Emulating the land surface with aiLand

Nina Raoult*, Ewan Pinnington, Gabriele Arduini, Souhail Boussetta, Matthew Chantry, Christoph Rüdiger

European Centre for Medium-Range Weather Forecasting, Reading, United Kingdom and Bonn, Germany; (*) nina.raoult@ecmwf.int

Motivation

To improve weather forecasts and climate projections, the Destination Earth programme is developing a **data-driven Earth system model**. A critical component of this effort is the **land surface**, which governs energy, water, and carbon exchanges with the atmosphere, and influences both short-term weather and long-term climate impacts. Here we present aiLand, the stand-alone emulator of ecLand, ECMWF's land surface scheme.

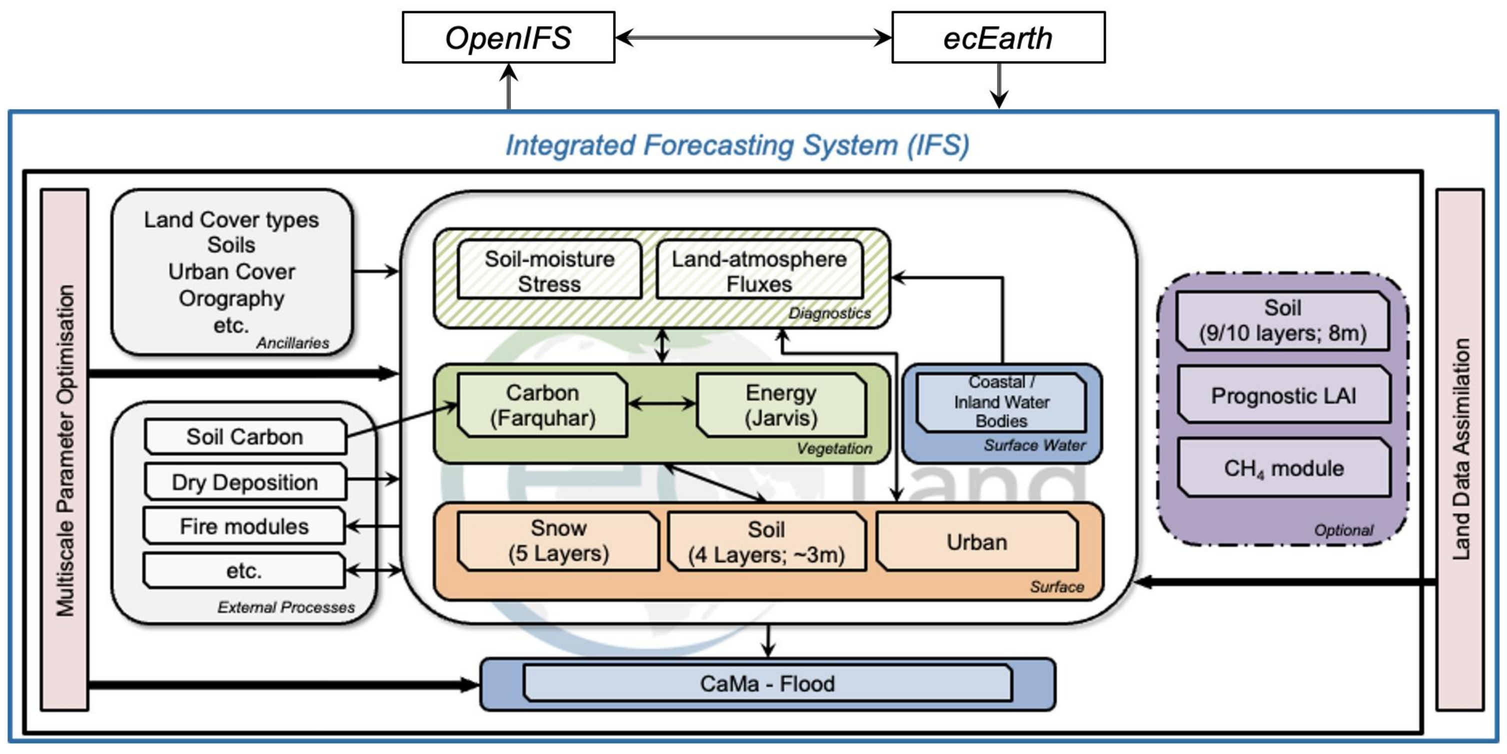


Fig: Schematic of ecLand, ECMWF's land surface scheme part of the Integrated Forecasting System

