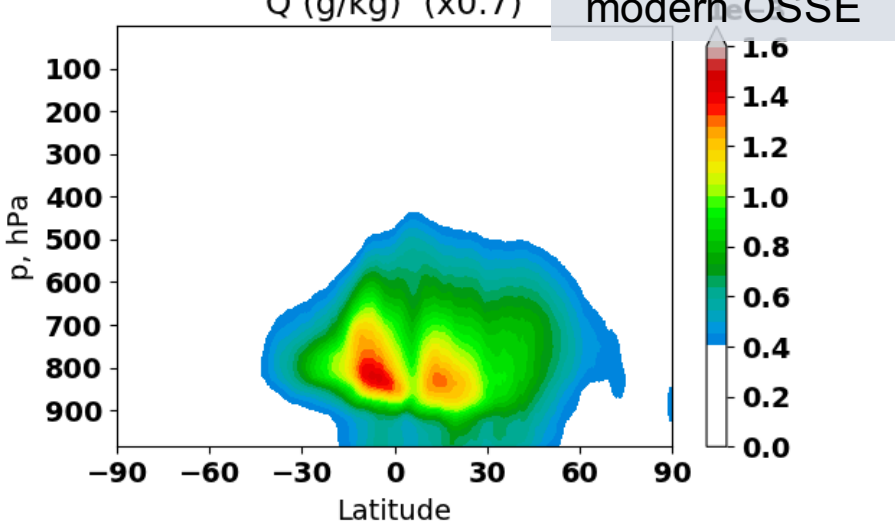
Dec 2010



historical OSSE

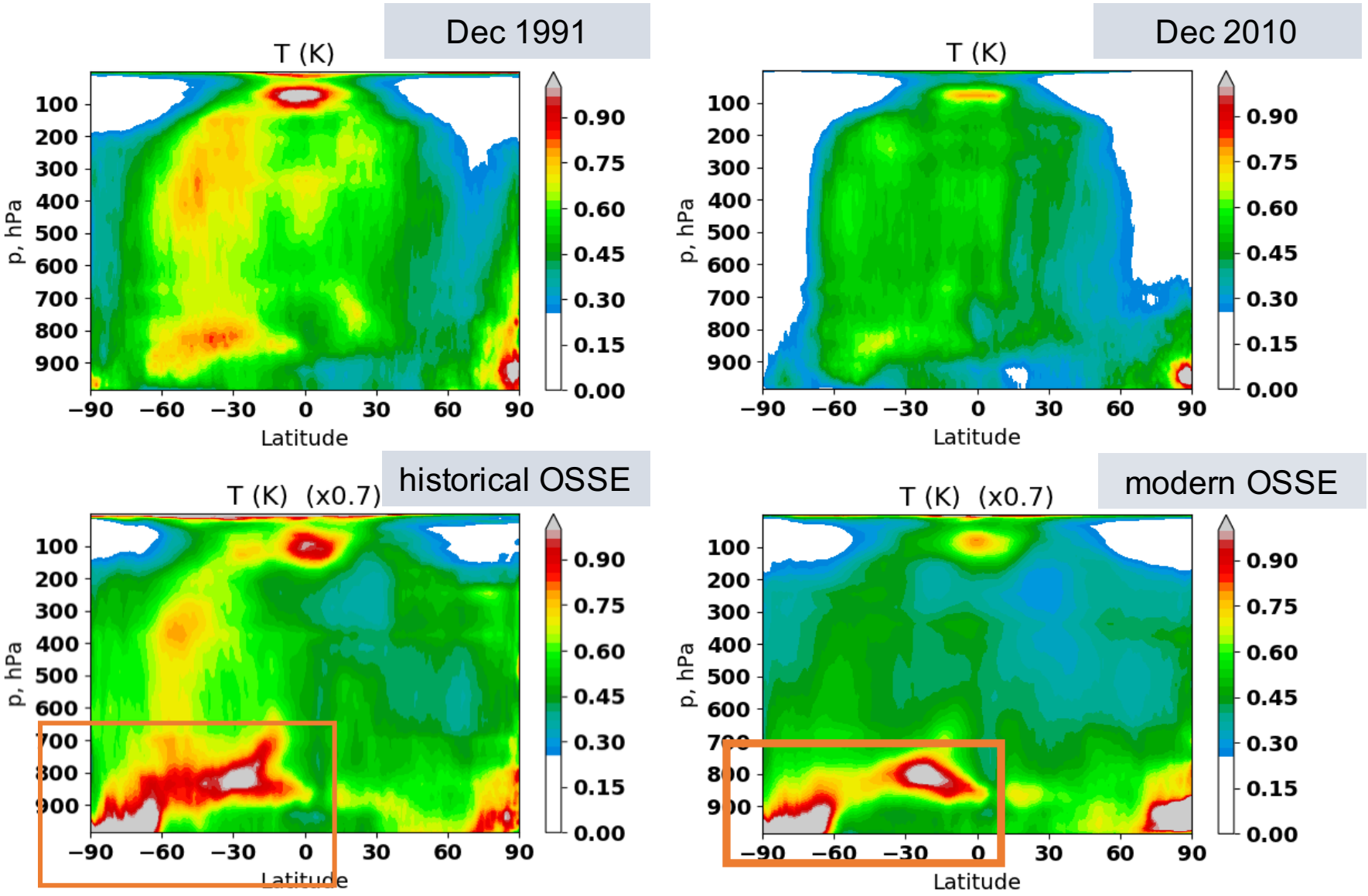


modern OSSE

