AI applications for the generation of a pan-European ecosystem reanalysis relying on the eLTER RI

Alexandre Belleflamme[1,3*], Daniel Caviedes-Voullieme[2,3], Christian Poppe Teràn[1,3], Bibi S. Naz[1,3], Harrie-Jan Hendricks Franssen[1,3], Stefan Kesselheim[4], Harry Vereecken[1,3]

(1) Institute of Bio- and Geosciences (Agrosphere, IBG-3), Research Centre Jülich, Germany; (2) Simulation and Data Laboratory Terrestrial Systems, Research Centre Jülich, Germany; (3) Centre for High-Performance Scientific Computing in Terrestrial Systems, Geoverbund ABC/J, Germany; (4) Jülich Supercomputing Centre, Research Centre Jülich, Germany

*E-mail: *a.belleflamme@fz-juelich.de ; web: www.fz-juelich.de/ibg/ibg-3/EN ; ORCID: 0000-0002-1664-3479*

CESOC Workshop on Large-Scale Deep Learning for the Earth System – 29-30.08.2024 – Bonn

What does the eLTER RI do?

Climate change, extreme weather events, and human activities increase pressure on ecosystems.

Sustainable management of ecosystems is important for preserving them and their biodiversity, but also for securing resources like food, feed, fibre, wood, and clean water.

eLTER RI will provide **services** based on a Europe-wide network of **ecological stations**.

➔ Analyse, monitor, predict state and evolution of ecosystems and derive actionable knowledge.

➔ We will build an ecosystem reanalysis, mainly based on eLTER observation data, to provide gap-free time series at pan-European scale on key variables of ecosystem functioning and services.

What is an ecosystem reanalysis?

Analogous to atmospheric reanalyses, but for ecosystem states and functions focussing on

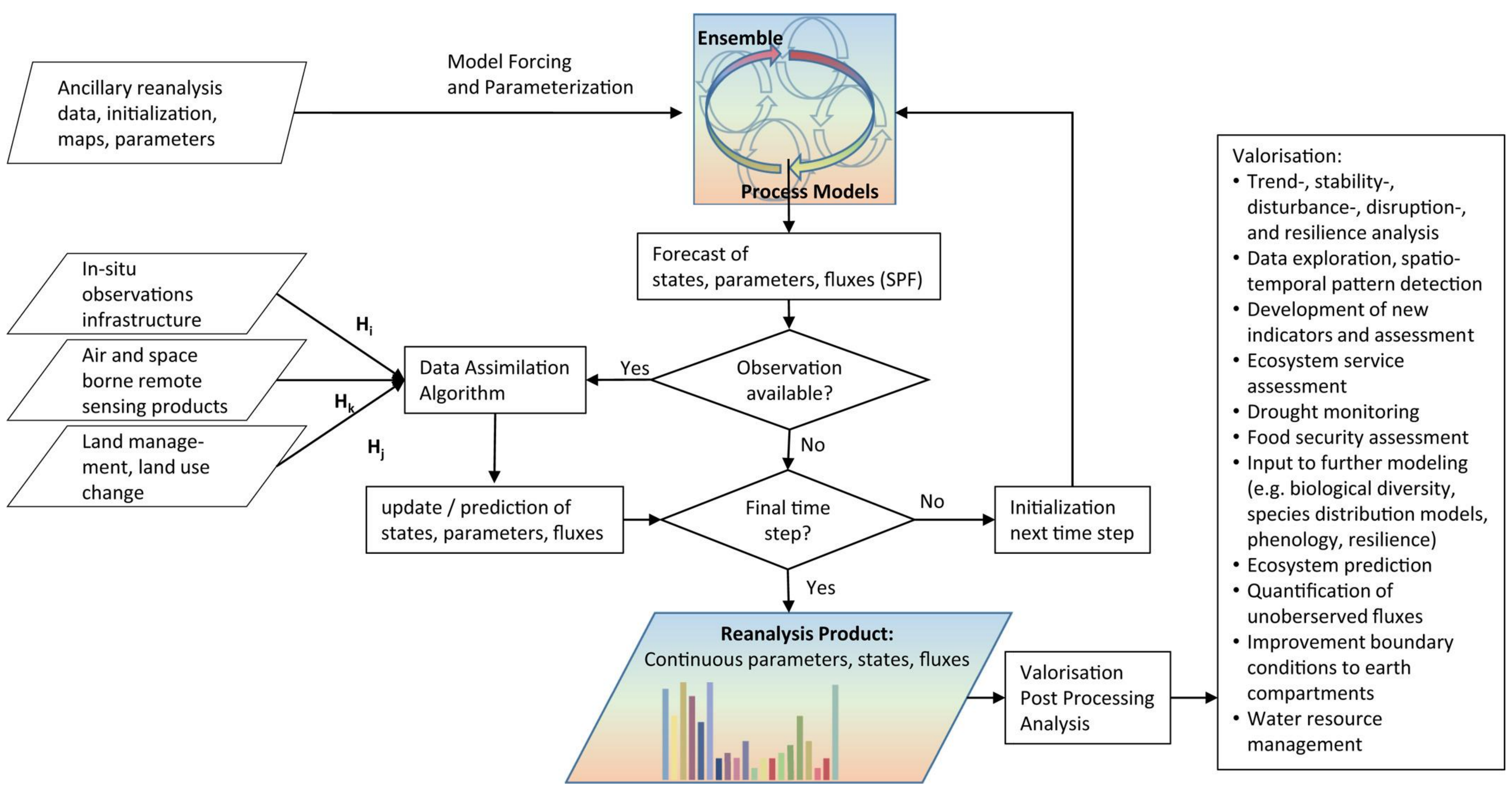
➔ **Surface and subsurface water cycle, biogeochemical variables, plant-water-soil interactions, energy and matter fluxes at the land surface, vegetation states and distribution, etc.**

