variational learning of sea surface current reconstructions from AIS data streams

Simon Benaïchouche

Eodyn

- funded in 2015
- exploit AIS data streams to produce current fields

AIS data streams

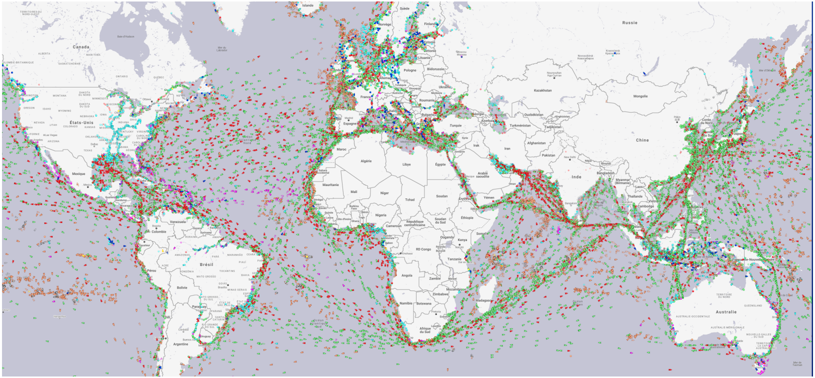


Figure: screenshot from http://www.marinetraffic.com

- mainly used to avoid collision
- an AIS message typically contains : MMSI number, GPS position, GPS speed, manoeuver status, and sometimes heading of boats

Linear model of observation

Speed over ground = speed on surface + current

we can derive for each pixel of the grid the following

$$\begin{pmatrix} cos(\theta_1) & 0 & ... & 0 & 1 & 0 \\ sin(\theta_1) & 0 & ... & 0 & 0 & 1 \\ ... & ... & ... & ... & ... & .. \\ 0 & 0 & ... & cos(\theta_n) & 1 & 0 \\ 0 & 0 & ... & sin(\theta_n) & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} V_S 1 \\ ... \\ U_1 \\ U_2 \end{pmatrix} = \begin{pmatrix} SOG_{1_1} \\ SOG_{2_1} \\ ... \\ SOG_{1_n} \\ SOG_{2_n} \end{pmatrix}$$

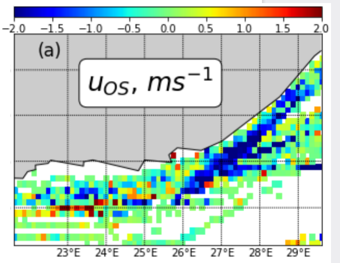
$Ax = y$ where $A \in \mathcal{M}^{2n \times n+2}$, $x \in \mathbb{R}^{2n+2}$ and $y \in \mathbb{R}^{2n}$.

least square estimator :

$$argminJ(x,y) = \|Ax - y\|_2^2 = (AA^t)^{-1}Ay \tag{1}$$

a ill-posed problem in the sense of Hadamard

- for 1 pixel : no existence and uniqueness of solutions under the linear model
- sparse observations



we need to add a regularization term :

$$X^* = \arg\min_X J(X, Y) + R(X) \tag{2}$$

a ill-posed problem in the sense of Hadamard

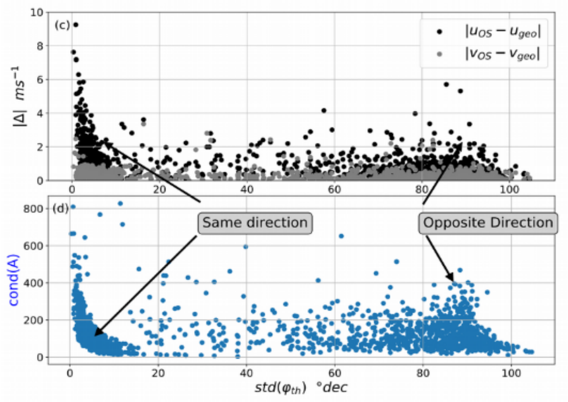
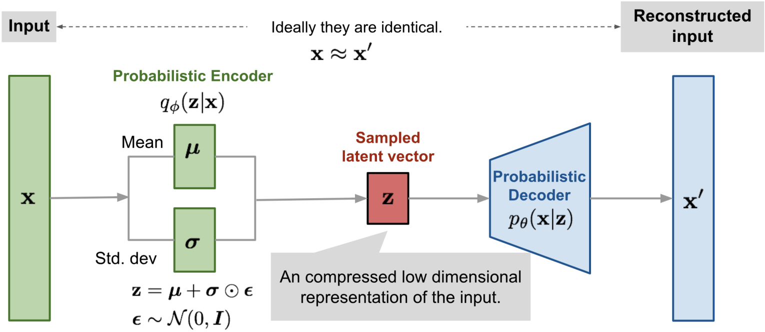


