Echo State Network for Dynamic Aperture prediction

Q. Bruant, B. Dalena, V. Gautard (CEA irfu and Paris-Saclay University)

F. Bugiotti, A. Guelfane, Y. Zuo, I. Cherkaoui, J. KPODJADAN (CentraleSupelec and Paris Saclay University)

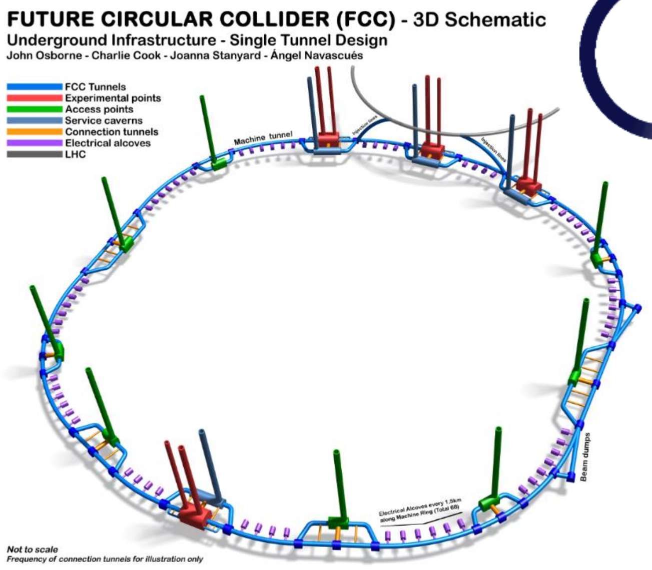
M. Casanova (CEA/IRFU & Poli Milano), L. Bonaventura (Politecnico di Milano), M. Giovannozzi (CERN)

Thanks to: J. Keintzel, M. Gael, Y. Ohnishi

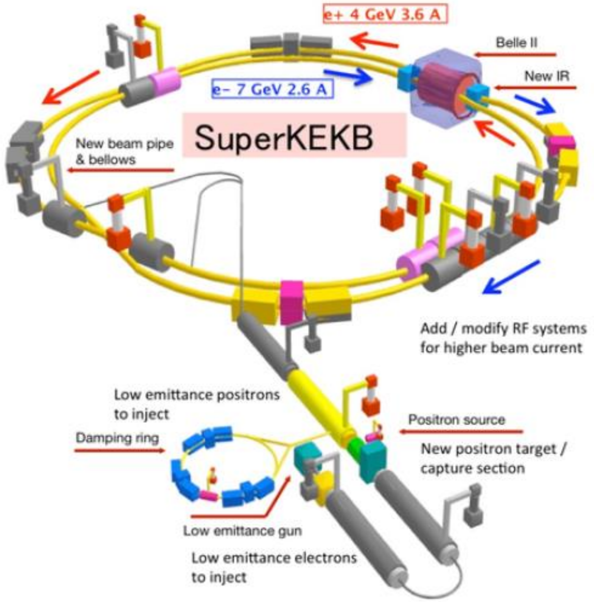
Outline

- Dynamic aperture
- ESN for DA prediction
- Results

FCC and SuperKEKB



FUTURE CIRCULAR COLLIDER



KEK

International FCC project (CERN as host lab)

Continue to study experimental high energy particle physics
Accelerator options: FCC-hh, FCC-ee, FCC-he

CDRs publié dans **European Physical Journal C (Vol 1) and ST (Vol 2-4)**

http://fcc-cdr.web.cern.ch/

Existing e+e- collider:
small size FCC-ee
proofs of principle of several design choices

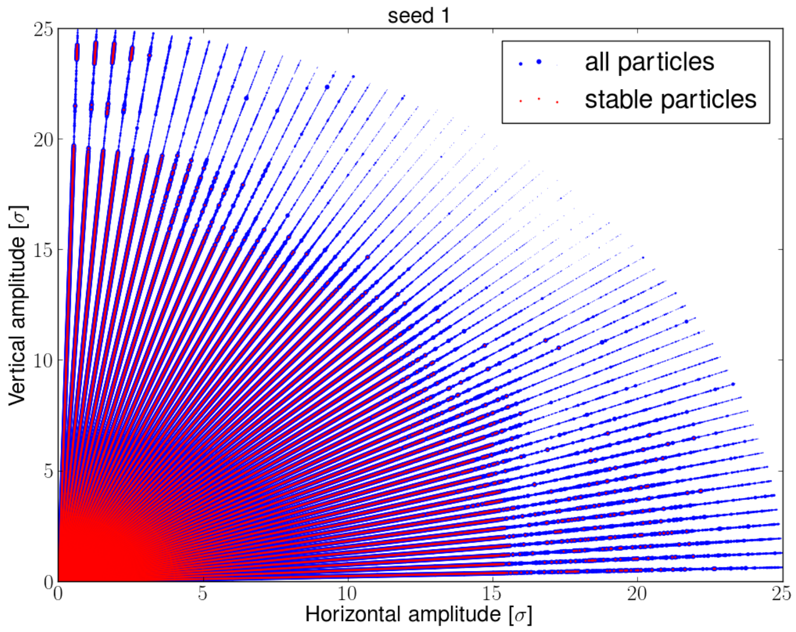
https://doi.org/10.1093/ptep/pts083

1. Dynamic Aperture (DA) prediction

Dynamic Aperture

Long-term Dynamic Aperture (DA):

- defines the region of stable motion of a particle in an accelerator
- is used to define tolerances on magnetic field quality
- is used to optimize the performance of the circular accelerators (beam losses, beam lifetime)
- corrects linear and non linear imperfections



Computed as the initial amplitude corresponding to particle lost after the 10^N revolutions in the accelerator (typically N=5,6)

