Lecture 3 on data assimilation:
Hybrid/ensemble variational methods and perspectives

Marc Bocquet

CEREA, joint lab École des Ponts ParisTech and EdF R&D, Université Paris-Est, France
Institut Pierre-Simon Laplace

(marc.bocquet@enpc.fr)

Outline

The classical smoother

▶ There are smoothing variants of the Kalman filter [Anderson and Moore 1979], the Kalman smoother which is used in the geosciences [Cohn et al. 1994].

▶ They have been adapted to the EnKF and variants [Evensen and Leeuwen 2000; Evensen 2009], [Cosme et al. 2012], [Bocquet and Sakov 2014], etc.

▶ Whatever the EnKF flavour, the update is of the form (without localisation)

$$\mathbf{E}_k^{\mathrm{a}} = \mathbf{E}_k^{\mathrm{f}} \mathbf{T}_k,$$

where $\mathbf{T}_k$ accounts for $\mathbf{y}_k$.

The classical smoother

▶ The ensemble update at t_l ($t_k \leqslant t_l$) given an observation vector at t_k is

$$\mathbf{E}_l^{\mathrm{a}} = \mathbf{M}_{l:k} \mathbf{E}_k^{\mathrm{a}} = \mathbf{M}_{l:k}(\mathbf{E}_k^{\mathrm{f}} \mathbf{T}_k) = (\mathbf{M}_{l:k} \mathbf{E}_k^{\mathrm{f}}) \mathbf{T}_k = \mathbf{E}_l^{\mathrm{f}} \mathbf{T}_k$$

where $\mathbf{M}_{l:k}$ is the tangent linear model from t_k to t_l.

▶ This can be generalised to the case $t_l \leqslant t_k$, where we need to define $\mathbf{M}_{k:l} = \mathbf{M}_{l:k}^{-1}$.

▶ Within a DAW $[t_0, \ldots, t_k, \ldots, t_K]$, we therefore have the smoothing update:

$$\mathbf{E}_k^{\mathrm{s}} = \mathbf{E}_k^{\mathrm{a}} \prod_{\substack{l=k}}^{l=K} \mathbf{T}_l.$$

The classical smoother

▶ Therefore, a simple variant parametrised by a lag L is based on two passes:

- one filtering pass with an EnKF. Requires to store the ensembles over the lag L,
- the second pass updates those ensembles using future observations (formula above).
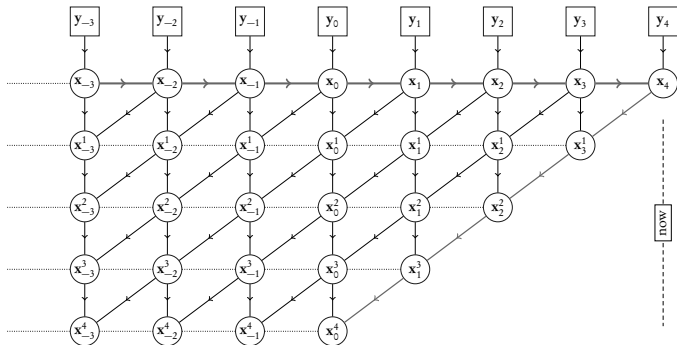
The classical smoother

▶ Synthetic experiment with the Lorenz-96 model, standard configuration.
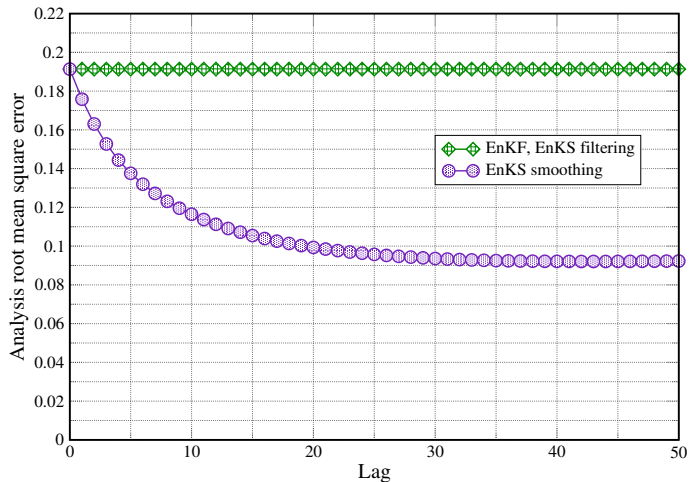
Asynchronous data assimilation for the EnKF

▶ How to simply and efficiently assimilate observations in between two update steps of the EnKF (linear order)? [Hunt, Kostelich, et al. 2007; Sakov, Evensen, et al. 2010]

▶ How to do so in presence of additive model error (linear order)?

$$\mathcal{J}(\mathbf{x}_0, \ldots, \mathbf{x}_K) = \|\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{a}}\|_{(\mathbf{P}_0^{\mathrm{a}})^{-1}}^2 + \sum_{k=0}^{K} \|\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)\|_{\mathbf{R}_k^{-1}}^2$$

$$+ \sum_{k=1}^{K} \|\mathbf{x}_k - \mathcal{M}_{k:k-1}(\mathbf{x}_{k-1})\|_{\mathbf{Q}_k^{-1}}^2 .$$

[Sakov and Bocquet 2018]

Asynchronous data assimilation for the EnKF

▶ Ensemble subspace representation, for $k = 1, \ldots, K$:

$$\mathbf{x}_0 = \mathbf{x}_0^a + \mathbf{X}_a^0 \mathbf{w}_0, \quad \mathbf{X}_a^0 (\mathbf{X}_a^0)^\top = \mathbf{P}_0^a, \quad \mathbf{X}_a^0 \mathbf{1} = 0,$$
$$\mathbf{x}_k = \mathcal{M}_{k:k-1}(\mathbf{x}_{k-1}) + \mathbf{X}_q^k \mathbf{w}_k, \quad \mathbf{X}_q^k (\mathbf{X}_q^k)^\top = \mathbf{Q}_k, \quad \mathbf{X}_q^k \mathbf{1} = 0.$$

▶ Compactification:

$$\mathbf{w} \equiv \mathrm{vec}\,(\mathbf{w}_0, \ldots, \mathbf{w}_K) \in \mathbb{R}^{N_\mathrm{w}} \quad \text{with} \quad N_\mathrm{w} = N_\mathrm{e} + \sum_{k=1}^K N_\mathrm{q}^k.$$

▶ Cost function in ensemble subspace [Desroziers, Camino, et al. 2014; Sakov and Bocquet 2018]:

$$J(\mathbf{w}) = \|\mathbf{w}\|^2 + \|\mathbf{y} - \mathcal{H}(\mathbf{x})\|_{\mathbf{R}^{-1}}^2.$$

