Early Classification and Transfer Learning Challenges in Large-scale Crop Type Mapping

Marc Rußwurm – Assistant Professor in Machine Learning for Remote Sensing

Time series and transfer learning workshop, 19th of October, Paris

WAGENINGEN
UNIVERSITY & RESEARCH

Outline

Importance of time series for large-scale vegetation monitoring

Early Time Series Classification

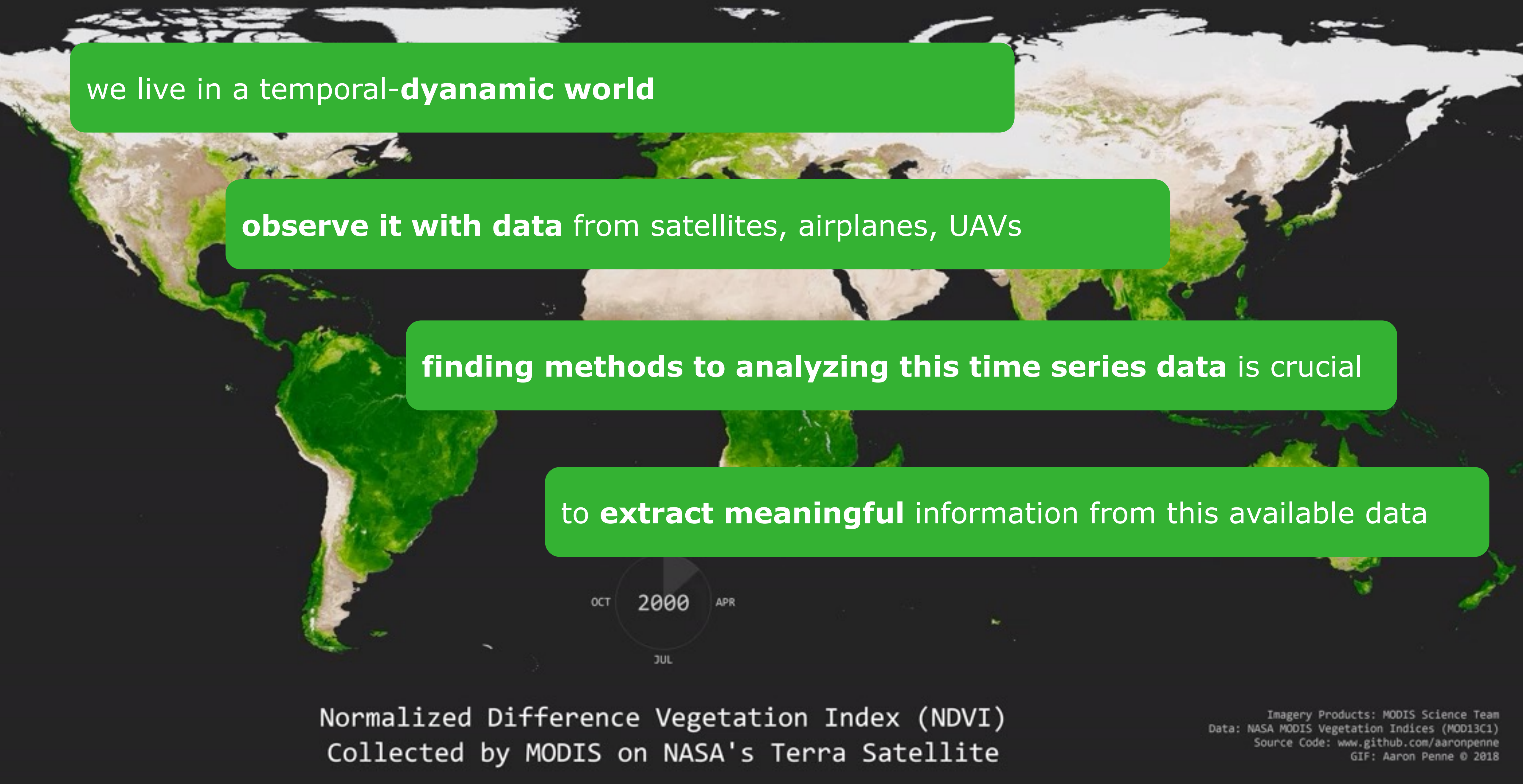
Transfer Learning and Challenges

we live in a temporal-**dyanamic world**

observe it with data from satellites, airplanes, UAVs

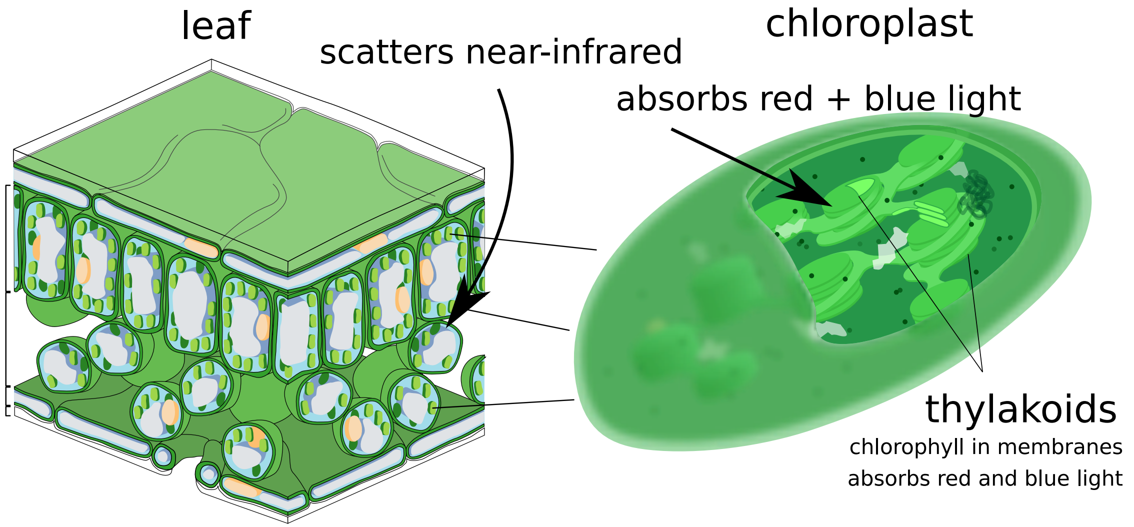
finding methods to analyzing this time series data is crucial