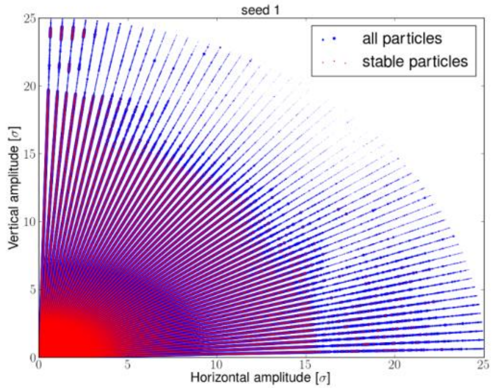
Particle tracking simulations:
From 1 day to one week or more…

Dynamic Aperture

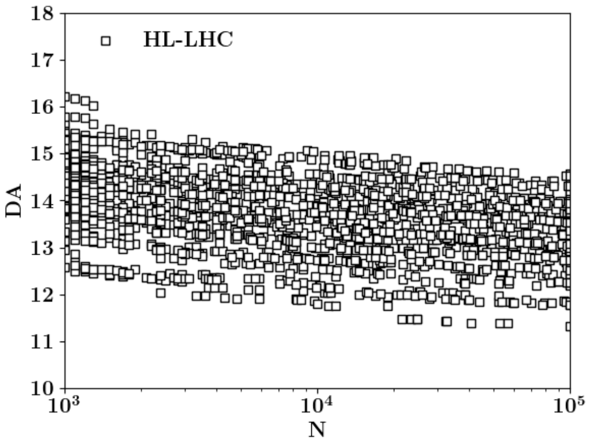
Bonded movement phase-space region

$$\mathcal{A}(N) = \int\int\int\int \chi(x_1, x_2, p_{x_1}, p_{x_2}) dx_1 dx_2 dp_{x_1} dp_{x_2}$$

$$\chi(x_1, x_2, p_{x_1}, p_{x_2}) = \begin{cases} 1 \text{ if the motion starting at } (x_1, x_2) \\ \text{with momentum } (p_{x_1}, p_{x_2}) \text{ is bounded after } N \text{ turns} \\ 0 \text{ else} \end{cases}$$



$$DA = \left(\frac{2\mathcal{A}(N)}{\pi^2}\right)^{\frac{1}{4}}$$



- Tracking simulations: initial particle amplitude of particle lost at 10^N revolutions for 60 different machine error configurations (seeds)
- Average over particle angles in x-y space: define a radius of the stable sphere (with r_s last stable radius at given angle and number of turns)

Analytical Scaling Laws and DA extrapolation

- Based on Nekhoroshev theorem

It proves the stability of a quasi integrable Hamiltonian system for finite but exponentially long time (N turns) on an open set of initial conditions (amplitudes)

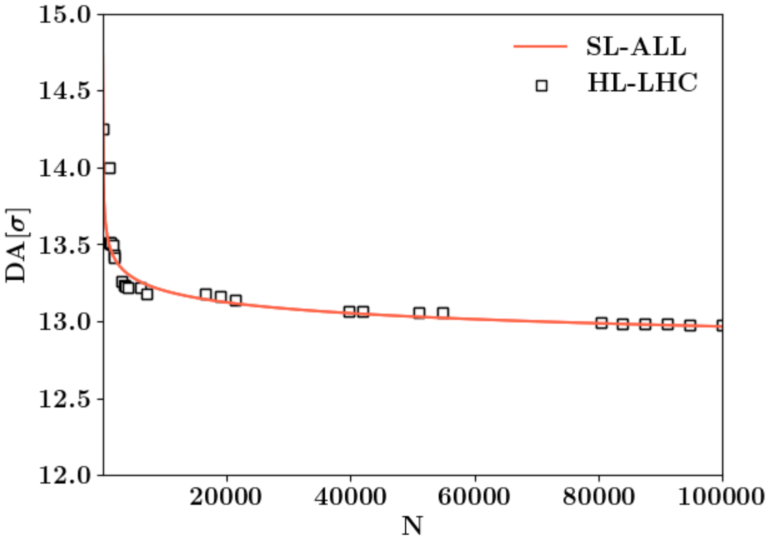
- Time evolution of DA (The Scaling Law):

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

[Phys. Rev. Acc. Beams 22 104003]

Two fitting parameters κ and ρ_*

It is used to fit (with Least Square Method) the DA data



- In order to improve the quality of the fit, **Gaussian Processes** are used to increase artificially the DA data

[https://www.mdpi.com/2078-2489/12/2/53]