Asynchronous data assimilation for the EnKF

▶ Linearisation (Gauss-Newton principle):

$$\mathbf{x} = \mathbf{x}^{\mathrm{f}} + \mathbf{X}\mathbf{w} + O(\|\mathbf{w}\|^2),$$

with

$$\mathbf{x}^{\mathrm{f}} \equiv \mathrm{vec}\left(\{\mathcal{M}_{k:0}(\mathbf{x}_0^{\mathrm{a}})\}_{k=0,\dots,K}\right),$$

and

$$\mathbf{X} \equiv \mathrm{vec}(\mathbf{X}_0, \dots, \mathbf{X}_K) \in \mathbb{R}^{(K+1)N_x \times N_{\mathrm{w}}},$$

$$\mathbf{X}_k \equiv \begin{cases} [\mathbf{X}_{\mathrm{a}}^0, \mathbf{0}], & k = 0 \\ [\mathbf{M}_{k:k-1}\mathbf{X}_{k-1}, \mathbf{X}_k^q, \mathbf{0}], & k = 1, \dots, K-1, \\ [\mathbf{M}_{k-1:k}\mathbf{X}_{k-1}, \mathbf{X}_k^q], & k = K. \end{cases}$$

[Sakov and Bocquet 2018]

Asynchronous data assimilation for the EnKF

▶ Cost function expansion:

$$J(\mathbf{w}) = \|\mathbf{w}\|^2 + \left\| \mathbf{y} - \mathcal{H}(\mathbf{x}^{\mathrm{f}}) - \mathbf{Y}\mathbf{w} + O(\|\mathbf{w}\|^2) \right\|_{\mathbf{R}^{-1}}^2 ,$$

where $\mathbf{Y} \equiv \mathrm{vec}\left(\{\mathbf{H}_k \mathbf{X}_k\}_{k=0,\ldots,K} \right)$.

▶ Linear order analysis (AEnKF-Q):

$$\mathbf{x}^{\star} = \mathbf{x}^{\mathrm{f}} + \mathbf{X}\mathbf{w}^{\star},$$

$$\mathbf{X}^{\star} = \mathbf{X}\mathbf{T}, \quad \mathbf{T} = \mathbf{D}^{-\frac{1}{2}}\mathbf{U},$$

$$\mathbf{w}^{\star} = \mathbf{D}^{-1}\mathbf{Y}^{\top}\mathbf{R}^{-1}\left[\mathbf{y} - \mathcal{H}(\mathbf{x}^{\mathrm{f}}) \right],$$

$$\mathbf{D} \equiv \mathbf{I} + \mathbf{Y}^{\top}\mathbf{R}^{-1}\mathbf{Y}.$$

▶ A linearised smoothing solution in presence of additive model error.

[Sakov and Bocquet 2018]

4D-ETKF

▶ In the absence of model error, the AEnKF-Q drastically simplifies. In particular

$$\mathbf{w} \in \mathbb{R}^{N_e}, \qquad \mathbf{X} \equiv \mathrm{vec}(\mathbf{X}_0, \mathbf{M}_{1:0}\mathbf{X}_0, \ldots, \mathbf{M}_{K:0}\mathbf{X}_0) \in \mathbb{R}^{(K+1)N_x \times N_e}.$$

▶ This yields the 4D-ETKF [Fertig et al. 2007] on an initial idea from [Hunt, Kalnay, et al. 2004].

▶ The 4D-ETKF also coincides with the first iteration of the outer loop of the IEnKS [Bocquet and Sakov 2014].

Outline

Hybrids

▶ Hybrid often refers to a combination of a variational method and of an EnKF.

▶ But, in practice, refers to hybridising a static error covariance matrix with a dynamical one sampled from an ensemble [Hamill and Snyder 2000], with the goal to use it in the analysis step of an EnKF. Effective covariance matrix:

$$\mathbf{B} = \alpha \mathbf{C} + (1-\alpha)\mathbf{X}_{\mathrm{f}}\mathbf{X}_{\mathrm{f}}^{\top},$$

where $\mathbf{C}$ is the static error covariance matrix, $\mathbf{X}_{\mathrm{f}}$ the matrix of the forecast ensemble anomalies, and $\alpha \in [0,1]$.

▶ In a stochastic EnKF, the updated ensemble can be obtained using several stochastically perturbed variational problems.

▶ Note that, as with the EnKF, it may be necessary to enforce localisation of the sample covariances:

$$\mathbf{B} = \alpha \mathbf{C} + (1-\alpha)\boldsymbol{\rho} \circ \left[\mathbf{X}_{\mathrm{f}}\mathbf{X}_{\mathrm{f}}^{\top}\right].$$

Ensemble of data assimilation

▶ Ensemble of data assimilations (EDA) stems from NWP centres that operate a 4DVar (Météo-France, the ECMWF, etc.)

▶ Idea: introduce dynamical errors that are absent in the traditional 4DVar.

▶ In order to build on the existing 4DVar systems, consider an ensemble of N_e 4D-Var analyses. Each analysis, indexed by i, uses a different first guess $\mathbf{x}_0^i$, and observations perturbed with $\boldsymbol{\epsilon}_k^i \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)$ to maintain statistical consistency. Hence, each analysis i carries out the minimisation of

$$\mathcal{J}_i^{\text{EDA}}(\mathbf{x}_0) = \frac{1}{2} \sum_{k=0}^{K} \left\| \mathbf{y}_k + \boldsymbol{\epsilon}_k^i - \mathcal{H}_k \circ \mathcal{M}_{k:0}(\mathbf{x}_0) \right\|_{\mathbf{R}_k^{-1}}^2 + \frac{1}{2} \left\| \mathbf{x}_0 - \mathbf{x}_0^i \right\|_{\mathbf{B}^{-1}}^2 .$$

This yields an updated ensemble, from which it is possible to assess a dynamical part of the error covariance matrix. Close to the idea of the hybrid EnKF-3DVar, but with a 4DVar scheme. [Raynaud et al. 2009; Raynaud et al. 2011; Berre et al. 2015; Bonavita, Raynaud, et al. 2011; Bonavita, Isaksen, et al. 2012; Jardak and Talagrand 2018].
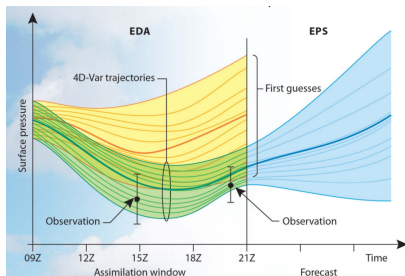
4DEnVar

▶ 4DEnVar has been developed by NWP centres with a 4D-Var in operation.
Primary goal: avoid maintaining the adjoint of the forecast model.

- The analysis is performed in the subspace spanned by the ensemble [Liu et al. 2008].

- The perturbations are usually generated stochastically, for instance using a stochastic EnKF [Liu et al. 2009; Buehner, Houtekamer, et al. 2010a].

- The 4DEnVar implementations usually come with an hybrid background as they have been developed in NWP centres with a 4D-Var

▶ Localisation is necessary and theoretically more challenging than with an EnKF [Bocquet 2016; Desroziers, Arbogast, et al. 2016].

