BRIDGING MECHANISTIC AND DATA-DRIVEN MODELS: PHYSICS-GUIDED SEQUENCE LEARNING FOR CROP GROWTH AND YIELD

Convergence Of Research in Digital Agriculture Leading Labs,
April 15th, 2026

Julian Agudelo, Vincent Guigue, Cristina Manfredotti,
Evelyne Lutton, Hadrien Piot

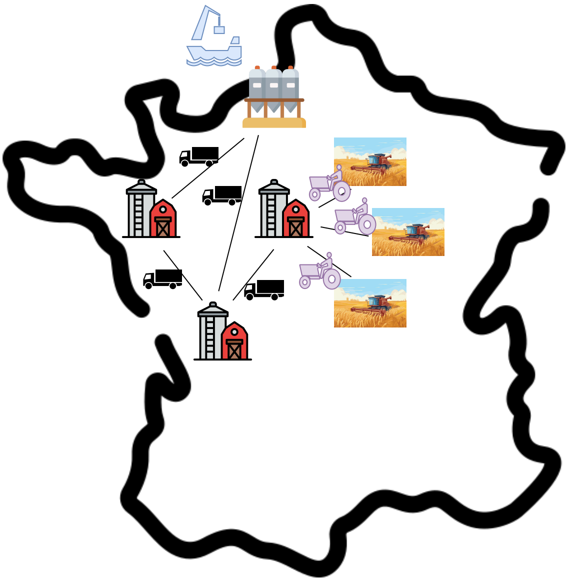
Introduction

Predicting yields...

- Applicable to **multiple species**

Predicting yields...

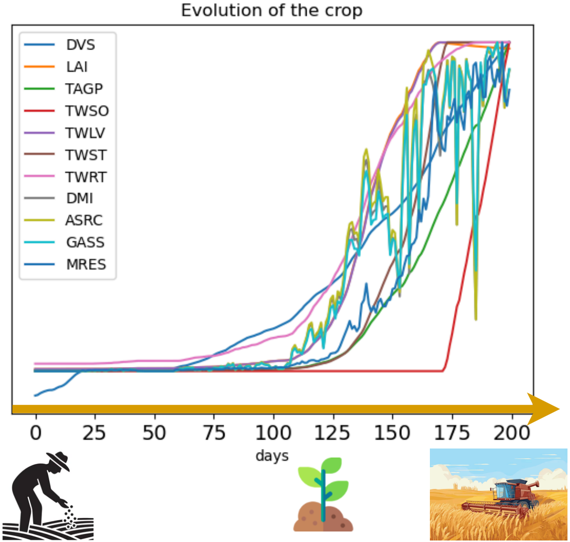
- Applicable to **multiple species**
- Enables optimization of the **logistics** chain



Predicting yields...

- Applicable to **multiple species**
- Enables optimization of the **logistics** chain
- Predicts the **entire plant life cycle** (enhancing explainability, trust, and performance)

Variable	Description
DVS	Crop development stage
LAI	Leaf area index
TAGP	Total above-ground biomass
TWSO	Storage organs biomass
TWLV	Leaves biomass
TWST	Stems biomass
TWRT	Roots biomass
DMI	Daily dry matter increase
ASRC	Net available assimilates
GASS	Assimilation rate (water stress)
MRES	Maintenance respiration rate



From Wofost to machine-learning

Mechanistic approaches propose this kind of solution...

But **high precision** = **very precise inputs**

- Weather
- Soil data
- Crop management plan
- Phylogenetic/phenological details of the seeds

⇒ We know field plots, yields...

We can retrieve weather & soil data.

WOFOST

From Wofost to machine-learning

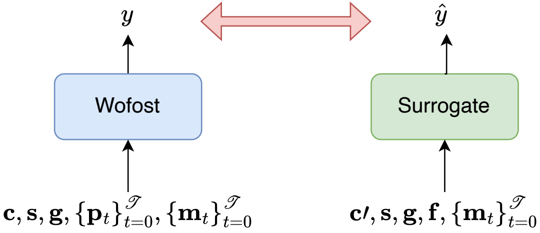
Mechanistic approaches propose this kind of solution...