AI applications

Goal

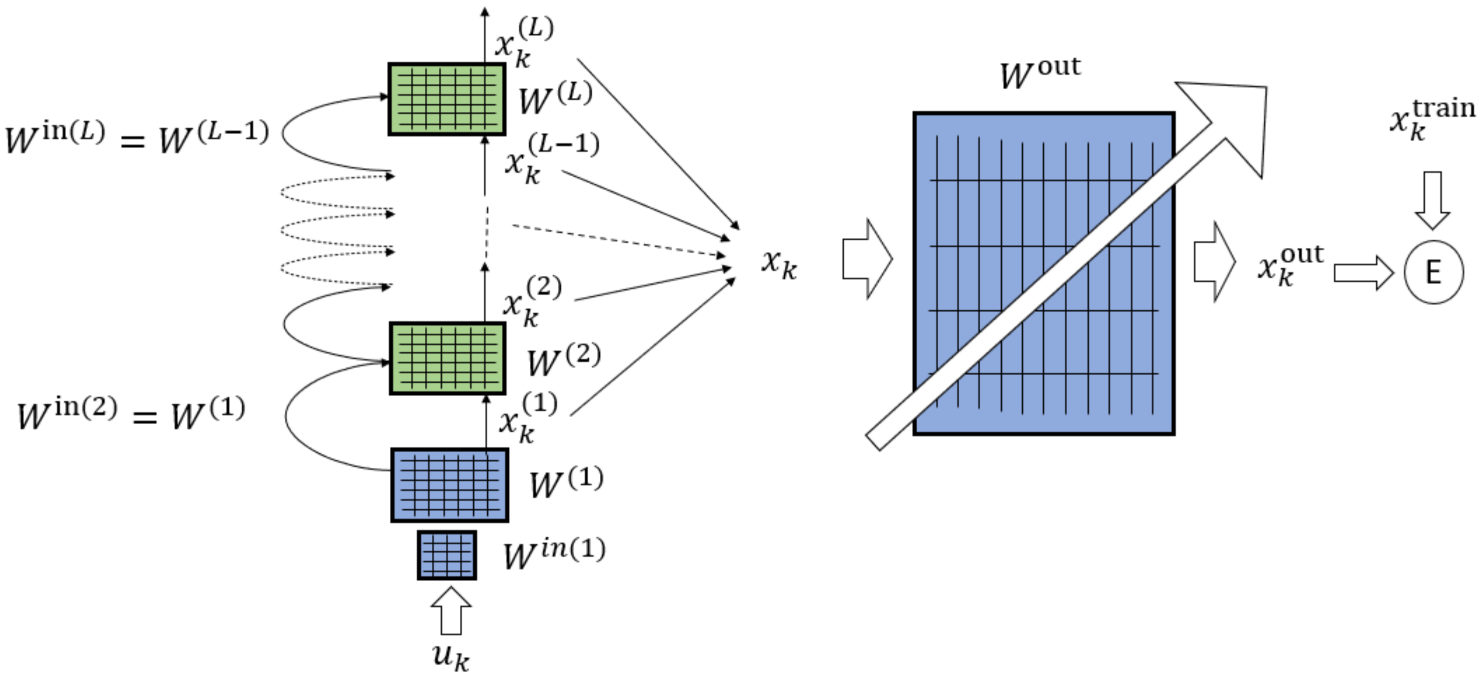
Predict the time evolution of Dynamic Aperture using tracking simulations for a limited number of turns and recurrent neural networks to generate (extrapolate) the numerical values at higher number of turns

- ✓ Fast surrogates models to speed up simulations of performance of the future Colliders
- ✓ Anomaly detection techniques and noise reduction techniques to improve Turn-by-Turn BPMs measurements

- ➢ Optimization of accelerator design and settings with Bayesian methods, reinforcement learning, etc?

2. ESN for DA prediction

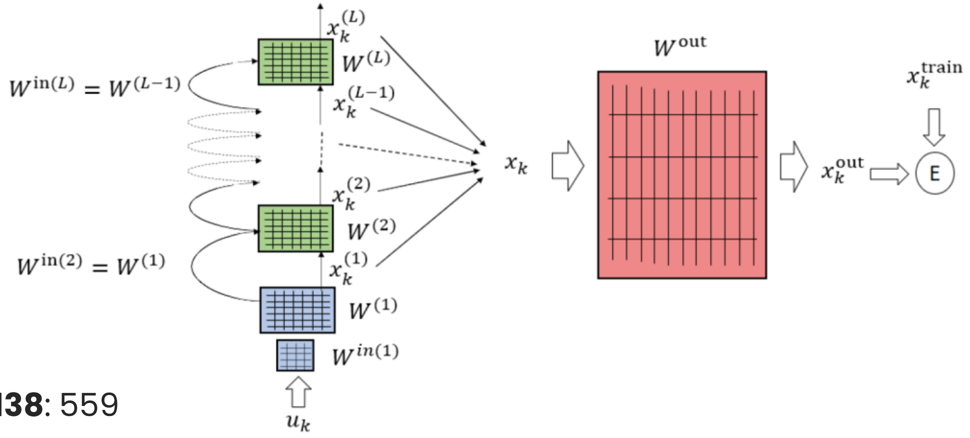
ECHO STATE NETWORKS (DeepESN)



- A class of recurrent neural network using the Reservoir Computing approach **[H. Jaeger 2001]**
- They have been proved to be an universal approximant of dynamical systems **[L. Grigoryeva and J.P. Ortega 2018]**
- The network hidden layer (the reservoir) is initialized randomly and not trained (no back propagation is required)
- The output layer only is trained (usually with linear regression) to compute the weight (W^{out}) that project the reservoir state into the predicted output

Echo State Networks (Deep ESN)

Reservoir Computing networks models and replicates the time evolution of Dynamic Aperture, allowing to speed−up tracking simulations for high energy hadron storage rings.



Deep Echo State Networks

M. Casanova, B.D. et al. , Eur. Phys. J. Plus (2023) **138**: 559

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

$$x_k^{out} = g(W^{out}[x_k, u_k])$$

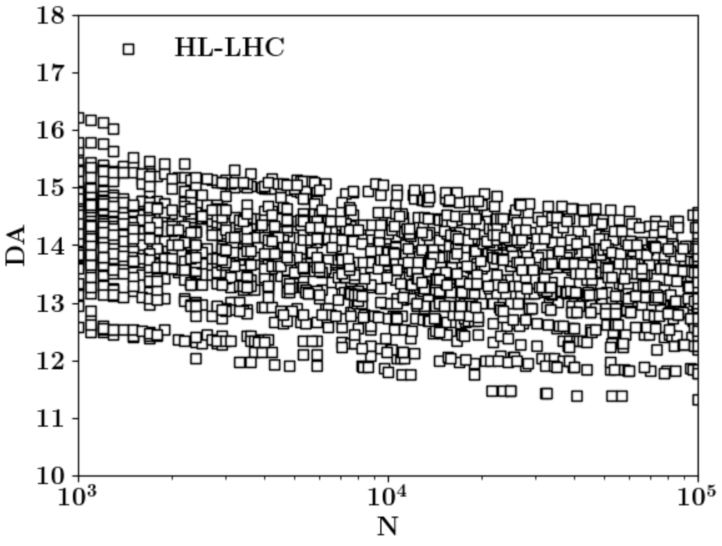
$$W^{out} = X^{target} X^T (X^T X + \beta I)^{-1}$$

