Echo State Network for Dynamic Aperture Prediction

Maxime CASANOVA

Supervised by

Barbara DALENA
Luca BONAVENTURA
Massimo *GIOVANNOZZI*

CONTEXT AND MOTIVATIONS

GOAL

Predicting the Dynamic Aperture (DA) for large number of turns N (10^7 turns) using an Echo State Network (ESN)

What is DA?

Region of stable motion of a particle after a certain number of turns in a circular accelerator.

Why do we need DA ?

In colliders, the possible sources of unstable motion are magnetic fields and elements placement imperfection -> reduction of the region of stable motion

DA is used to define tolerance on magnetic field quality and non linear correction schemes

CONTEXT AND MOTIVATIONS

Difficulties

Estimating DA for large number of turns is very computationally time consuming
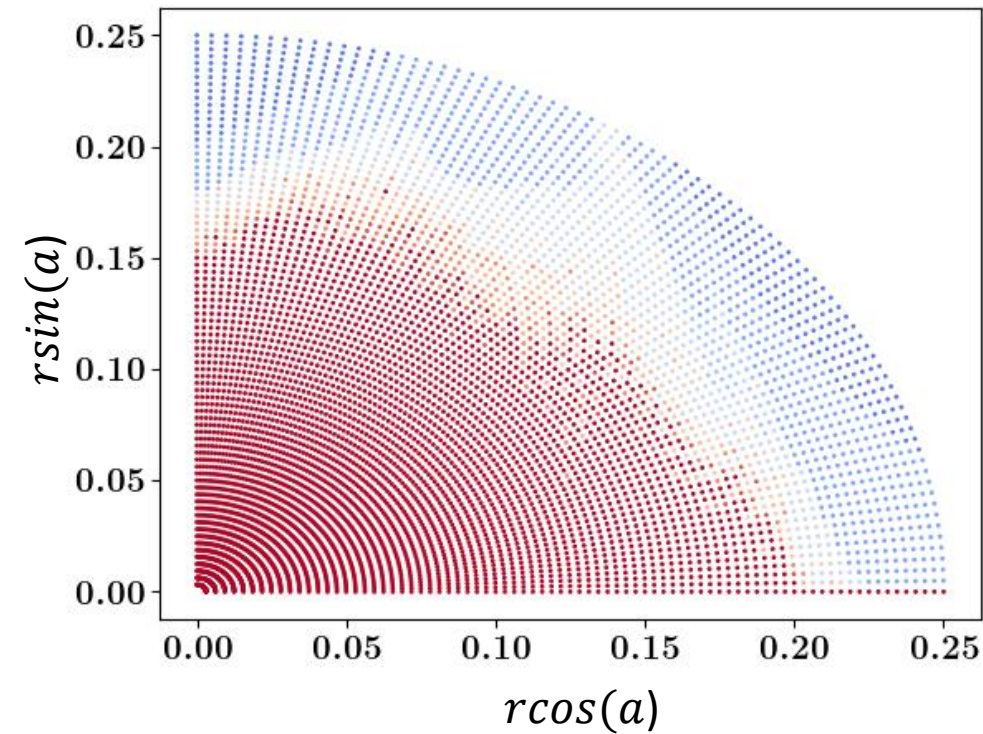
Solutions

- 0) Using Analytical models (SL)
- 1) Using ESN
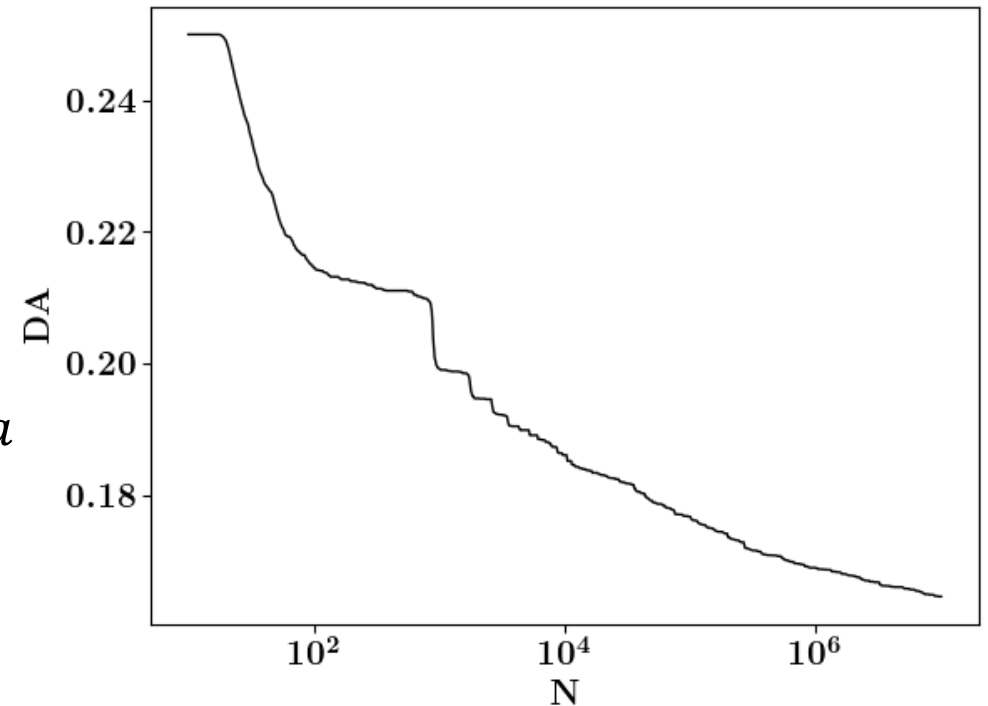- 2) Combining analytical models and ESN (SL-ESN)