But **high precision = very precise inputs**

- Weather
- Soil data
- Crop management plan
- Phylogenetic/phenological details of the seeds

⇒ We know field plots, yields...

We can retrieve weather & soil data.



$\mathbf{c}$	Crop-specific features
$\mathbf{s}$	Soil characteristics
$\mathbf{g}$	Initial conditions
$\{\mathbf{p}_t\}_{t=0}^{\mathcal{T}}$	Agronomic management practices
$\{\mathbf{m}_t\}_{t=0}^{\mathcal{T}}$	Meteorological observations
$\mathbf{f}$	Farm id ⇒ profile
$\mathbf{c}'$	Crop id ⇒ profile

From Wofost to machine-learning

Mechanistic approaches propose this kind of solution...

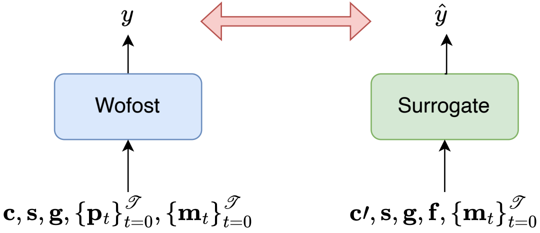
But **high precision = very precise inputs**

- Weather
- Soil data
- Crop management plan
- Phylogenetic/phenological details of the seeds

⇒ We know field plots, yields...

We can retrieve weather & soil data.

WOFOST

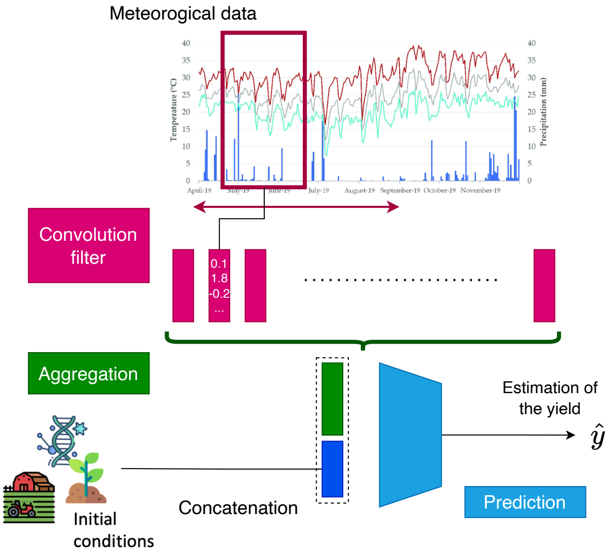


⇒ Surrogate = less input requirements, differentiable, easier to calibrate, faster

Modeling:
Building the surrogate

How to deal with heterogeneous input? (1)

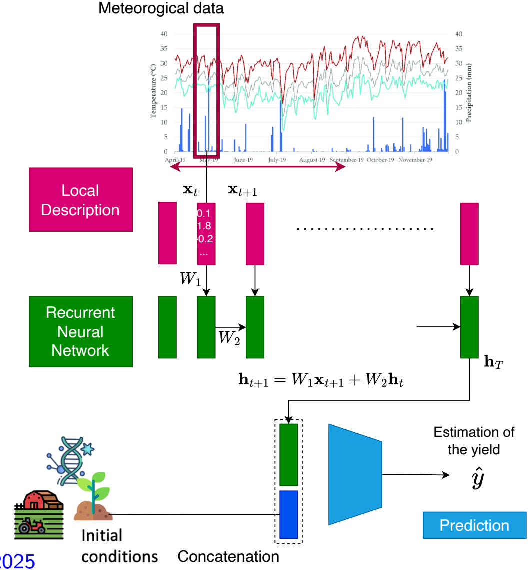
1. **Convolution filter** = describing each time step Detect pattern
2. **Average / Max** = aggregating global information
3. **Concatenation** = Mixing different modalities
4. **Linear or non linear mapping / Multilayer-Perceptron** = Prediction