Architecture

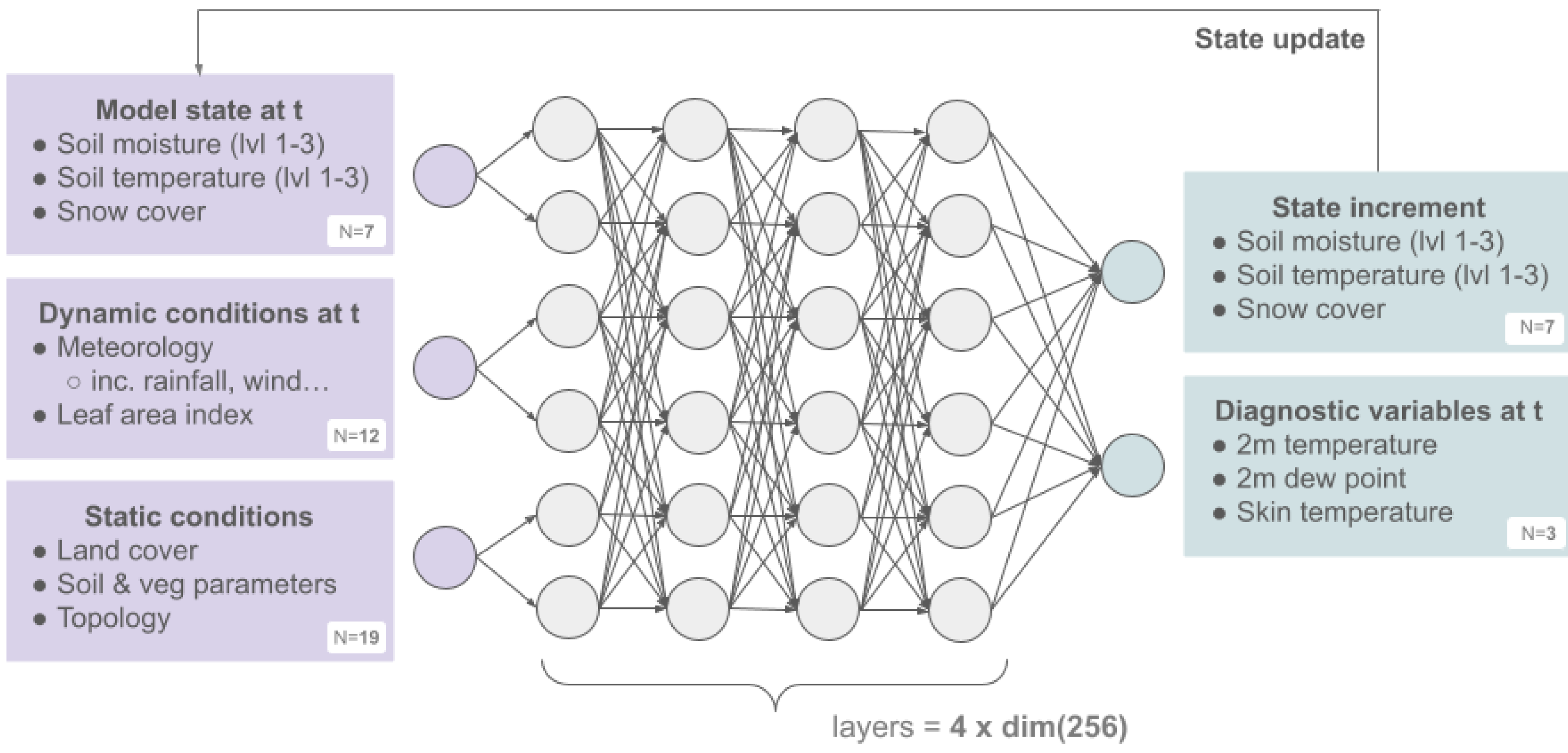
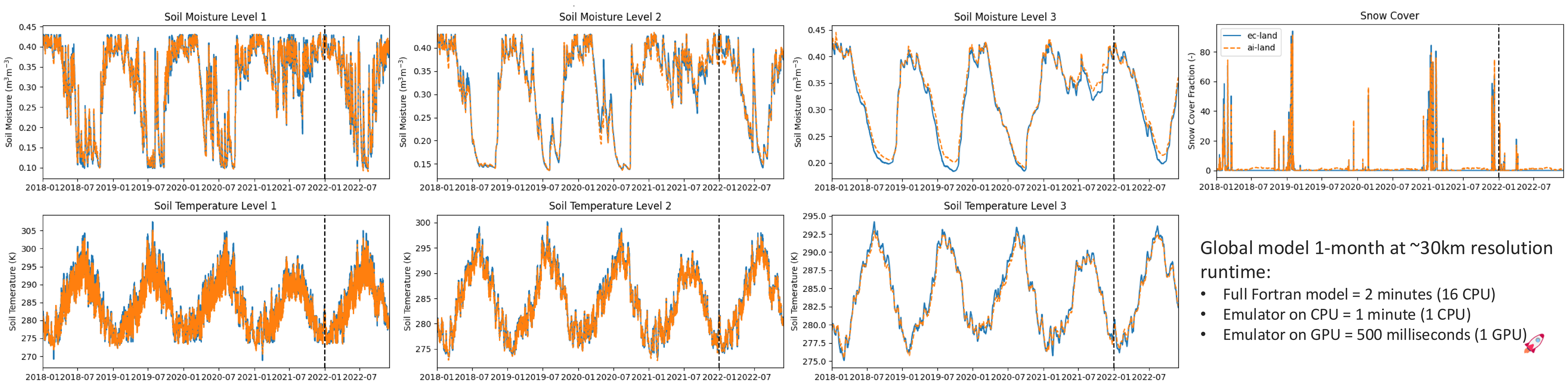