Why ESN

Computationally more efficient than the others standards NNs and proved to be an universal approximant for dynamical systems
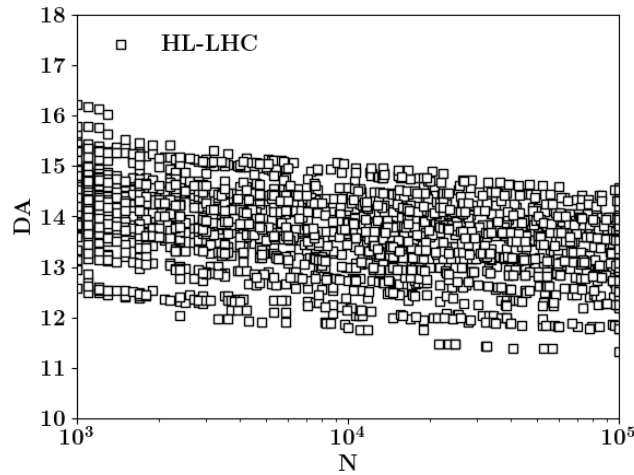
DYNAMIC APERTURE ESTIMATION
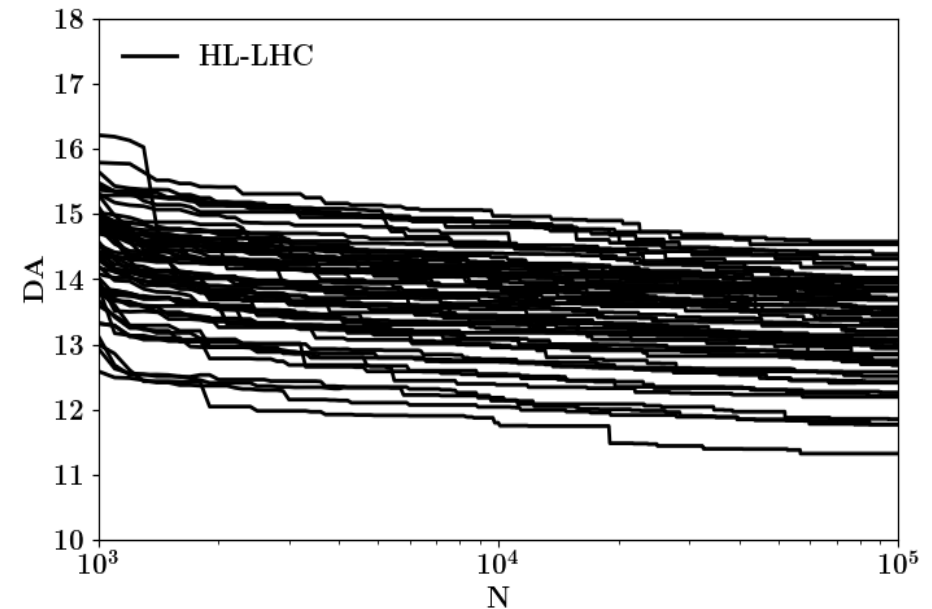

$$DA = \int_0^{\pi/2} r(a, N)\, da$$

where $r(\alpha, N)$ denotes the last value of r whose orbit is bounded after N turns

HL-LHC DATA

DA evaluated until 10^5 turns N from a realistic model of the High Luminosity LHC using SixTrack simulations

Stepwise function

60 seed corresponding to a different machine configuration (different randomly distributed magnetic field error)

4D HENON MAP

The 4D Hénon Map is a simplified model allowing to study the DA evolution to a larger number of turns than the one obtained using the realistic model of the HL-HLC

$$\begin{pmatrix} x_1^{(n+1)} \\ p_{x1}^{(n+1)} \\ x_2^{(n+1)} \\ p_{x2}^{(n+1)} \end{pmatrix} = \text{L} \begin{pmatrix} x_1^{(n)} \\ p_{x1}^{(n)} + \left(x_1^{(n)}\right)^2 - \left(x_2^{(n)}\right)^2 + \mu \left(\left(x_1^{(n)}\right)^3 - 3\left(x_2^{(n)}\right)^2 x_1^{(n)}\right) \\ x_2^{(n)} \\ p_{x2}^{(n+1)} - 2x_1^{(n)} x_2^{(n)} + \mu \left(\left(x_2^{(n)}\right)^3 - 3\left(x_1^{(n)}\right)^2 x_2^{(n)}\right) \end{pmatrix}$$

Model the effects of a sextupole and octupole

5

4D HENON MAP

$$L = \begin{pmatrix} R(w_{x1}^{(n)}) & 0 \\ 0 & R(w_{x2}^{(n)}) \end{pmatrix}, R \text{ 2D rotations}$$

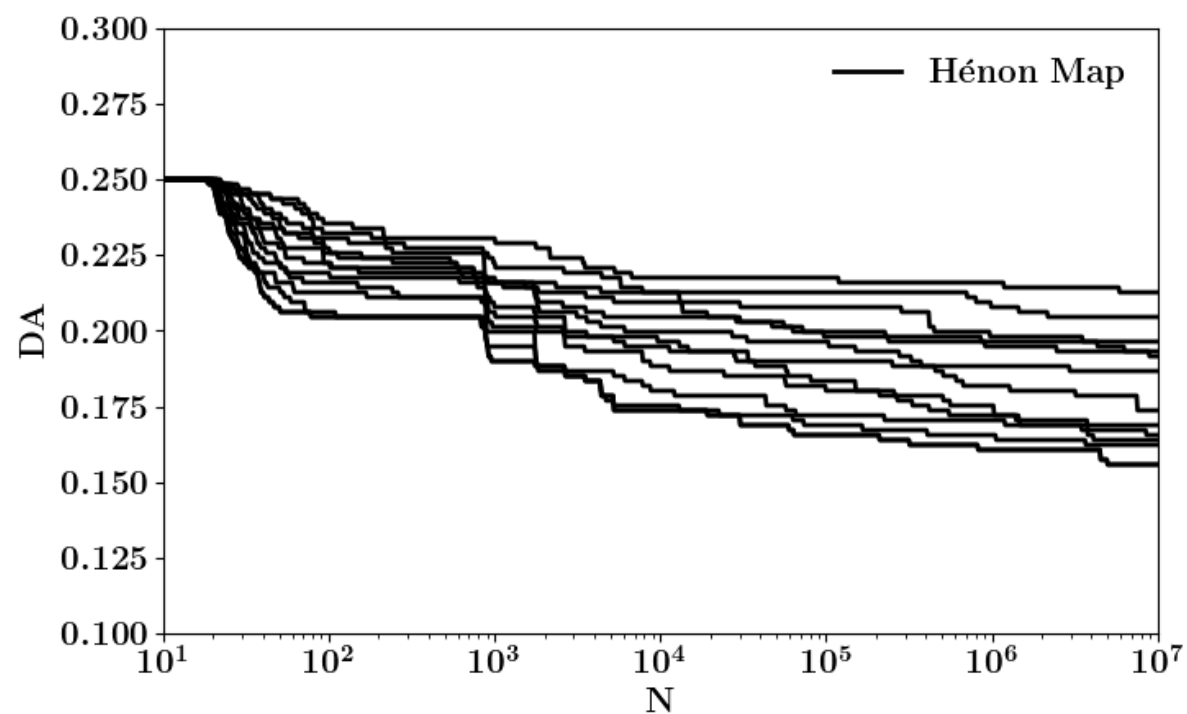The linear frequencies $w_{x1}^{(n)}$ and $w_{x2}^{(n)}$ vary such that

$$w_{xi}^{(n)} = w_{xi0}\left(1 + \varepsilon \sum_{k=1}^{m} \varepsilon_k \cos(\Omega_k n)\right), i = 1,2$$

where ε denotes the tune modulation

4D HENON MAP DATA

12 cases (different ε & $\mu = 0$)

DA evaluated until 10^7 turns N

$\varepsilon = 5.10^{-4}$

$\varepsilon = 60.10^{-4}$

A SCALING LAW MODEL

What is it ?

Analytical model based on the **Nekhoroshev theorem** used to extrapolate the DA

It provides an estimate for the number of turns $N(r)$ for which the orbit of an initial condition of amplitude r remains bounded

Scaling Law

$$DA^{SL} = \rho_* \left(\frac{\kappa}{2e}\right) \frac{1}{\ln(N)^{\kappa}}$$

Two fitting parameters κ and ρ_*

How to perform the fit ?

