Singular vector-based thinning of satellite data

Peter Bauer, Roberto Buizza, Carla Cardinali and Jean-Noël Thepaut

European Centre for Medium-Range Weather Forecasts

Outline

1. Satellite data thinning
2. Singular vector-based adaptive observation strategy
3. Singular vector-based satellite data thinning
1. Satellite data thinning

In NWP a wide range of satellite observations is used to constrain the analysis. Currently about 95% of the data employed in the 4D-Var system at ECMWF originates from satellites and about 90% is assimilated as radiances.

Data thinning is applied for two main reasons:

- **reduction of the effects of spatial observation error correlation**, caused by the fact that the errors of neighbouring observations may be correlated due to instrument errors, if recorded with the same device, and due to observation operator errors.
- **reduction of data volume**

At ECMWF spatial thinning is applied to all satellite observations. **Thinning and observational quality control remove about 95% of the data**.

In the current operational system thinning is performed at 1.25°.

1. The scope of this work

The scope of this work is to assess whether singular vectors (SVs) can be used to identify areas where extra satellite data can be used to reduce analysis uncertainty and the forecast error. **SV-based thinning** has been applied to the Southern Hemisphere extratropics (SH) for 2 seasons, using the ECMWF 4D-Var data assimilation.

4D-Var has been run in 5 configurations with different satellite data coverage:

- **low**- EXP and **high-density** EXP-HI coverage over the whole globe
- **SV-thinning** higher density only SV-target regions **EXP-SV** and **EXP-CLI**
- **Random thinning**, higher density only in randomly-defined regions **EXP-RND**

The EXP-SV target regions have been defined by the **SVs computed routinely for the ECMWF Ensemble Prediction System**.

The EXP-CLI target region has been defined by the **SV seasonal average distribution of the previous year**.
Outline

1. Satellite data thinning
2. Singular vector-based adaptive observation strategy
3. Singular vector-based satellite data thinning

2. Using singular vectors to target observations

The idea of using singular vectors (SVs), i.e. the perturbations with the fastest growth rate over a finite time interval, to identify regions where to take extra observations is not new.

Buizza & Montani (1999) used SVs to identify regions over the Atlantic where to take extra observations to reduce the forecast error over Europe.

- In some cases errors could be reduced by up to 13%.


(i) observations taken in SV-target areas are more valuable than in Random areas.
(ii) it is important that daily set of singular vectors are used to compute the target areas.
(iii) the value of targeted observations depends on the region, the season and the baseline observing system (data-rich or data-poor).
2. Singular vectors

Singular vectors define perturbations with the fastest growth over a finite time interval. In our study, growth is measured using a total-energy metric (see Palmer et al 1998 for a discussion of SV sensitivity to metric choices).

The problem of the computation of the directions of maximum growth of a time evolving trajectory reduces to the computation of the singular vectors of \( K = E_0^{1/2} E E_0^{1/2} \), i.e. to solving the following eigenvalue problem:

\[
E_0^{-1/2} L^* E E_0^{-1/2} \mathbf{v} = \sigma^2 \mathbf{v}
\]

where:
- \( E_0 \) and \( E \) are the initial and final time metrics
- \( L(t,0) \) is the linear propagator, and \( L^* \) its adjoint
- The trajectory is time-evolving trajectory
- \( t \) is the optimization time interval

2. SV-based target area definition

SV-based targeted regions have been defined as in Buizza & Montani (1999). 48-hour singular vectors routinely computed over the SH to defined the ECMWF Ensemble Prediction System (EPS) initial conditions, have been used to define the target regions. The final-time norm has been set to be either the North-American or the European verification region.