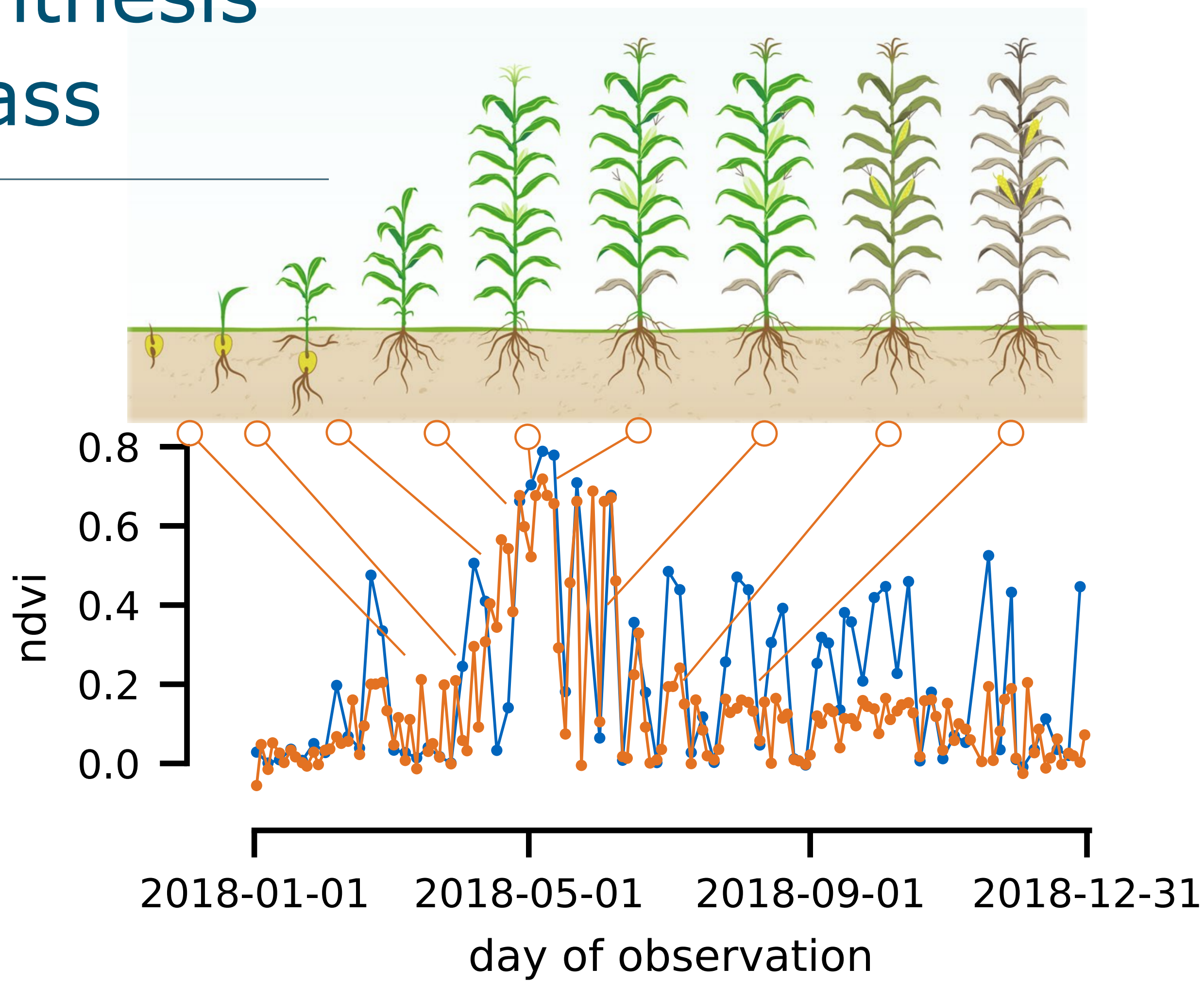
to **extract meaningful** information from this available data