Using the least square method

LEAKY ECHO STATE NETWORK

We recall the discrete time dynamic of the leaky ESN using the Euler explicit method with T time steps of size Δt:

$$x_{k+1} = \left(1 - a\frac{\Delta t}{c}\right) x_k + \frac{\Delta t}{c} f\left(W^{in} u_k + W x_k\right)$$

$$x_k^{out} = g(W^{out}[x_{k+1}, u_k])$$

where a denotes the leaking rate, c a global time constant, f a sigmoid function, g the output activation function, W^{in} the input weight matrix, W the reservoir matrix, W^{out} the output weight matrix, u the ESN input and x^{out} the ESN output
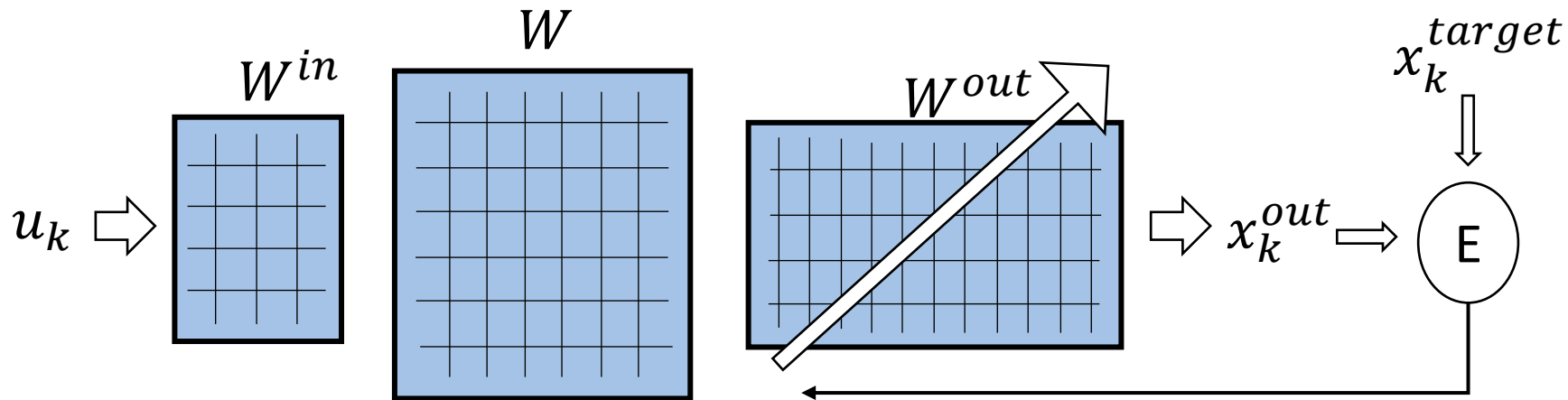
W^{in} and W are randomly initialized

LEAKY ECHO STATE NETWORK

Training W^{out} is the only trained matrix using Ridge regression with β regularization

$$W^{out} = X^{target} X^T (XX^T + \beta I)$$

$$X = \begin{vmatrix} u_0 & \text{------} & u_{T-1} \\ x_1 & \text{------} & x_T \end{vmatrix}$$

where X^{target} contains the training data x^{target} and X the concatenation of the ESN input u and internal state x

SIMULATION SETUP

Fixed Hyperparameters

Reservoir size $N_r = 50$

Burn-in $BI = 0$

Leaking rate $a = 0.01$

Spectral radius $\rho = 0.99$ (satisfy the ESP)

Sparisty $s = 0$ (matrices full dense)

-> The regularization parameter β is the only FREE hyperparameter found by the validation process for each input

Validation data

ESN : uses the DA data direclty from the HL-LHC and Henon map models

SL-ESN : uses the data obtained by fitting the scaling law SL

TRAINING/VALIDATION/TEST DATA

Training data   Compute W^{out}

Validation data   Tune only the β hyperparameter (the others are fixed acording previous studies)

Test data   Evaluate the model for new data

HL-LHC

Train + Validation : $[10, 5.10^4]$ turns

Test : $[5.10^4, 1.10^5]$ turns

4D Hénon Map

Train + Validation : $[10, 5.10^4]$ turns

Test : $[5.10^4, 1.10^7]$ turns

VALIDATION PROCEDURE

GOAL

Find the best β which minimize the MSE in the **Validation set**

Step 1

Generate randomly $N_w = 30$ different pair of input and reservoir weight matrices (W^{in}, W)

$(W^{in}, W)^1$ $(W^{in}, W)^2$ $(W^{in}, W)^{30}$
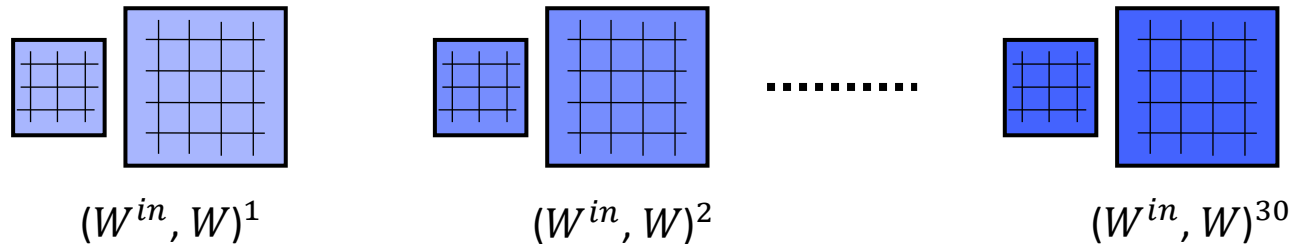
Step 2

Generate a list of β of size N_β of different values of β

$$\beta = \left(\beta^1, \ldots, \beta^{N_\beta}\right)$$

VALIDATION PROCEDURE

Step 3

Compute the MSE^{val} in the validation set for each β and every pairs (W^{in}, W)

$$M = \begin{pmatrix} MSE^{val,\beta^1}_{(W^{in}, W)^1} & \cdots & MSE^{val,\beta^1}_{(W^{in}, W)^{30}} \\ \vdots & \ddots & \vdots \\ MSE^{val,\beta^{N_\beta}}_{(W^{in}, W)^1} & \cdots & MSE^{val,\beta^{N_\beta}}_{(W^{in}, W)^4} \end{pmatrix}$$

Step 4

For each β, compute the mean of each rows of M

$$m_{mean} = (m^{\beta^1}_{mean}, m^{\beta^2}_{mean}, \dots, m^{\beta^{N_\beta}}_{mean})^T$$

Step 5

Choose the minimum of m_{mean} and select the corresponding β. If the minimum is at the $-$ith row, then we will select β^i

14

EVALUATION PROCEDURE

GOAL

Given β^i found in the validation method, evaluate the ESN in the **Test set**

Step 1

Given β^i, store the evaluated dynamic of the ESN x^{out} for each pairs (W^{in}, W)

