Data assimilation for chaotic geophysical dynamics

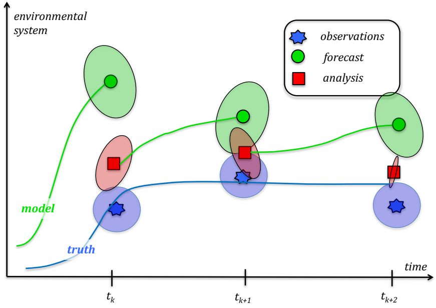
A *very* brief overview

Alberto Carrassi[1,2], **Colin Grudzien**[1,3] and **Marc Bocquet**[4]

(1): Nansen Environmental and Remote Sensing Center, Norway
(2): Geophysical Institute, University of Bergen, Norway
(3): Dept. of Statistics, University of Nevada in Reno, USA
(4): CEREA, joint lab École des Ponts ParisTech and EdF R&D, IPSL, France

With contributions from and acknowledgment to:
A. Apte (ICTS, India), **L. Descamps** (Meteo France), **M. Ghil** (ENS/UCLA, FR/USA), **K. Gurumoorthy** (ICTS, India), **C.K.R.T. Jones**, (U. of North Carolina in Chapel Hill), **L. Palatella** (Lecce, IT), **O. Talagrand** (LMD, FR), **A. Trevisan** (ISAC, IT), **S. Vannitsem** (RMI, BE), **F. Uboldi** (ARIANET, IT), **D. Zupanski** (CIRA, USA), **M. Zupanski** (CIRA, USA).

Data Assimilation: *What and Why?*



- Large family of methods to perform state/parameter/model estimation by combining (taking the best of) models and data.
- The quantities of interest are **probability density functions** (PDFs).
- They quantify the uncertainty on the estimate.
- The goal is to sequentially estimate the conditional PDF $p(\mathbf{x}|\mathbf{y})$, the posterior.
- The PDFs are evolved in time and updated at analysis times using **Bayes' rule**.

$$p(\mathbf{x}_{K:0}|\mathbf{y}_{K:1}) \propto p(\mathbf{x}_0) \prod_{k=1}^{K} p(\mathbf{y}_k|\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{x}_{k-1})$$

DA methods in geosciences: *key challenges*

- Huge models, $m \geq 10^9$ & massive dataset, $d \geq 10^7$ daily obs (yet not enough!) $\Longrightarrow$ Quest for computationally affordable solutions.

- Fully Bayesian DA very difficult (*curse of dimensionality*).

- Gaussian/Linear hypotheses allow to derive computationally tractable methods $\Longrightarrow$ **Kalman filter/smoother**.

- And their Monte Carlo, (still Gaussian) nonlinear approx, **ensemble Kalman filter/smoother**.

- The transition density, $p(\mathbf{x}_k|\mathbf{x}_{k-1})$, approximated by running an ensemble of N trajectories.

- EnKF highly rank deficient in geophysical applications: $N = \mathcal{O}(10^2) \ll m$. Yet it works!

DA methods for chaotic dynamics: *key challenges*

- Atmosphere and ocean, are examples of chaotic dissipative dynamics $\Longrightarrow$ Highly state-dependent error growth.
- DA must track and incorporate this flow-dependency in the quantification of the uncertainty (*i.e.* error covariance).
- Dissipation induces an "effective" dimensional reduction $\Longrightarrow$ The error dynamics is confined to a subspace of much smaller dimension, $n_0 \ll m$: the **unstable subspace**
- The existence of the underlying *unstable-stable splitting of the phase space* expected to have enormous impact on DA.

Motivations

1. Is there any fingerprint of the unstable subspace on the fate of (En)KF and (En)KS?
2. Can dynamical properties be used to design computationally cheap DA strategies?

