Learning variational data assimilation models with uncertainty quantification

Institute : Laboratoire des Sciences du Climat et de l'Environnement

Joint work with P.Naveau and R.Fablet

LSCE

Introduction

Based on previous work focusing on learning jointly dynamical model and solver in a variational data assimilation framework (see *Fablet et al., 2021*), we extend this approach to account for uncertainties.

Variational data assimilation

4DVar objective function

$$J(x) = \sum_{i=1}^{n} ||Hx_i - y_i||_R^2 + \sum_{i=1}^{n} ||x_i - x_i^{(f)}||_Q^2$$

(see *Tremolet, 2008*)

- x estimated state of the system with x at time t_i : x_i
- Observation at time t_i : y_i
- Forecast of the numerical system : $\Phi(x) = x^{(f)}$
- R covariance matrix of observation error
- Q covariance matrix of model error

Main question

Could we use our knowledge in 4DVar optimization to approximate the distribution $x|y$ instead of a pointwise estimate?

ELBO formulation

Evidence lower bound : *(Hoffman & Johnson,2016)*

$$\log p(y) \geqslant \mathbf{E}_{x \sim q_\theta} \log \left(\frac{p(x,y)}{q_\theta(x)}\right)$$

Maximum when :

$$q_\theta \sim p(x|y)$$

ELBO formulation

Evidence lower bound :

$$\log p(y) \geqslant \mathbf{E}_{x \sim q_\theta} \log \left(\frac{p(x,y)}{q_\theta(x)} \right)$$

$$\iff$$

$$\log p(y) \geqslant \mathbf{E}_{x \sim q_\theta} \log \left(p(y|x) \right) - D_{KL}{}^{1}(q_\theta || p_x).$$

[1] $\mathrm{D}_{KL}(q||p) = \mathbf{E}_{x \sim q} \log \left(\frac{q}{p} \right)$

Full gaussian example

Assumption :

$$\begin{aligned}(y|x) &\sim \mathcal{N}(Hx, R)\\ x &\sim \mathcal{N}(\mu^*, \Sigma^*)\end{aligned}$$

Gaussian parametrization of q:

$$q_\theta = q_{(\mu,\Sigma)} \sim \mathcal{N}(\mu, \Sigma)$$

Full gaussian example : explicit ELBO fromulation

$$\mathrm{E}_{x \sim q_\theta} \log\left(p(y|x)\right) \quad - \quad D_{KL}(q_\theta || p_x)$$

Full gaussian example : explicit ELBO fromulation

$$\mathrm{E}_{x \sim q_\theta} \log\left(p(y|x)\right) \quad - \quad D_{KL}(q_\theta || p_x)$$

$$-\frac{1}{2}\left(tr(R^{-1}\Sigma) + \log(|R|) + ||y - H\mu||_R^2\right)$$

Full gaussian example : explicit ELBO fromulation

$$\mathrm{E}_{x \sim q_\theta} \log\left(p(y|x)\right) \quad - \quad D_{KL}(q_\theta || p_x)$$

$$-\frac{1}{2}(tr(R^{-1}\Sigma + \log(|R|) + ||y - H\mu||_R^2)$$

Observation term

Full gaussian example : explicit ELBO formulation

$$\mathrm{E}_{x\sim q_\theta} \log\left(p(y|x)\right) \qquad - \qquad D_{KL}(q_\theta||p_x)$$

Full gaussian example : explicit ELBO formulation

$$\mathrm{E}_{x\sim q_\theta} \log\left(p(y|x)\right) \quad - \quad D_{KL}(q_\theta||p_x)$$

$$\frac{1}{2} \times (tr(\Sigma^{*-1}\Sigma) + log(\frac{|\Sigma^*|}{|\Sigma|}) + ||\mu^* - \mu||^2_{\Sigma^*})$$

Full gaussian example : explicit ELBO formulation

$$\mathrm{E}_{x \sim q_\theta} \log\left(p(y|x)\right) \quad - \quad D_{KL}(q_\theta || p_x)$$

$$\frac{1}{2} \times \underbrace{\left(tr(\Sigma^{*-1}\Sigma) + log(\frac{|\Sigma^*|}{|\Sigma|}) + ||\mu^* - \mu||^2_{\Sigma^*}\right)}_{g(\mu,\Sigma)}$$

In general, g is not known

Rewriting trick

We use the following trick. Let's introduce Φ such as :

$$g(\mu, \Sigma) = ||\Phi(\mu, \Sigma) - (\mu, \Sigma)||^2$$

Why?