Nyamane et al, 2024
Harnessing deep learning for meteorological drought forecasts in the Northern Cape, South Africa

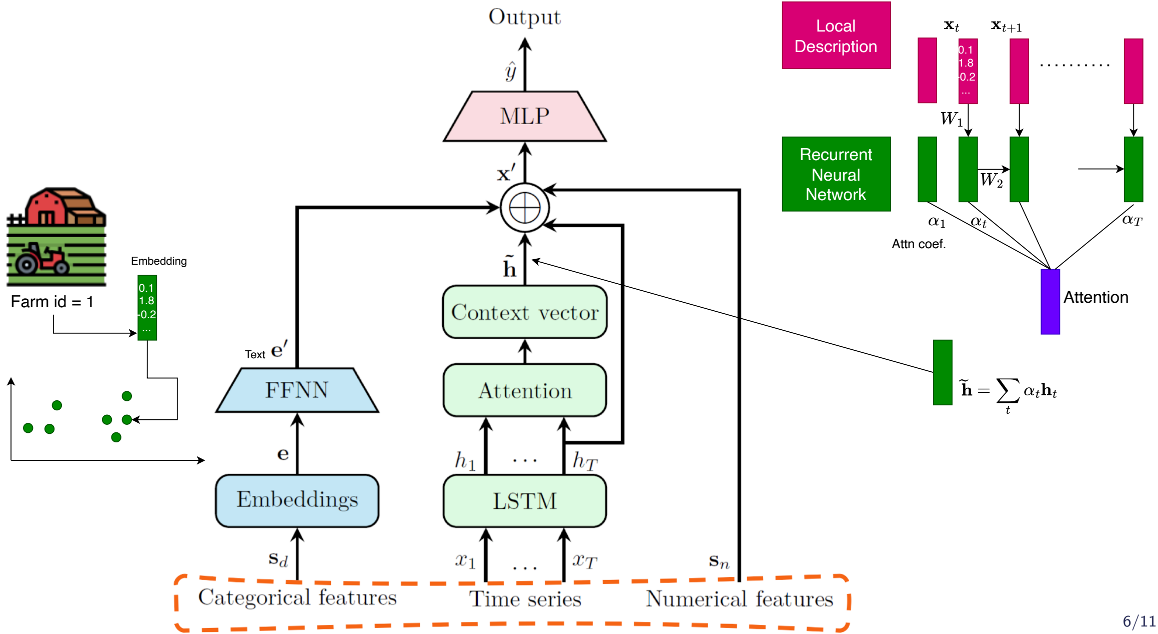
How to deal with heterogeneous input? (2)

1. **Recurrent unit** =
   Aggregate local information
   Aggregation = pattern detection
2. **Concatenation** =
   Mixing different modalities
3. **Linear or non linear mapping / Multilayer-Perceptron** =
   Prediction