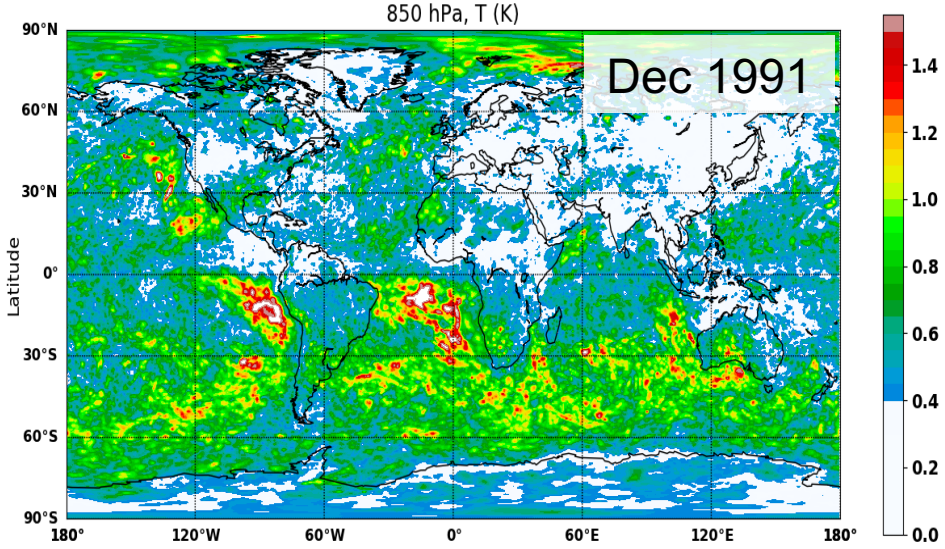


Std stream differences

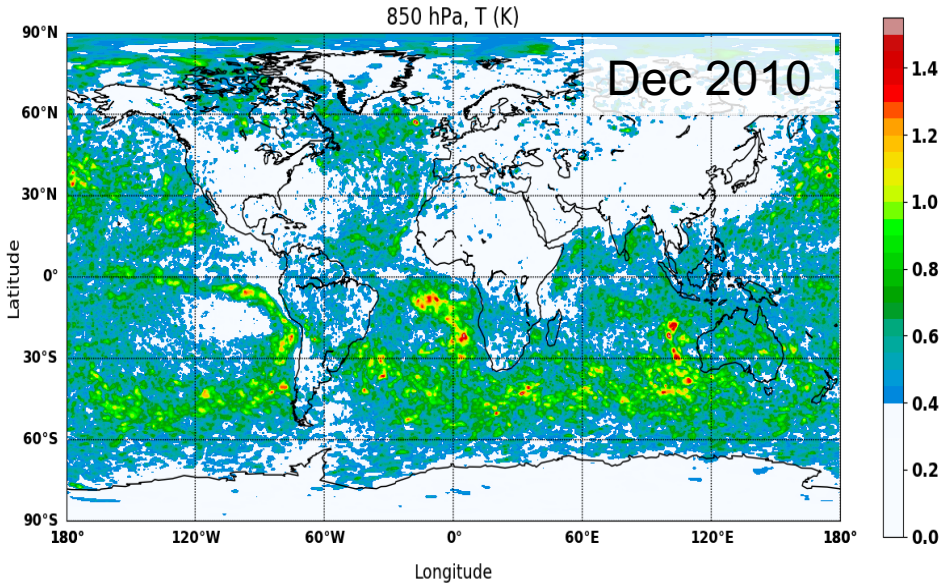
Std analysis error (3DVar-OSSE)