Once the 50 leading SVs have been computed, the following total-energy based localization function has been computed:

\[
f(x) = \frac{1}{N_{SV}} \sum_{j=1}^{N_{SV}} \frac{\sigma_j}{\sigma_1} TE_j(x)
\]

For each case, the SV-target area is defined by the \( N_{gp} (=480) \) grid points \( x_g \) with the largest \( f(x) \).
2. Impact of tg-obs

SV-target (SV)  
Random (RD)  
Verification region
Target Areas 120 grid points

Atlantic  
2% SV increase forecast error compared to ~1.5% RANDOM

Pacific  
4% SV increase forecast error compared to ~0.5% RANDOM

Extra-Tropical transitions  
The error increase is ~6-13%

Data-Rich Ocean

Average Winter 2003/04 RMSE started from analyses generated without (SEAOUT) and with all (SEAIN) data over the ocean, and with data only in SV-based (SVIN) or random (RDIN) areas

Results indicate that if the baseline system is data void over the ocean, using obs in the SV-target areas (red lines) has a larger impact than using obs in random areas (green lines)

In other words, it is better to observe the SV-targeted areas than Random areas
3. Experiments list

This plot shows the density of AMSU-A channel 9 data for the case of 2008/12/14@00UTC for the different experiments:

- EXP-HI: global thinning to 0.625°
- EXP: global thinning to 1.25° (i.e. ope)
- EXP-SV: EXP but with SV thinning 0.625°
- EXP-CLI: EXP but with SVcli thinning 0.625°
- EXP-RND: EXP but with random thinning 0.625°

Target areas occupy same fraction (15%) of the Southern hemisphere. The SV-based climatology was derived from the mean 2007 SV-areas.

Experiments have been run for JAS08 and D08JF09.

Forecasts from these experiments have been verified against EXP-HI analyses (~83 cases per season, i.e. without first 7 days to avoid spin-up)
3. The ECMWF 4D-Var data-assimilation system

The following version of the ECMWF data-assimilation system has been used in our experiments:

- 12-hour 4-dimentional Variational Data Assimilation System (4D-VAR)
- Resolution:
  - Outer loops at T₅₁ L₆₀
  - Inner loops at T₁₅₉ L₆₀ resolution (50 iter) and T₉₅ L₆₀ (70 iter, reduced phys)

From each analysis, 10-day T₅₁ L₆₀ forecasts have been run.

3. Eady index

The comparison of seasonal average Eady indices can be used to understand seasonal variations in the target regions. The Eady index (Hoskins and Valdes 1990) as measure of baroclinic instability: the static stability N and the wind shear have been computed using the 300- and 1000-hPa potential temperature and wind:

\[ \sigma_E = 0.31 \frac{f du}{N dz} \]

- In JAS (cold season) the Eady index is higher (perturbations grow faster) and the area with high Eady index extends more towards the Equator than in DJF (warm season).
- \( \sigma_E(JAS08) \) and \( \sigma_E(JAS07) \) are rather different, with \( \sigma_E(JAS08) \) showing higher values towards the pole.
- \( \sigma_E(D08JF09) \) and \( \sigma_E(D07JF08) \) are rather similar.
3. Average SV-loc and Eady index: JAS08 Cold S.

The top plot shows the JAS08 zonal-average location of the SV-CLI, SV and RND target regions. The bottom panel shows the zonal average Eady index in JAS07 and JAS08

- **RND- and SV-target** regions cover on average a similar area
- **CLI-target** region peaks more towards the Equator (i.e. away from icy/cloudy regions where satellite data are rejected)

- The difference between the SV- and CLI-target regions is reflected in the differences in the Eady indices, with $\sigma_E$ (JAS07) characterized by one peak at $\sim 40^\circ$S while $\sigma_E$ (JAS08) shows two peaks and $\sim 35^\circ$S at $\sim 60^\circ$S

---

2. Average SV-loc and Eady index: D08JF09 Warm S.

The top plot shows the D08JF09 zonal-average location of the SV-CLI, SV and RND target regions. The bottom panel shows the zonal average Eady index in D07JF08 and D08JF09

- The **RND- SV- and CLI-target** regions cover on average different areas, with SV- and CLI- shifted more towards the Equator. The RND-target region is the one shifted more towards the icy/cloudy regions close to the pole

- $\sigma_E$ (D08JF09) and $\sigma_E$ (D07JF08) peak at about 50°S where also the SV- and CLI-target regions peak
3. Degree of Freedom for Signal (DFS)

The DFS measures the amount of information extracted from the observations during the assimilation process (Cardinali et al 2004). DFS depends on the observation and the background accuracies and model used as space/time propagator.