At **high spatial resolution** (e.g. 3km) to account for the spatial heterogeneity of the land surface.

With the **land surface model** CLM (Community Land Model) at pan-European scale.

Model-data fusion with eLTER observations, remote sensing products, atmospheric reanalysis, etc.

Several computational and methodological obstacles
➔ **AI-based applications**



Flow chart on creation and valorization of an Earth system reanalysis product, model-data-fusion scheme, uncertainty propagation, and time-step iterative data assimilation. From Baatz et al. (2021)[1]

AI 1 – Model-data fusion and parameter estimation

Challenges

Classical data assimilation requires the computing of a huge amount (>100) of ensemble members in parallel.

The same yields for the estimation of the parameterization uncertainty.

AI application

We plan to use a non-parametric machine learning model (e.g., Gaussian process regression or eXtreme Gradient Boosting) in combination with data assimilation (Cleary et al., 2021)[2].

This allows us to minimise the model error of CLM due to parameterization uncertainty, and it avoids systematic errors due to the integration of less precise, globally defined standard parameter values.

➔ **Reduce computing resources needs**
➔ **Reduce parameterization errors**
➔ **Increase ensemble size**

AI 2 – Spin-up and initial conditions

Challenges

High-resolution initial conditions are needed.

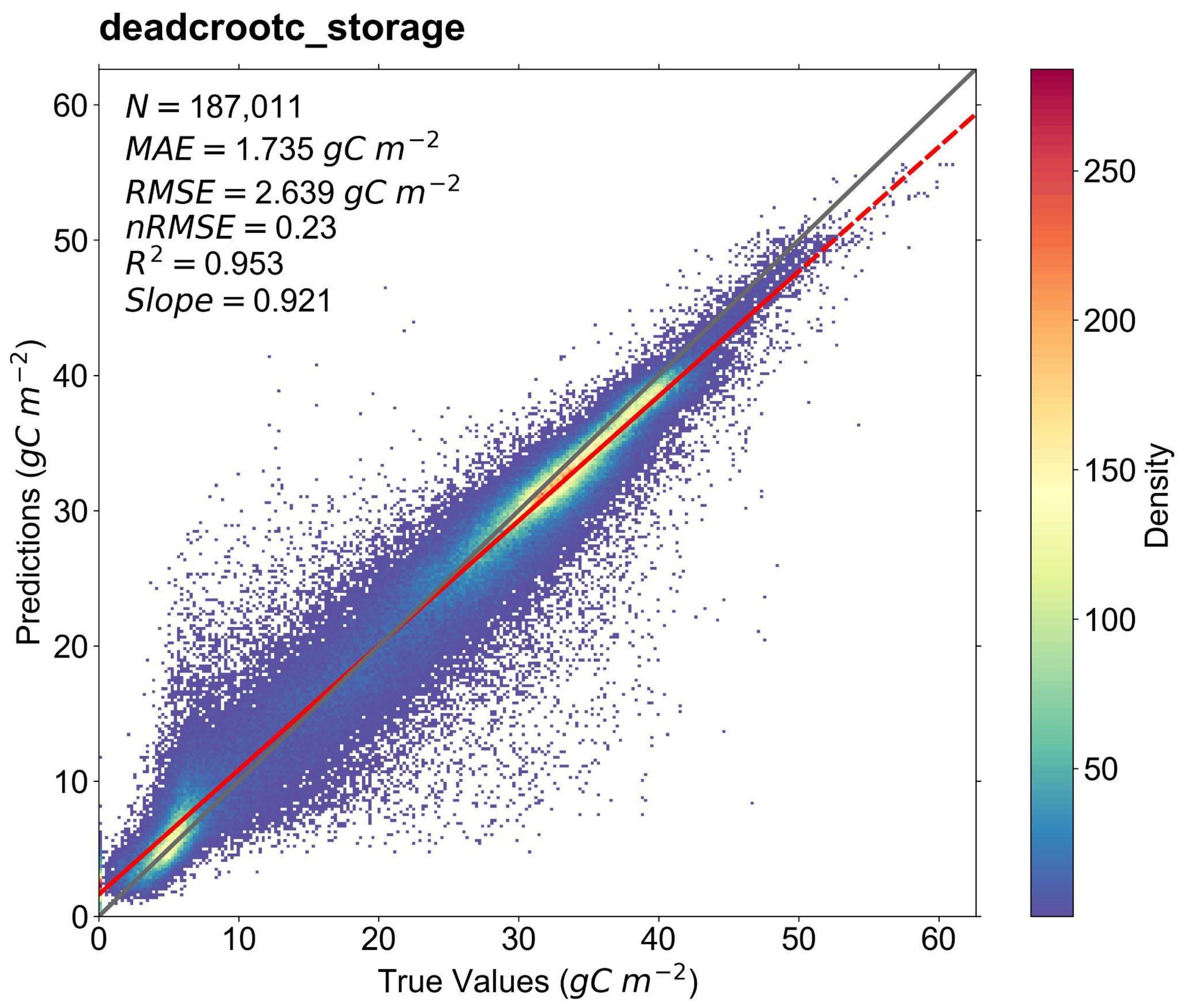
The slow carbon pool needs a spin-up of several centuries to reach equilibrium.

AI application

We plan to build a surrogate model for CLM based on a 3D sliding window (SWIN) Transformer (Bi et al., 2022)[3].

This allows us to determine the initial conditions for the carbon storage and cycle. It can even account for different climate change projections and scenarios.

➔ **Reduce computing resources needs**