▶ Many variants of the 4DEnVar exists, depending on the availability of the adjoint models and how the perturbations are generated [Buehner, Houtekamer, et al. 2010a; Buehner, Houtekamer, et al. 2010b; M. Zhang and F. Zhang 2012; Clayton et al. 2013; Poterjoy and F. Zhang 2015]. Several 4DEnVar systems are now operational or on the verge or being so [Buehner, Morneau, et al. 2013; Gustafsson et al. 2014; Desroziers, Camino, et al. 2014; Lorenc et al. 2015; Kleist and Ide 2015; Buehner, McTaggart-Cowan, et al. 2015; Bannister 2017].

▶ A further recent sophistication is to create an EDA of 4DEnVar in order to generate the perturbations [Bowler et al. 2017; Arbogast et al. 2017].

Outline

1. A selection of smoothers
   - Classical smoother
   - Asynchronous EnKF
   - 4D-ETKF

2. Hybrid and ensemble variational techniques
   - Hybrid techniques
   - Ensemble of data assimilation
   - 4DEnVar

3. **The iterative ensemble Kalman smoother**
   - **Theory**
   - **Numerical experiments**

4. Machine learning and data assimilation
   - Emergence of machine learning in data assimilation
   - Numerical example
   - Challenges

5. References

Iterative ensemble Kalman filter/smoother: origin

▶ The iterative extended Kalman filter [Wishner et al. 1969; Jazwinski 1970] IEKF
▶ The iterative extended Kalman smoother [Bell 1994] IEKS

Much too costly + needs the TLM and the adjoint $\longrightarrow$ ensemble methods

▶ The iterative ensemble Kalman filter [Sakov, Oliver, et al. 2012; Bocquet and Sakov 2012] IEnKF
▶ The iterative ensemble Kalman smoother [Bocquet and Sakov 2014] IEnKS

TLM and adjoint free if need be, nonlinear analysis scheme, Gaussian consistent updating of the perturbations

▶ About terminology: The IEnKS does both filtering and smoothing.
As in [Bell 1994], and as for 4D-Var, smoothing primarily means that the analysis is smoothed out over a time-window. The distinction between filtering and smoothing is not as clear cut as in the classical (non-iterative) EnKS.

Iterative ensemble Kalman filter/smoother: contributions

▶ Jazwinksi and Bell' ideas transposed to ensemble methods: a chain of papers:

- [Zupanski 2005]: Nonlinear variational analysis in an EnKF framework.

- [Gu and Oliver 2007]: Initial idea.

- [Yang et al. 2012]: RIP: a closely related idea.

- [Sakov, Oliver, et al. 2012]: The key idea.

- [Bocquet and Sakov 2012]: Correction of the bundle and transform schemes.

- [Bocquet and Sakov 2014]: 4D analysis scheme + richer cycling (overlapping DAWs).

- [Bocquet 2016]: Localisation of the scheme, covariant localisation.

- [Fillion, Bocquet, and Gratton 2018]: Quasi-static variant.

- [Sakov and Bocquet 2018; Sakov, Haussaire, et al. 2018; Fillion, Bocquet, Sakov, et al. 2018]:
  Extension to model error (IEnKF-Q and AEnKF-Q).

▶ The IEnKS is a sleek archetype for most 4D-EnVar methods.
Derived from Bayes' rule, it helps identify the fundamental approximations implicitly made in 4D EnVar variants.

The IEnKS: the cycling

▶ L: length of the data assimilation window,

▶ S: shift of the data assimilation window in between two updates.
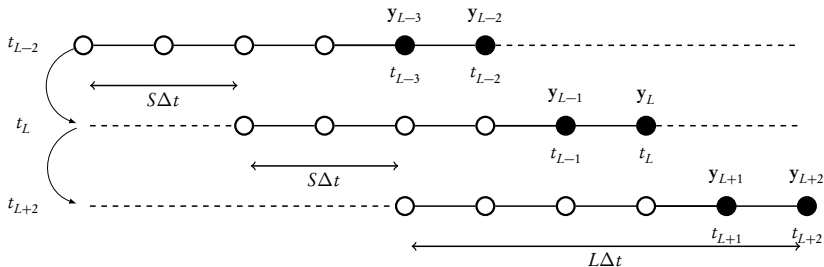
$$\boxed{\text{Variational analysis in ens. space}} \rightarrow \boxed{\text{Posterior ens. generation}} \rightarrow \boxed{\text{Ens. forecast}}$$

The IEnKS: a variational standpoint

▶ Analysis IEnKS cost function in state space $p(\mathbf{x}_0|\mathbf{y}_{L:-\infty}) \propto \exp(-\mathcal{J}(\mathbf{x}_0))$:

$$\mathcal{J}(\mathbf{x}_0) = \frac{1}{2} \sum_{k=1}^{L} \|\mathbf{y}_k - \mathcal{F}_{k:0}(\mathbf{x}_0)\|^2_{\beta_k \mathbf{R}_k^{-1}} + \frac{1}{2}\|\mathbf{x}_0 - \overline{\mathbf{x}}_0\|_{\mathbf{P}_0^{-1}},$$

where $\mathcal{F}_{k:0} = \mathcal{H}_k \circ \mathcal{M}_{k:0}$ and $\{\beta_0, \beta_1, \ldots, \beta_L\}$ weight the observations within the window.

▶ Reduced scheme in ensemble space, $\mathbf{x}_0 = \overline{\mathbf{x}}_0 + \mathbf{X}_0 \mathbf{w}$, where $\mathbf{X}_0$ is the ensemble anomaly matrix:

$$J(\mathbf{w}) = \mathcal{J}(\overline{\mathbf{x}}_0 + \mathbf{X}_0 \mathbf{w}).$$

▶ IEnKS cost function in ensemble subspace:

$$J(\mathbf{w}) = \frac{1}{2}\sum_{k=1}^{L}\|\mathbf{y}_k - \mathcal{F}_{k:0}(\overline{\mathbf{x}}_0 + \mathbf{X}_0\mathbf{w})\|^2_{\beta_k \mathbf{R}_k^{-1}} + \frac{1}{2}\|\mathbf{w}\|^2.$$

The IEnKS: minimisation schemes

▶ As a variational reduced method, one can use Gauss-Newton [Sakov, Oliver, et al. 2012], Levenberg-Marquardt [Sakov, Oliver, et al. 2012], quasi-Newton techniques, trust region, etc., minimisation schemes.