Deterministic linear case: *behavior of the KF and KS*

(Some) key **analytic results** (without controllability):

- **Collapse of the uncertainty**: KF error covariance asymptotically in the span of the unstable-neutral backward Lyapunov vectors (BLVs$^{\mathrm{u}}$) [Gurumoorthy *et al* 2017]

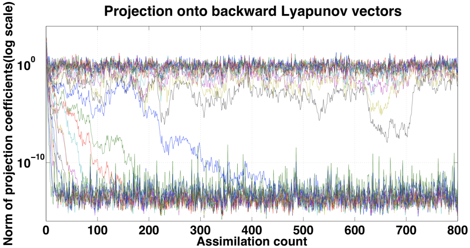
- **Convergence of the covariance**: Low rank, n_0, KF covariance, initialized in the span of BLVs$^{\mathrm{u}}$, converges to the true KF one

$$\lim_{k\to\infty} ||\mathbf{P}_k - \hat{\mathbf{P}}_k|| = 0$$

if the unstable-neutral subspace is observed [Bocquet *et al* 2017]. Warning: *neutral modes are tricky!*

- Likewise demonstrated for Kalman smoother [Bocquet & Carrassi 2017].

$m = 40$, $n_0 = 14$



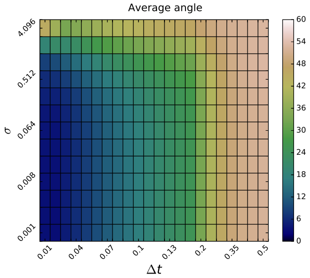
KF/KS reduced rank surrogates based on BLVs are possible.

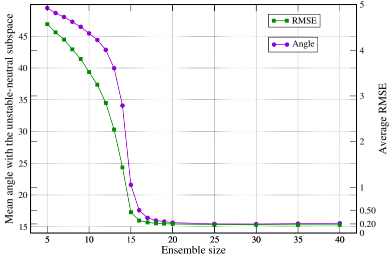
Deterministic nonlinear case: *behavior of the EnKF and EnKS*

▶ Asymptotic rank of EnKF covariances related to multiplicity and strength of unstable Lyapunov exponents (LEs) [Carrassi *et al* 2009; Gonzalez-Tokman & Hunt 2013].

▶ When the EnKF/EnKS ensemble subspace recovers the unstable subspace the unknown system state is estimated with high accuracy (sudden drop of RMSE) [Bocquet & Carrassi, 2017].

- Lorenz 96 model, $m = 40$, $n_0 = 14$
- **Left** - Angle Unstable/Ensemble subspaces *vs* $(\Delta t^{obs}, \sigma^{obs})$.
- **Right** - EnKF RMSE (green) and Angle (purple) *vs* N.



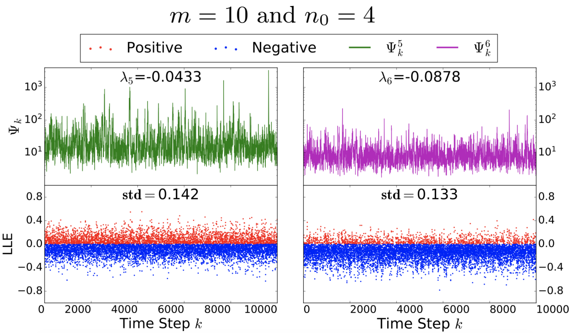


Nonlinear systems, with "weakly nonlinear" error dynamics, need only n_0 members!

Error in stochastic models: *What the role of the instabilities?*

$$\mathbf{x}_k = \mathcal{M}_{k:k-1}(\mathbf{x}_{k-1}) + \boldsymbol{\eta}_k, \qquad \boldsymbol{\eta}_k \in \mathcal{N}(\mathbf{0}, \mathbf{Q}_k)$$

- By re-introducing perturbations, the error covariance is generally full rank (as $\mathbf{Q}_k$).
- Asymptotic uncertainty in the stable BLVs **no longer zero, but still bounded.**
- However, the bounds, $\boldsymbol{\Psi}^i_k$, depend on [Grudzien *et al* 2018a]
  1. the model error size (*i.e.* $||\mathbf{Q}||$),
  2. the variance of the local LEs (LLEs).