This figure shows the average DFS in JAS08 Cold S for satellite radiances:

\[ \text{EXP-HI} \text{ 70\% more DFS than EXP} \]

33\% IASI-AMSU-A  21\% AIRS  7\% HIRS  6\% Others

3. DFS and N_{obs}

The top panel shows N_{obs} difference:

\[ r_{N_{obs}} = \frac{N_{obs}(EPS-tg) - N_{obs}(EXP)}{N_{obs}(EXP)} \]

The bottom panel shows DFS difference:

\[ r_{DFS} = \frac{DFS(EPS-tg) - DFS(EXP)}{DFS(EXP)} \]

(a) \( N_{obs} \) is lower in JAS08 (cold SH season) than D08JF09 because more satellite data are rejected over sea ice and/or cloudy regions.

(b) DFS is higher in JAS08 than in D08JF09 because in cold season perturbations grow faster (see differences in Eady index between JAS and DJF) and observations have a larger impact on the analyses.
3. DFS and $N_{obs}$

In the 2007 studies, we found that DFS was always higher for EXP-SV compared to RD. Observed structures in SV-areas analyzed (c) in D08JF09 DFS(RND) > DFS(SV) in agreement with the previous studies. By contrast, in JAS08 DFS(RND) > DFS(SV) DA system was NOT able to extract information from the obs located in the SV-area.

1) B static $\rightarrow$ Flow-dependent
2) Observation error larger than signal

3. Forecast error growth in JAS08 and D08JF09

Forecast error over SH grows faster in JAS08 (cold season) than in D08JF09 (warm season).

This is consistent with Eady index statistics, that showed that perturbation growth is higher in JAS.

The impact on forecast error is, on average, small, with the curves of the four experiments laying one on top of the other.
3. Impact on forecast error: JAS08

This figure shows the JAS08 average impact on forecast error:
\[
\frac{\text{rmse}(\text{EPS} - \text{tg}) - \text{rmse}(\text{EXP})}{\text{rmse}(\text{EXP})}
\]

- EXP-Hi has the lowest error
- The difference between EXP-SV and EXP-RND is small, in agreement with DFS statistics
- In this season, the 2-3 day forecast error decreases by ~1-2% if extra satellite obs are added in target regions covering 15% of SH
- The effect of the extra obs is smaller compared to DJF, and detectable only up to fc day 2. This indicates that initial-condition errors dominates only for ~2 days, when other sources of forecast errors become important

3. Impact on forecast error: D08JF09

This figure shows the D08JF09 average impact on the forecast error:
\[
\frac{\text{rmse}(\text{EPS} - \text{tg}) - \text{rmse}(\text{EXP})}{\text{rmse}(\text{EXP})}
\]

- The difference between EXP-SV and EXP-RND is larger, in agreement with DFS statistics that showed that more information was extracted from obs located in SV-areas
- In this season, the 2-3 day forecast error decreases by ~2-4% if extra satellite obs are added in target regions covering 15% of SH
- The effect of the extra obs is larger compared to JAS, and detectable for longer. This indicates that initial condition errors have a larger influence for a longer time
Conclusions

The ECMWF current operational setting (1.25°) is too conservative and satellite data density could be increased. 0.625° thinning will be considered for operational implementation.

Volume increase can be kept to affordable levels if satellite data density is increased only in a selective way. This study has shown that over the SH, singular vector-based satellite data thinning is a strategy that can lead to better analyses and lower forecast errors.

Results are seasonal dependent:

- In JAS08 the impact of the extra data is small. DFS indicates that less information was extracted from the obs located in the SV-area than in other regions. Eady index statistics indicate that perturbation grow faster, and this could explain why other sources of forecast errors becomes as important as initial condition.
- In D08JF09, all targeting experiments (EXP-SV/CLI/RND) perform better than EXP for a longer period. Of the three targeting experiments, EXP-SV gives the best results, very similar to EXP-HI after 72h. DFS of obs located in SV-areas is larger than the DFS of obs located in other regions. Perturbation growth is slower (Eady index) and initial condition errors are dominant for a longer time period.
- The day-2 average forecast error reduction of ~4% detected for Z500 over SH is equivalent to a predictability gain of ~1 hour.

Bibliography on SV-based targeting