Gauss-Newton scheme (the Hessian is approximate)

$$\mathbf{w}^{(p+1)} = \mathbf{w}^{(p)} - \widetilde{\mathcal{H}}_{(p)}^{-1} \nabla J_{(p)},$$

$$\mathbf{x}_0^{(p)} = \mathbf{x}_0^{(0)} + \mathbf{X}_0 \mathbf{w}^{(p)},$$

$$\nabla J_{(p)} = -\sum_{k=1}^{L} \mathbf{Y}_{k,(p)}^{\mathrm{T}} \beta_k \mathbf{R}_k^{-1} \left(\mathbf{y}_k - \mathcal{F}_{k:0}(\mathbf{x}_0^{(p)}) \right) + \mathbf{w}^{(p)},$$

$$\widetilde{\mathcal{H}}_{(p)} = \mathbf{I}_{\mathrm{e}} + \sum_{k=1}^{L} \mathbf{Y}_{k,(p)}^{\mathrm{T}} \beta_k \mathbf{R}_k^{-1} \mathbf{Y}_{k,(p)},$$

$$\mathbf{Y}_{k,(p)} = [\mathcal{F}_{k:0}]'_{|\mathbf{x}_0^{(p)}} \mathbf{X}_0.$$

The IEnKS: computing the sensitivities

▶ Sensitivities $\mathbf{Y}_{k,(p)}$ computed by ensemble propagation without TLM and adjoint [Gu and Oliver 2007; Liu et al. 2008; Buehner, Houtekamer, et al. 2010a]

▶ First option [Sakov, Oliver, et al. 2012]: the transform scheme. The ensemble is preconditioned before its propagation using the ensemble transform

$$
\mathbf{T}_{k,(p)} = \left(\mathbf{I}_{\mathrm{e}} + \sum_{k=1}^{L} \beta_k \mathbf{Y}_{k,(p)}^{\top} \mathbf{R}_k^{-1} \mathbf{Y}_{k,(p)} \right)^{-1/2},
$$

obtained at the previous iteration. The inverse transformation is applied after propagation.

▶ Second option [Bocquet and Sakov 2012]: the bundle scheme. It simply mimics the action of the tangent linear by finite difference:

$$
\mathbf{Y}_{k,(p)} \approx \frac{1}{\varepsilon} \mathcal{F}_{k:0} \left(\mathbf{x}^{(p)} \mathbf{1}^{\top} + \varepsilon \mathbf{X}_0 \right) \left(\mathbf{I}_{\mathrm{e}} - \frac{\mathbf{1}\mathbf{1}^{\top}}{N_{\mathrm{e}}} \right).
$$

▶ We found that the transform and the bundle variants perform almost equally well on the tested cases, provided the nonlinearity is not too strong.

The IEnKS: ensemble update and the forecast step

▶ Perturbation update: same as the ETKF:

$$\mathbf{E}_0^\star = \mathbf{x}_0^\star \mathbf{1}^\top + \sqrt{N_e - 1}\mathbf{X}_0 \widetilde{\mathcal{H}}_\star^{\frac{1}{2}} \mathbf{U} \quad \text{where} \quad \mathbf{U1} = \mathbf{1}.$$

(Alternative updates for state-space/stochastique/hybrid formulations.)

▶ Forecast: propagate the updated ensemble from t_0 to t_S:

$$\mathbf{E}_S = \mathcal{M}_{S:0}(\mathbf{E}_0).$$

The IEnKS: A prototype for nonlinear 4DEnVar schemes

▶ The IEnKS offers a sleek prototype of nonlinear four-dimensional ensemble variational methods. It can be derived from Bayes'rule.

▶ Because $\mathcal{H} \longrightarrow \mathcal{H} \circ \mathcal{M}$, the MLEF can be seen as a subcase of the IEnKS.

▶ 4D-ETKF can be seen as a subcase of the IEnKS when using the one iteration and the ensemble transform sensitivity estimation.

▶ 4DEnVar: can also be seen as a subcase of the IEnKS (when using the ensemble transform sensitivity estimation).

The IEnKS: single vs multiple data assimilation (1/2)

▶ More than 4D-Var, the IEnKS cycling questions the cycling of DA schemes: how to chain the data assimilation windows? How to best overlap time-windows?
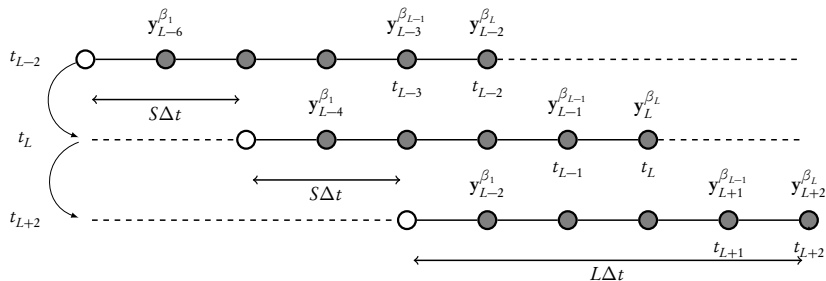
▶ SDA IEnKS: The observation vectors are assimilated once and for all. Exact scheme.

▶ MDA IEnKS: The observation vectors are assimilated several times and pondered with weights β_k within the window.

The IEnKS: single vs multiple data assimilation (2/2)

▶ Two flavors of Multiple Data Assimilation:

- The splitting of observations: Following the partition $1 = \sum_{k=1}^{L} \beta_k$, the observation vector $\mathbf{y}$ with prior error covariance matrix is split into L partial observation $\mathbf{y}^{\beta_k}$, with prior error covariance matrix $\beta_k^{-1} \mathbf{R}_k$.
  It is a consistent approach in the Gaussian/linear limit, and one hopes it is still so in nonlinear conditions [Emerick and Reynolds, 2012].

- The multiple assimilation of each observation with its original weights. It is (heuristically) correct but the filtering/smoothing pdf (essentially) becomes a power of the searched pdf!

▶ [Optional] An extra step in the analysis:

- MDA IEnKS does not immediately yield the *filtering* pdf.

- To approach the correct filtering/smoothing pdf, one needs an extra step, that we called the balancing step which reweights the observations within the data assimilation window, and perform a final analysis.

Application to the Lorenz-96 model

▶ Chaotic low-order model: Lorenz-96, $N_x = 40$, $N_e = 20$, $\Delta t = 0.05$, $\mathbf{R} = \mathbf{I}$.

