Explorative Analysis and Visualization of Large Information Spaces

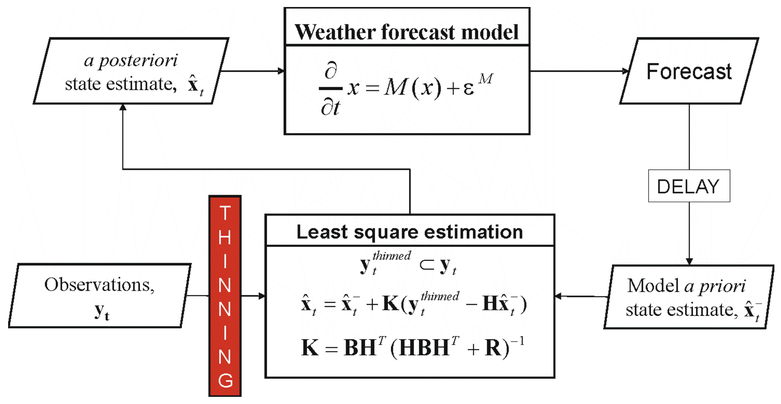
Adaptive thinning of weather observations

Vladimir Bondarenko

Universität Konstanz

1 Introduction

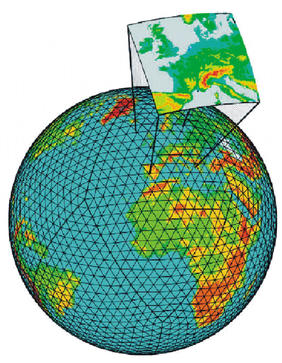
Weather forecast as an initial value problem



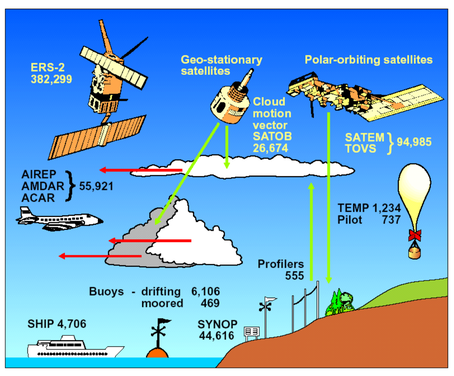
- The initial weather state estimate is an average of *a priory* model estimate on a model grid and irregularly distributed observations.
- $\mathbf{H}$ is a linearized interpolation operator between the model grid and observation positions: $\mathbf{y} = \mathbf{H}\mathbf{x}$
- The gain matrix $\mathbf{K}$ takes into account the *a priory* estimate-error, $\mathbf{B}$, and observation-error, $\mathbf{R}$, covariances.

The model grid and observation systems

Triangular grid:
N = 163 842
A_{min} = 2407 km^2
A_{max} = 3756 km^2
Δ_{av} = 59.9 km
Δ_{min} = 55.1 km
Δ_{max} = 65.8 km



$\dim(x) \approx 10^7$



$\dim(y) \approx 10^5 - 10^7$

Observed variable	Observation system
Temperature, pressure, humidity	Synoptic stations, buoys, airplanes.
Vertical temperature profiles, wind	Radiosondes, polar satellites.
Horizontal wind	Geostationary satellites.

Problem statement

The initial value estimation with a full set of observations is:

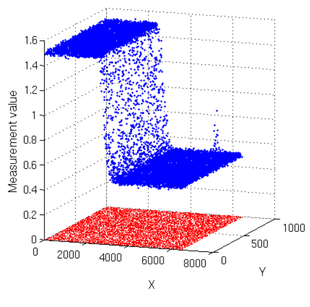
1. Expensive: due to the extremely large sizes of observation and state vectors
2. Inefficient: due to observation redundancy in observation-dense areas
3. Suboptimal: due to poorly known error statistics

The remedy: observation thinning.

2 Observation thinning

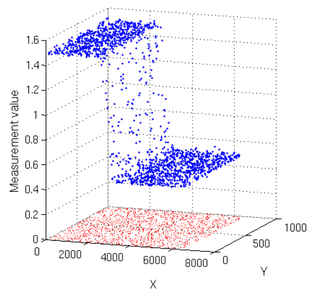
Observation thinning is a sampling procedure that maps a complete observation set onto one of its subsets.
The thinning methods can be divided into non-adaptive and adaptive.