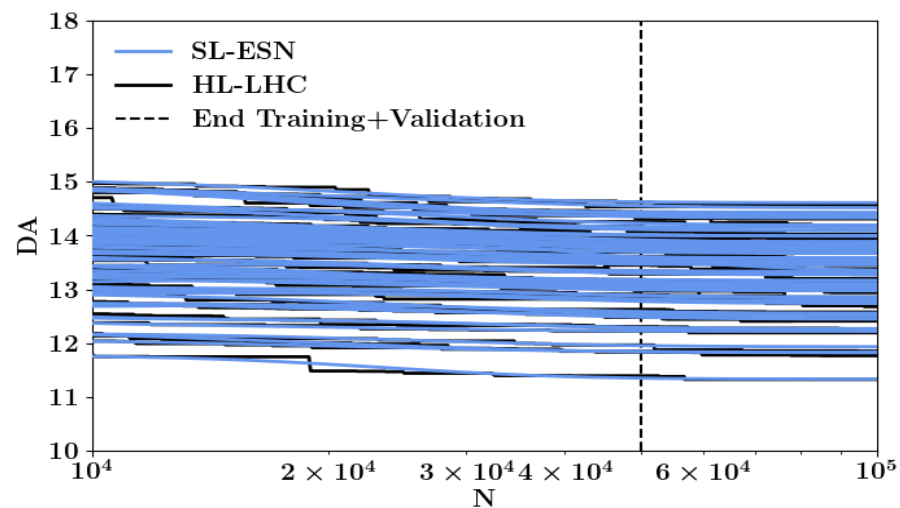$$x^{out} = (x_1^{out}, x_2^{out}, \dots, x_{30}^{out})^T$$

Step 2

Take the mean x_{mean}^{out} of x^{out}

Step 3

Compute standard deviation σ of x_{mean}^{out}

RESULTS HL-LHC
(60 seeds)
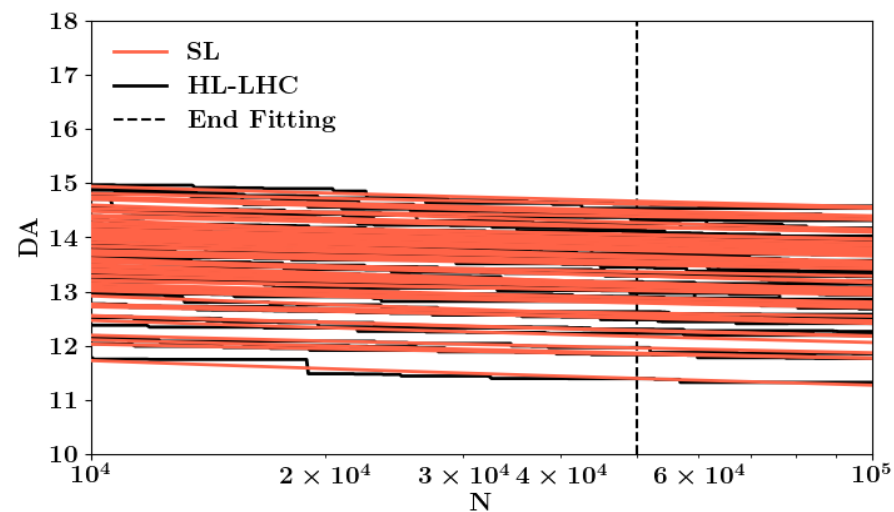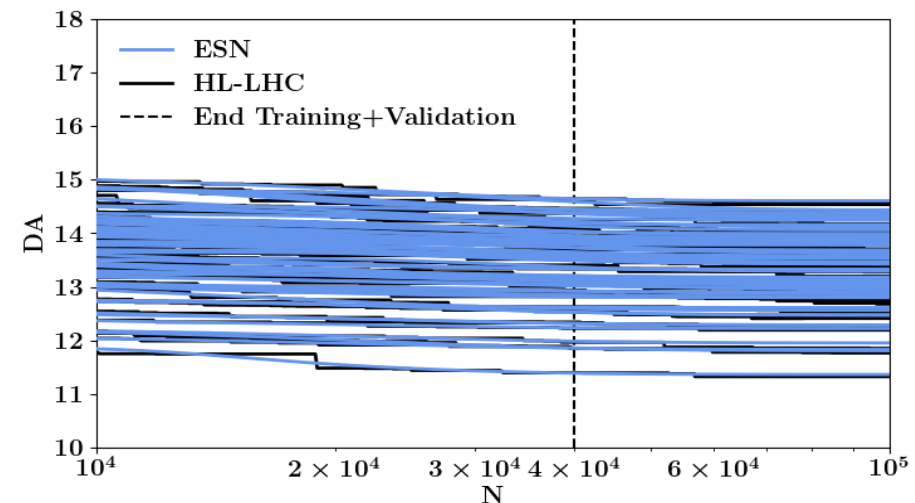
RESULTS HL-LHC