Figure: source : presentation by Clément Legoff

A priori knowledge

- current fields are structured objects $\subset \mathcal{M}$ (manifold hypothesis)
- X_T and X_{T+1} should be "close"

generative modeling : VAEs



generative modeling : VAEs

VAE (kingma et al 2014) learns both : parsimonious representations of data and a lower bound on data likelihood.

let $\mathrm{X} = \{x_i\}$ a dataset of i.i.d samples seen as realization of random process involving latent variables, z_i, i.e $p_\theta(z|x)$.

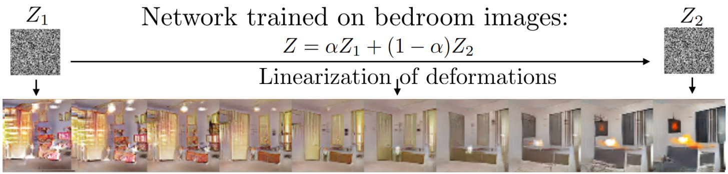
we train VAE by minimizing the following loss

$$\mathcal{L}(x, \theta) = E_{q_\theta(z|x)}[log(p_\theta(x_i|z)] - KL(q_\theta(z|x_i)||p(z))$$

$-\mathcal{L}(x, \theta)$ is a lower bound on data likelihood

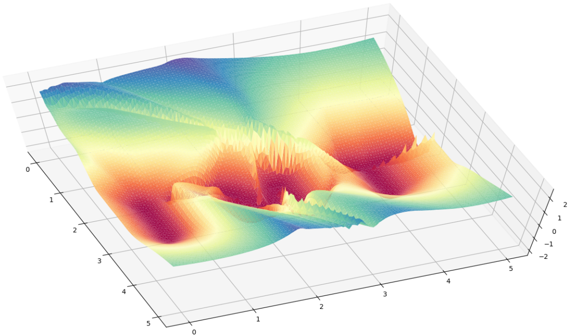
with $p(z) \sim \mathcal{N}(0, I_k)$, the KL term enforce latent variable to be centered on the unit ball.

Variational auto-encoders



Example : graph of reconstruction loss associated to an AE

Encoding "IMT" in 1d using a MLP



generative models as non-linear dictionary for inverse problems solving

Idea : restrict the possible solutions to the image of the decoder.
associated optimization problems :

$$Z^* = \arg\min J(\Phi(Z), Y) + R(Z, \Phi(Z)) \tag{3}$$

we define a projection operator as the result of a gradient descent process:
$z_0 = 0, z_{k+1} = z_k - \lambda \nabla (J(\phi(Z_k), Y) + R(Z, \Phi(Z))$
i.e : we're looking at stable points of :

$$\begin{cases} z'(t) & = & -(J(\Phi(Z), Y) + R(Z, \Phi(Z))) \\ z(0) & = & z_0 \end{cases} \tag{4}$$

Remark : if Φ_1 and Φ_2 denotes two C^1 NN with $Im(\Phi_1) = Im(\Phi_2)$.
gradient flows can be $\neq$ but shares the same zeros.

Regularizing effect : solution of (3) invariant under a class of deformations :

example with PCA decoder:

min J(Proj(X),Y) = min J($\Phi(z)$,Y). and we have :

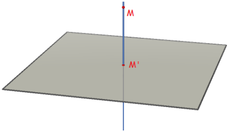


Figure: projection on affine subspace

$I(x) = \{f / Proj(f(x)) = Proj(x)\}$ is big (restriction to invertible tranformation = group structure)

$O_x = \{y = f(x) / f \in I(x)\}$ is the affine subspace $\perp$ $Im(\Phi)$ which contains x.

regularization by dimension reduction : non linear case

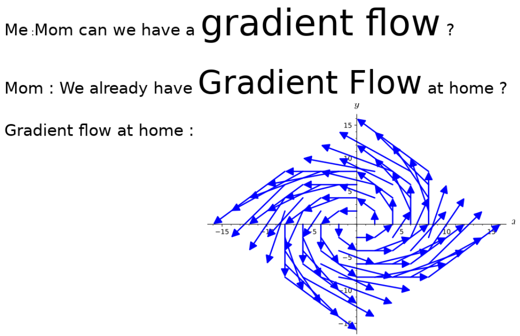
suppose Diff Φ Lischitz, associated flow is differentiable but its differential is not full rank ($\leq$ dim latent space).