- Common reformulation in ML regularization techniques
- Analogy with the dynamical term of variational cost
- It implies a dynamical evolution of μ and Σ

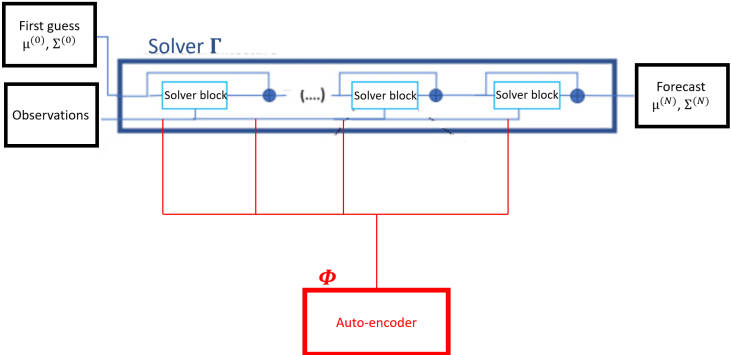
New variational cost

$$U_{\hat{\Phi}(y,\mu,\Sigma)} = ||y - H\mu||^2 + ||\hat{\Phi}(\mu,\Sigma) - (\mu,\Sigma)||^2$$

NN framework

Operator and solver

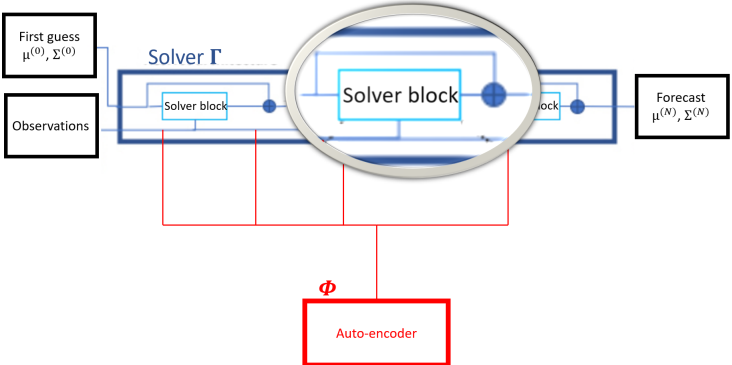
- Dynamical operator Φ : Auto-encoder or Gibbs Energy NN
- Solver Γ: iterative gradient-based inversion algorithm to minimize previously defined variational cost



NN framework

Operator and solver

- Dynamical operator Φ : Auto-encoder or Gibbs Energy NN
- Solver Γ: iterative gradient-based inversion algorithm to minimize previously defined variational cost



Block cell

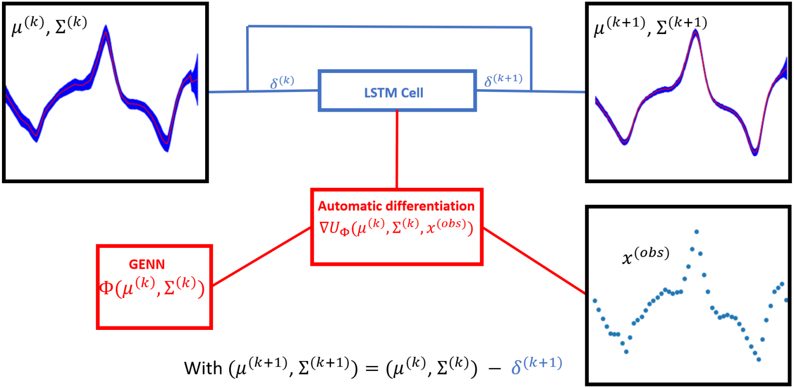


Figure: Iteration of the solver

Learning setting

Let us denote by $\Psi_{\Phi,\Gamma}(\mu^{(0)}, \Sigma^{(0)}, y)$ the resulting model.

An entropy criterion (see *Bocquet et al., 2020*)

If the dataset comprises true states $x_1, x_2, ..., x_N$, we can consider the following learning loss : $L = \sum_n -ln(P_{\Psi_{\Phi,\Gamma}(\mu_n^{(0)},\Sigma_n^{(0)},y_n)}(x_n))$

i.e :

$L = \sum_n \frac{1}{2}({}^t(x_n - \mu_n^{(N)}(\Sigma_n^{(N)})^{-1}(x_n - \mu_n^{(N)}) + \ln(\det(\Sigma_n^{(N)}))$

Studied datasets

Datasets

- Auto-regressive linear models
- Lorenz 63
- Danube discharge measurement network

Our model has been tested both in prediction and reconstruction.

AR model

We simulate a dataset which satisfies a linear dynamics of the form:

$$\begin{cases} X_t & = AX_{t-1} + BX_{t-2} + \eta_t \\ Y_t & = X_t + \epsilon \end{cases}$$

We studied two cases :

- State independent model noise $\eta_t \propto \mathcal{N}(0\ ,\ I)$
- State dependent model noise $\eta_t \propto CX_{t-1}\mathcal{N}(0\ ,\ I)$

AR model

Score for forecasted steps :

Method	Type of model error	MSE	Entropy	Known model and errors
4DvarnetSto	State independent	$4.78\ 10^{-4}$	-2.38	No
	State dependent	$3.19\ 10^{-3}$	-1.45	No
Kalman Filter	State independent	$4.48\ 10^{-4}$	-2.29	Yes
	State dependent	$1.58\ 10^{-3}$	-1.47	Yes