(60 seeds)
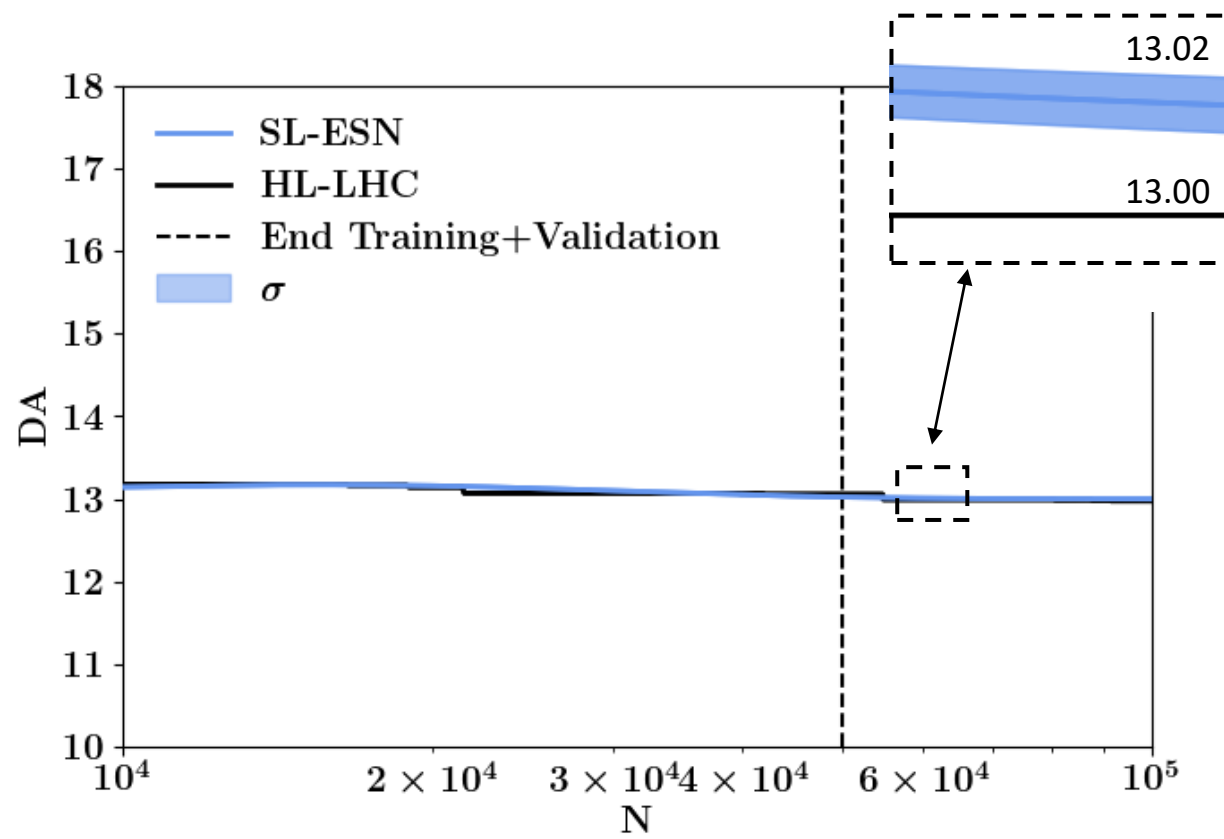
The prediction given by the SL-ESN is in average 25% better
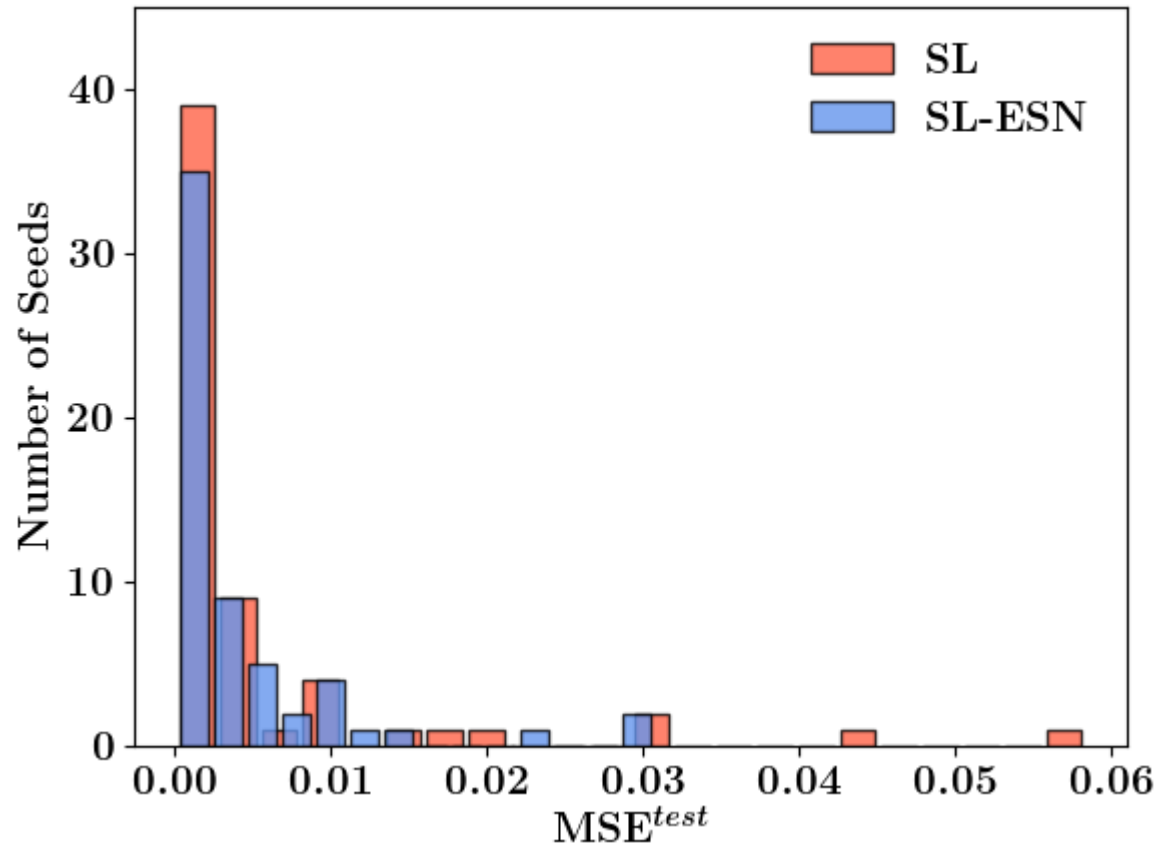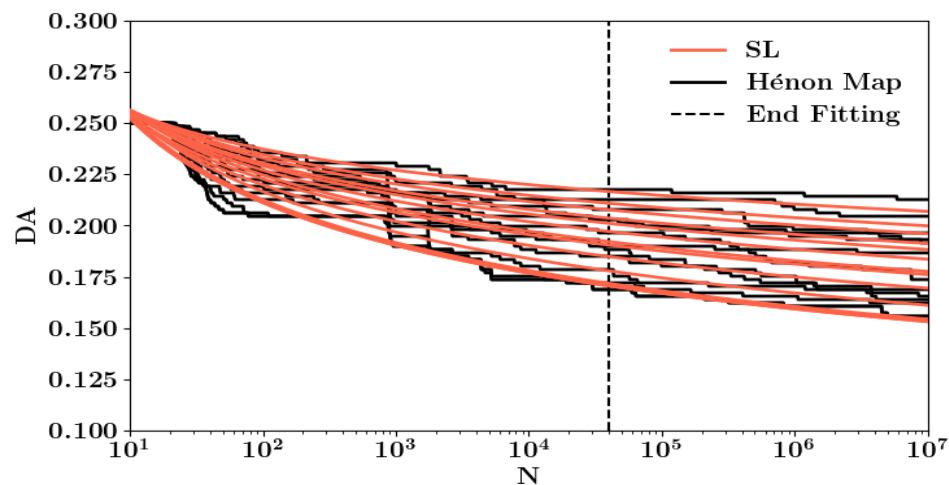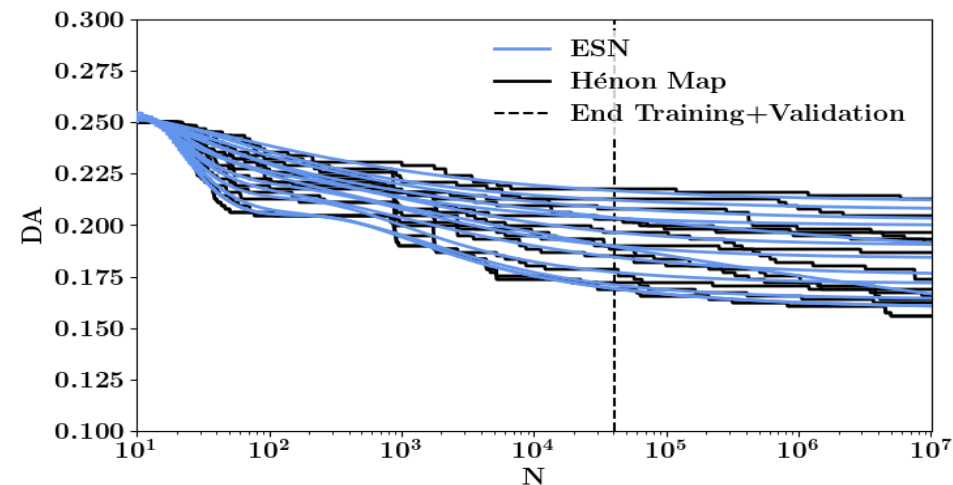
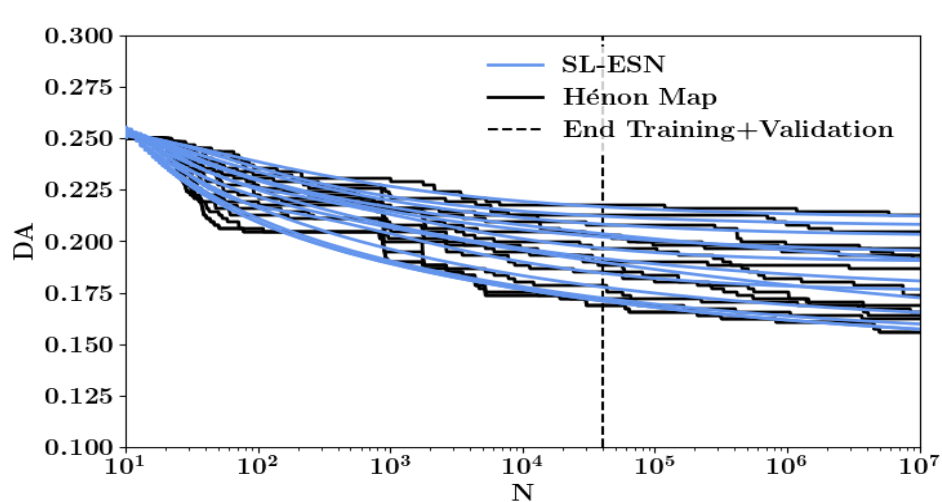The