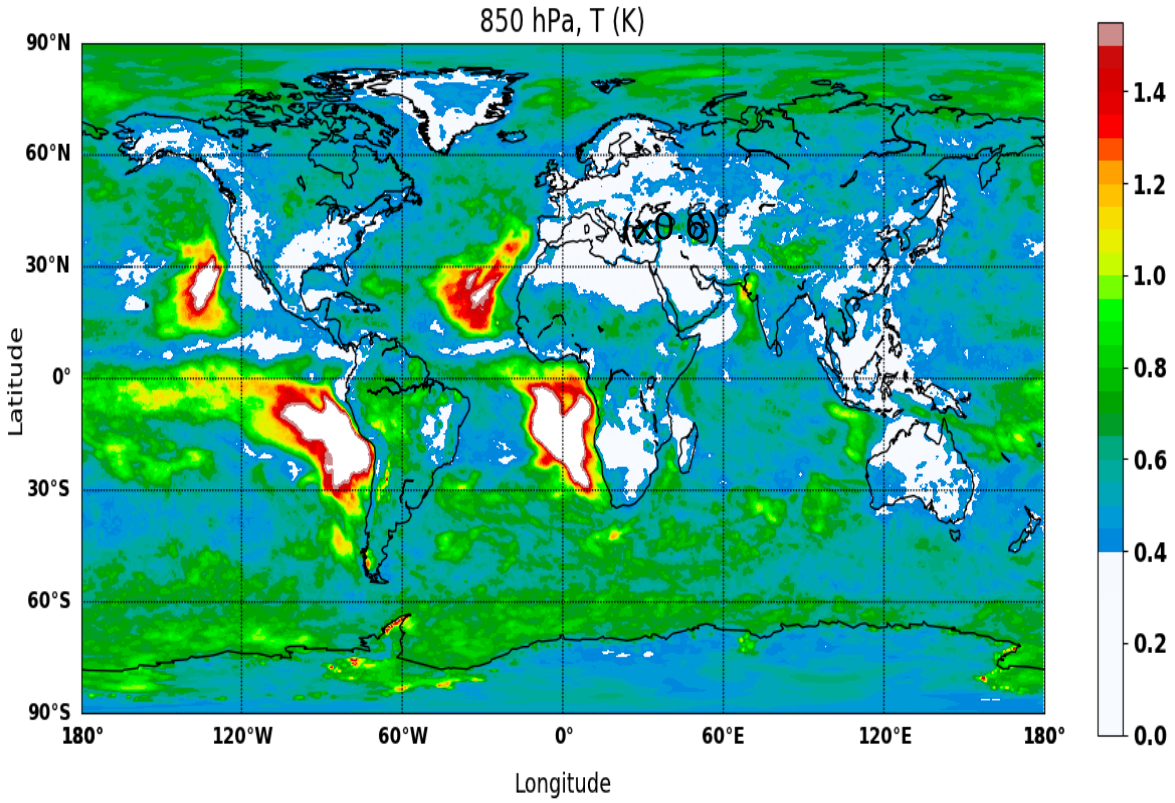
Stream 1&2 difference



Stream 3&4 difference

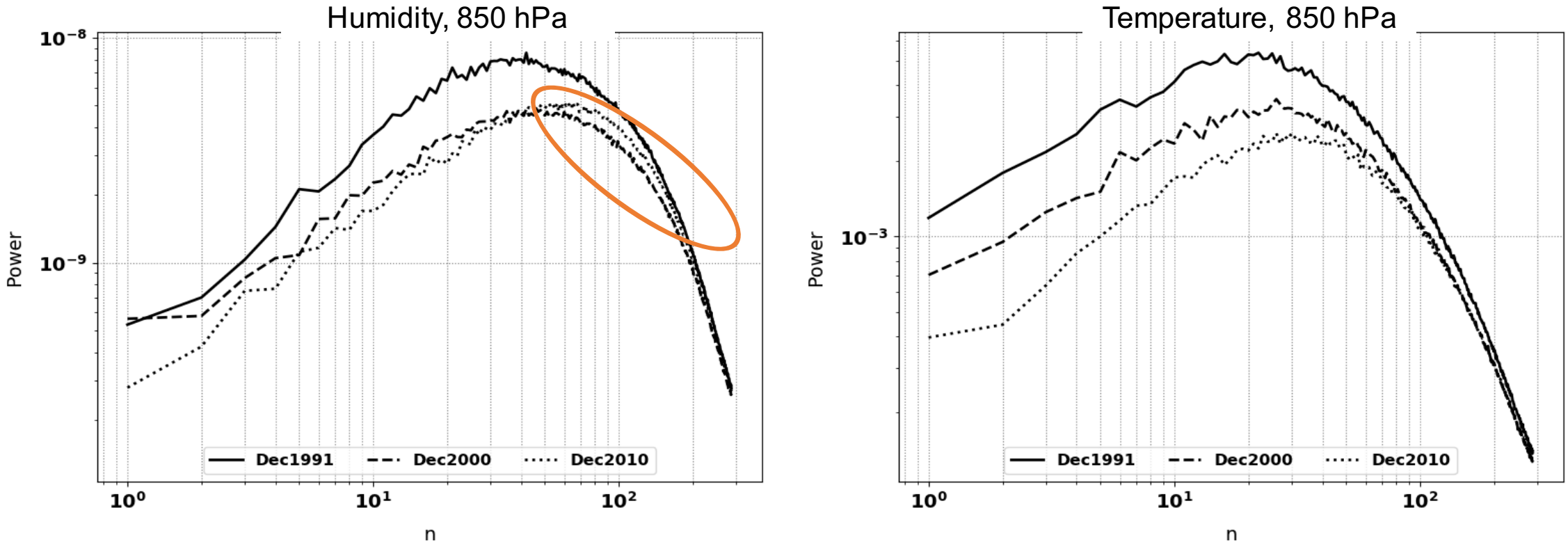


Analysis Error std (OSSE)



Differences in model physics between the Nature Run and forecast model result in large magnitude errors in temperature in regions of low stratiform marine clouds.

