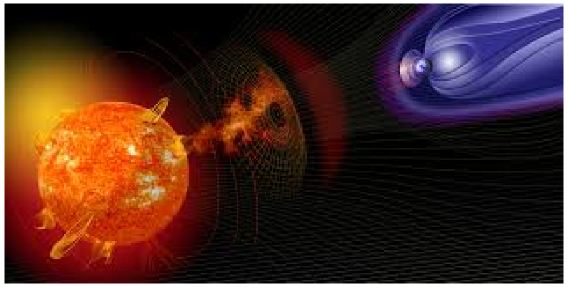
A machine learning approach to solar wind speed forecasting from solar images



Joint project with CWI (https ://projects.cwi.nl/mlspaceweather/)

- Mandar Chandorkar (Phd, CWI) ⟶ More details in poster
- Mhamed Hajaiej (Md, École Polytechnique)
- Enrico Camporeale (CIRES, University of Colorado)
- Cyril Furtlehner (INRIA, LRI, Université Paris-Sud)
- Bala Poduval (University of New Hampshire)
- Michèle Sebag (CNRS, LRI, Université Paris-Sud)

Overview

- Forecasting solar wind speed at L_1 from solar/heliospheric data : a complex regression problem.
- Feature extraction from input data : deep auto-encoders.
- Dynamic time lag regression problem : a Bayesian approach.
- Medium scale experiments and perspectives.

Problem statement

Forecasting the solar wind speed recorded at L1 from solar/heliospheric observations.

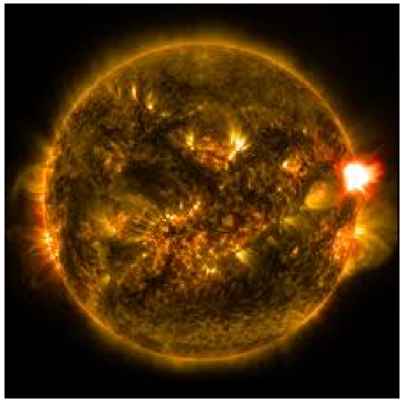
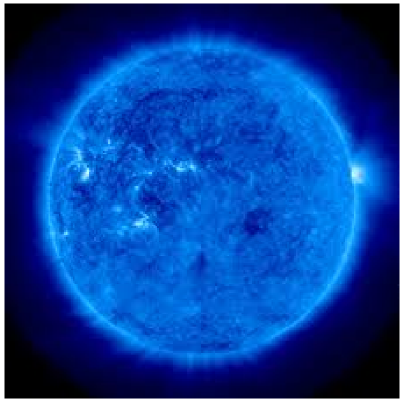
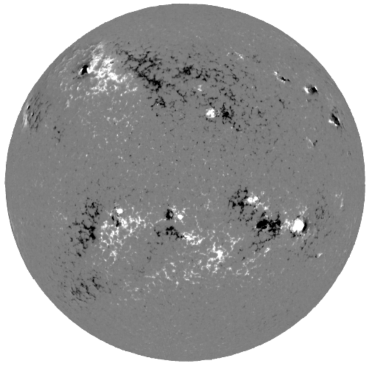
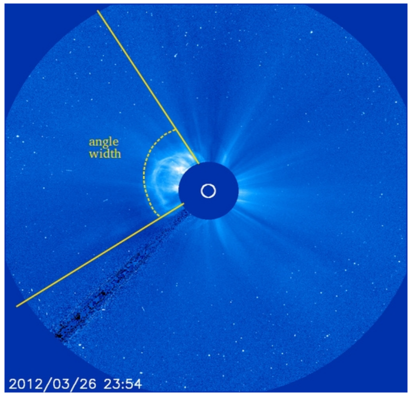
Complex regression problem in two respects :

- (i) Badly conditioned input-output problem :
  Large dimension of input signal : full scale images $d = 512^2 \times \#channels$
  Scalar output (SW speed).
  $\longrightarrow$ Dilution of the cause in the input signal (bad SNR).

Problem statement

Input data consists of various solar images channels :

different wavelengths, magnetograms, LASCO images of CME.

Problem statement

Forecasting the solar wind speed recorded at L1 from solar/heliospheric observations.

Complex regression problem in two respects :

- (i) Badly conditioned input-output problem :
  Large dimension of input signal : full scale images $d = 512^2 \times \#channels$
  Scalar output (SW speed).
  $\longrightarrow$ Dilution of the cause in the input signal (bad SNR).

- (ii) stochastic non-constant time lag in a range of 1 to 5 days :
  yet another factor of 10 to 100 in the input signal.

Problem statement

Forecasting the solar wind speed recorded at L1 from solar/heliospheric observations.

Complex regression problem in two respects :

- (i) Badly conditioned input-output problem :
  Large dimension of input signal : full scale images $d = 512^2 \times \#channels$
  Scalar output (SW speed).
  $\longrightarrow$ Dilution of the cause in the input signal (bad SNR).

- (ii) stochastic non-constant time lag in a range of 1 to 5 days :
  yet another factor of 10 to 100 in the input signal.

$\longrightarrow$ *Brute force model might have little chance to succeed.*

Feature extraction from input data : deep auto-encoders

Is it possible to compress the source and how much ?