Fig: Multi-layered perceptron architecture used in aiLand

We tested various machine learning methods (MLP, LSTM, XGBoost, Graph NN). We chose the MLP because it offers a good balance between model complexity and performance, is easily differentiable for integration into data assimilation and parameter estimation systems, and exploits the independent, column-based structure of ecLand.

Overall emulator performance



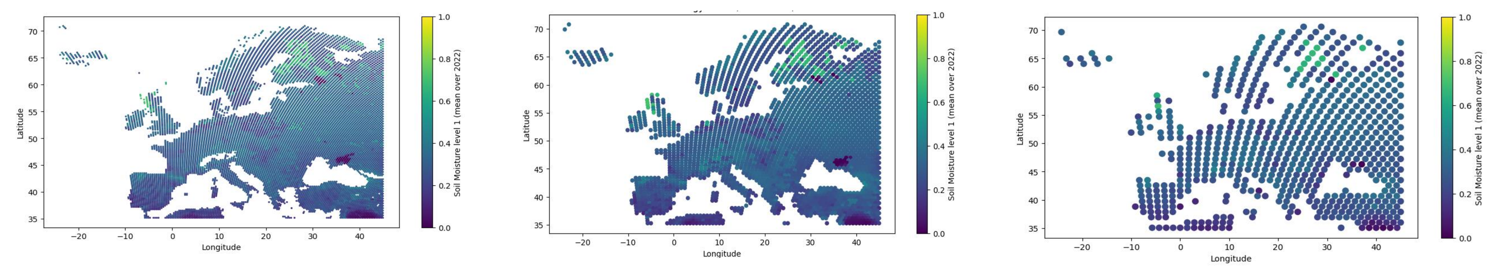
Global model 1-month at ~30km resolution runtime:

- Full Fortran model = 2 minutes (16 CPU)
- Emulator on CPU = 1 minute (1 CPU)
- Emulator on GPU = 500 milliseconds (1 GPU)

Fig: ecLand vs aiLand (48.01 N, 5.18 E)

Model transferability

As a point model, aiLand is **spatial resolution agnostic;**



Trained at ~25 km **Evaluated** over ~50 km **Evaluated** over ~100 km

This also means aiLand can be **evaluated spatially** on domains not included in the training.