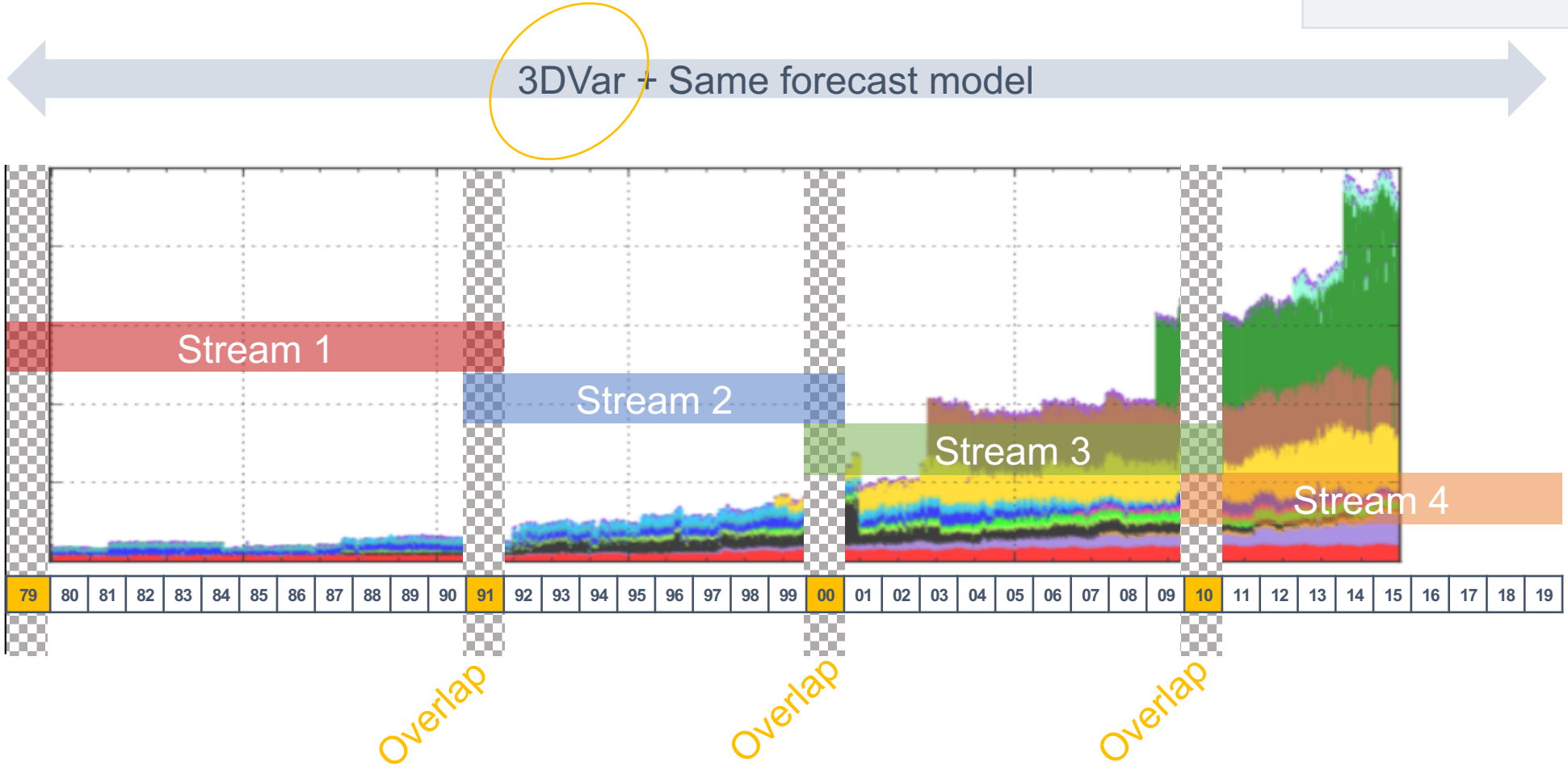
Spectra of stream differences



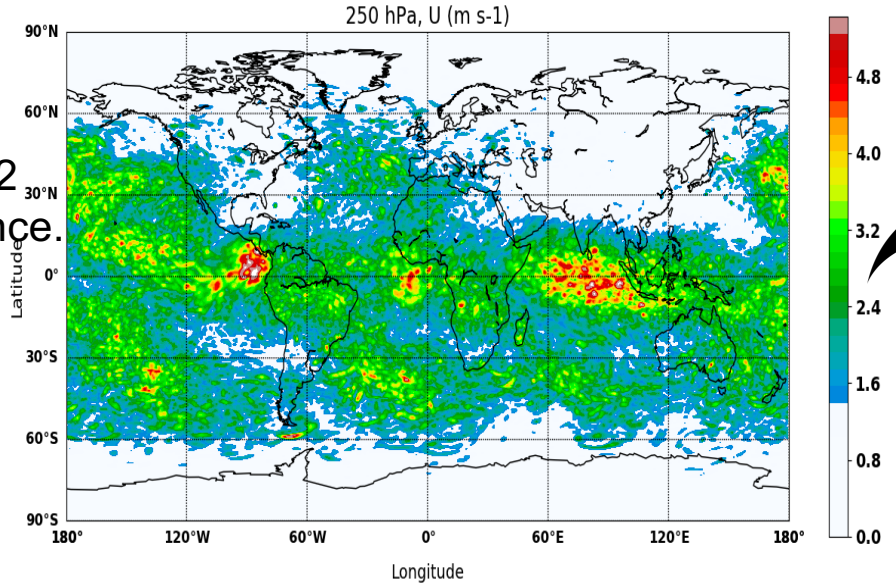
Spectra of the temporal variance of stream differences for temperature (right) and specific humidity (left). Abscissa represents spectral wave number. Statistics computed for the last month of each of the three overlap periods: Dec 1991 (solid), Dec 2000 (dashed), and Dec 2010 (dotted).