- $I(x) = \{f / Proj(f(x)) = Proj(x)\}$ contains diffeomorphisms.
- displacement along O_x can be done using infinitesimal displacement alongside tangent plane $\perp$ $Im(Diff(\Phi))$.

Remark : it's not really a projection if $Im(\Phi)$ is a non-convex set...

sensibility to latent representation

We compute a gradient flow on a latent space, not on the projected manifold



there is a way to construct an algorithms invariant by training for two neural networks that satisfies $Im(\Phi_1) = Im(\Phi_2)$ using Gram-schmidt algorithm on tangent planes (slow).

Resnet and neural ODEs

$$x_{k+1} = x_k + \Delta t NN(x) \tag{5}$$

interpretation of residual networks architectures as numerical integration scheme (euler, rk4...) of the following cauchy problems.

$$\left\{ \begin{array}{rcl} x'(t) & = & NN(x(t)) \\ x(0) & = & x_0 \end{array} \right. \tag{6}$$

learning as an optimal control problem : recent approach allows us to compute the adjoint.

joint learning of representation and dynamical systems while solving the targeted inverse problem:
minimize the loss :

$$U^*, \theta^* = \arg\min_{X} J(U, Y) + R(U, \theta) \quad (7)$$

$$w.r.t$$

$$\left\{ \begin{array}{rcl} U(., t) & = & \Phi(\tilde{Z}(t)) \\ Z'(t) & = & f(Z(t)) \\ Z(0) & = & Z_0 \end{array} \right. \quad (8)$$

and

$$R(U) := \lambda_U \mathcal{L}_{VAE}(U, \theta) + \lambda_V \mathcal{L}_{VAE}(V, \theta)$$

V denotes an external dataset of representations, for the experiment we used the assimilated product GLORYS.

proposed approach : how to learn the dynamical system ?

current fields time-series stated as decoded latent variable that follow an ODE. this system can not be seen as autonomous, it depends from unobservered states.

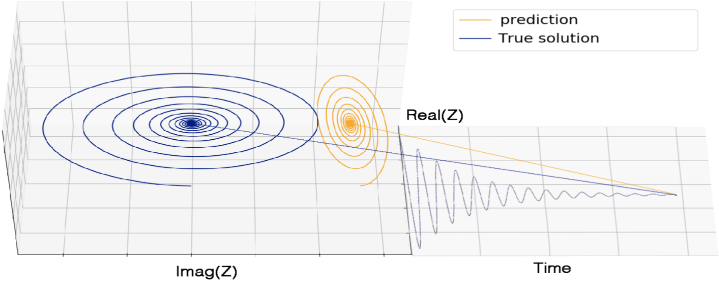


Figure: Learning Latent Dynamics for Partially-Observed Chaotic Systems

training

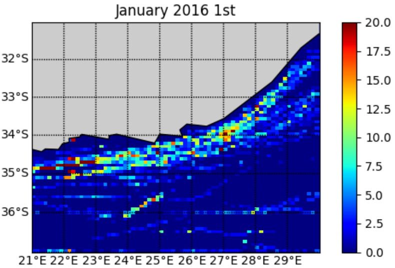


Figure: AIS message density on

- focus on aghulas current. year 2016, 4 millions of AIS messages.
- External representations : GLORYS time series of current field from 1993 to july 2015.

time integration windows : 8 days.

latent space dim = 50 + 10.

original dimension of the optimization problem : $1.2 * 10^5$

results

validation with a dataset of drifting buyos. 3 periods :

- a first period from 01/01 to 03/20 (summer)
- a second period from 04/09 to 06/28 (transition autumn-winter)
- a third period from 06/29 to 09/16 (winter)