L63 model

Reconstruction with only the first variable observed once every eight time steps for two different settings :

Standard L63 dynamics :

$$\begin{cases} \frac{dx}{dt} & = & \sigma(y-x) \\ \frac{dy}{dt} & = & \rho x - y - xz \\ \frac{dz}{dt} & = & xy - \beta z \end{cases}$$

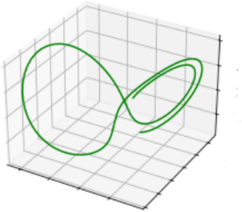
Stochastic L63 (*Chapron et al., 2018*) :

$$\begin{cases} dX & = & (\sigma(Y-X) - \frac{4}{2\Gamma}X)dt \\ dY & = & (\rho X - Y - XZ - \frac{4}{2\Gamma})dt \\ & & + \frac{\rho - Z}{\Gamma^{\frac{1}{2}}} dB_t \\ dZ & = & (XY - \beta Z - \frac{8}{2\Gamma}Z)dt \\ & & + \frac{Y}{\Gamma^{\frac{1}{2}}} dB_t \end{cases}$$

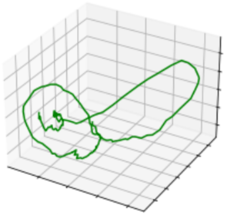
L63 model

L63 experiments :
Reconstruction with only the first variable observed once every eight time steps for two different settings :

Standard L63 dynamics :



Stochastic L63 (*Chapron et al., 2018*) :



L63 model

Method	Stochastic	MSE	Entropy
4DvarnetSto	No	0.45	-4.60
	Yes	3.51	-1.42
EnKF with first variable observed	No	1.40	-0.48
	Yes	23.8	4.9
EnKF with two first variables observed	No	0.40	8.13
	Yes	4.44	-1.85
EnKF with all variables observed	No	0.38	44.7
	Yes	2.60	-2.47

Danube river network dataset

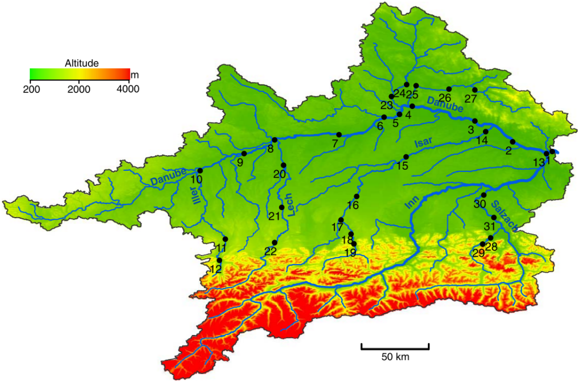


Figure: 31 gauging stations on the Danube river network (*Asadi et al., 2015*), with 50 years of daily measurements (1960-2010)

Discharge reconstruction task

Setting :
Reconstruction task for which observations are available every 4 days for only 15 stations

Visualizaition

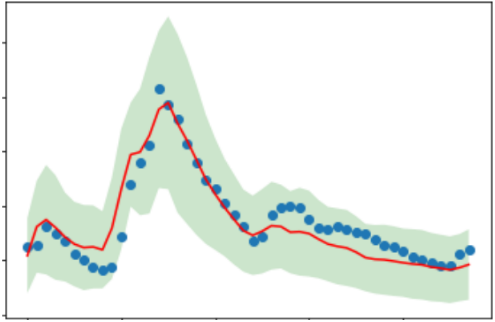


Figure: Hidden observation (blue dots), estimated mean (red curve) and 95% confidence interval

Conclusion and perspectives

- Based on 4DVar-like variational cost infered from ELBO maximization, we have been able to give the best gaussian approximation of $(x|y)$
- No prior knowledge on the dynamic is required, neither on the error
- This framework can be extended to other parametric distribution, especially heavier-tailed distribution.
- More complex type of noise can be simulated to evaluate our method for highly non-gaussian distribution estimation

References

- P. Asadi, A. C. Davison, and S. Engelke.
  Extremes on river networks
  The Annals of Applied Statistics, 2015.

- R.Fablet
  End-to-end learning for variational assimilation models and solvers
  Journal of Advances in Modeling Earth Systems, 2021.

- Y.Trémolet
  Model error estimation in 4DVar
  QJRMS, 2007.

- B. Chapron, P. Derian, E. Mémin and V.Resseguier
  Large-scale flows under location uncertainty: a consistent stochastic framework
  QJRMS, 2018.

References

M. Bocquet, J. Brajard, A. Carrassi, and L. Bertino
Bayesian inference of chaotic dynamics by merging data assimilation, machine learning and expectation-maximization.
arXiv preprint, 2020

M. D. Hoffman M. J. Johnson
Elbo surgery: yet another way to carve up the variational evidence lower bound.
NIPS (Vol. 1, No. 2), 2016