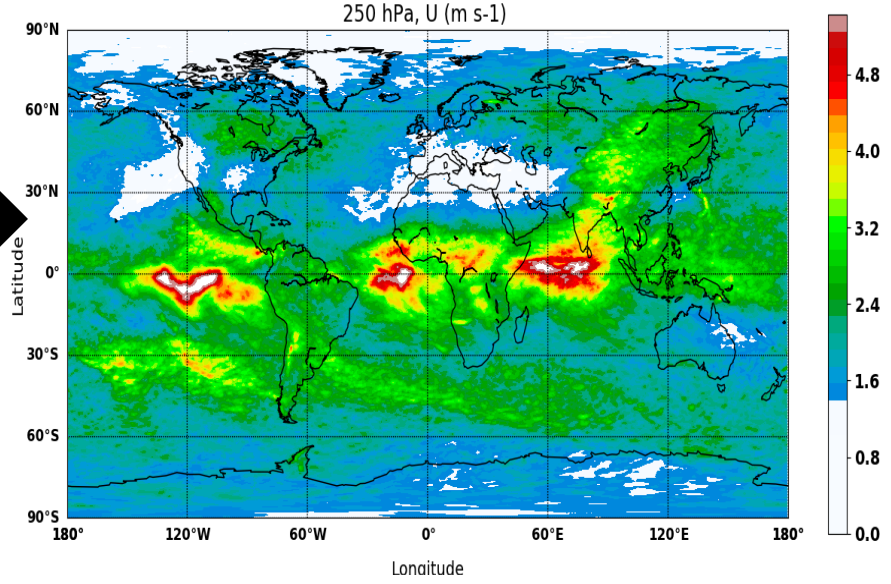
MERRA-2 Production streams

$$0 \leq S(x_1 - x_2) \leq 2\,S(e)$$

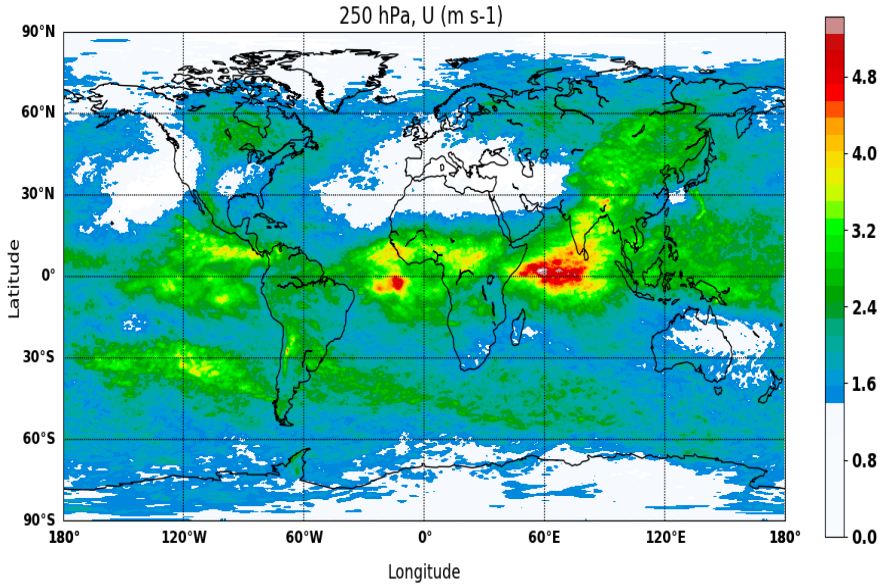


Std MERRA-2 Stream difference. Dec, 2010





Analysis error 3DVar OSSE

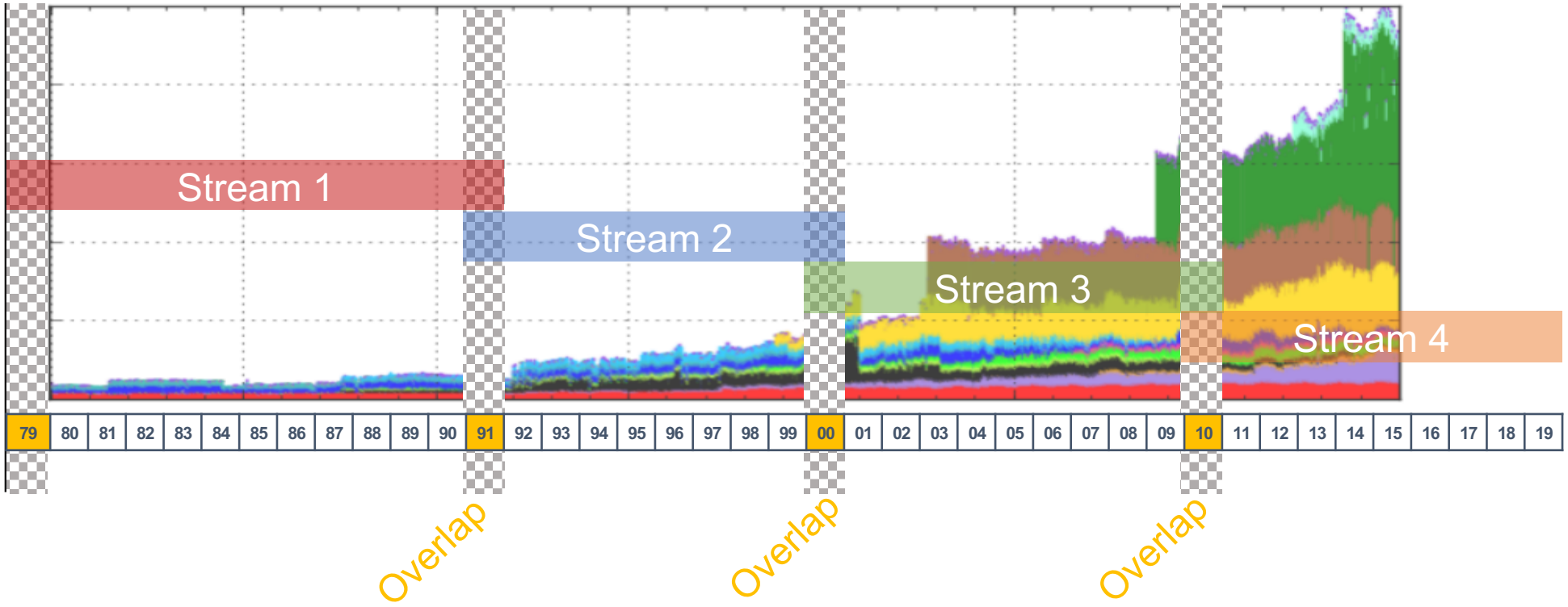


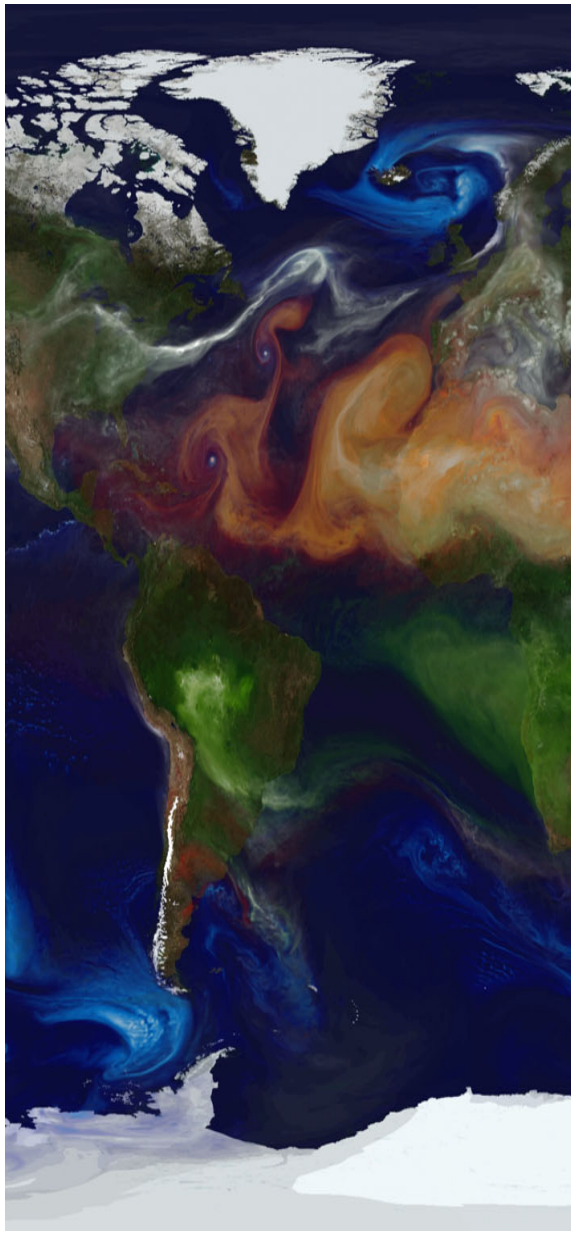
Analysis error 4DEnVar OSSE

?

GEOS-R21C

Choice of Production streams should take into account the historical observing system timeline





Summary

- Constraint:

Convergence of parallel stream is tightly linked to the constraint the data and the data assimilation is applying on the cycling of the streams during the overlaps. The stronger the data constraint and the smaller the analysis error, the closer the overlapping streams will be.

- Practicality:

Choice of transition periods needs to consider the timeline of the historical observing system.

- Future:

The observing system has come a long way, but there's still room for improvement. Future observing systems should target areas of (still) large analysis error. Data recovery and reprocessing efforts are critical in reanalysis. Better use of past observing systems as in all-sky DA or accounting for inter-channel correlations may also help.

Improvements in the data assimilation algorithm as well as forecast models are still needed to improve future reanalysis products.