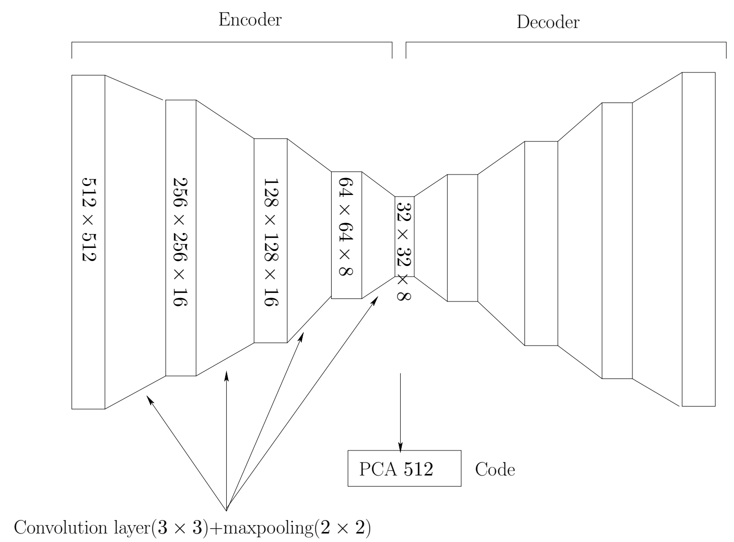
Non-supervised approach with deep auto-encoder (AE), variational auto-encoder (VAE)

Risk : loosing the relevant information w.r.t. the target output (solar wind speed)

To be considered as a pre-training step.

Feature extraction from input data : deep auto-encoders

Convolution network based auto-encoders



Loss function : reweighted pixel values ($y_i \in [0, 1]$) distance

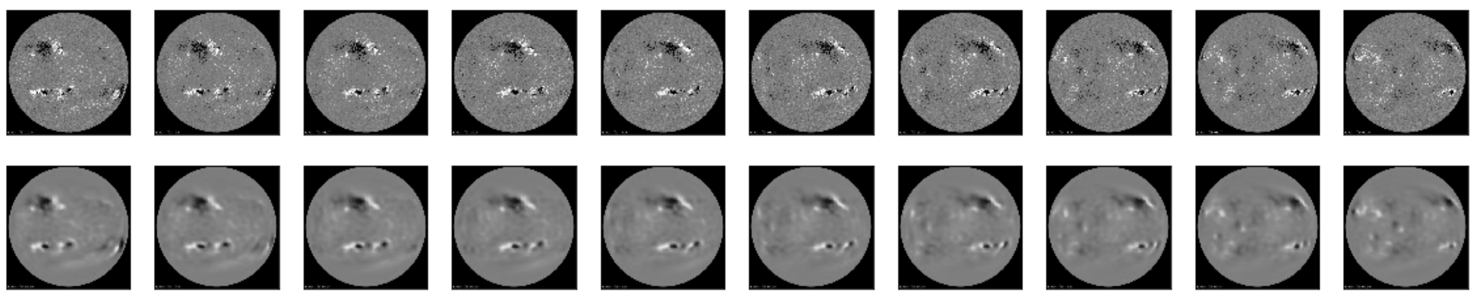
$$\mathcal{L}(\mathbf{y}^{\text{in}}, \mathbf{y}^{\text{out}}) = \sum_{i \in Pixels} \frac{(y_i^{\text{in}} - y_i^{\text{out}})^2}{(y_i^{\text{in}} + \epsilon)^\alpha} + \frac{(y_i^{\text{in}} - y_i^{\text{out}})^2}{(1 - y_i^{\text{in}} + \epsilon)^\alpha}$$

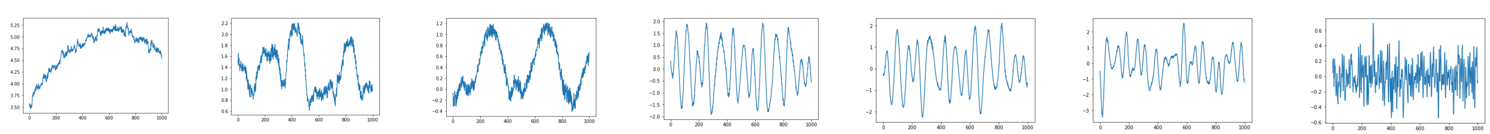
Feature extraction from input data : deep auto-encoders

Dimensional reduction of magnetograms

Example 1 :

Convolution network + PCA : $512^2 \longrightarrow 512$



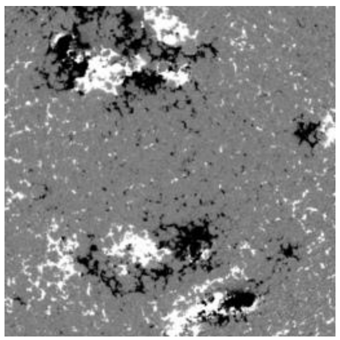


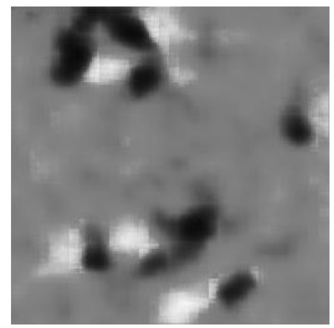
0 1 2 5 6 10 300

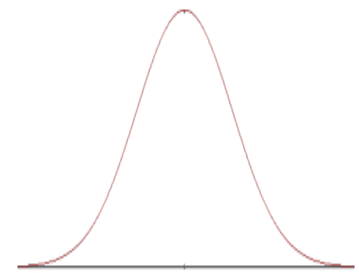
Dimensional reduction of magnetograms

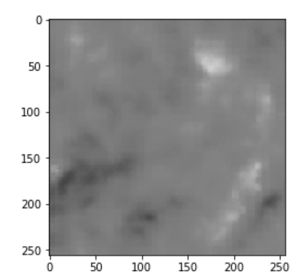
Example 2 :