- If model error is large and/or the LLEs have high variance, the bounds will be impractically large.

In stochastic systems it is *necessary* to include weakly stable BLVs of high variance.

Error in stochastic models: *The upwelling effect*

- Will the *necessary* increase $\mathbf{N} = \mathbf{n_0} \to \mathbf{n_0} + \mathbf{n_{ws}}$ also be *sufficient*?

- To answer this, write the model propagator in the basis of the BLVs using the recursive QR decomposition

$$\mathbf{M}_k = \mathbf{E}_k \mathbf{U}_k \mathbf{E}_k^{\mathrm{T}}, \quad \mathbf{E}_k = (\mathbf{E}_k^{\mathrm{f}}\ \mathbf{E}_k^{\mathrm{u}}) \ \text{ with } \ \mathbf{U}_k = \begin{pmatrix} \mathbf{U}_k^{\mathrm{ff}} & \mathbf{U}_k^{\mathrm{fu}} \\ 0 & \mathbf{U}_k^{\mathrm{uu}} \end{pmatrix}$$

and partition the error into **filtered**/**unfiltered** variables $\boldsymbol{\epsilon}_k = \mathbf{E}_k^{\mathrm{f}} \boldsymbol{\epsilon}_k^{\mathrm{f}} + \mathbf{E}_k^{\mathrm{u}} \boldsymbol{\epsilon}_k^{\mathrm{u}}$

- The error in the filtered space (“seen” by DA) is given recursively by [Grudzien *et al* 2018b]

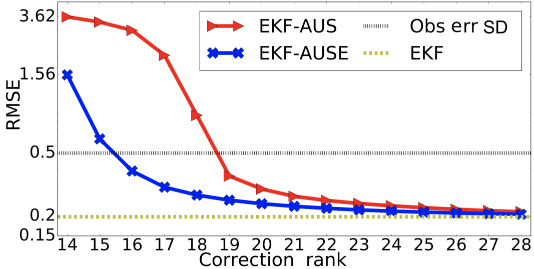
$$\boldsymbol{\epsilon}_{k+1}^{\mathrm{f}} = (\mathbf{U}_{k+1}^{\mathrm{ff}} - \mathbf{U}_{k+1}^{\mathrm{ff}} \mathbf{K}_k \mathbf{H}_k \mathbf{E}_k^{\mathrm{f}}) \boldsymbol{\epsilon}_k^{\mathrm{f}} - \mathbf{U}_{k+1}^{\mathrm{ff}} \mathbf{K}_k \boldsymbol{\epsilon}_k^{\mathrm{obs}} + \boldsymbol{\eta}_k^{\mathrm{f}} + (\mathbf{U}_{k+1}^{\mathrm{fu}} - \mathbf{U}_{k+1}^{\mathrm{ff}} \mathbf{K}_k \mathbf{H}_k \mathbf{E}_k^{\mathrm{u}}) \boldsymbol{\epsilon}_k^{\mathrm{u}}$$

- **The terms in black** correspond to the usual KF-like recursion and highlight the stabilizing effect of the DA [Carrassi *et al* 2008b].

- **The terms in red** disappear when the filtered subspace is the entire state space ($n = m$).