▶ Comparison of 4D-Var $S = 1$, EnKS $S = 1$, SDA IEnKS $S = 1$, SDA IEnKS $S = L$, and MDA IEnKS $S = 1$ [Bocquet and Sakov 2013].
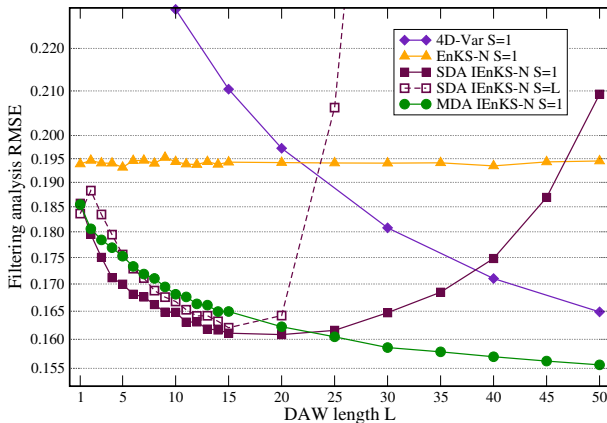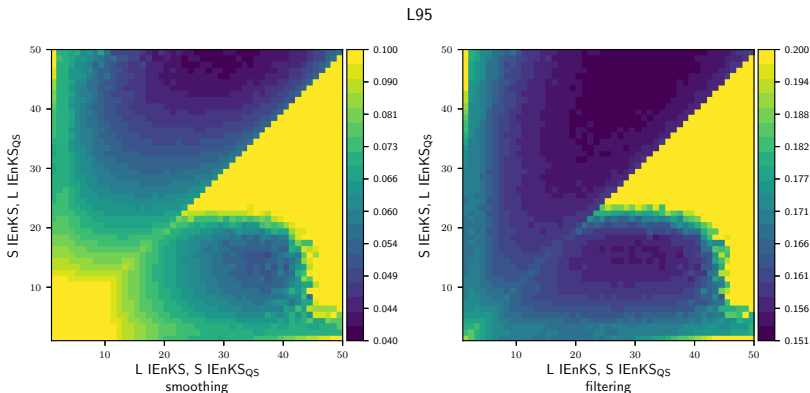
Application to the Lorenz-96 model
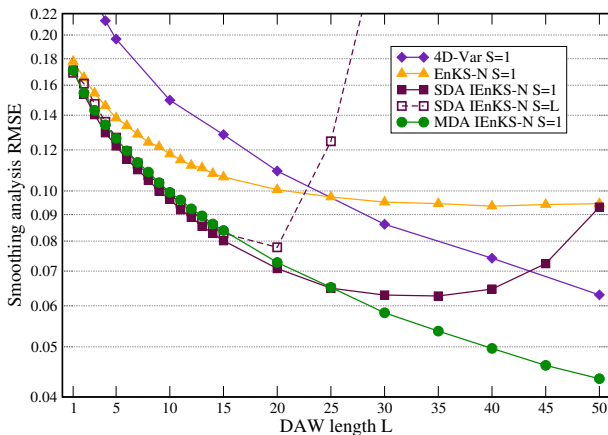
▶ Systematic outperformance of the IEnKS over the EnKF, 4D-Var, EnKS in all regimes for chaotic models (L63, L96, NEDYM, L96-T, L96-GRS, 2D forced turbulence, QG model) in perfect model conditions [Bocquet and Sakov 2013].

Quasi-static IEnKS

▶ Idea inspired from [Pires et al. 1996]: Assimilate progressively the observations within the DAW so that the cost function gradually changes and that one can track its global minimum. This only changes the initial point of the minimisation.

▶ Extension to the IEnKS is natural [Fillion, Bocquet, and Gratton 2018], and, to a large extend, avoids the need for MDA.

L95

Outline

Emergence of machine learning techniques

▶ Machine learning and data assimilation are both part of estimation theory.
They are glorified regression techniques.

▶ They both intent to make generalisations/predictions.

▶ They share standpoints (Bayesian statistics) and some techniques (regression, optimisation).

▶ However,

- Data assimilation uses large dataset and costly models.
- Machine learning uses huge/large datasets and cheap parametrised models.

Emergence of machine learning techniques
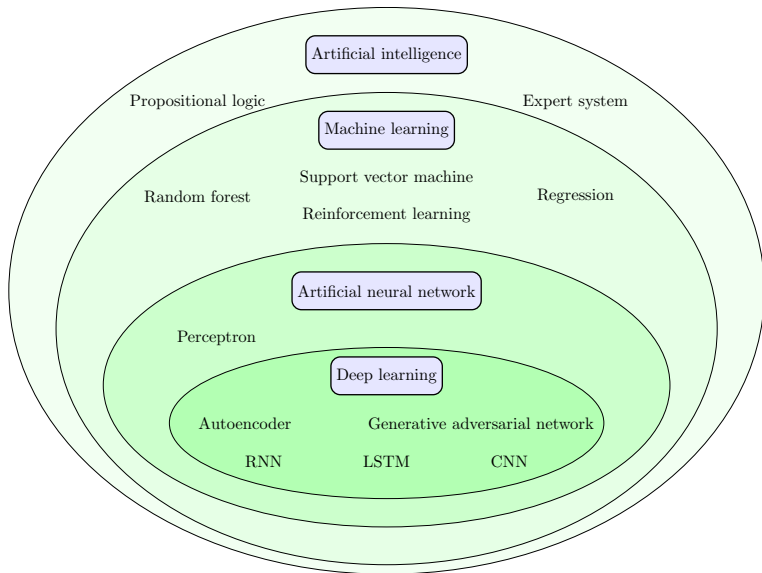
Emergence of machine learning techniques

▶ Why this ML tsunami?

- New sparse representations of data that yields better and numerically affordable optimisation.

- Relies on deep learning libraries (Tensorflow, Pytorch, Keras, etc.) powered by Google, Facebook, Microsoft, Nvidia, etc.

▶ Why this ongoing ML hype in geophysical data assimilation?

- Huge success of deep learning in vision, speech recognition and AI in general. This makes it fashionable in geophysics.

- Forces to reconsider difficult questions (model error). Gives an alibi to reconsider those questions!

- Some of the application in vision, speech can be straightforwardly extrapolated to geosciences.

What can ML bring to NWP and data assimilation?

▶ Advanced quality control of observations and forecasts.

▶ Emulate, build surrogate models for subpart of the main forecast model, for instance subgrid scale parametrisations, microphysics, convection parametrisations, etc.

▶ Improvement of existing DA schemes, especially ensemble-based methods. Substitute for the analysis, refinement and regularisation of existing DA schemes.

▶ Bias correction, residual model error correction with application to forecasting and re-analysis.

▶ Generate tangent linear and adjoint of emulated components of the model.

▶ Postprocessing, downscaling: advanced and nonlinear statistical adaptation and correction, downscaling, feature detection, feature extraction.

[Dueben and Bauer 2018; Reichstein et al. 2019; Bolton and Zanna 2019] and many others.

ODE representation for the surrogate model

▶ Ordinary differential equations (ODEs) representation of the surrogate dynamics

$$\frac{d\mathbf{x}}{dt} = \boldsymbol{\phi}_{\mathbf{A}}(\mathbf{x}), \qquad \boldsymbol{\phi}_{\mathbf{A}}(\mathbf{x}) = \mathbf{A}\mathbf{r}(\mathbf{x}),$$

where
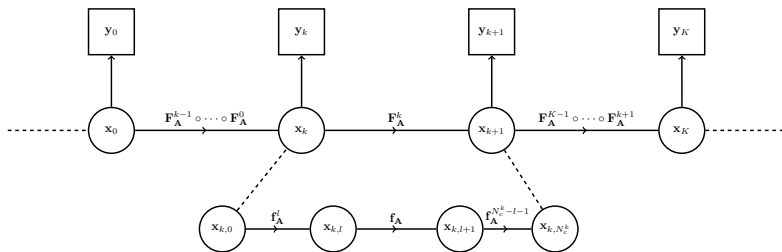
- **A** is a matrix of coefficients of size $N_x \times N_p$,
- $\mathbf{r}(\mathbf{x})$ is a vector of nonlinear regressors of size N_p.