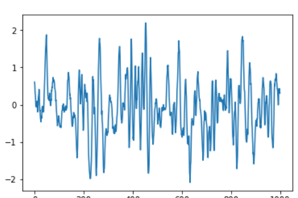
Variational autoencoder : $256^2 \longrightarrow 90$

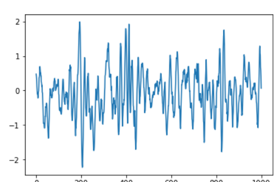


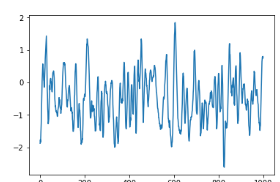








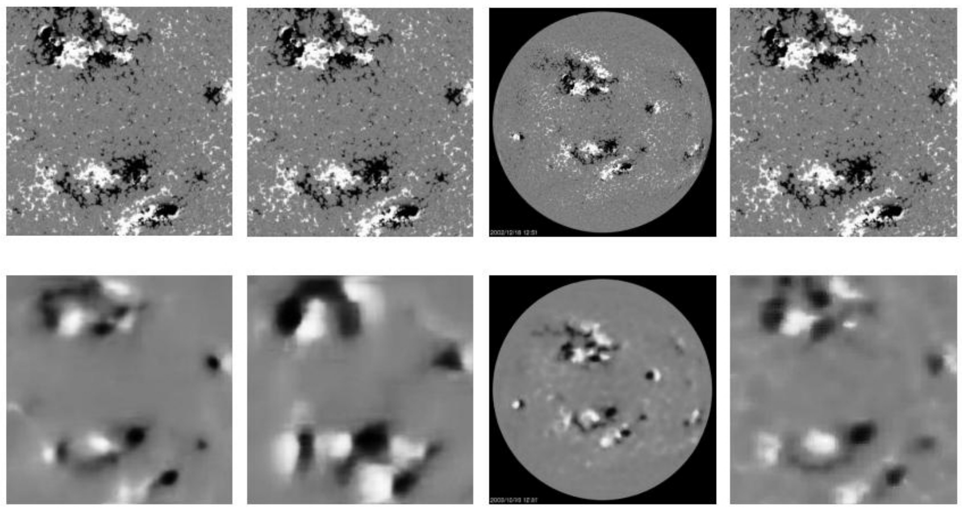




0 1 80

Dimensional reduction of magnetograms

Visual comparison :



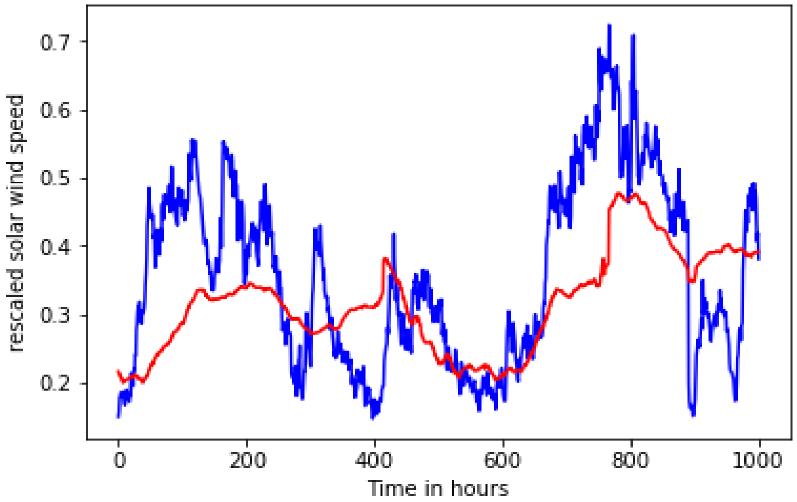
Input image

Output image

Conv AE	Conv AE	Conv AE	VAE
$(256^2 \to 512)$	+ dense layer $(256^2 \to 64)$	+ PCA $(512^2 \to 512)$	$(256^2 \to 90)$

Solar wind first prediction Test

Input : sequences of magnetogram's (SOHO) AE features + LASCO coronagraph(NOAA) for $[t_0 - 4\text{days}, t_0]$
Output : Solar wind at L_1 and $t_0 + 1\text{day}$ (OMNI)



	Correlation	RMSE (kms^{-1})
Random predictor	0	165
Mean predictor	–	139
Magnetogram AE	0.05	193
Halo CME	0.14	–
Magnetogram Sequence AE	0.236	121
Magn. Seq. AE + Halo CME	**0.335**	**120**

Dynamic time lag regression

Goal : jointly infer the time lag and predict the solar wind speed at L_1

Motivation :

- reducing the dimension of the input signal
- increasing model's interpretability

Remark : the time-lag plays the role of a latent variable (never observed)

Barely discussed problem in ML.