Vegetation relies on photosynthesis to convert light to biomass

photosynthesis takes place in plant leafs within chloroplasts

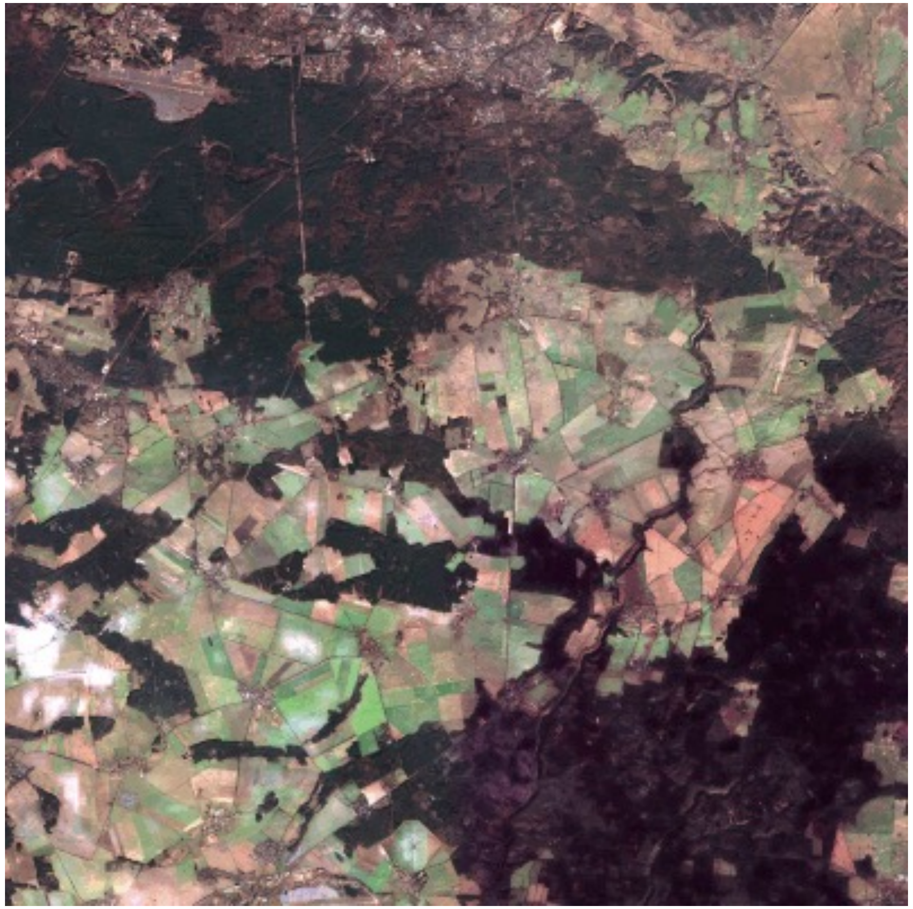


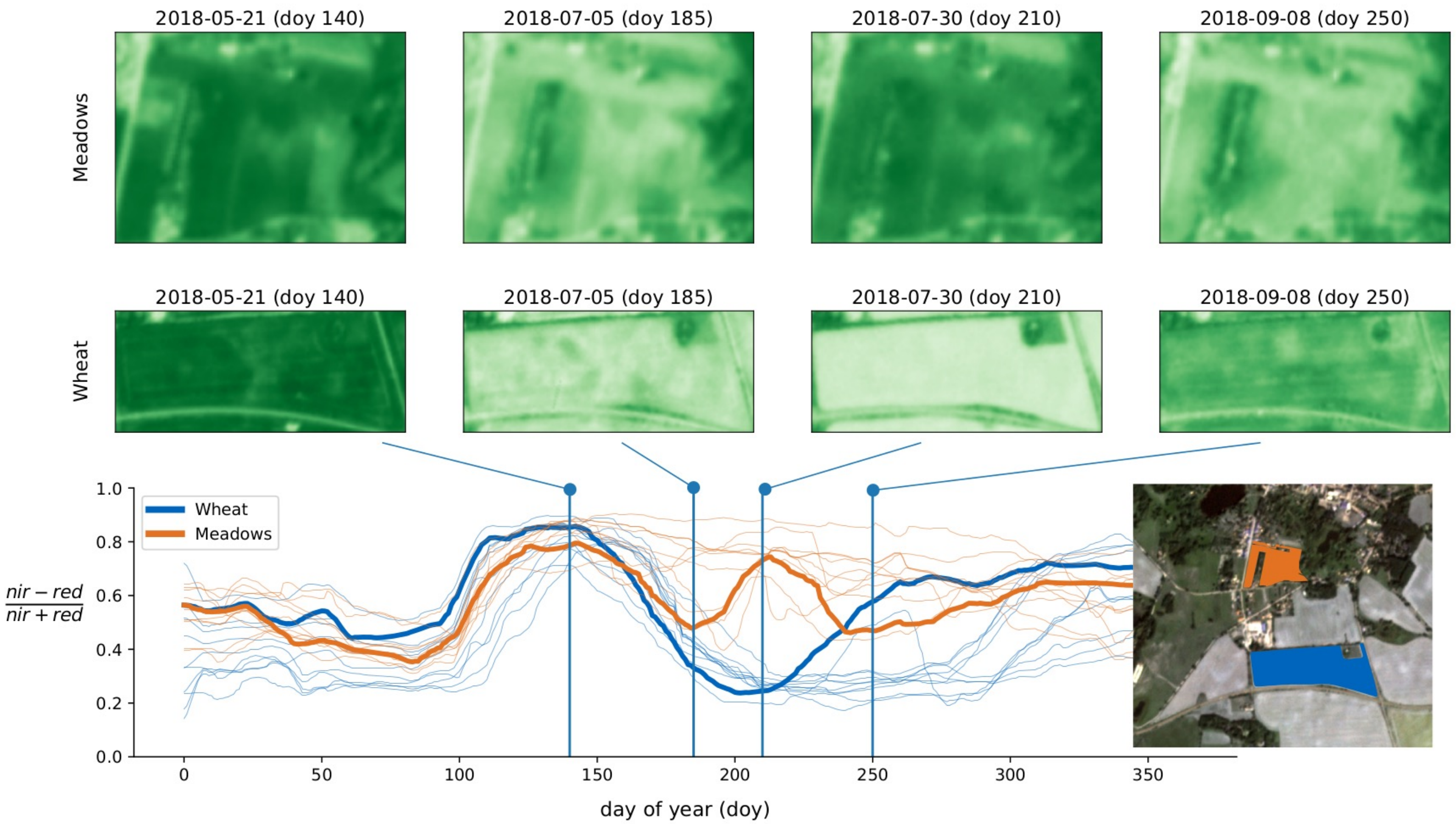


we can measure this red and near-infrared reflectance from satellites in regular time intervals