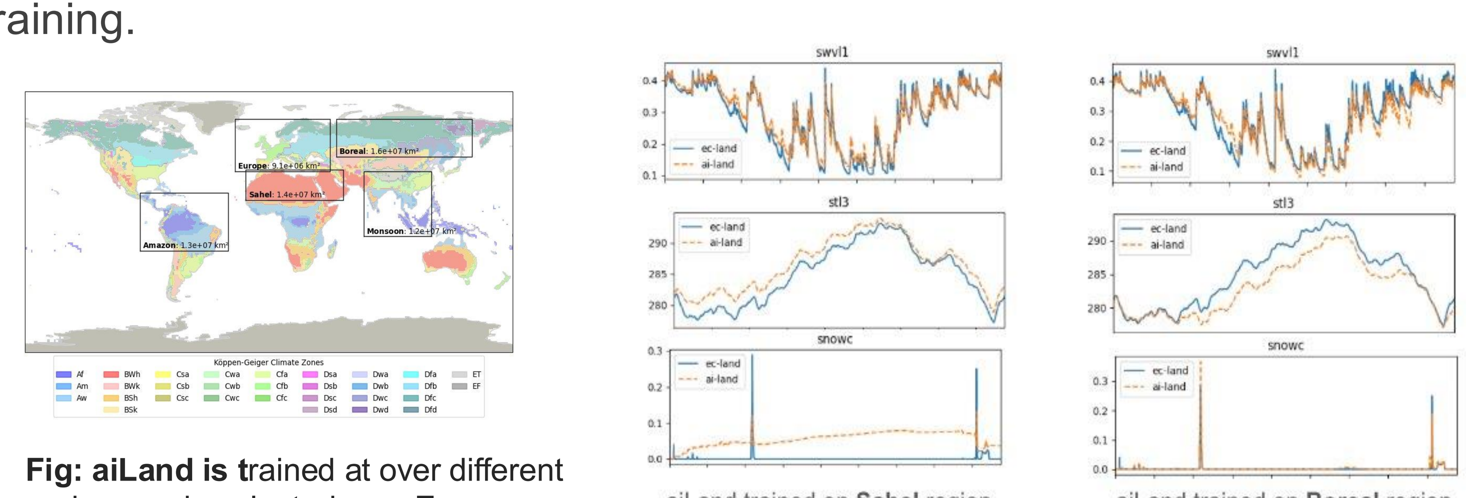


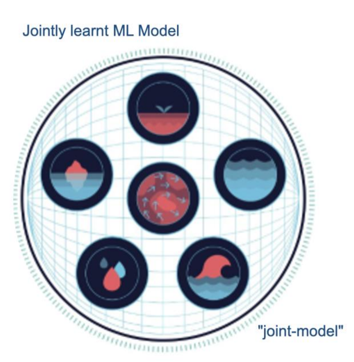
Fig: aiLand is trained at over different regions and evaluated over Europe.

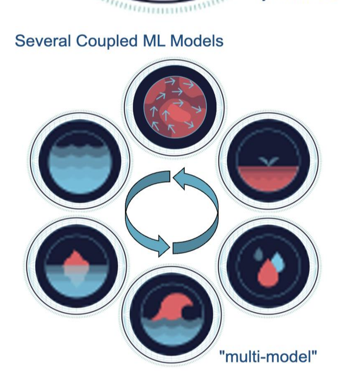
aiLand trained on **Sahel** region aiLand trained on **Boreal** region

Coupling to the other components

ML model coupling: as part of a broader effort to evaluate coupling strategies at ECMWF, we are comparing:

- **Joint Model:** All components trained together. Land variables like soil moisture and soil temperature are already included in AIFS; snow depth and cover are currently being added.
- **Multi-Model:** Components trained separately. We are exploring direct coupling of **aiLand** to a version of the AIFS with surface variables.