Available methods have too restrictive hypothesis

- linear methods are not adapted.
- Dynamic time warping (DTW), assume monotonicity, one to one cause-effect mapping

Our proposal : a Bayesian combination of experts.

Dynamic time lag regression

Deterministic formulation of the problem :

$$y(t+\tau(t)) = f[x(t)]$$
$$\tau(t) = g[x(t)]$$

with

$$f : \mathcal{X} \to \mathbb{R}, \qquad \text{and} \qquad g : \mathcal{X} \to \mathbb{R}^+,$$

- $x(t) \in \mathbb{R}^d$, $d \gg 1$, input data containing the hidden cause
- $y(t) \in \mathbb{R}$ scalar, the effect
- $\tau(t) \in \mathbb{R}^+$, the time-lag between cause and effect

Underlying assumptions :

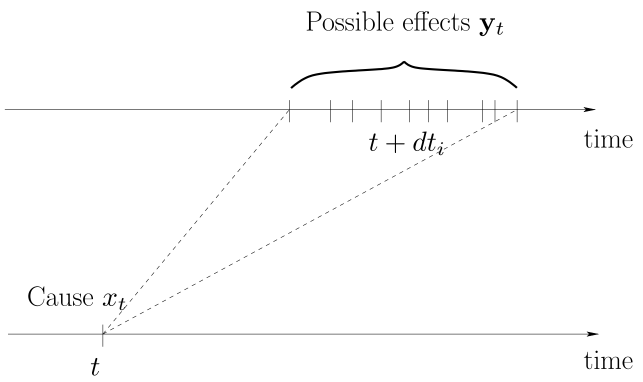
- both effect and time-lag are caused solely by $x(t)$.
- τ is possibly many to one, aggregation of causes to be specified later.

Dynamic time lag regression

Probabilistic relaxation : a Bayesian combination of experts

Discretization of the problem : time $t \in \Omega$ is now discreet and

- $\{dt_i, i \in \mathcal{T}\}$: set of possible time-lag
- $(x_t, \mathbf{y}_t)$: input-output pairs
  with $\mathbf{y}_t = \{y_{t+dt_i}, i \in \mathcal{T}\}$



Assume a given set of predictors $\{\hat{y}_i(x), i \in \mathcal{T}\}$ for $\mathbf{y}_t$ given $x_t = x$. Define a set $\{\tau_i(x), i \in \mathcal{T}\} \in \{0,1\}^{|\mathcal{T}|}$ of stochastic binary variables :

$$\tau_i(x_t) = (y_{t+dt_i} \textit{ is caused by } x_t) \in \{0,1\}.$$

Gaussian mixture :

$$P\big[\mathbf{y}_t | x_t = x\big] = \sum_{\tau} \hat{p}\big(\tau_1, \ldots, \tau_n | x\big)\, \mathcal{N}\big(\hat{\mathbf{y}}(x), \sigma(\tau)\big)$$

Dynamic time lag regression

Model specifications :

$$P\big[\mathbf{y}_t|x_t = x\big] = \sum_{\tau} \hat{p}\big(\tau_1,\ldots,\tau_n|x\big)\prod_{i\in T}\mathcal{N}\big(\hat{y}_i(x),\sigma_i(\tau)\big)$$

- $\hat{p}\big(\tau_1,\ldots,\tau_n|x\big)$: joint probability measure of time-lagged effects caused by x

Our present choice : $\sum_{i\in\mathcal{T}}\tau_i = 1$ one single effect per cause

set of probability weights : $\{\hat{p}_i(x), i\in\mathcal{T}\}$, $\hat{p}_i(x) = \hat{p}(t_{j\neq i},\tau_i = 1)$.

- Variance of predictors :

$$\sigma_i(\tau)^2 = \frac{\sigma^2}{1+\alpha\tau_i} \quad \text{with} \quad \begin{cases}\sigma^2 : \text{ default variance of } \; y_{t+dt_i} - \hat{y}_i(x_t) \\ \alpha > 0 : \text{ reduced variance of the good predictor}\end{cases}$$

Dynamic time lag regression

Parameters of the model :

Two sets of functions of the input x and two meta parameters have to be learned :

- The predictors $\hat{\mathbf{y}}(x) = \{\hat{y}_i(x), i \in \mathcal{T}\}$
- The probability weights $\hat{\mathbf{p}}(x) = \{\hat{p}_i(x), i \in \mathcal{T}\}$
- σ^2
- α

Learning criterion : The loss function is given by the

$$\mathcal{L}[\mathbf{x}, \mathbf{y} | \hat{\mathbf{y}}, \hat{\mathbf{p}}, \sigma, \alpha] \qquad \text{Log likelihood of the data } (\mathbf{x}, \mathbf{y})$$

Learning strategy :

- $\hat{\mathbf{y}}$ and $\hat{\mathbf{p}}$ are modelled by means of coupled neural nets
- σ and α are optimized in an outer loop