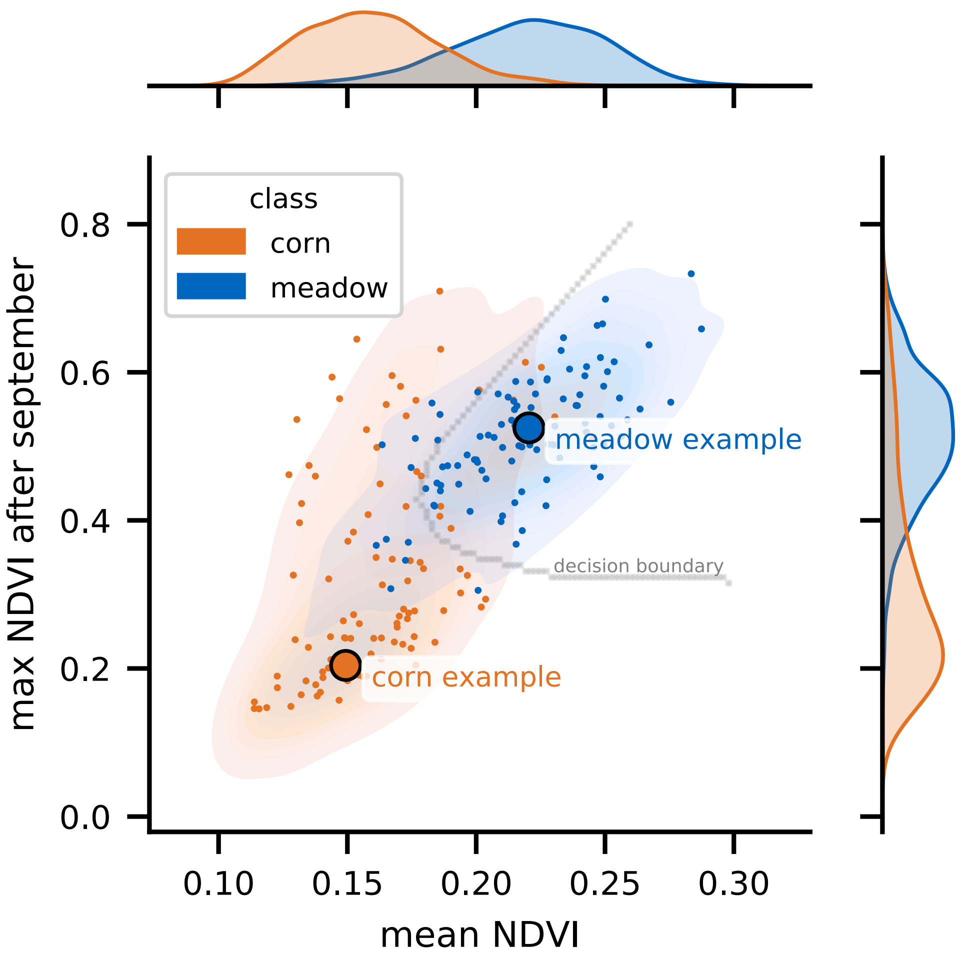
WAGENINGEN UNIVERSITY & RESEARCH

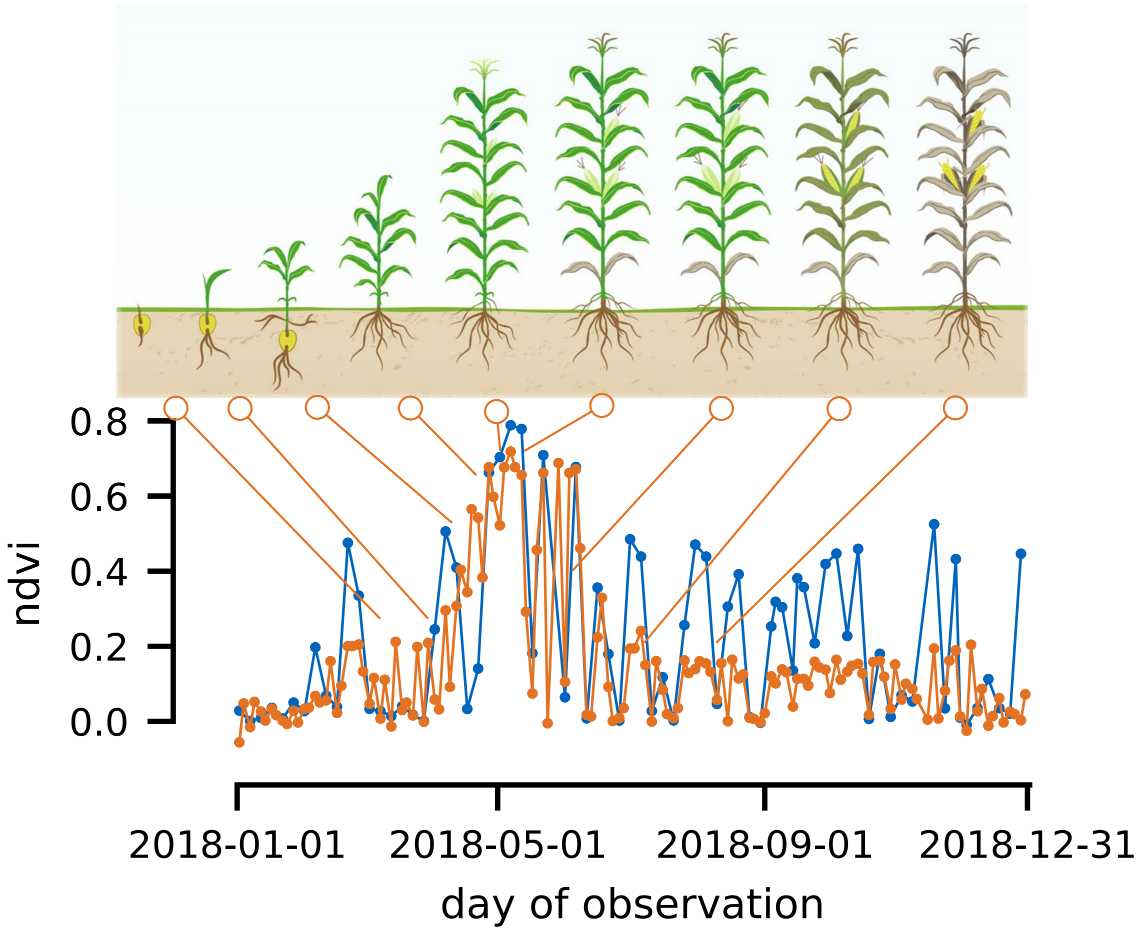
Crop Type Mapping





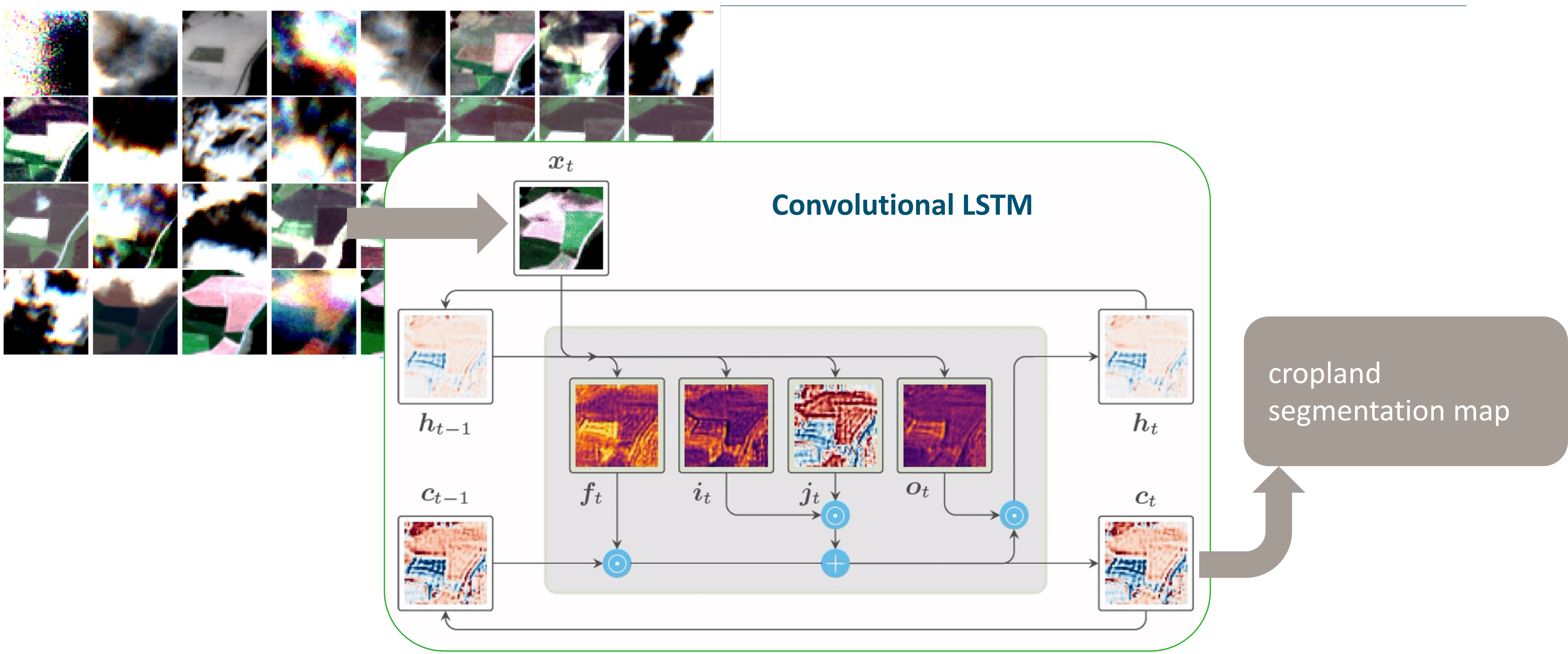
Crop Type Classification





WAGENINGEN
UNIVERSITY & RESEARCH

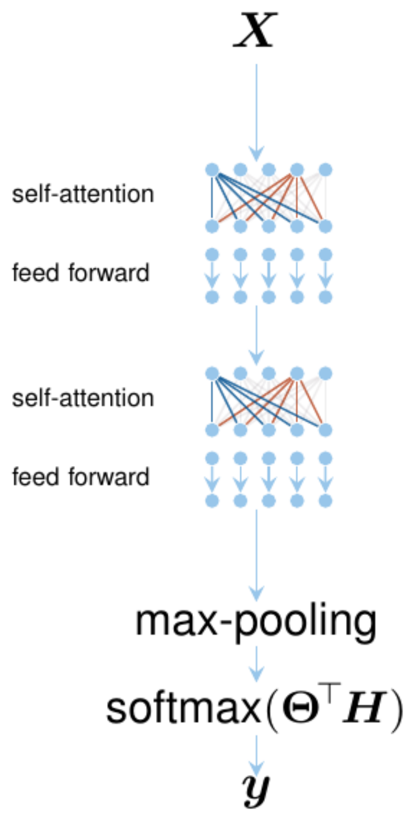
Long-Short Term Memory (LSTM) Time Series Classifiers