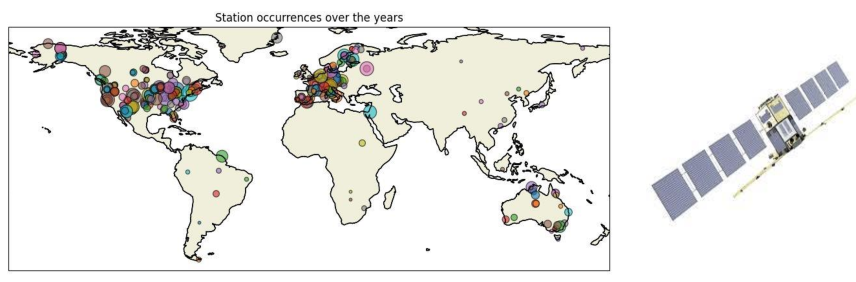


Physical model coupling: we are exploring how to interface **aiLand** with the IFS as part of ongoing work with the new **GT4Py** dynamical core—a modern, Python-based framework optimized for GPU computing.

Improving the physical model

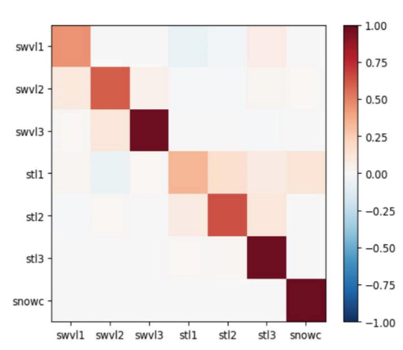
Developing an ML model is not done in isolation—an important question is how it can provide feedback to improve the physical model. For example,

➢ By fine tuning against **observations**, can we **learn biases** in the physical model?



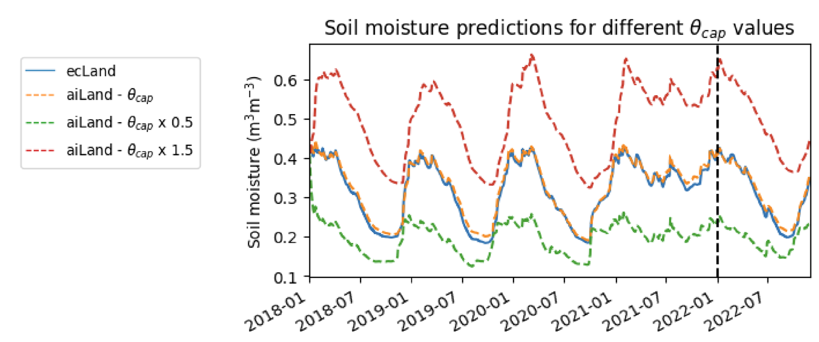
We aim to leverage both in situ and satellite observations.
- For example, can latent and sensible heat, learned as diagnostic variables, be fine-tuned using FLUXNET data?
- What is the impact on predicted soil moisture when Level 1 brightness temperatures are included in the training?

➢ Can we exploit the **differentiability** of aiLand for land model **data assimilation**?



aiLand's Jacobian shows plausible sensitivities between different variables
- Can these sensitivities enhance the accuracy of state updates during data assimilation?
- How do these sensitivities affect the propagation of uncertainty through the system?

➢ With more information on parameter sensitivities, can we use the emulator for **parameter estimation**?



aiLand already captures the expected soil moisture response to changes in field capacity.
- If we train aiLand using a large parameter perturbation experiment, can we reduce the model's parametric uncertainty?
- Can the trained model generalise to unseen parameter combinations or extreme climate conditions?

Acknowledgements

This work is funded by Destination Earth. Further acknowledge our colleagues at ECMWF for developing and maintaining ecLand (formerly HTESSEL). This work also leverages tools from the anemoi ecosystem, specifically the **anemoi-datasets** package to format and catalogue the training data.

anemoi ecLand

Funded by the European Union **Destination Earth** implemented by ECMWF esa EUMETSAT