➔ **Increase number of scenarios**



Approximation of initial states (here: dead root carbon storage across Europe) with ML. Chen et al., in preparation.

AI 3 – Downscaling and spatial interpolation

Challenges

Mechanistic models have practical limitations to further increase their spatial resolution,

while the land surface is characterised by a very high, small scale heterogeneity.

AI application

We will use a Convolutional Neural Network, which learns the relationships between land surface features, their heterogeneity, and the target variables (e.g., C and N states and fluxes).

In addition, we will use a diffusion model able to downscale these variables at field scale. This could be complemented by a physical Loss for enhanced precision guidance (Lu & Xu, 2024)[4].

This will allow us to interpolate between scales and eventually to reach relevant scales for decision support (< 1km).

➔ **Downscale beyond mechanistic models capabilities**
➔ **Account for high spatial heterogeneity**

References

[1] Baatz, R., Hendricks Franssen, H. J.,Euskirchen, E., Sihi, D., Dietze, M.,Ciavatta, S., et al. (2021). Reanalysis in Earth system science: Toward terrestrial ecosystem reanalysis. Reviews of Geophysics, 59, e2020RG000715, doi: 10.1029/2020RG000715

[2] Cleary, E., Garbuno-Inigo, A., Lan, S., Schneider, T., Stuart, A.M. (2021). Calibrate, emulate, sample. Journal of Computational Physics, 424, 109716, doi: 10.1016/j.jcp.2020.109716

[3] Bi, C., Hu, N., Zou, Y., Zhang, S., Xu, S., Yu, H. (2022). Development of deep learning methodology for maize seed variety recognition based on improved Swin Transformer. Agronomy, 12, 8, doi: 10.3390/agronomy12081843

[4] Lu, M., Xu, X. (2024). TRNN: an efficient time-series recurrent neural network for stock price prediction. Information Sciences, 657, 119951, doi: 10.1016/j.ins.2023.119951