Table: Reconstruction performance evaluate from independent in situ data

Data	**Method**	**summer**	**autumn**	**winter**
Satellite altimetry	OI	0.1580	0.1374	**0.0513**
AIS	OI	0.1041	0.1739	0.2017
AIS	VAE-NODE networks	**0.0609**	**0.1148**	0.0616

plots

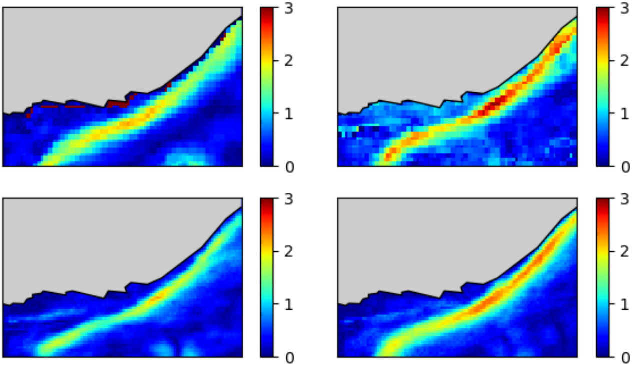


Figure: Reconstructed velocity fields for January 16th 2016 for the altimetry (top, right) and AIS baselines (top left) and the two configurations of the proposed framework with (bottom right) and without (bottom left) the use of GLORYS data in the training phase.

daily RMSE on a 180 day assimilation

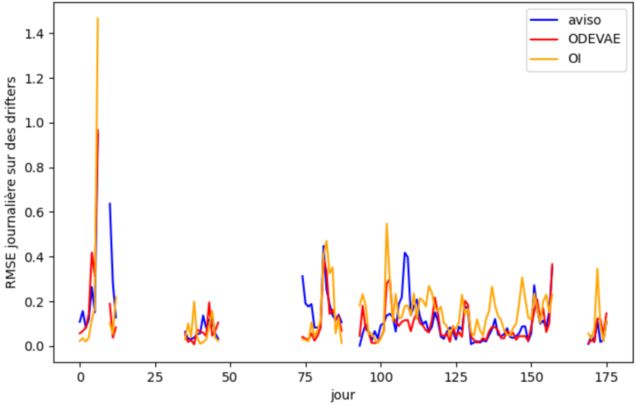
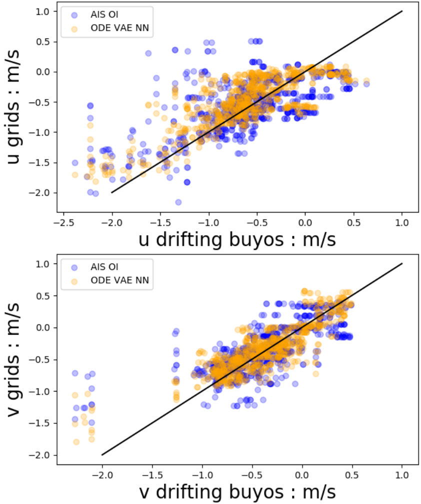


Figure: comparison between RMSE with drifting buyos

validation on drifters



the neural ODE framework for time interpolation

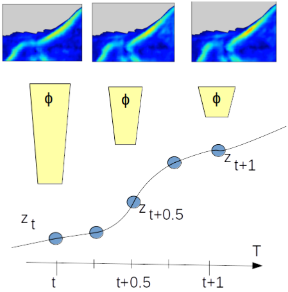


Figure: non-linear interpolation in the latent space using the Rk4 integrator

Interpolation	MSE
24h	0.1148
12h	0.1091
6h	0.1016
3h	0.0919

Future work

- work on assimilation error
- VAE produces "blurry" images =¿ multi-scale models ?
- using a Hemlotz decomposition on the decoder : $\Phi = curl(f) + \nabla g$ to assimilate divergent free field (such as geostrophic current) and perform sensor fusions

references

- Auto-encoding variational bayes *Kingma et al*
- Neural Ordinary differential equations *Ricky T. Q. Chen et al*
- Learning Latent Dynamics for Partially-Observed Chaotic Systems *S. Ouala et al*
- End-to-end learning of energy-based representations for irregularly-sampled signals and images *R. fablet et al*
- Group invariant scattering *S. Mallat*