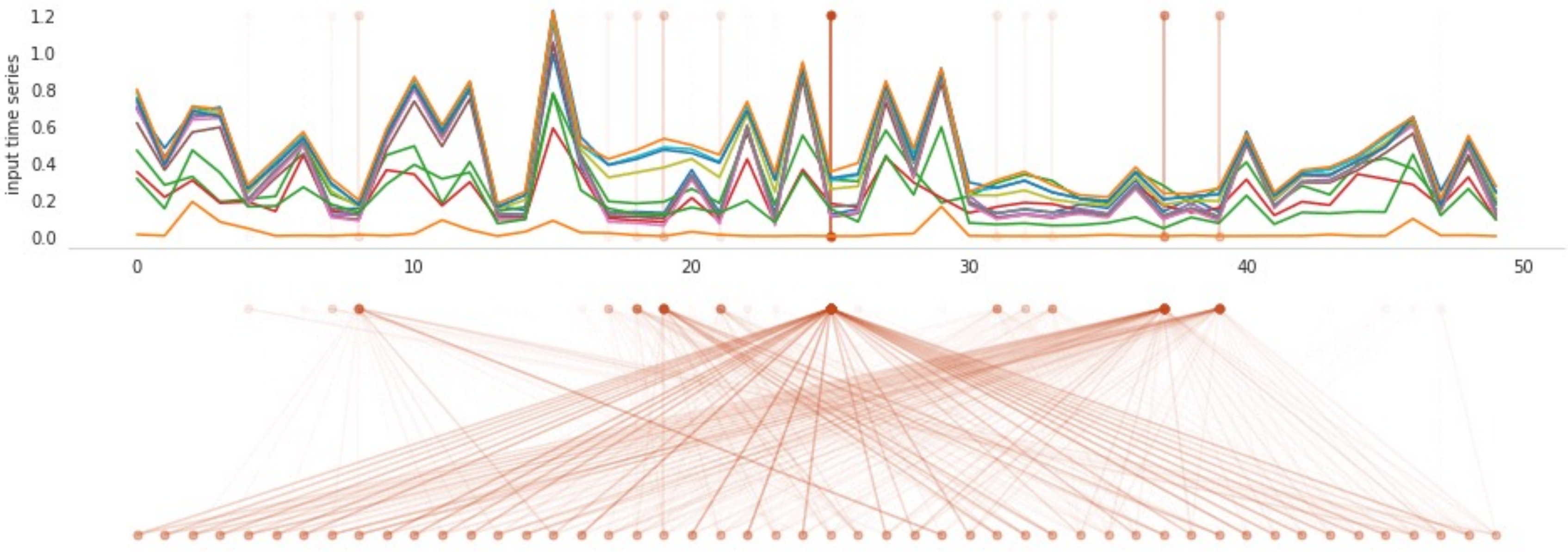


WAGENINGEN UNIVERSITY & RESEARCH

Rußwurm, M., & Körner, M. (2018). Multi-temporal land cover classification with sequential recurrent encoders. ISPRS International Journal of Geo-Information, 7(4), 129.

Self-Attention in (small) Transformer Models





WAGENINGEN UNIVERSITY & RESEARCH

Rußwurm, M., & Körner, M. (2020). Self-attention for raw optical satellite time series classification. *ISPRS journal of photogrammetry and remote sensing*, *169*, 421-435.

Crop Type information in France

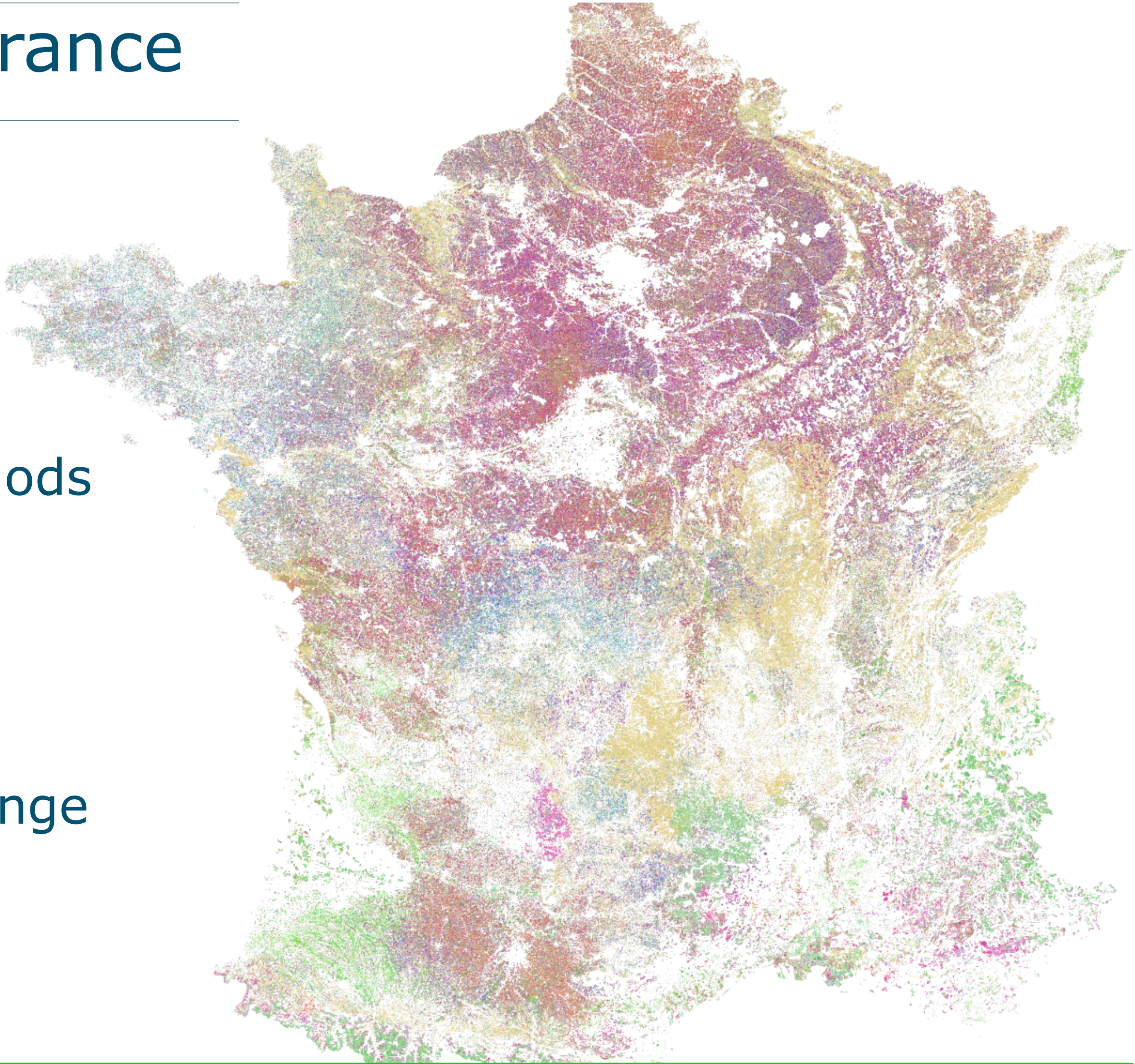
Europe:

Monitoring european crop subsidy
(€386.6 billion 2021-2027)

- great data basis to study methods

Globally:

- monitoring food security
- estimate impact of climate change on agricultural production



WAGENINGEN
UNIVERSITY & RESEARCH

10 million field geometries and cultivated crops published each year

With enough data, all deep learning models do well

Bavaria, Germany

(c) Overall accuracy metric 23-class dataset

acc.	RF	LSTM-RNN	Transformer	DuPLO	MS-ResNet	TempCNN
pre	0.83	$0.85^{\pm 0.01}$	$0.85^{\pm 0.02}$	$\mathbf{0.86}$	$0.83^{\pm 0.02}$	$\mathbf{0.86}^{\pm 0.00}$
raw	0.71	$\mathbf{0.81}^{\pm 0.01}$	$0.80^{\pm 0.02}$	0.79	$0.79^{\pm 0.03}$	$0.79^{\pm 0.00}$

(d) Overall accuracy metric 12-class dataset

acc.	RF	LSTM-RNN	Transformer	DuPLO	MS-ResNet	TempCNN
pre	0.91	$\mathbf{0.92}^{\pm 0.01}$	$\mathbf{0.92}^{\pm 0.03}$	$\mathbf{0.92}$	$0.91^{\pm 0.01}$	$\mathbf{0.92}^{\pm 0.02}$
raw	0.80	$\mathbf{0.90}^{\pm 0.00}$	$0.89^{\pm 0.01}$	0.87	$0.87^{\pm 0.01}$	$0.83^{\pm 0.04}$

Brittany, BreizhCrops

	shallow	convolution				recurrence		attention
FRH04	RF	TempCNN	MS-ResNet	InceptionTime	OmniscCNN	LSTM	StarRNN	Transformer
overall accuracy	0.78	0.79	0.77	0.77	0.73	**0.80**	0.79	**0.80**
average accuracy	0.54	0.55	0.54	0.53	0.52	0.57	0.56	**0.58**
weighted f-score	0.77	0.79	0.77	0.77	0.72	**0.80**	0.79	**0.80**
kappa-metric	0.71	0.73	0.70	0.70	0.65	0.74	0.73	**0.75**
FRH01, 02, 04								
overall accuracy	$0.78^{\pm 0.02}$	$0.80^{\pm 0.01}$	$0.77^{\pm 0.02}$	$0.73^{\pm 0.04}$	$0.77^{\pm 0.05}$	$0.80^{\pm 0.02}$	$0.80^{\pm 0.01}$	$\mathbf{0.81}^{\pm 0.01}$
average accuracy	$0.54^{\pm 0.01}$	$0.57^{\pm 0.01}$	$0.57^{\pm 0.03}$	$0.52^{\pm 0.01}$	$0.55^{\pm 0.03}$	$0.57^{\pm 0.01}$	$0.56^{\pm 0.00}$	$\mathbf{0.59}^{\pm 0.01}$
weighted f-score	$0.77^{\pm 0.02}$	$0.80^{\pm 0.01}$	$0.76^{\pm 0.01}$	$0.69^{\pm 0.08}$	$0.75^{\pm 0.06}$	$0.80^{\pm 0.03}$	$0.80^{\pm 0.01}$	$\mathbf{0.81}^{\pm 0.01}$
kappa-metric	$0.71^{\pm 0.03}$	$0.74^{\pm 0.01}$	$0.71^{\pm 0.01}$	$0.66^{\pm 0.05}$	$0.70^{\pm 0.07}$	$0.75^{\pm 0.03}$	$0.74^{\pm 0.02}$	$\mathbf{0.76}^{\pm 0.02}$

WAGENINGEN
UNIVERSITY & RESEARCH

Rußwurm, M., Pelletier, C., Zollner, M., Lefèvre, S., & Körner, M. (2020). BREIZHCROPS: A TIME SERIES DATASET FOR CROP TYPE MAPPING. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 43, 1545-1551.

Whats’s left: two **challenges** in this field

1: timely in-season classification

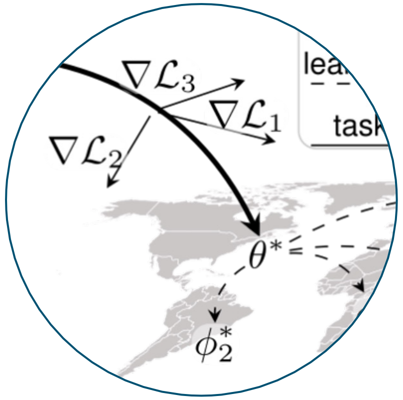
early time series classification

Applicable to other problems, such as

- deforestation alerts
- disaster mapping (change detection)