[Bocquet, Brajard, et al. 2019; J. Brajard et al. 2019]

Integration scheme and cycling

▶ Compositions of integration schemes:

$$\mathbf{x}_{k+1} = \mathbf{F}_\mathbf{A}^k(\mathbf{x}_k) \quad \text{where} \quad \mathbf{F}_\mathbf{A}^k \equiv \mathbf{f}_\mathbf{A}^{N_c^k} \equiv \underbrace{\mathbf{f}_\mathbf{A} \circ \ldots \circ \mathbf{f}_\mathbf{A}}_{N_c^k \text{ times}},$$

Bayesian analysis of the problem

▶ Bayesian view on state and model estimation:

$$p(\mathbf{A}, \mathbf{x}_{0:K}|\mathbf{y}_{0:K}) = \frac{p(\mathbf{y}_{0:K}|\mathbf{x}_{0:K}, \mathbf{A})p(\mathbf{x}_{0:K}|\mathbf{A})p(\mathbf{A})}{p(\mathbf{y}_{0:K})}.$$

▶ Data assimilation cost function assuming Gaussian error statistics and Markovian dynamics:

$$\mathcal{J}(\mathbf{A}, \mathbf{x}_{0:K}) = \frac{1}{2}\sum_{k=0}^{K}\|\mathbf{y}_k - \mathbf{H}_k(\mathbf{x}_k)\|_{\mathbf{R}_k^{-1}}^2 + \frac{1}{2}\sum_{k=1}^{K}\left\|\mathbf{x}_k - \mathbf{F}_\mathbf{A}^{k-1}(\mathbf{x}_{k-1})\right\|_{\mathbf{Q}_k^{-1}}^2 - \ln p(\mathbf{x}_0, \mathbf{A}).$$

⟶ Allows to handle partial and noisy observations.

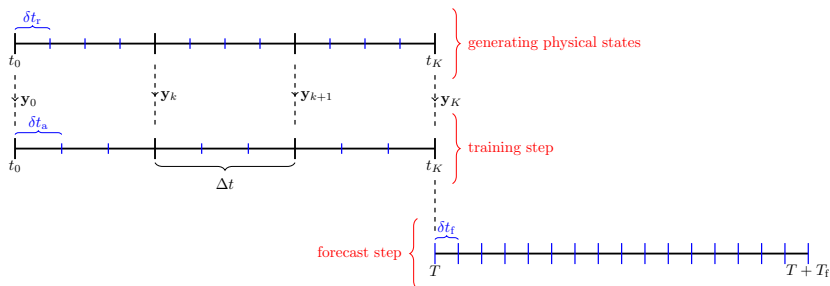▶ Typical machine learning cost function with $\mathbf{H}_k = \mathbf{I}_k$ in the limit $\mathbf{R}_k \longrightarrow \mathbf{0}$:

$$\mathcal{J}(\mathbf{A}) \approx \frac{1}{2}\sum_{k=1}^{K}\left\|\mathbf{y}_k - \mathbf{F}_\mathbf{A}^{k-1}(\mathbf{y}_{k-1})\right\|_{\mathbf{Q}_k^{-1}}^2 - \ln p(\mathbf{y}_0, \mathbf{A}).$$

[Abarbanel et al. 2018; Bocquet, Brajard, et al 2019; Hsieh and Tang 1998]

Experiment plan

▶ The reference model, the surrogate model and the forecasting system

▶ Metrics of comparison:

- Model: ODE coefficients norm $\|\mathbf{A}_{\mathrm{a}} - \mathbf{A}_{\mathrm{r}}\|_{\infty}$
- RMSE between the reference and the surrogate forecasts as a function of the lead time (averaged over many initial conditions).

Identifiable model and perfect observations

▶ Inferring the dynamics from dense & noiseless observations of identifiable models

- The Lorenz 63 model (L63, 3 variables):

$$\frac{\mathrm{d}x_0}{\mathrm{d}t} = \sigma(x_1 - x_0),$$

$$\frac{\mathrm{d}x_1}{\mathrm{d}t} = \rho x_0 - x_1 - x_0 x_2,$$

$$\frac{\mathrm{d}x_2}{\mathrm{d}t} = \rho x_0 x_1 - \beta x_2,$$

$\longrightarrow \|\mathbf{A}_a - \mathbf{A}_r\|_\infty \sim 10^{-13}$ almost perfect reconstruction at machine precision.

- The Lorenz 96 model (L96, 40 variables)

$$\frac{\mathrm{d}x_n}{\mathrm{d}t} = (x_{n+1} - x_{n-2})x_{n-1} - x_n + F,$$

$\longrightarrow \|\mathbf{A}_a - \mathbf{A}_r\|_\infty \sim 10^{-13}$ close to perfect reconstruction at machine precision.

Example of dynamics reconstruction

▶ Inferring the dynamics from dense & noiseless observations of a non-identifiable model

The Lorenz 96 model (40 variables). Surrogate model based on an RK2 scheme. Analysis of the modelling depth as a function of N_c.
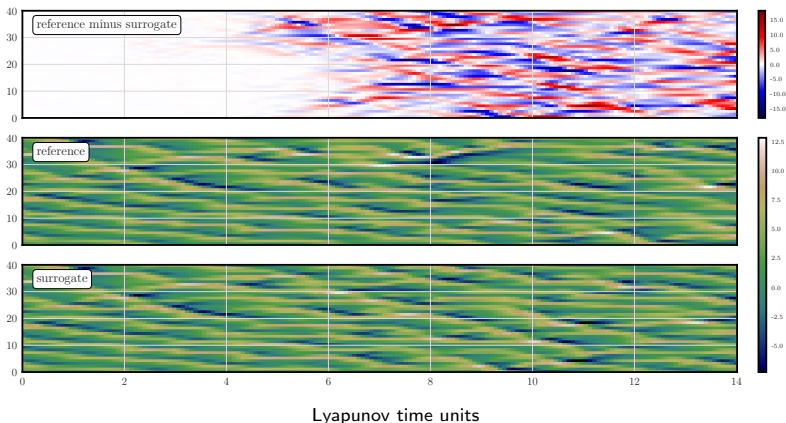
Lyapunov time units

Example of dynamics reconstruction

▶ Inferring the dynamics from dense & noiseless observations of a non-identifiable model

The Kuramoto-Sivashinski (KS) model (continuous, 128 variables).

$$\frac{\partial u}{\partial t} = -u\frac{\partial u}{\partial x} - \frac{\partial^2 u}{\partial x^2} - \frac{\partial^4 u}{\partial x^4},$$
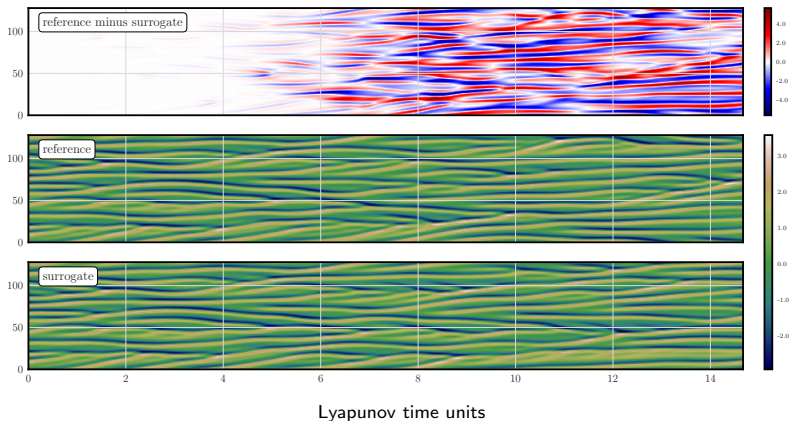