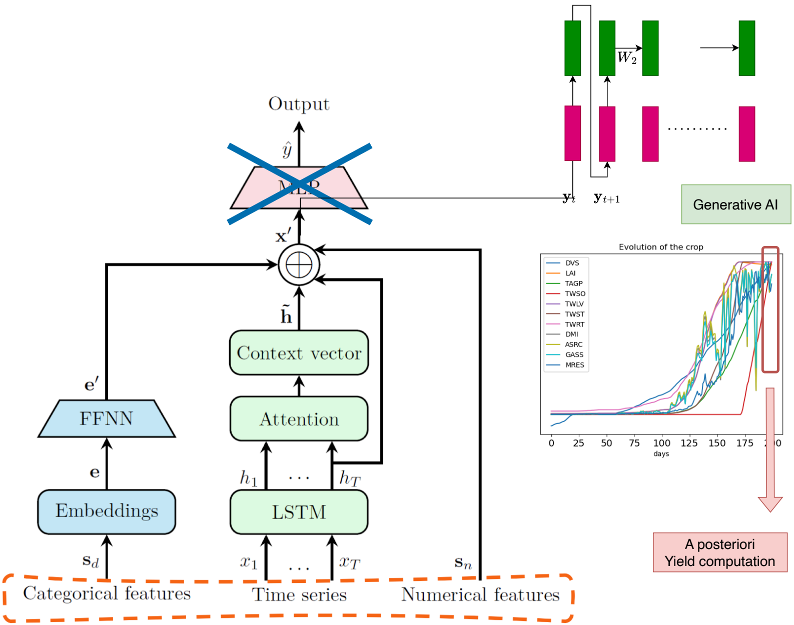


Minixhofer et al,
ICML Workshop on Tackling Climate Change with ML 2025
Droughted: A dataset and methodology for drought forecasting spanning multiple climate zones.

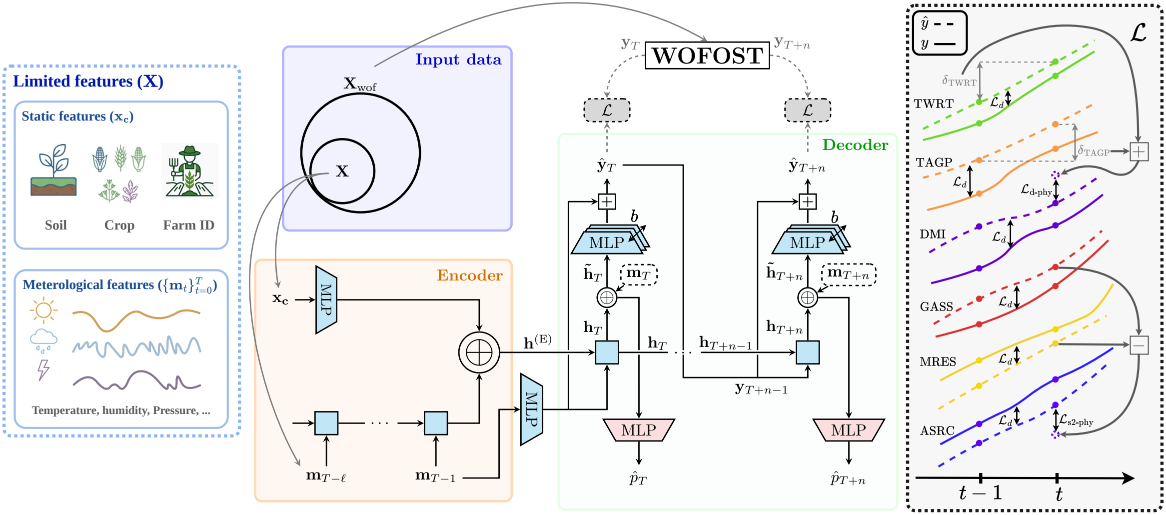
Direct yield prediction



Sequence to sequence: generating the crop growth process



Sequence to sequence: generating the crop growth process



Details of the losses

General curve fitting:

$$\mathcal{L}_{\mathrm{d}} = \mathrm{MSE}\big(\{\mathbf{y}_t\}_{t=T}^{\mathcal{T}}, \{\hat{\mathbf{y}}_t\}_{t=T}^{\mathcal{T}}\big) \tag{1}$$

(s) Internal biomass consistency:

$$\mathcal{L}_{\mathrm{s1\text{-}phy}} = \big((\hat{y}_t^{\mathsf{TWSO}} + \hat{y}_t^{\mathsf{TWLV}} + \hat{y}_t^{\mathsf{TWST}}) - \hat{y}_t^{\mathsf{TAGP}}\big)^2 \tag{2}$$

(d) Relationship : available assimilates / maintenance respiration rate:

$$\mathcal{L}_{\mathrm{s2\text{-}phy}} = \big((\hat{y}_t^{\mathsf{GASS}} - \hat{y}_t^{\mathsf{MRES}}) - \hat{y}_t^{\mathsf{ASRC}}\big)^2 \tag{3}$$