Based on saddle point equations

$$\Big(\frac{\partial \mathcal{L}}{\partial \hat{\mathbf{y}}(x)}, \frac{\partial \mathcal{L}}{\partial \hat{\mathbf{p}}(x)}, \frac{\partial \mathcal{L}}{\partial \sigma}, \frac{\partial \mathcal{L}}{\partial \alpha}\Big) = 0$$

Dynamic time lag regression

Log likelihood :

- the data : $\{(x, \mathbf{y})\}_{\rm data}$
- the parameters : $\theta = (\hat{\mathbf{y}}, \hat{\mathbf{p}}, \sigma, \alpha)$

Conditional probability of predictor $\hat{y}_i(x)$ being the right one given $(x, \mathbf{y})$:

$$q_i(x, \mathbf{y}) \stackrel{\rm def}{=} P(\tau_i = 1 | x, \mathbf{y}) = \frac{\hat{p}_i(x)}{Z(x, \mathbf{y}|\theta)}\; e^{-\frac{\alpha}{2\sigma^2}(y_i - \hat{y}_i(x))^2}$$

The log likelihood is given in closed form

$$\mathcal{L}\big[\{(x, \mathbf{y})\}_{\rm data}|\theta\big] = -\log(\frac{\sigma^{|T|}}{1+\alpha}) - \mathbb{E}_{\rm data}\Big[\sum_{i\in T}\frac{1}{2\sigma^2}\big(y_i - \hat{y}_i(x)\big)^2 - \log\big(Z(x, \mathbf{y}|\theta)\big)\Big],$$

$$\text{with} \qquad Z(x, \mathbf{y}|\theta) = \sum_{i\in\mathcal{T}} \hat{p}_i(x)\; e^{-\frac{\alpha}{2\sigma^2}(y_i - \hat{y}_i(x))^2}$$

Dynamic time lag regression

Saddle point equations :

Two statistical quantities :

$$\sigma_0^2 = \frac{1}{|T|}\mathbb{E}_{data}\Big(\sum_{i\in T}(y_i - \hat{y}_i(x))^2\Big) \qquad C_1[\mathbf{q}] = \frac{1}{\sigma_0^2}\mathbb{E}_{data}\Big(\sum_{i\in T} q_i(x,\mathbf{y})(y_i - \hat{y}_i(x))^2\Big),$$

representing mean variance of predictor and relative error of model-weighted predictors
The saddle point relations read :

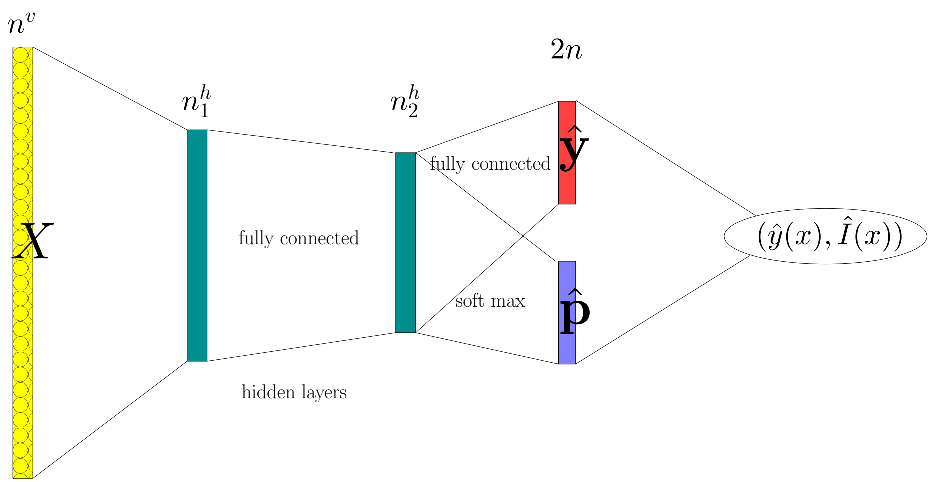
$$\frac{\sigma^2}{\sigma_0^2} = \frac{|T| - C_1[\mathbf{q}]}{|T|-1} \qquad \hat{y}_i(x) = \frac{\mathbb{E}_{data}\Big[y_i\big(1+\alpha q_i(x,\mathbf{y})\big)\Big|x\Big]}{\mathbb{E}_{data}\Big[1+\alpha q_i(x,\mathbf{y})\Big|x\Big]}$$

$$\alpha = \frac{|T|}{|T|-1}\frac{1-C_1[\mathbf{q}]}{C_1[\mathbf{q}]} \qquad \hat{p}_i(x) = \mathbb{E}_{data}\big[q_i(x,\mathbf{y})|x\big].$$

Dynamic time lag regression

Practical implementation :

Initialization of α and σ
$it \longleftarrow 0$;
while *it* $<$ *max* **do**
    while *epoch* **do**
        $\theta \longleftarrow Optimize(\mathcal{L}(\theta, \alpha, \sigma^2))$;
    end
    $\sigma^2 \longleftarrow \sigma_0^2 \frac{|T| - C_1[\mathbf{q}]}{|T| - 1}$;
    $\alpha \longleftarrow \frac{|T|}{|T|-1} \frac{1 - C_1[\mathbf{q}]}{C_1[\mathbf{q}]}$;
end
Result: Model parameters $\theta = (\hat{\mathbf{y}}, \hat{\mathbf{p}})$,
hyper-parameters α, σ^2



Predicted time-lag index :

$$\hat{I}(x) = \underset{i}{\operatorname{argmax}}\big(\hat{p}_i(x)\big)$$

Dynamic time lag regression

Linear stability analysis :

There is a degenerate saddle point at $(\hat{p}_i(x) = 1/|T|, \alpha = 0, \sigma^2 = \sigma_0^2)$.
Insufficiently specialized predictors $\hat{y}_i$ may drive the system to this point.