Error in stochastic models: *The upwelling effect*

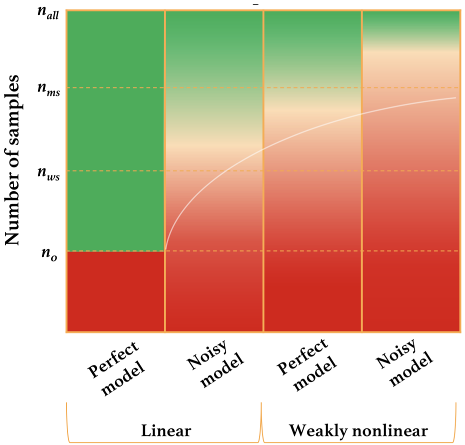
- When $n < m$, they represent the **dynamical upwelling** of the unfiltered error into the filtered variables [Grudzien *et al* 2018b].
- It moves uncertainty from unfiltered to filtered subspace, *i.e.* from the stabler to the unstable subspace.
- This phenomenon **occurs whenever** $n < m$, but is **exacerbated by stochastic noise**.
- Leads to underestimating the error in the (En)KF $\Rightarrow$ Need for **inflation** to prevent divergence.



- **EKF** solves the *full-rank* recursion.
- **EKF-AUS** solves the *low-rank* recursion **without upwelling** (black terms only).
- **EKF-AUSE** solves the *low-rank* recursion **with upwelling** (black+red terms).

EnKF/KS with chaotic systems: *A summary on the required ensemble size*

- Illustration of the minimum number of ensemble members to achieve filter accuracy.
- Different **model scenarios** are given in the *x-axis*.
- The number of members (samples) is given in the *y-axis* with:
  - n_0: number of unstable-neutral BLVs.
  - n_{ws}: number of unstable-neutral BLVs + weakly stable.
  - n_{ms}: number of unstable-neutral BLVs + weakly stable + more stable.
  - n_{all}: number of member = model dimension (*i.e.* full-rank filter).



Grudzien *et al* 2018b

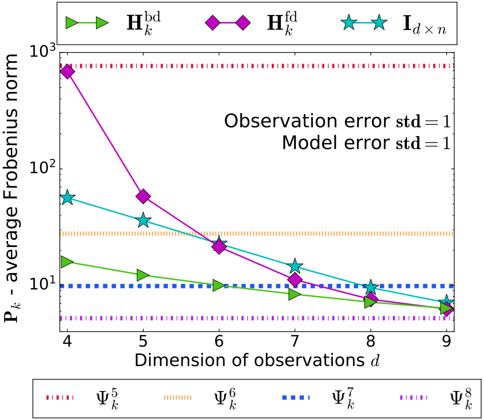
Assimilation in the unstable subspace - AUS

- These properties are at the basis of the **assimilation in the unstable subspace** (**AUS**, by *A. Trevisan & Collaborators*), where the unstable subspace is explicitly used in the DA to:

  - parametrize the description (both temporally and spatially) of the uncertainty in the state estimate (*i.e.* the covariance) $\Longleftrightarrow$ Acting on $\mathbf{K}$ [Trevisan *et al* 2010; Trevisan & Palatella 2011; Palatella & Trevisan 2015]
  - design of the observational network (types, distribution, frequency) $\Longleftrightarrow$ Acting on the operator $\mathcal{H}$ [Trevisan & Uboldi, 2004; Carrassi *et al* 2007]
  - or both [Carrassi *et al* 2008a; 2008b]

Projecting the data in the unstable subspace - Conditions for filter stability

KF with data projected on a subspace of dimension d compared to a full KF.



[Grudzien *et al* 2018a]

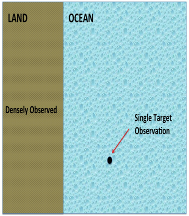
- $\mathbf{H}^{\mathrm{fd}}$ - Observe within the subspace of the d leading FLV $\Rightarrow$ Satisfy a weaker necessary condition [Frank and Zhuk 2018]
- $\mathbf{H}^{\mathrm{bd}}$ - Observe within the subspace of the d leading BLV. $\Rightarrow$ Satisfy a stronger sufficient condition [Bocquet *et al* 2017]
- $\mathbf{I}_{d\times n}$ - Observe the first d components.
- Projected observations based on dynamics was studied earlier by [Law *et al* 2016].