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 , x_k is the concatenation of all $x^{(l)}_k$ and x^{out} the ESN output

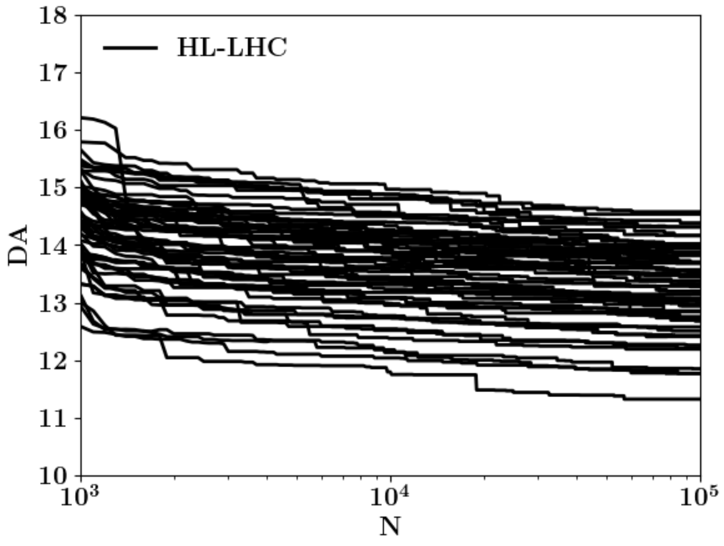
W^{in} and $W^{(l)}$ are randomly initialized

We impose the spectral radius (ρ) of the matrix $\frac{\Delta t}{c}|W| + \left(1 - a\frac{\Delta t}{c}\right) I < 1$ $\Rightarrow$ it guarantees the sufficient condition of the ESP

DATA preparation and splitting

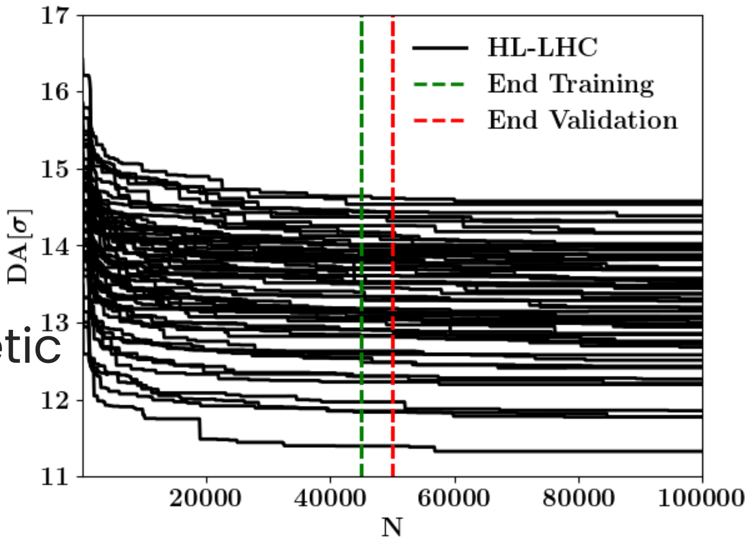


Stepwise constant function



Data splitting:

DA(N) [10, 10^4] train
DA(N)]10^4, 5 10^4] validation
DA(N)]5 10^4, 10^5] test



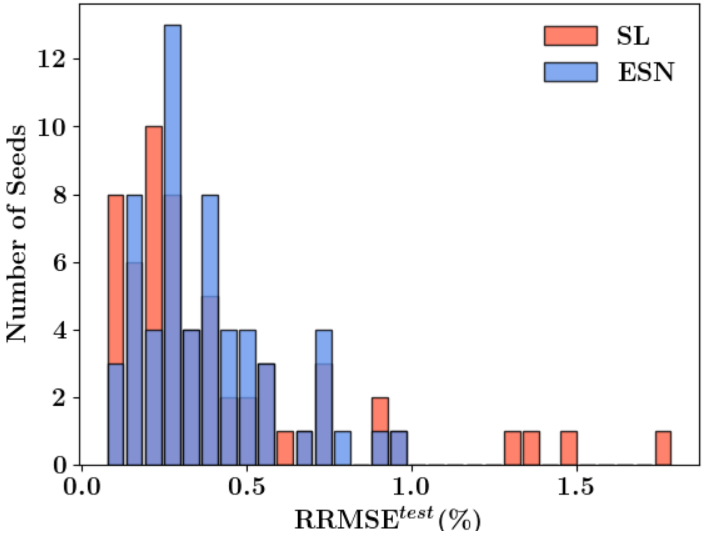
- High-Luminosity LHC, simulated
- 60 configurations → 60 distributions in magnetic lattice
- Nominal parameters

3. Results

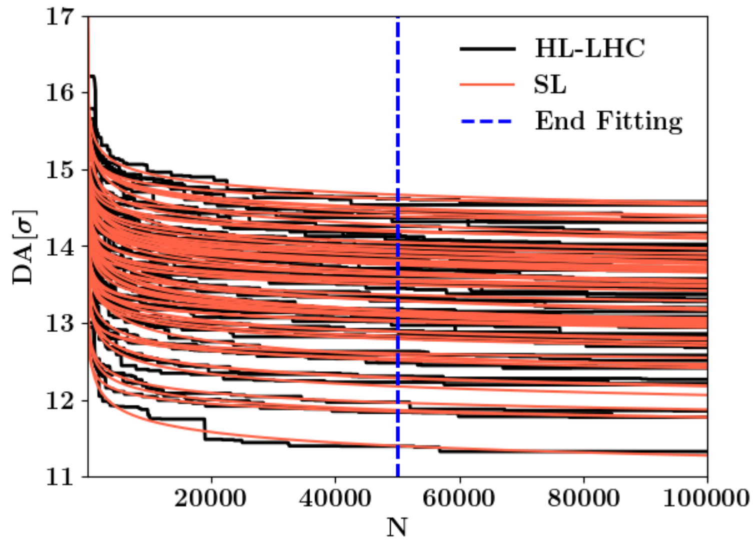
DA ESN prediction and comparison with Scaling Laws (SL)