The Hessian involves additional statistical observables $\big(q_i(x, \mathbf{y}) = P(\tau_i = 1|x, \mathbf{y})\big)$:

$$C_2[\mathbf{q}] = \frac{1}{\sigma_0^4}\mathbb{E}_{data}\Big[\sum_{i\in T} q_i(x,\mathbf{y})\Big(\Delta y_i^2(x) - \sum_{j\in T} q_j(x,\mathbf{y})\Delta y_j^2(x)\Big)^2\Big],$$

mean diversity of predictors

$$u_i[x, \mathbf{q}] = \frac{1}{\sigma_0^2}\mathbb{E}_{data}\Big[q_i(x,\mathbf{y})\big(\Delta y_i^2(x) - \sum_{j\in T} q_j(x,\mathbf{y})\Delta y_j^2(x)\big)\Big| x\Big],$$

individual relative error

Dynamic time lag regression

Linear stability analysis :

Partial stability condition (at frozen $\hat{\mathbf{p}}$) :

$$C_2[\mathbf{q}] < 2C_1^2[\mathbf{q}] + \mathcal{O}(\frac{1}{|T|}).$$

Main instability at the degenerate fixed point :

$$d\hat{\mathbf{p}}(x) \propto -|\mathbf{u}(x)|^2\mathbf{u}(x)$$

rewards predictors with lowest errors by increasing their weights
drives the system toward other solutions, like solutions of the form :

$$\hat{p}_i(x) = \delta_{iI(x)},$$

Note : Consistency between highest log likelihood and $\mathcal{L}_2$ loss optimality.

Experiments

Synthetic data :Constant acceleration model

- input data : $x_t \in \mathbb{R}^{10}$ obeying *Stochastic Langevin Dynamics* ;

$$x_{t+1} = (1-\eta)x_t + \mathcal{N}(0, a^2) \qquad (\eta = 0.02, a^2 = 0.7)$$

- time-lag : width of the time lag interval $|T| = 20$.

$$v_t = k||x_t||^2 + c \qquad (k = 5, c = 100)$$

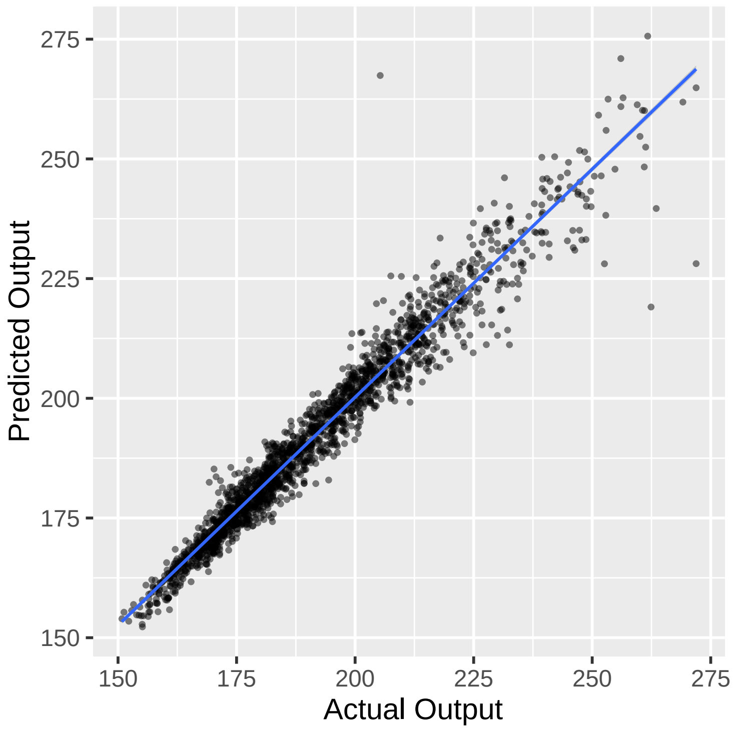
$$\tau(x_t) = \frac{\sqrt{v_t^2 + 2ad} - v}{a}, \qquad (a = 5, d = 1000)$$