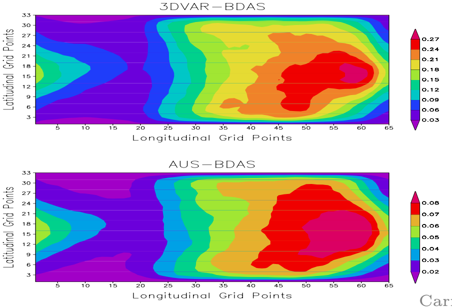
AUS and Target Observations

TARGET OBSERVATION STRATEGY: **Breeding on the Data Assimilation System BDAS**

- Quasi-geostrophic atmospheric model (Rotunno and Bao, 1996 MWR)
- Perfect model setup - Observation Dense area (1-20 Longitude) - Target Area, one obs between 21-64 Longitude

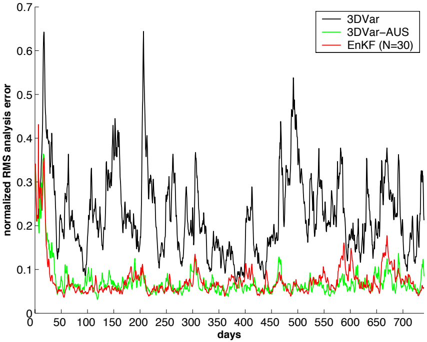
Experiment	Ocean Obs Type/Positioning/Assimilation	RMSE
LO	-	0.462
FO	vert.Prof/fixed(in the max(err))/3DVar	0.338
RO	vert.Prof/random/3DVar	0.311
3DVar-BDAS	vert.Prof/BDAS/3DVar	0.184
AUS-BDAS	temp.1-Level/BDAS/AUS	0.060





Carrassi *et al* 2007

Hybrid 3DVar-AUS: *Enhancing the performance of a 3DVar using AUS*

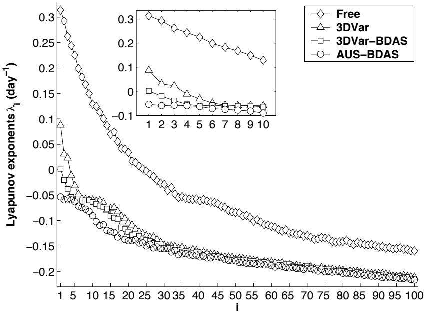


Carrassi *et al* 2008a

- ▶ QG model on a β-plane; network of randomly distributed obs (vertical soundings).
- ▶ 3DVar-AUS: (1) AUS assimilates the obs located on the unstable mode; (2) 3DVar process the remaining obs.
- ▶ 3DVar-AUS comparable to EnKF with only one unstable mode $\Rightarrow$ Reduced computational cost and implementation on a pre-existing 3DVar scheme.

Does stabilization improve estimation?

Quasi-Geostrophic Model (Rotunno and Bao, 1996)



Experiment	RMSE
Free	1
3DVar	0.321
3DVar-BDAS	0.163
AUS-BDAS	0.058

▶ DA "always" provides a stabilizing effect (*e.g.* compare 3DVar with free system Lyapunov spectrum) but ...

▶ if the DA is designed to kill the instabilities, the estimation error is efficiently reduced

Carrassi *et al* 2008b

Conclusions

- We have shown that the (En)KF/(En)KS in deterministic dynamics naturally project the uncertainty on the unstable-neutral subspace $\Rightarrow N = n_0$ members are sufficient.
- These properties are at the basis of the **assimilation in the unstable subspace** (**AUS**, by *A. Trevisan & Collaborators*), where the unstable subspace is explicitly used in the DA process.
- AUS has been successfully applied to deterministic atmospheric, oceanic and traffic models [see Palatella *et al* 2013 for a review].
- In stochastic dynamics we have shown that weakly stable modes of high variance must be be included.
- Furthermore we have demonstrated the existence of an *upwelling* of uncertainty from unfiltered-to-filtered subspace that motivates the need for multiplicative inflation.
- All has been done within a Gaussian framework $\Rightarrow$ Can the unstable subspace be used to develop efficient fully Bayesian (*Particle Filters*) methods? Maybe... [see Maclean & Van Vleck 2019 - Next talk.]