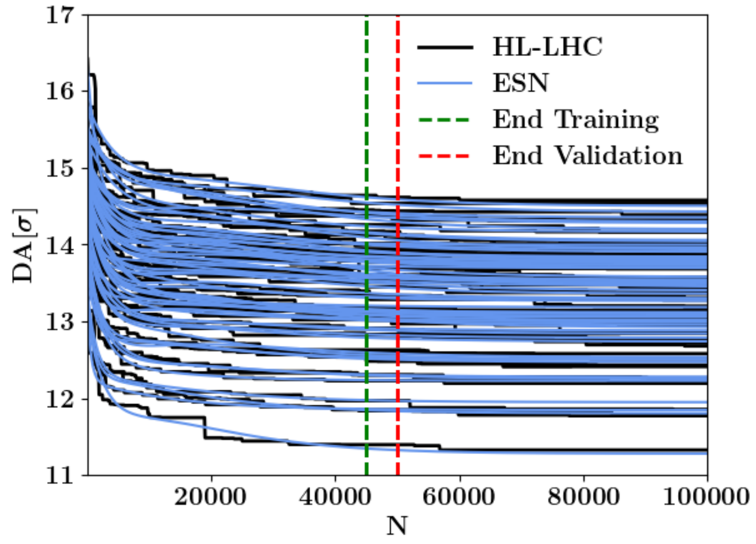
- **ESN** are able to give **physical predictions** for all seeds
- **ESN** predictions are better for the **outlier** seeds than **SL** fit (max MSE ~50% lower)
- **ESN** predictions are on average comparable to **SL** fit

$$\text{RRMSE}^{\text{test}} = 100\sqrt{\frac{\sum_{k=1}^{k_{\text{test}}}(x_{\text{mean},k}^{\text{out}} - x_k^{\text{test}})^2}{\sum_{k=1}^{k_{\text{test}}}(x_k^{\text{test}})^2}}$$



	Mean	Max	Min	Std
ESN	0.37	0.94	0.06	0.20
SL	0.42	1.78	0.07	0.35



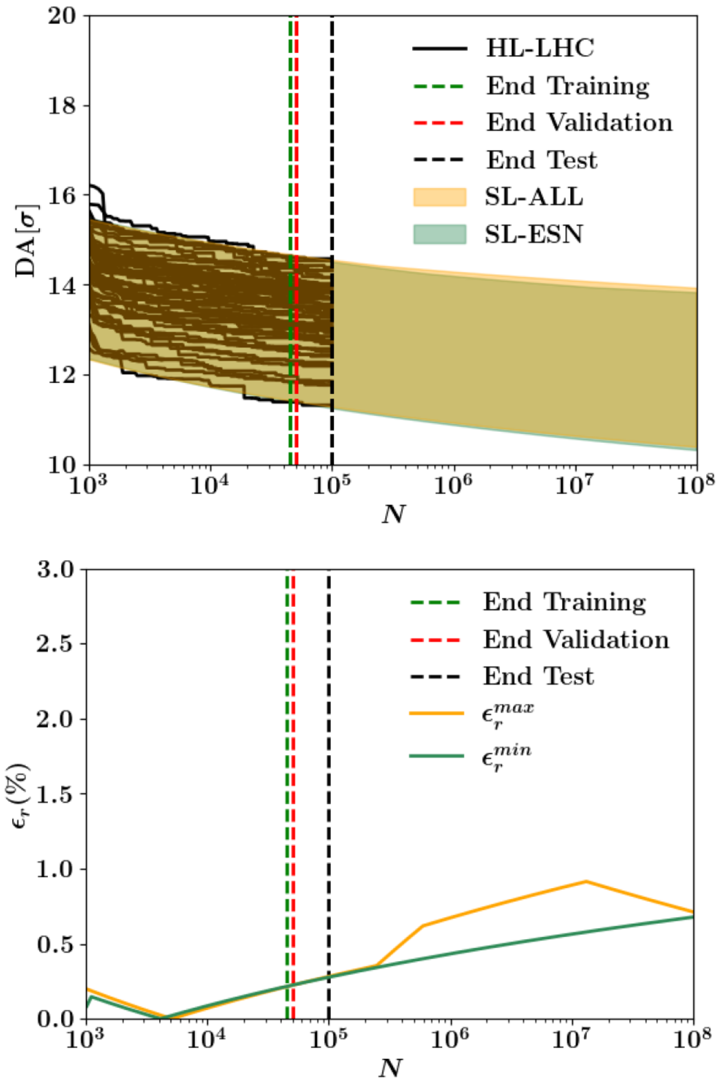


Forecast beyond the data set (HL-LHC)

We fit SL on ESN prediction up to 10^5 turns and we forecast until 10^8 turns (**SL-ESN**).
We compare the results with SL fit on tracking data and extrapolation up to 10^8 turns (**SL-ALL**).