Lyapunov time units

Example of dynamics reconstruction

▶ Inferring the dynamics from dense & noiseless observations of a non-identifiable model

The Kuramoto-Sivashinski (KS) model (continuous, 128 variables).

$$\frac{\partial u}{\partial t} = -u\frac{\partial u}{\partial x} - \frac{\partial^2 u}{\partial x^2} - \frac{\partial^4 u}{\partial x^4},$$
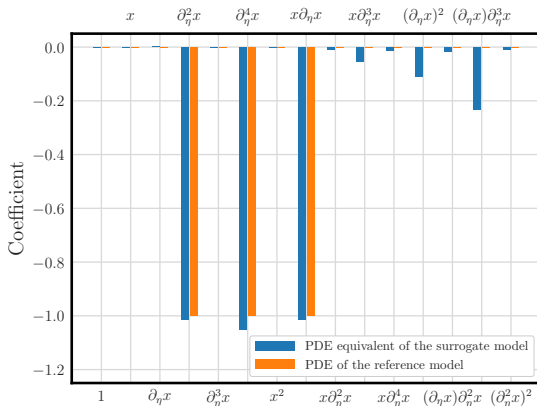
Challenges

▶ Difficulties to extract physical knowledge from tuned NNs.

▶ How to enforce physical constraints (energy, vorticity conservation laws) in NN architectures?

▶ How to enforce local and global symmetries?

▶ Tuning statistical hyperparameters is difficult and costly ($\sim$ DA).

▶ Nonlinear optimisation issues (underlying cost function may have local minima: not as safe as in sequential DA).

▶ How to interface NNs (mostly developed in python) with legacy codes (mostly developed in fortran, at best in C++)?

[Dueben and Bauer 2018]

References I

[1] H. D. I. Abarbanel, P. J. Rozdeba, and S. Shirman. "Machine Learning: Deepest Learning as Statistical Data Assimilation Problems". In: *Neural Computation* 30 (2018), pp. 2025–2055. eprint: https://doi.org/10.1162/neco_a_01094.

[2] B. D. O. Anderson and J. B. Moore. *Optimal Filtering*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc, 1979, p. 357.

[3] E. Arbogast, G. Desroziers, and L. Berre. "A parallel implementation of a 4DEnVar ensemble". In: *Q. J. R. Meteorol. Soc.* 143 (2017), pp. 2073–2083.

[4] R. N. Bannister. "A review of operational methods of variational and ensemble-variational data assimilation". In: *Q. J. R. Meteorol. Soc.* 143 (2017), pp. 607–633.

[5] B. M. Bell. "The iterated Kalman smoother as a Gauss-Newton method". In: *SIAM J. Optim.* 4 (1994), pp. 626–636.

[6] L. Berre, H. Varella, and G. Desroziers. "Modelling of flow-dependent ensemble-based background-error correlations using a wavelet formulation in 4D-Var at Météo-France". In: *Q. J. R. Meteorol. Soc.* 141 (2015), pp. 2803–2812.

[7] M. Bocquet. "Localization and the iterative ensemble Kalman smoother". In: *Q. J. R. Meteorol. Soc.* 142 (2016), pp. 1075–1089.

[8] M. Bocquet, J Brajard, et al. "Data assimilation as a learning tool to infer ordinary differential equation representations of dynamical models". In: *Nonlin. Processes Geophys.* 26 (2019), pp. 143–162.

[9] M. Bocquet and P. Sakov. "An iterative ensemble Kalman smoother". In: *Q. J. R. Meteorol. Soc.* 140 (2014), pp. 1521–1535.

[10] M. Bocquet and P. Sakov. "Combining inflation-free and iterative ensemble Kalman filters for strongly nonlinear systems". In: *Nonlin. Processes Geophys.* 19 (2012), pp. 383–399.

[11] M. Bocquet and P. Sakov. "Joint state and parameter estimation with an iterative ensemble Kalman smoother". In: *Nonlin. Processes Geophys.* 20 (2013), pp. 803–818.

[12] T. Bolton and L. Zanna. "Applications of Deep Learning to Ocean Data Inference and Subgrid Parameterization". In: *Journal of Advances in Modeling Earth Systems* 11 (2019), pp. 376–399.

[13] M. Bonavita, L. Isaksen, and E. Hólm. "On the use of EDA background error variances in the ECMWF 4D-Var". In: *Q. J. R. Meteorol. Soc.* 138 (2012), pp. 1540–1559.

[14] M. Bonavita, L. Raynaud, and L. Isaksen. "Estimating background-error variances with the ECMWF ensemble of data assimilation system: some effects of ensemble size and day-to-day variability". In: *Q. J. R. Meteorol. Soc.* 137 (2011), pp. 423–434.

[15] N. E. Bowler et al. "Inflation and localization tests in the development of an ensemble of 4D-ensemble variational assimilations". In: *Q. J. R. Meteorol. Soc.* 143 (2017), pp. 1280–1302.

References II

[16] J. Brajard et al. "Combining data assimilation and machine learning to emulate a dynamical model from sparse and noisy observations: a case study with the Lorenz 96 model". In: *Geosci. Model Dev. Discuss.* 2019 (2019), pp. 1–21.

[17] M. Buehner, P. L. Houtekamer, et al. "Intercomparison of Variational Data Assimilation and the Ensemble Kalman Filter for Global Deterministic NWP. Part I: Description and Single-Observation Experiments". In: *Mon. Wea. Rev.* 138 (2010), pp. 1550–1566.

[18] M. Buehner, P. L. Houtekamer, et al. "Intercomparison of Variational Data Assimilation and the Ensemble Kalman Filter for Global Deterministic NWP. Part II: One-Month Experiments with Real Observations". In: *Mon. Wea. Rev.* 138 (2010), pp. 1567–1586.

[19] M. Buehner, R. McTaggart-Cowan, et al. "Implementation of Deterministic Weather Forecasting Systems based on Ensemble-Variational Data Assimilation at Environment Canada. Part I: The Global System". In: *Mon. Wea. Rev.* 143 (2015), pp. 2532–2559.

[20] M. Buehner, J. Morneau, and C. Charette. "Four-dimensional ensemble-variational data assimilation for global deterministic weather prediction". In: *Nonlin. Processes Geophys.* 20 (2013), pp. 669–682.

[21] A. M. Clayton, A. C. Lorenc, and D. M. Barker. "Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office". In: *Q. J. R. Meteorol. Soc.* 139 (2013), pp. 1445–1461.

[22] S. E. Cohn, N. S. Sivakumaran, and R. Todling. "A Fixed-Lag Kalman Smoother for Retrospective Data Assimilation". In: *Mon. Wea. Rev.* 122 (1994), pp. 2838–2867.

[23] E. Cosme et al. "Smoothing problems in a Bayesian framework and their linear Gaussian solutions". In: *Mon. Wea. Rev.* 140 (2012), pp. 683–695.

[24] G. Desroziers, E. Arbogast, and L. Berre. "Improving spatial localization in 4DEnVar". In: *Q. J. R. Meteorol. Soc.* 142 (2016), pp. 3171–3185.