Bibliography

[1] Bocquet, M., K. Gurumoorthry, A. Apte, A. Carrassi, C. Grudzien, and C.K.R.T. Jones, 2017. Degenerate Kalman filter error covariances and their convergence onto the unstable subspace. SIAM/ASA J. of Uncertainty Quantification 5, 304-333
[2] Bocquet M. and A. Carrassi, 2017. Four-dimensional ensemble variational data assimilation and the unstable subspace. Tellus A 69, 1304504
[3] Carrassi, A., A. Trevisan and F. Uboldi, 2007. Adaptive observations and assimilation in the unstable subspace by Breeding on the Data Assimilation System. Tellus, 59, 101-113
[4] Carrassi, A., M. Ghil, A. Trevisan and F. Uboldi, 2008a. Data assimilation as a nonlinear dynamical system problem: Stability and convergence of the prediction-assimilation system. Chaos. 1, 023112
[5] Carrassi, A., A. Trevisan, L. Descamps, O. Talagrand and F. Uboldi, 2008. Controlling instabilities along a 3DVar analysis cycle by assimilating in the unstable subspace: a comparison with the EnKF. Nonlin. Proc. Geophys. 15, 503-521
[6] Carrassi, A., S. Vannitsem D. Zupanski and M. Zupanski, 2009. The Maximum Likelihood Ensemble Filter performances in chaotic systems. Tellus. 61, 587-600
[7] Grudzien, C., A. Carrassi and M. Bocquet, 2018a. Asymptotic forecast uncertainty and the unstable subspace in the presence of additive model error. SIAM/ASA J. Uncertainty Quantification. 6(4), 1335- 1363.
[8] Grudzien,C., A. Carrassi and M. Bocquet, 2018b. Chaotic dynamics and the role of covariance inflation for reduced rank Kalman filters with model error. Nonlin. Proc. Geophys., 25, 633-648
[9] Gurumoorthry, K., C. Grudzien, A. Apte, A. Carrassi and C.K.R.T. Jones, 2017. Rank deficiency of Kalman error covariance matrices in linear perfect model. SIAM J. on Control and Optimization 55, 741-759
[10] Maclean, J. and Erik S. Van Vleck, 2019. Projected data assimilation. arxiv.org/abs/1902.04212
[11] Palatella, L., A., Carrassi and A. Trevisan, 2013. Lyapunov vectors and Assimilation in the Unstable Subspace: theory and applications. J. Phys. A: Math. Theor. 46, 254020
[12] Palatella, L. and Trevisan, A., 2015. Interaction of Lyapunov vectors in the formulation of the nonlinear extension of the Kalman filter, Physical Rev. E, 91, 042 905.
[13] Trevisan, A. and Uboldi, F. 2004. Assimilation of standard and targeted observations within the unstable subspace of the observation-analysis-forecast cycle, J. Atmos. Sci., 61, 103–113.
[14] Trevisan, A., D'Isidoro, M., and Talagrand, O., 2010. Four-dimensional variational assimilation in the unstable subspace and the optimal subspace dimension, Q. J. Roy. Meteorol. Soc., 136, 487–496.
[15] Trevisan, A. and Palatella, L., 2011. On the Kalman Filter error covariance collapse into the unstable subspace, Nonlin. Processes Geophys., 18, 243-250.

Get a Ph.D. at the University of Nevada, Reno

- **Ph.D.s** in **Mathematics** (pure or applied) and **Statistics & Data Science**.
- **R1 Research University**.
- **Over 27 research faculty**, more than 10 hired since 2015 and growing.
- **SIAM Student Chapter** Adviser: **Colin Grudzien.**
- **ASA Student Chapter** Adviser: **Mihye Ahn.**

https://www.unr.edu/math