



HAL
open science

Curriculum learning for data-driven modeling of dynamical systems

Onofrio Semeraro, Michele Alessandro Bucci, Alexandre Allauzen, Sergio Chibbaro, Lionel Mathelin

► **To cite this version:**

Onofrio Semeraro, Michele Alessandro Bucci, Alexandre Allauzen, Sergio Chibbaro, Lionel Mathelin. Curriculum learning for data-driven modeling of dynamical systems. Dynamics Days Europe 2023, Sep 2023, Napoli, Italy. hal-04407784

HAL Id: hal-04407784

<https://hal.science/hal-04407784>

Submitted on 21 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Contributed ID number: ... (to be assigned by DynamicsDaysEurope2023)

ORAL

Curriculum learning for data-driven modeling of dynamical systems

The reliable prediction of the temporal behaviour of complex systems is key in numerous scientific fields. This strong interest is however hindered by modelling issues: often, the governing equations describing the physics of the system under consideration are not accessible or, when known, their solution might require a computational time incompatible with the prediction time constraints. The blooming of machine learning tools, boosted by large availability of data, computational time and new methodologies such as deep learning, lead to a renewed interest in approximating complex systems in a generic functional format and informing it from available observations, as illustrated by the numerous successful examples of the last years. However, generalizability of the models and margins of guarantee are examined mainly by relying on prior knowledge of the physics. Also, the importance of the chosen data is often overlooked. These issues can be considered from a different viewpoint by adopting a curriculum learning strategy. This concept has been developed and successfully applied in robotics and control of systems; curriculum learning consists of structuring the training dataset starting from simple samples towards more complex ones in order to favour convergence and generalization: here, we apply this idea for the learning of complex dynamical systems in a systematic way, and illustrate it using the Lorenz system. First, leveraging insights from the ergodic theory, we assess the amount of data sufficient for a-priori guaranteeing a faithful model of the physical system based on Long-Short-Term Neural Networks, and identifying the conditions for which the available amount of data allows generalization or lead to over-fitting. Second, keeping the same architecture and without any additional constraint, we thoroughly investigate the impact of the training set and its structure on the quality of long-term predictions, by considering entropy as a metric of complexity of the dataset; we show how an informed design of the training set based on the analysis of the entropy significantly improves the resulting models in terms of generalizability, and provide insights on the amount and the choice of data. Finally, we will focus on good-practices inspired by this framework for an effective data-driven modelling.

Onofrio Semeraro

LISN-CNRS, Université Paris-Saclay, 91440, Orsay (FR)
onofrio.semeraro@universite-paris-saclay.fr

Alessandro Bucci

TAU-Team, INRIA Saclay, LISN, Université Paris-Saclay, 91190,
Gif-sur-Yvette (FR)
bucci.malessandro@gmail.com

Alexandre Allauzen

LAMSADE, Université Paris Dauphine, 75016, Paris (FR)

alexandre.allauzen@dauphine.psl.eu

Sergio Chibbaro
Université Paris-Saclay, LISN-CNRS, 91440, Orsay (FR)
sergio.chibbaro@universite-paris-saclay.fr

Lionel Mathelin
LISN-CNRS, Université Paris-Saclay, 91440, Orsay (FR)
lionel.mathelin@lisn.upsaclay.fr

Topic 1: **Machine and Deep Learning**

Topic 2: **Chaos**

SESSION: _____