[25] G. Desroziers, J.-T. Camino, and L. Berre. "4DEnVar: link with 4D state formulation of variational assimilation and different possible implementations". In: *Q. J. R. Meteorol. Soc.* 140 (2014), pp. 2097–2110.

[26] P. D. Dueben and P. Bauer. "Challenges and design choices for global weather and climate models based on machine learning". In: *Geosci. Model Dev.* 11 (2018), pp. 3999–4009.

[27] G. Evensen. *Data Assimilation: The Ensemble Kalman Filter*. Second. Springer-Verlag Berlin Heildelberg, 2009, p. 307.

[28] G. Evensen and P. J. van Leeuwen. "An Ensemble Kalman Smoother for Nonlinear Dynamics". In: *Mon. Wea. Rev.* 128 (2000), pp. 1852–1867.

[29] E. J. Fertig, J. Harlim, and B. R. Hunt. "A comparative study of 4D-VAR and a 4D Ensemble Kalman Filter: perfect model simulations with Lorenz-96". In: *Tellus A* 59 (2007), pp. 96–100.

[30] A. Fillion, M. Bocquet, and S. Gratton. "Quasi static ensemble variational data assimilation: a theoretical and numerical study with the iterative ensemble Kalman smoother". In: *Nonlin. Processes Geophys.* 25 (2018), pp. 315–334.

References III

[31] A. Fillion, M. Bocquet, P. Sakov, et al. "An iterative ensemble Kalman smoother in presence of additive model error". In: *SIAM/ASA J. Uncertainty Quantification* 0 (2018). Submitted, pp. 0-0.

[32] Y. Gu and D. S. Oliver. "An Iterative Ensemble Kalman Filter for Multiphase Fluid Flow Data Assimilation". In: *SPE Journal* 12 (2007), pp. 438–446.

[33] N. Gustafsson, J. Bojarova, and O. Vignes. "A hybrid variational ensemble data assimilation for the HIgh Resolution Limited Area Model (HIRLAM)". In: *Nonlin. Processes Geophys.* 21 (2014), pp. 303–323.

[34] T. M. Hamill and C. Snyder. "A Hybrid Ensemble Kalman Filter-3D Variational Analysis Scheme". In: *Mon. Wea. Rev.* 128 (2000), pp. 2905–2919.

[35] W. W. Hsieh and B. Tang. "Applying Neural Network Models to Prediction and Data Analysis in Meteorology and Oceanography". In: *Bull. Amer. Meteor. Soc.* 79 (1998), pp. 1855–1870.

[36] B. R. Hunt, E. Kalnay, et al. "Four-dimensional ensemble Kalman filtering". In: *Tellus A* 56 (2004), pp. 273–277.

[37] B. R. Hunt, E. J. Kostelich, and I. Szunyogh. "Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter". In: *Physica D* 230 (2007), pp. 112–126.

[38] M. Jardak and O. Talagrand. "Ensemble Variational Assimilation as a Probabilistic Estimator. Part I: The linear and weak non-linear case". In: *Nonlin. Processes Geophys. Discuss.* 2018 (2018), pp. 1–39.

[39] A. H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, New-York, 1970, p. 376.

[40] D. T. Kleist and K. Ide. "An OSSE-based evaluation of hybrid variational-ensemble data assimilation for the NCEP GFS, Part II: 4DEnVar and hybrid variants". In: *Mon. Wea. Rev.* 143 (2015), pp. 452–470.

[41] C. Liu, Q. Xiao, and B. Wang. "An Ensemble-Based Four-Dimensional Variational Data Assimilation Scheme. Part I: Technical Formulation and Preliminary Test". In: *Mon. Wea. Rev.* 136 (2008), pp. 3363–3373.

[42] C. Liu, Q. Xiao, and B. Wang. "An Ensemble-Based Four-Dimensional Variational Data Assimilation Scheme. Part II: Observing System Simulation Experiments with Advanced Research WRF (ARW)". In: *Mon. Wea. Rev.* 137 (2009), pp. 1687–1704.

[43] A. C. Lorenc et al. "Comparison of hybrid-4DEnVar and hybrid-4DVar data assimilation methods for global NWP". In: *Mon. Wea. Rev.* 143 (2015), pp. 2012–229.

[44] C. Pires, R. Vautard, and O. Talagrand. "On extending the limits of variational assimilation in nonlinear chaotic systems". In: *Tellus A* 48 (1996), pp. 96–121.

References IV

[45] J. Poterjoy and F. Zhang. "Systematic Comparison of Four-Dimensional Data Assimilation Methods With and Without the Tangent Linear Model Using Hybrid Background Error Covariance: E4DVar versus 4DEnVar". In: *Mon. Wea. Rev.* 143 (2015), pp. 1601–1621.

[46] L. Raynaud, L. Berre, and G. Desroziers. "An extended specification of flow-dependent background error variances in the Météo-France global 4D-Var system". In: *Q. J. R. Meteorol. Soc.* 137 (2011), pp. 607–619.

[47] L. Raynaud, L. Berre, and G. Desroziers. "Objective filtering of ensemble-based background-error variances". In: *Q. J. R. Meteorol. Soc.* 135 (2009), pp. 1177–1199.

[48] M Reichstein et al. "Deep learning and process understanding for data-driven Earth system science". In: *Nature* 566 (2019), pp. 195–204.

[49] P. Sakov and M. Bocquet. "Asynchronous data assimilation with the EnKF in presence of additive model error". In: *Tellus A* 70 (2018), p. 1414545.

[50] P. Sakov, G. Evensen, and L. Bertino. "Asynchronous data assimilation with the EnKF". In: *Tellus A* 62 (2010), pp. 24–29.

[51] P. Sakov, J.-M. Haussaire, and M. Bocquet. "An iterative ensemble Kalman filter in presence of additive model error". In: *Q. J. R. Meteorol. Soc.* 144 (2018), pp. 1297–1309. eprint: arXiv:1711.06110.

[52] P. Sakov, D. S. Oliver, and L. Bertino. "An iterative EnKF for strongly nonlinear systems". In: *Mon. Wea. Rev.* 140 (2012), pp. 1988–2004.

[53] R. P. Wishner, J. A. Tabaczynski, and M. Athans. "A Comparison of Three Non-Linear Filters". In: *Automatica* 5 (1969), pp. 487–496.

[54] S.-C. Yang, E. Kalnay, and B. Hunt. "Handling Nonlinearity in an Ensemble Kalman Filter: Experiments with the Three-Variable Lorenz Model". In: *Mon. Wea. Rev.* 140 (2012), pp. 2628–2646.

[55] M. Zhang and F. Zhang. "E4DVar: Coupling an ensemble Kalman filter with four-dimensional variational data assimilation in a limited-area weather prediction model". In: *Mon. Wea. Rev.* 140 (2012), pp. 587–600.

[56] M. Zupanski. "Maximum Likelihood Ensemble Filter: Theoretical Aspects". In: *Mon. Wea. Rev.* 133 (2005), pp. 1710–1726.