(d) Dry matter available for partitioning in day $t-1$ to be consistent with the total biomass augmentation between days $t-1$ and t:

$$\mathcal{L}_{\mathrm{d\text{-}phy}} = \big(\hat{y}_{t-1}^{\mathsf{DMI}} - (\hat{y}_t^{\mathsf{TWRT}} - \hat{y}_{t-1}^{\mathsf{TWRT}}) - (\hat{y}_t^{\mathsf{TAGP}} - \hat{y}_{t-1}^{\mathsf{TAGP}})\big)^2 \tag{4}$$

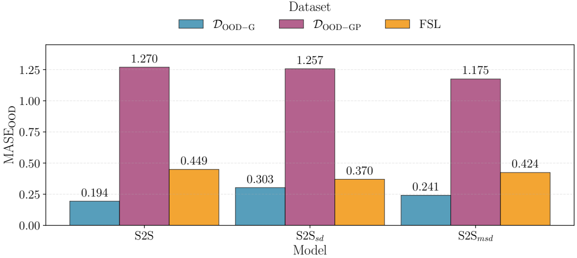
(m) + by-design monotonic constraint on several curves

Results

Comparative Results / ablation study

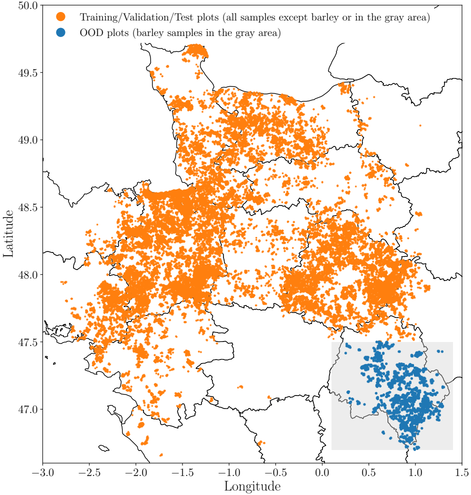
Model	$\overline{\textbf{MASE}}$	$\textbf{MAE}_{\textbf{yield}}$	$\Delta\overline{\textbf{MASE}}_{\textbf{OOD}}$	$\overline{\textbf{V}_{mon}}$	$\textbf{V}_{biom}$	$\textbf{V}_{ass}$	$\textbf{V}_{mat}$
MLP	-	2079.236	-	-	-	-	-
MLP-GRU	-	**1935.846**	-	-	-	-	-
S-S2S	0.287	1445.836	**0.380**	0.091	-	-	-
S-S2S$_m$	**0.266**	**1134.727**	0.564	**0**	-	-	-
S2S	0.143	671.954	7.881	0.182	10.254	2.252	9.738
S2S$_s$	0.161	722.312	7.633	0.141	4.519	0.660	12.830
S2S$_d$	0.140	573.741	7.264	0.192	12.756	2.890	2.060
S2S$_{sd}$	0.139	480.627	8.043	0.180	**1.356**	**0.510**	2.210
S2S$_m$	**0.131**	549.320	9.763	**0**	24.401	2.069	**1.114**
S2S$_{ms}$	0.141	495.730	8.255	**0**	134.119	0.890	1.509
S2S$_{md}$	0.212	2183.760	**5.514**	**0**	1297.219	2.581	5.470
S2S$_{msd}$	0.150	**474.619**	6.833	**0**	69.199	0.702	4.003

Out-Of-Domain Evaluation



- OOD-G: Geographic
  ⇒ OK (easy setting)
- OOD-GP: Geographic + species
  ⇒ KO
- OOD-GP + physical constraints
  ⇒ keep good performances

What happens in a new context?
Across new regions and species?



Conclusion & perpectives

- Equivalent performances to wofost + faster
- Nice OOD performances
- New opportunities:
  e.g. what was the optimal date for seeding? ⇒ gradient descent

Perspectives

- **Calibration:** how to preserve physical modeling while incorporating new ground truth information?
- **Simulation:** intensive simulation & latent space exploration ⇒ opportunity to model climate change
- **HCI / textual data** chat with the model to refine prediction & make it available for everyone
- **Logistics:** incorporate predictions into logistic simulations