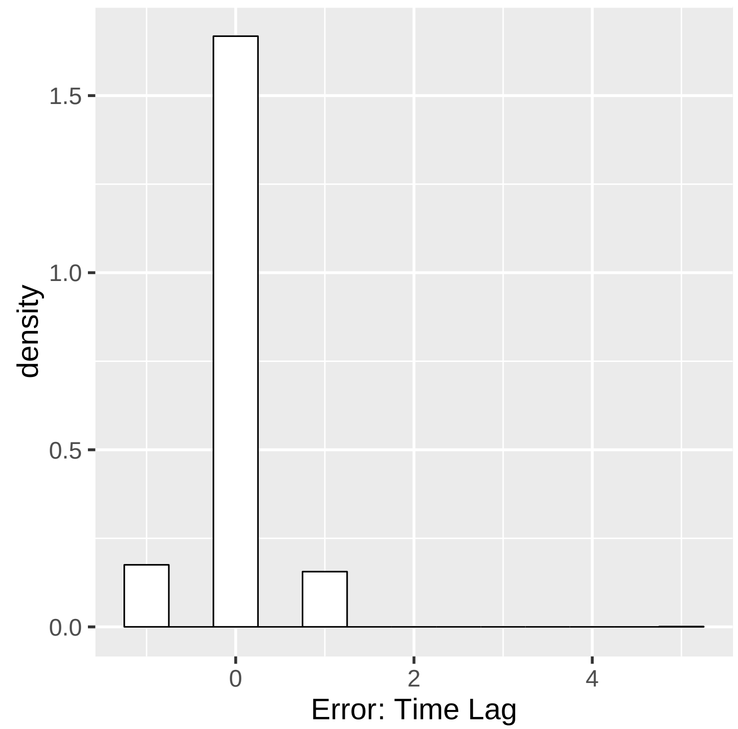
- output data : $y_t \in \mathbb{R}^+$ function of the norm x_t :

$$y_{t+\tau(x_t)} = k||x_t||^2 + c + a\tau(x_t).$$

Experiments

Synthetic data :





Experiments

Solar wind prediction at L_1, > 2 days ahead :

- Input data at t_0 in $\mathbb{R}^{375}$ thanks to various sources (GONG, OMNI) :

$$x_t = \big(\log(\mathbf{f}_S), \mathbf{B}_{cp}, v_{27}, SSN, F10.7\big) \in \mathbb{R}^{180} \times \mathbb{R}^{180} \times \mathbb{R}^{|T|} \times \mathbb{R}^{+} \times \mathbb{R}^{+}$$

  Corresponding to Flux tube expansion from CSSS model,radial magnetic fiels strength, recorded solar speed wind 27 days prior,sun spot number, measured radio flux

- Output data $\mathbf{y}_t$: solar wind speed rank $(t_0 + 2days, t_0 + 5days)$ with time-lag discretization : $|T| = 12$.
- Data : 9 well separated Carrington rotations
- Cross validated experiments : 8 CR for learning $+1$ CR for testing done 9 times.

Experiments

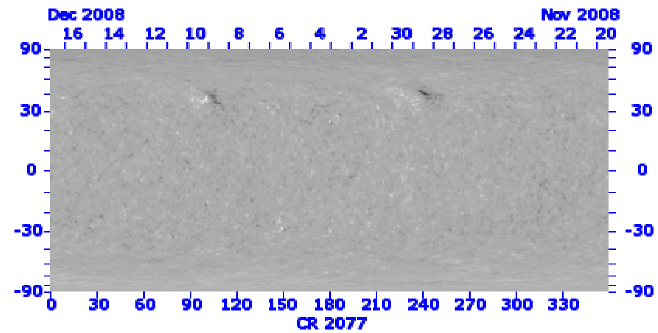
Flux tube expansion from CSSS model *(Zhao & Hoeksema,1995, Poduval & Zhao 2014 ; Poduval 2016)*

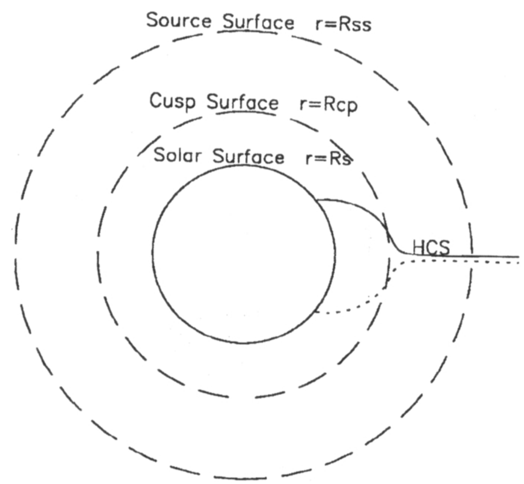
$$\mathrm{FTE}(\Phi) = \frac{R^2_{\mathrm{phot}} B_{\mathrm{phot}}}{R^2_{\mathrm{cp}} B_{\mathrm{cp}}} \in \mathbb{R}^{180}$$

with Φ the Carrington longitude.