predictions given by the ESN is in average 2% lower
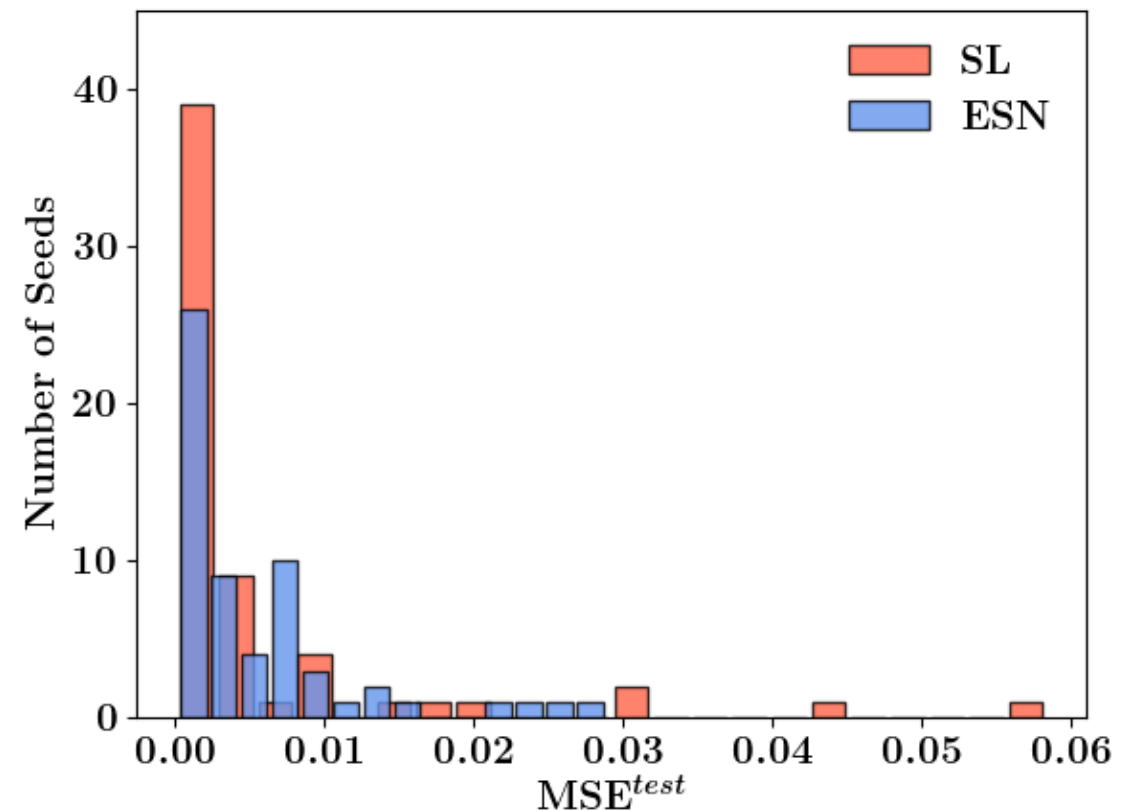
RESULTS HL-LHC
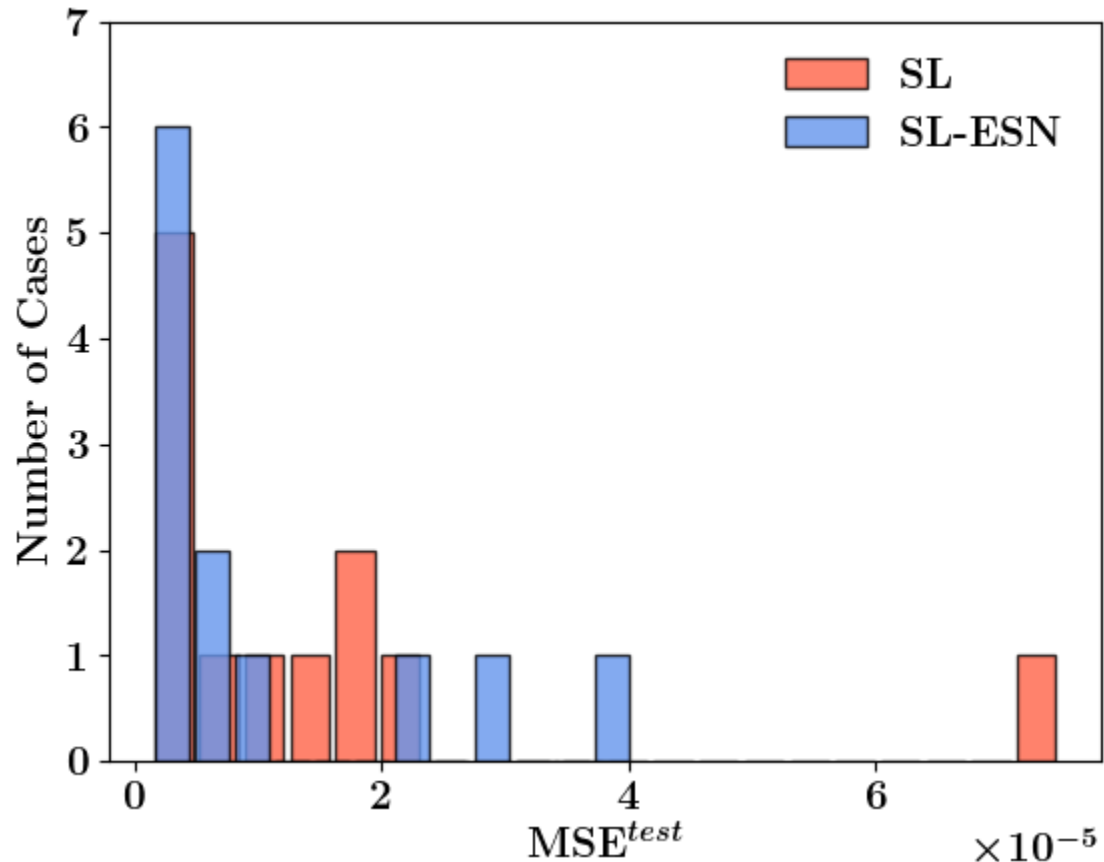## (1st seed)
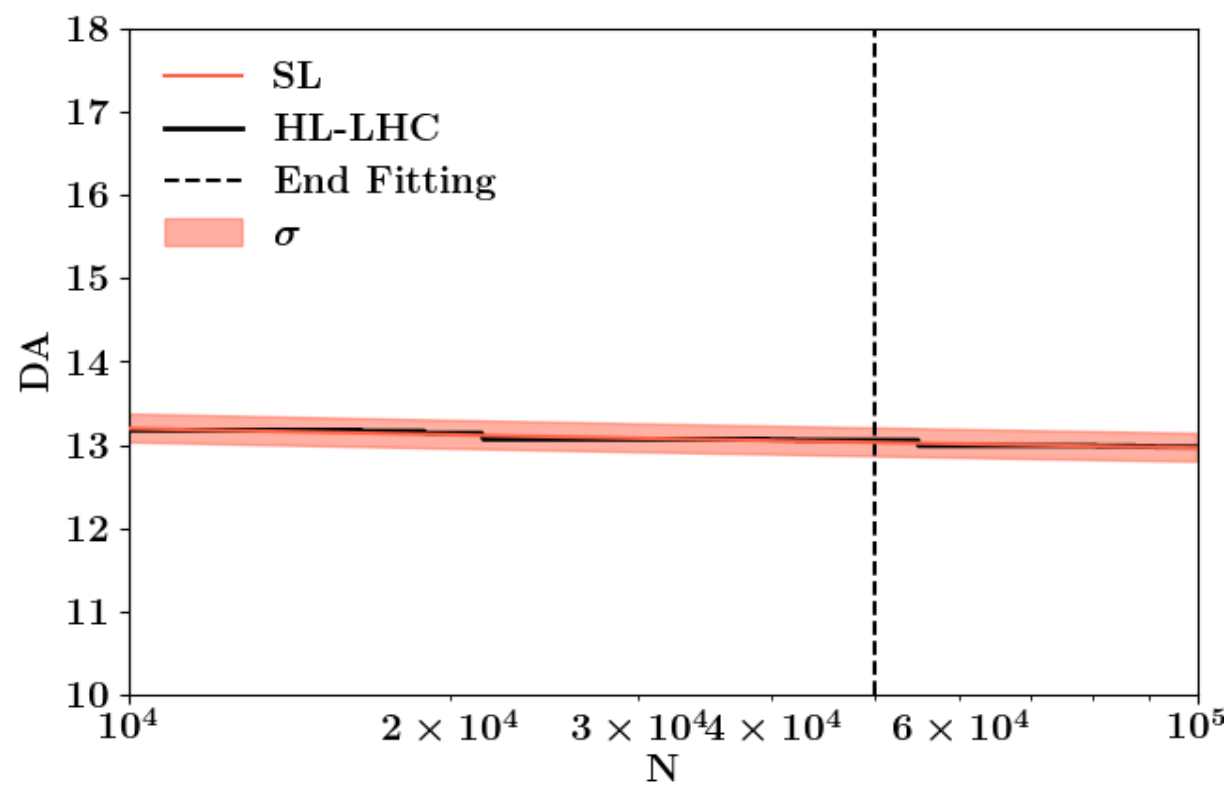
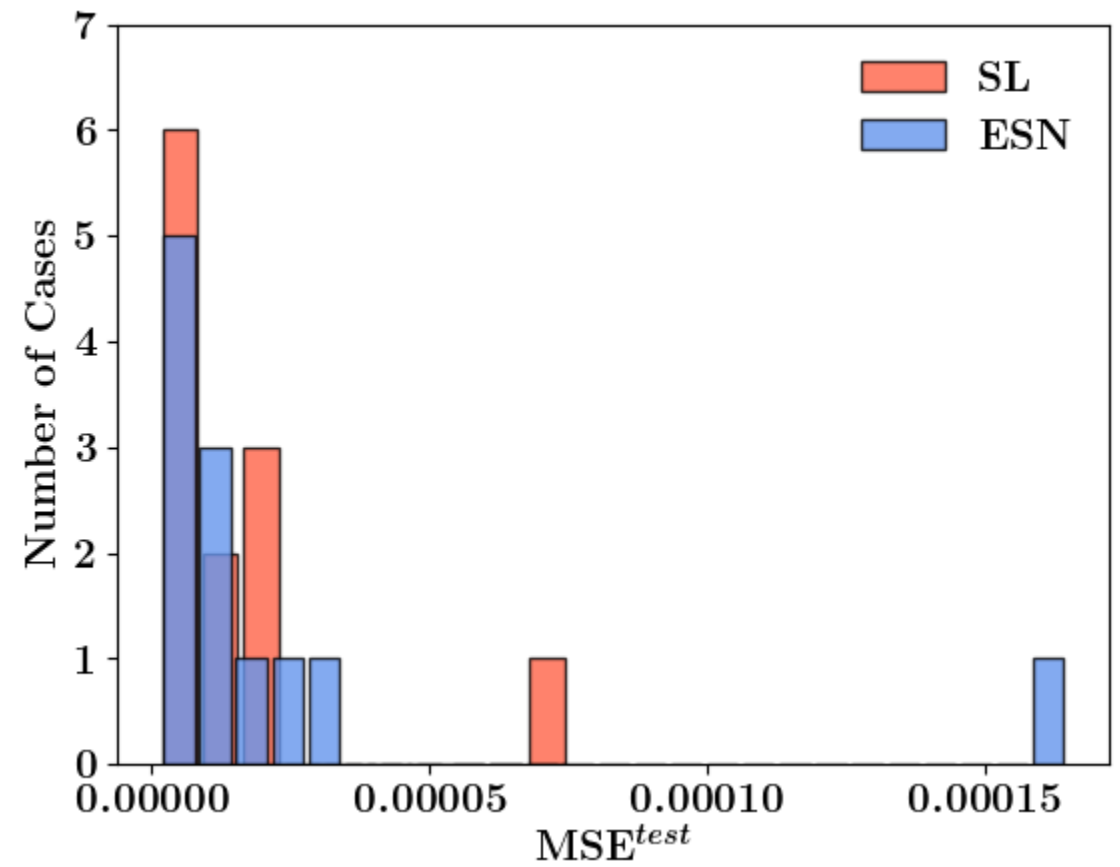# RESULTS 4D Hénon Map
(sextupole $\mu = 0$)