The envelope defined by the error on the minimum and on the maximum DA value between the **SL-ALL** and **SL-ESN** is below 1%.

$$\varepsilon_r^{min,max} = \left(\frac{DA_{SL-ALL}^{min,max} - DA_{SL-ESN}^{min,max}}{DA_{SL-ALL}^{min,max}} \right)$$



ESN prediction robustness

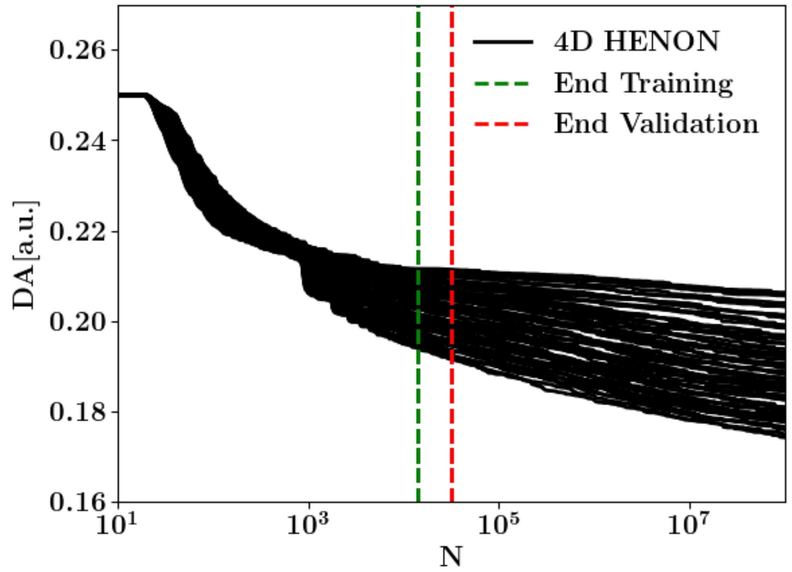
4D Henon map

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

With $\mathrm{L} = \begin{pmatrix} R(w_{x1}^{(n)}) & 0 \\ 0 & R(w_{x2}^{(n)}) \end{pmatrix}$ w linear frequencies

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

ε allow for tune modulation



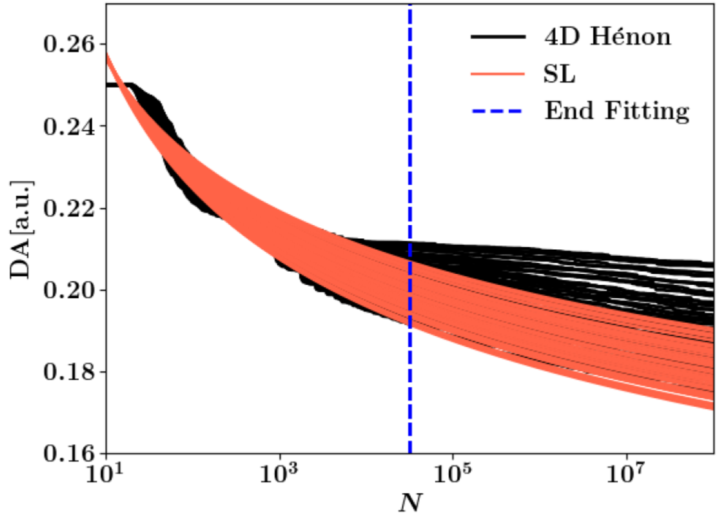
Henon map data splitting

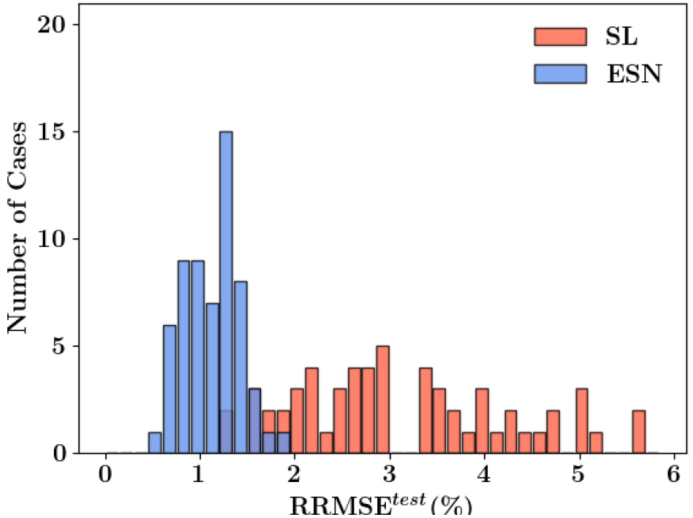
Henon Map results

- Tuning of the ESN hyper-parameters on Henon Map data

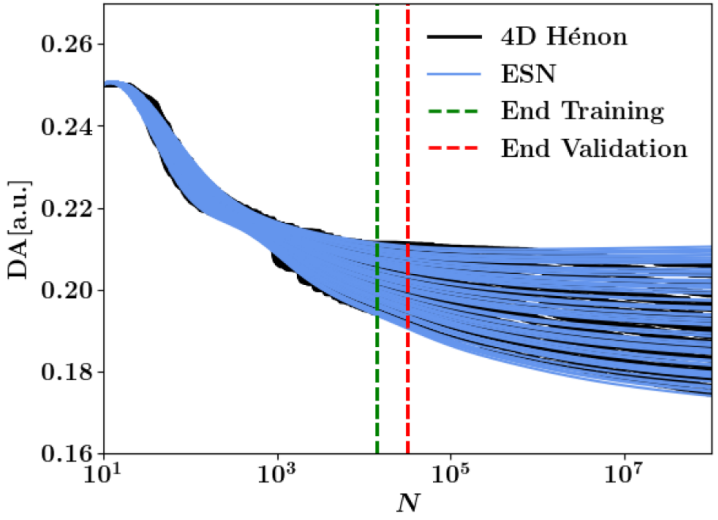
N_r	β	ρ	a	BI	L	Δt	f	s
20	9.10^{-6}	0.99	1	0	1	0.004	tanh	0