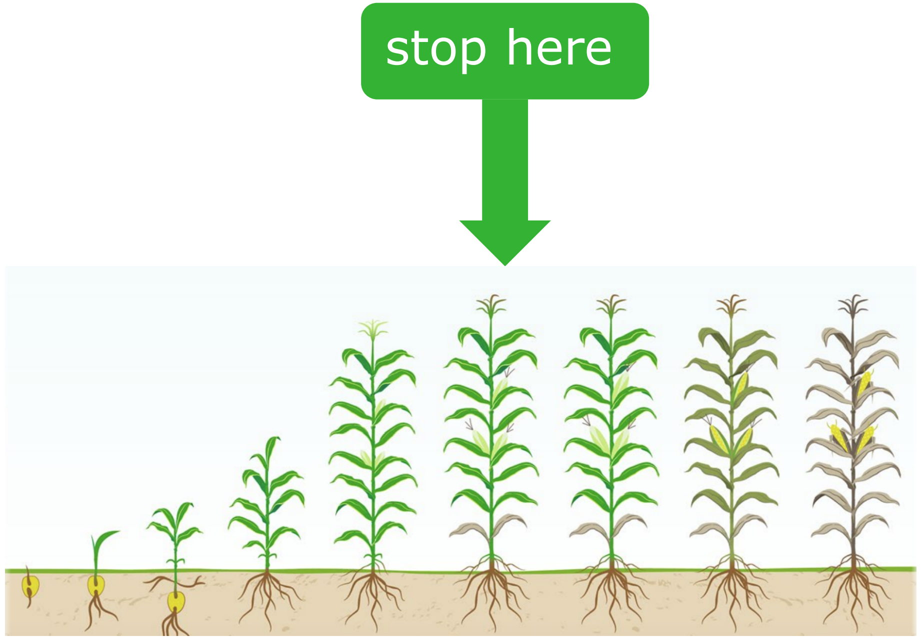
2: transfer learning across regions

domain shift in real-world data



Applicable beyond time series,

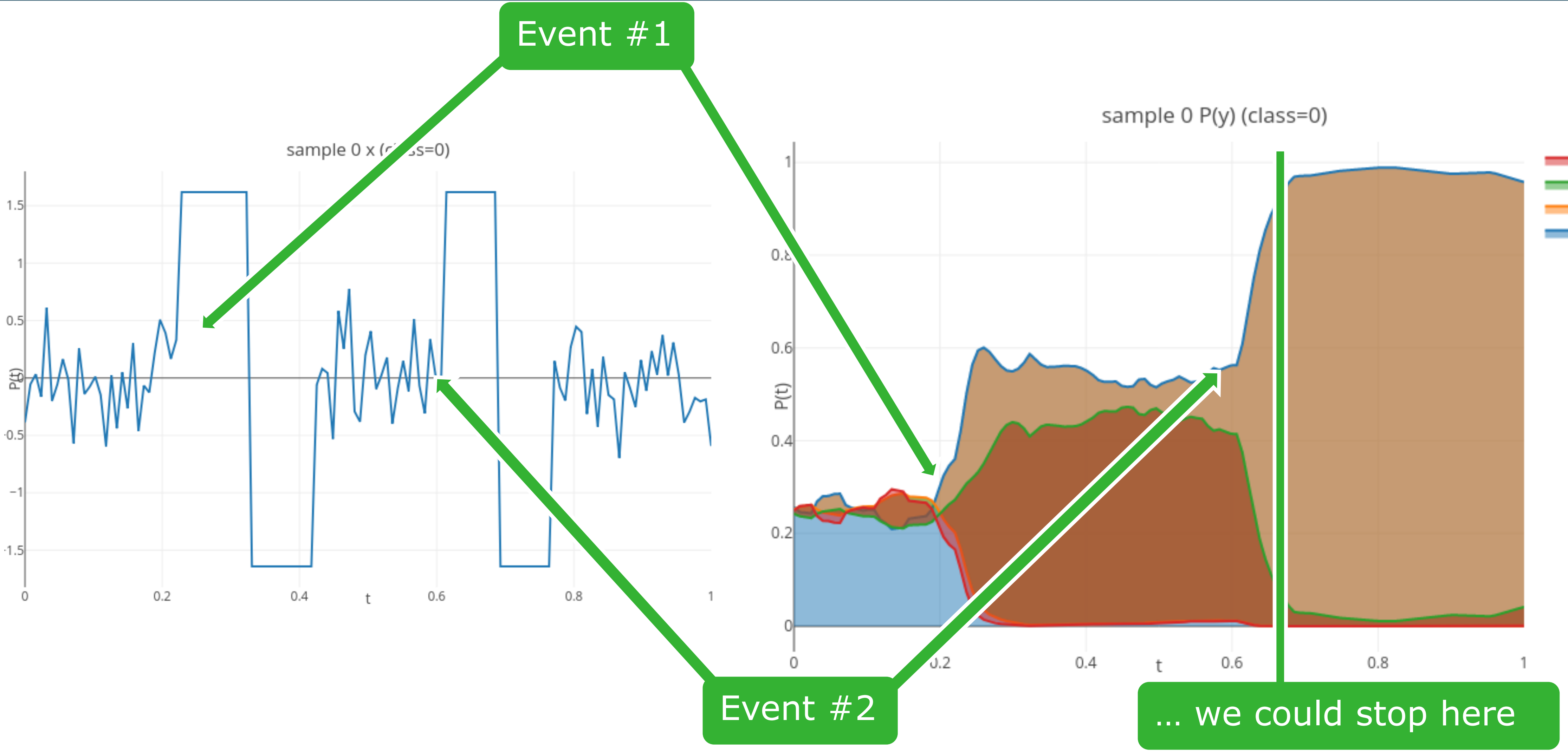
- land cover maps
- generalizability reliability beyond the training regions/years



1 Early Time Series Classification

What is Early Classification

TwoPatterns dataset from the Hexagon ML UCR Time Series Classification Archive, Dau et al., 2018



Hoang Anh Dau, Eamonn Keogh, Kaveh Kamgar, Chin-Chia Michael Yeh, Yan Zhu, Shaghayegh Gharghabi , Chotirat Ann Ratanamahatana, Yanping Chen, Bing Hu, Nurjahan Begum, Anthony Bagnall , Abdullah Mueen, Gustavo Batista, & Hexagon-ML (2019). *The UCR Time Series Classification Archive.*

Early Classification and Selective Prediction

Selective Prediction: The ability to to **abstain** from a **decision** when lacking confidence [1]

- uncertainty based methods through Monte-Carlo Dropout or Model Ensembles
- architecture based methods through a selection head of abstention logit

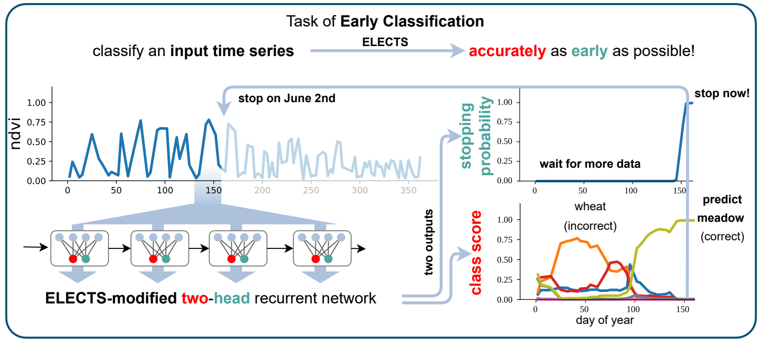
Early Classification: The **accurate** classification of time series as **early** as possible [2]