Full 2D observation set



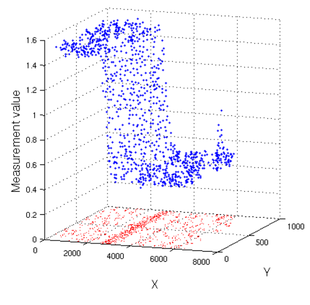
Large gradient and small scale feature

Non-adaptive thinning



Uniform spatial observation distribution

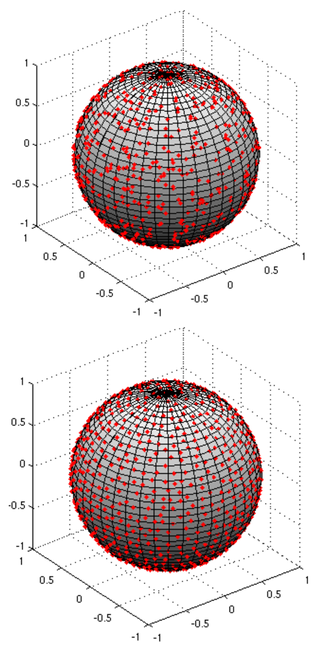
Adaptive thinning



Non-uniform gradient preserving distribution

3 Thinning algorithms

Non-adaptive thinning



Random thinning

- Sampling points are distributed uniformly over the sphere

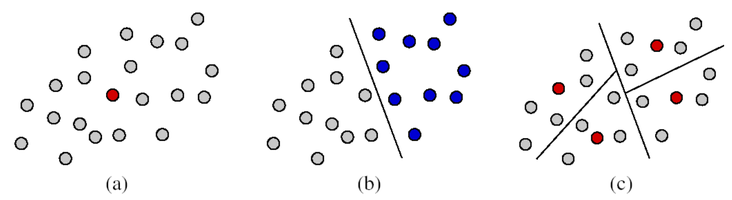
Poisson-disk thinning
Sampling points are distributed uniformly with a minimum distance constrained maintained.
Algorithms:

- Dart-throwing,
- Farthest-point sampling,
- Hammersley or Halton point sequences.

Adaptive thinning

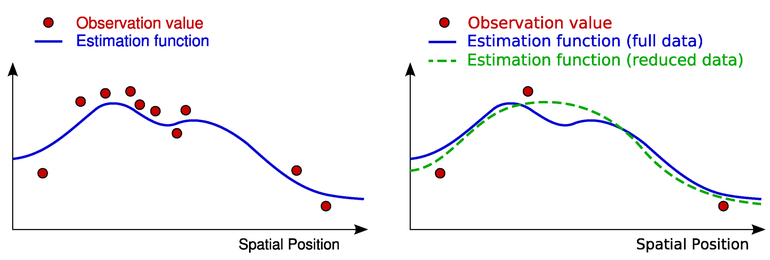
- Top-down clustering [2]

a) k-Means clustering.
b) PCA cluster splitting.
c) The set of centroids is the reduced observation set.



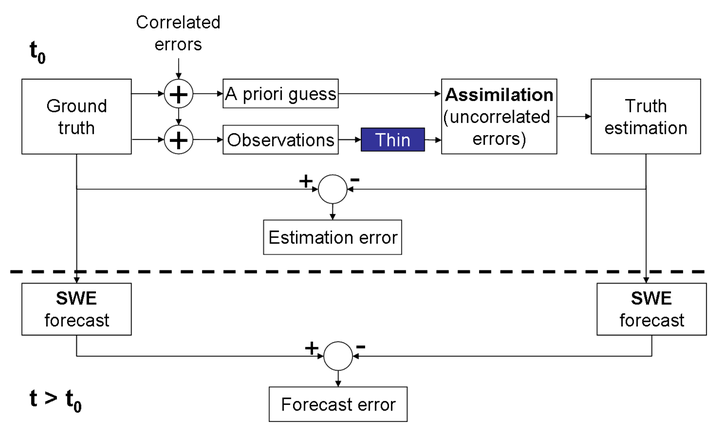
- Estimation error analysis [2]

– Iterative bottom-up procedure.
– Heuristic estimation function.
– The observation minimizing the expected estimation error is thinned out.

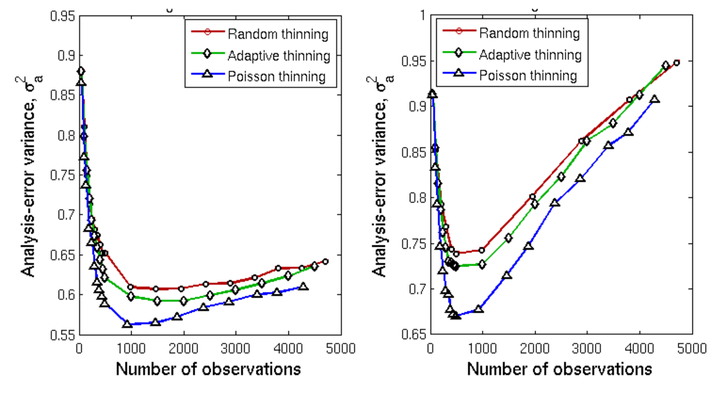


4 Algorithms evaluation

Experimental evaluation framework [1]



- **Ground truth**: Gaussian stationary homogeneous random process.
- **Error statistics**: isotropic exponentially decaying auto-correlation function.
- **Suboptimal estimation**: Observation errors are assumed to be uncorrelated.
- **Forecast error:** Forecast is simulated as solution of nonlinear PDE (Shallow Water Equations).



- **Suboptimal estimation**: large estimation error at dense observation sets due to the neglected observation-error correlation.
- **Non-adaptive vs. adaptive thinning**: Poisson-disk thinning is better than adaptive methods in terms of estimation MSE due to its better error-decorrelation properties.

References

[1] V. Bondarenko, T. Ochotta, D. Saupe, and W. Wergen. The interaction between model resolution, observation resolution and observation density in data assimilation: a two-dimensional study. *Proceedings 11th Symposium on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans, and Land Surface*, 2007.

[2] T. Ochotta, C. Gebhardt, D. Saupe, and W. Wergen. Adaptive thinning of atmospheric observations in data assimilation with vector quantization and filtering methods. *Quarterly Journal of the Royal Meteorological Society*, 131:3427–3437, October 2005.