RESULTS 4D Hénon Map
(sextupole $\mu = 0$)

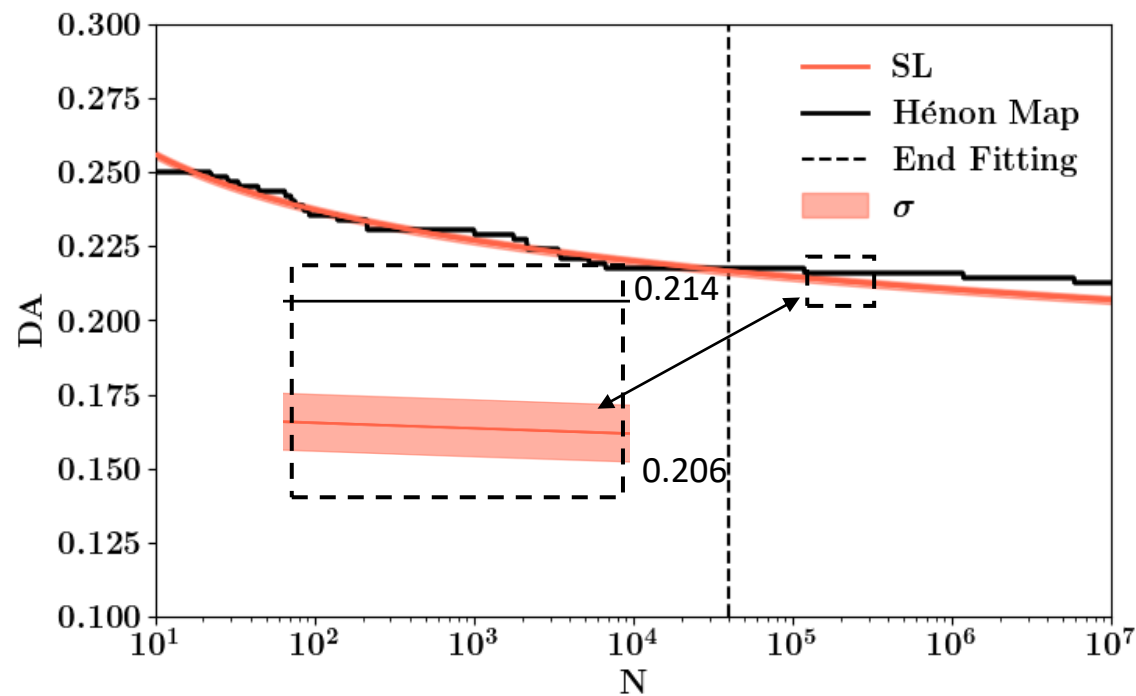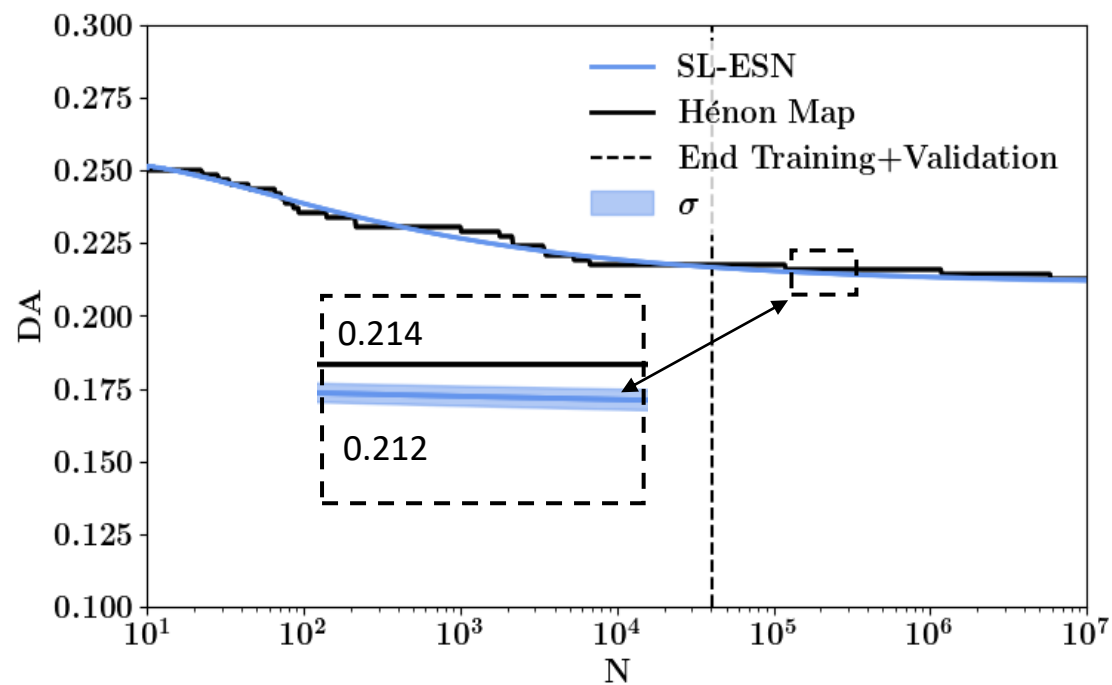The predictions given by the SL-ESN is in average 29% better