$$R_{\mathrm{cp}} \simeq 2.5\ R_{\mathrm{sun}}$$

$$R_{\mathrm{ss}} \simeq 15\ R_{\mathrm{sun}}$$





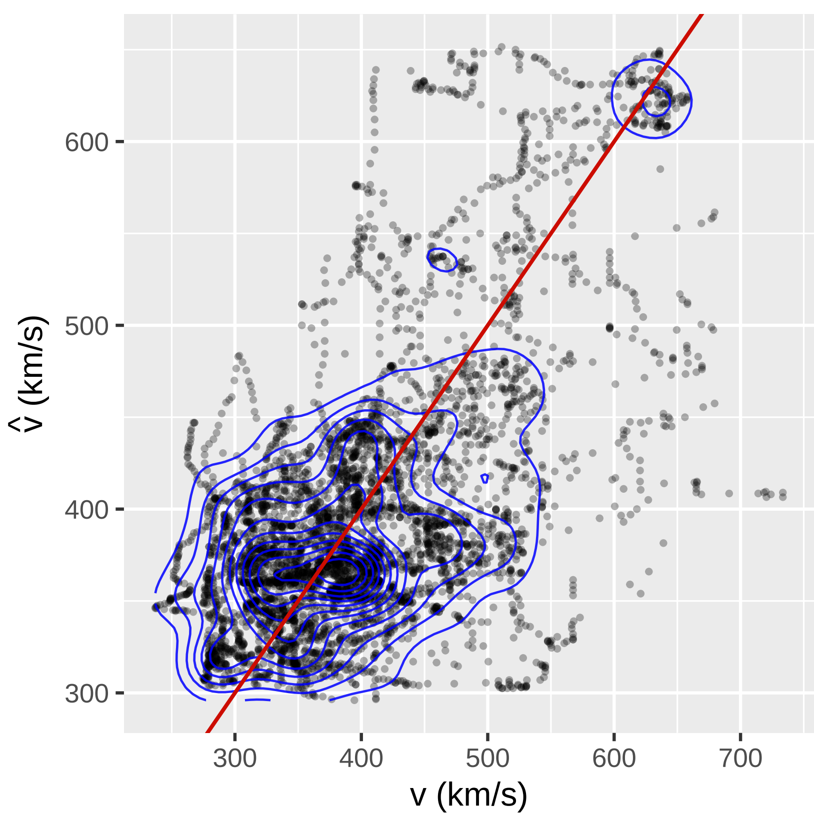
Wang & Sheeley (1990)

speed $\mathrm{km s}^{-1}$	FTE
> 750	< 4.5
$650 - 750$	$4.5 - 8$
$550 - 650$	$8 - 10$
$450 - 550$	$10 - 20$
< 450	> 20

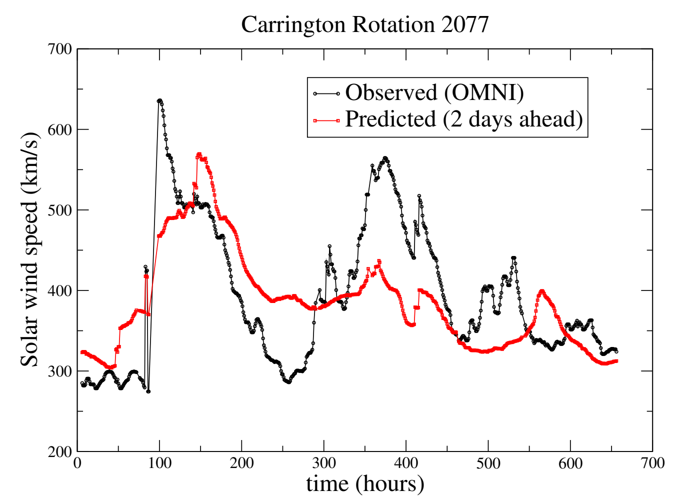
based on photospheric synoptic maps (source GONG)

Experiments

Solar wind prediction at L_1, > 2 days ahead :



	Pearson Corr.	Rank Corr.	MAE(kms^{-1})	RMSE(kms^{-1})
Fixed TL predictor	0.41	0.034	66.44	84.53
Random TL predictor	0.59	0.51	58.15	76.46
DTLR predictor	**0.61**	**0.52**	**56.0**	**73.0**



Model	M.A.E	R.M.S.E
Ensemble Median (WSA, Reiss et. al, 2019)	62.24	74.86
Persistence (4 days)	130.48	161.99
Persistence (27 days)	66.54	78.86
Fixed Lag Baseline	67.33	80.39
DTLR	**54.41**	**65.18**

Conclusion & Perspectives

- DTLR : an interesting ML problem, motivated by space weather forecasting but more general.

- Our Bayesian approach is based on a minimal model : possible refinements ahead.

- The neural net architecture is also minimal : should combine and pre-train with AE.

- more experiments needed, to extend and select the relevant input information

Dynamic time lag regression

Consistency of the predictor :

Given learned predictors $\hat{y}_i(x)$ and weights $\hat{p}_i(x)$, the predicted value is

$$\hat{y}(x) = \hat{y}_{\hat{I}(x)}(x)$$

with $\hat{I}(x)$ the predicted time lag index.

Call $\mathcal{L}_2(\hat{y}, \hat{I}) = \mathbb{E}_{data}\Big[\big(y_{\hat{I}(x)} - \hat{y}(x)\big)^2\Big]$ the "natural" loss.

Proposition 1. *The optimal predictor w.r.t. $\mathcal{L}_2$ is given by*

$$y^\star(x) = \hat{y}_{I^\star(x)}(x) \qquad \textit{with} \qquad I^\star(x) = \operatorname*{argmax}_i \big(\hat{p}_i(x)\big).$$

Note : $\mathcal{L}_2$ is not usable for training (non-continuous) nor for validation (incomplete).