- ESN model performs better than SL on Henon Map data





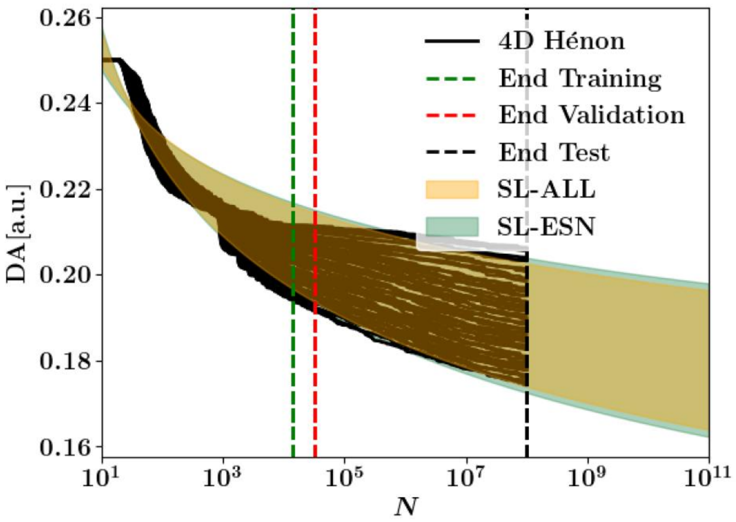
	Mean	Max	Min	Std
ESN	1.13	1.89	0.59	0.28
SL	3.17	5.85	1.25	1.18

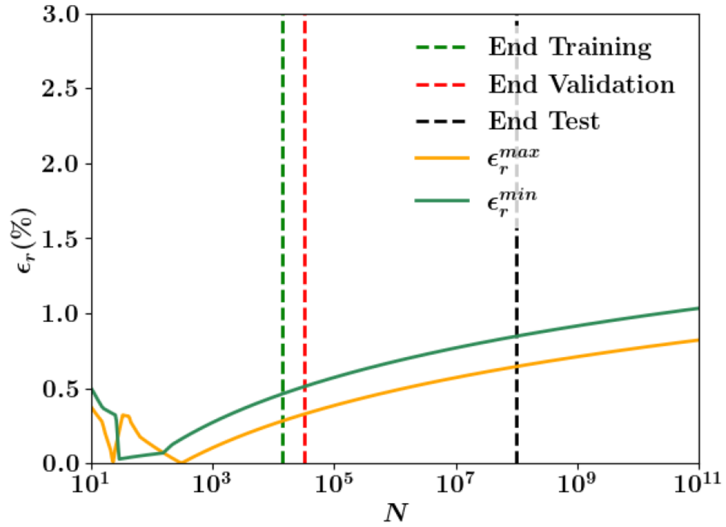


Forecast beyond the data set (Henon Map)

We fit SL on ESN prediction up to 10^8 turns and we forecast until 10^{11} turns (**SL-ESN**).
We compare the results with SL fit on tracking data and extrapolation up to 10^{11} turns (**SL-ALL**).

The envelope defined by the error on the minimum and on the maximum DA value between the **SL-ALL** and **SL-ESN** is below 1%.





$$\varepsilon_r^{min,max} = \left(\frac{DA_{SL-ALL}^{min,max} - DA_{SL-ESN}^{min,max}}{DA_{SL-ALL}^{min,max}} \right)$$

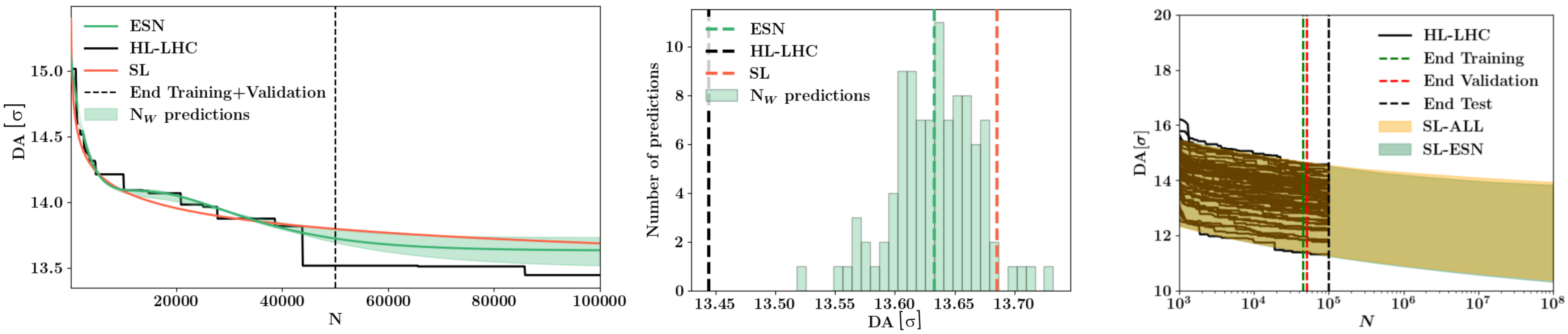
Ensemble approach

Training and validation:

- Generate randomly N_w =100 different pairs of input and reservoir matrices (W_{in}, W)
- Consider a set H of *hyper-parameters* (Nr, L, BI ρ, β, Δt)
- Train and validate the MSE on the validation data set for H and (W_{in}, W) pairs
- For each H and each seed, compute the **mean RRMSE over the 100** (W_{in}, W) pairs
- Choose the H set value which minimize the mean RRMSE in the validation data, on average over the seeds

Prediction:

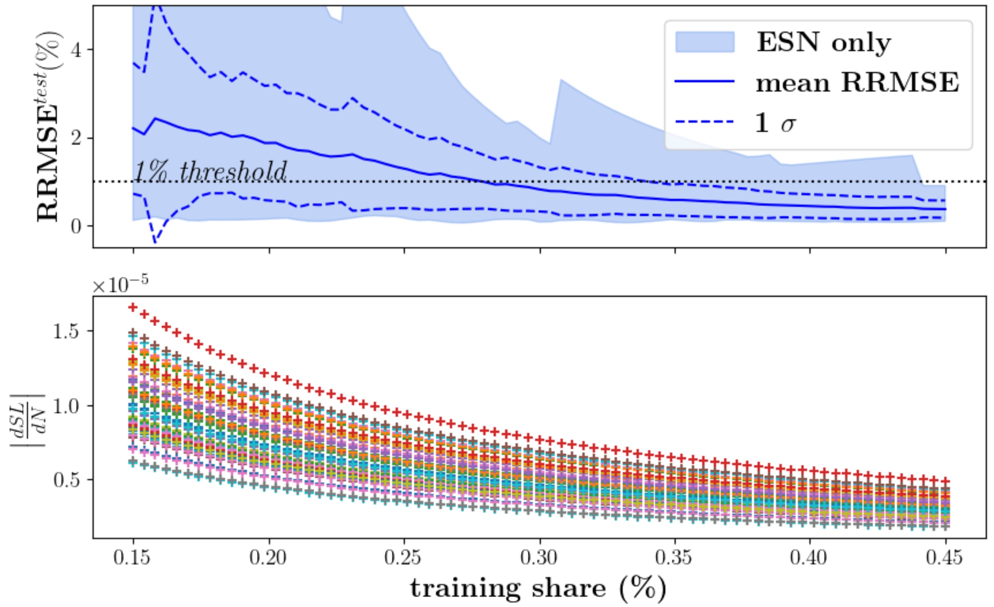
- For the selected H, compute the ESN output for the test data, for the 100 (W_{in}, W) pairs
- Compute the **mean of the ESN output over the 100** (W_{in}, W) pairs for each seed



Data splitting and hyper parameters search

Can we automatize and speed up as much as possible the determination of the data splitting and of the hyper-parameters ?

Data splitting



$$\frac{dDA}{dN}(N) = -\frac{bk}{N}\mathcal{B}^{-k}\ln\left(\frac{N}{N_0}\right)^{-k-1}$$

Bayesian search of the optimum hyper-parameters of the Deep ESN.

Python hyperopt library results:

Set #	L	N_r	BI	dt	β	ρ	Time(s)	$RRMSE^{test}_{max}$	$RRMSE^{test}_{mean}$
1	2	20	2	0.171	1.387	1.293	14801.7	2.261	0.925
2	2	15	2	0.138	0.379	0.537	18039.4	1.518	0.499
3	1	30	2	0.111	0.618	0.298	7557.7	1.901	0.552
4	1	20	2	0.025	1.877	1.259	6249.9	1.336	0.449
5	1	10	2	0.432	0.561	0.181	10438.5	2.674	1.112

Table: Set of optimized hyperparameters *via* hyperopt

L	N_r	BI	dt	β	ρ	$RRMSE^{test}_{max}$	$RRMSE^{test}_{mean}$
1	20	0	0.009	0.0224	0.19	0.892	0.365

Table: "by-hand" optimized set of parameters

Conclusion and Outlook

Long term DA can be predicted by Echo State Network (ESN)

- Combination of the scaling law and the ESN is the best approach.
- Tracking performed after 5e4 turns can be avoided by replacing it with the predictions of the ESN, gaining a factor 20 in CPU time.
- The partition of available data into training, validation, and test data sets could be obtained using an appropriate algorithm?
- Investigate the possibility of using ESN to improve the modelling of beam lifetime and luminosity evolution.
- The predictive power of ESN could be applied to other indicators of chaos ?

Perspectives

Fully automatize the optimum hyper parameters and the data splitting (training/validation/test) determination:

- Exploiting derivatives of the scaling law
- Bayesian search of hyper parameters

SOME REFERENCES

- E. Todesco and M. Giovannozzi, "Dynamic aperture estimates and phase-space distortions in nonlinear betatron motion" Phys. Rev. E53, 4067, 1996 https://link.aps.org/doi/10.1103/PhysRevE.53.4067
- D. P. Anderson, "BOINC: a System for Public-Resource Computing and Storage", Proc. of the 5thIEEE/ACM International Workshop on Grid Computing, Pittsburgh, USA,2004. J. Barranco et al., "LHC@Home: a BOINC-based volunteer computing infrastructure for physics studies at CERN",OpenEngineering7378, 2017
- A. Bazzani, M. Giovannozzi et al., "Advances on the modeling of the time evolution of dynamic aperture of hadron circular accelerators" Phys. Rev. ST Accel. Beams, 22, 104003, (2019) https://link.aps.org/doi/10.1103/PhysRevAccelBeams.22.104003
- M. Giovannozzi et al., "Machine Learning Applied to the Analysis of Nonlinear Beam Dynamics Simulations for the CERN Large Hadron Collider and Its Luminosity Upgrade", Information 12, 53, 2021 https://www.mdpi.com/2078-2489/12/2/53
- **[H. Jaeger 2001]** H. Jaeger, "The echo state approach to analysing and training recurrent neural networks", GMD-Report 148, German National Research Institute for Computer Science, 2001
- M. Lukosevicius and H. Jaeger, Reservoir computing approaches to recurrent neural network training. Computer Science Review, 3, 127--149, 2009
- J. Pathak et al., "Using Machine learning to replicate chaotic attractors, and calculate lyapunov exponent from data", chaos27, 121102 (2017)
- **[L. Grigoryeva and J.P. Ortega 2018]** L. Grigoryeva and J. P. Ortega. «Echo state networks are universal». In: Neural Networks 108 (2018)
- A.Hart and J. Hook and J. Dawes, Embedding and approximation theorems for echo state networks, Neural Networks 128, 234--247, 2020

THANK YOU