The predictions given by the ESN is in average 27% lower

RESULTS 4D Hénon Map
(sextupole $\mu = 0$)

$\varepsilon = 5.10^{-4}$

CONCLUSION

For the cases studied,

     - In average the prediction given by SL-ESN is 25% better than with SL

     - The maximal prediction error is almost two times lower with SL-ESN than with SL

What Next ?

- How to compare scaling law with ESN
- Use ESN to produce at lower cost longer time series to be used to fit the scaling law
- Physics Informed ESN adding a non linear term in the cost function during training
- Studying the effects of the octupole for the 4D Hénon Map $(\mu \neq 0)$
- Compare computational cost of sparse and full dense matrix

REFERENCES

Lyudmila Grigoryeva and Juan-Pablo Ortega. «Echo state networks are universal». In: *Neural Networks* 108 (2018)

E. Todesco and M. Giovannozzi. «Dynamic aperture estimates and phase- space distorsions in nonlinear betatron motion». In: *Physical review. E,* (1996)

M. Titze M. Giovannozzi C.E. Montanari. «Dynamic aperture estimates for 4D and 6D non-linear motion and beam intensity evolution models». (2022)

Mantas Lukoševičius and Herbert Jaeger. «Reservoir computing approaches to recurrent neural network training». In: Comput*er Science Review* (2009)