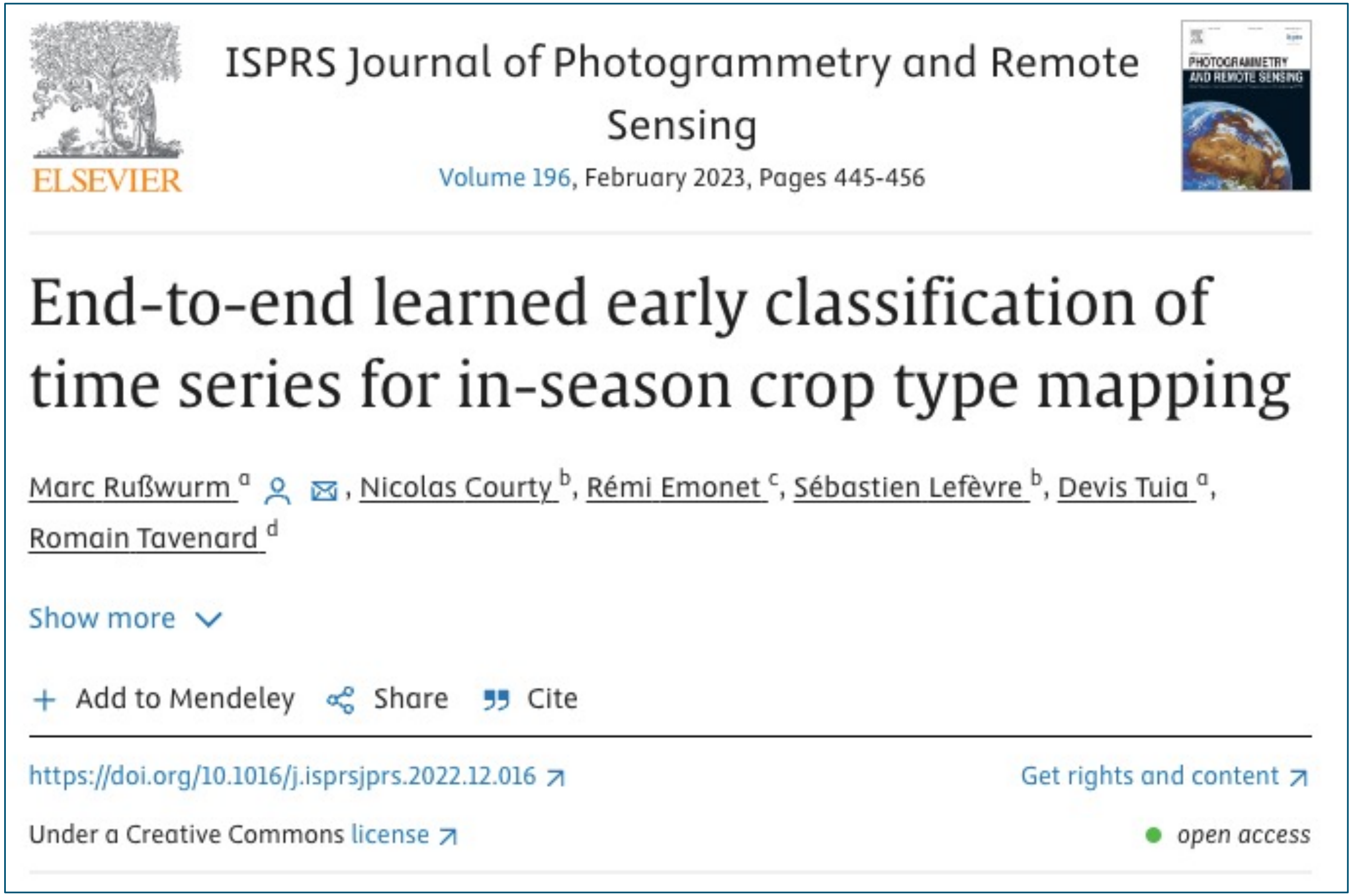
- is selective prediction in a time series context: we abstain from taking a decision and wait for more data

[1] Leo Feng, Mohamed Osama Ahmed, Hossein Hajimirsadeghi, and Amir H Abdi. Towards better selective classification. In The Eleventh International Conference on Learning Representations, 2023
[2] Ashish Gupta, Hari Prabhat Gupta, Bhaskar Biswas, and Tanima Dutta. Approaches and applications of early classification of time series: A review. IEEE Transactions on Artificial Intelligence

End-to-end learned early classification of time series

ELECTS-modified LSTM





ELSEVIER

ISPRS Journal of Photogrammetry and Remote Sensing

Volume 196, February 2023, Pages 445-456

End-to-end learned early classification of time series for in-season crop type mapping

Marc Rußwurm [a], Nicolas Courty [b], Rémi Emonet [c], Sébastien Lefèvre [b], Devis Tuia [a], Romain Tavenard [d]

Show more

+ Add to Mendeley Share Cite

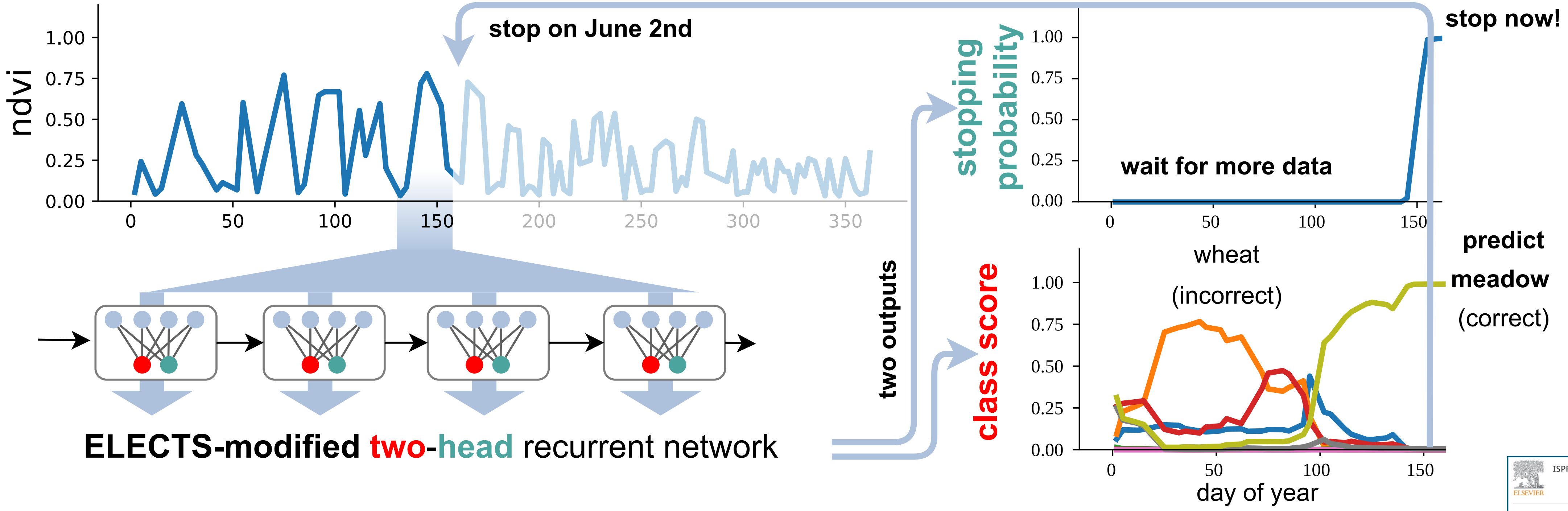
https://doi.org/10.1016/j.isprsjprs.2022.12.016

Get rights and content

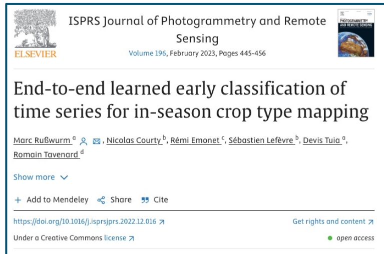
End-to-end learned early classification of time series for in-season crop type mapping

Task of **Early Classification**

classify an **input time series** —ELECTS→ **accurately** as **early** as possible!



WAGENINGEN UNIVERSITY & RESEARCH



ISPRS Journal of Photogrammetry and Remote Sensing

Volume 196, February 2023, Pages 445-456

End-to-end learned early classification of time series for in-season crop type mapping

Marc Rußwurm [a], Nicolas Courty [b], Rémi Emonet [c], Sébastien Lefèvre [b], Devis Tuia [a], Romain Tavenard [d]

Show more

Add to Mendeley Share Cite

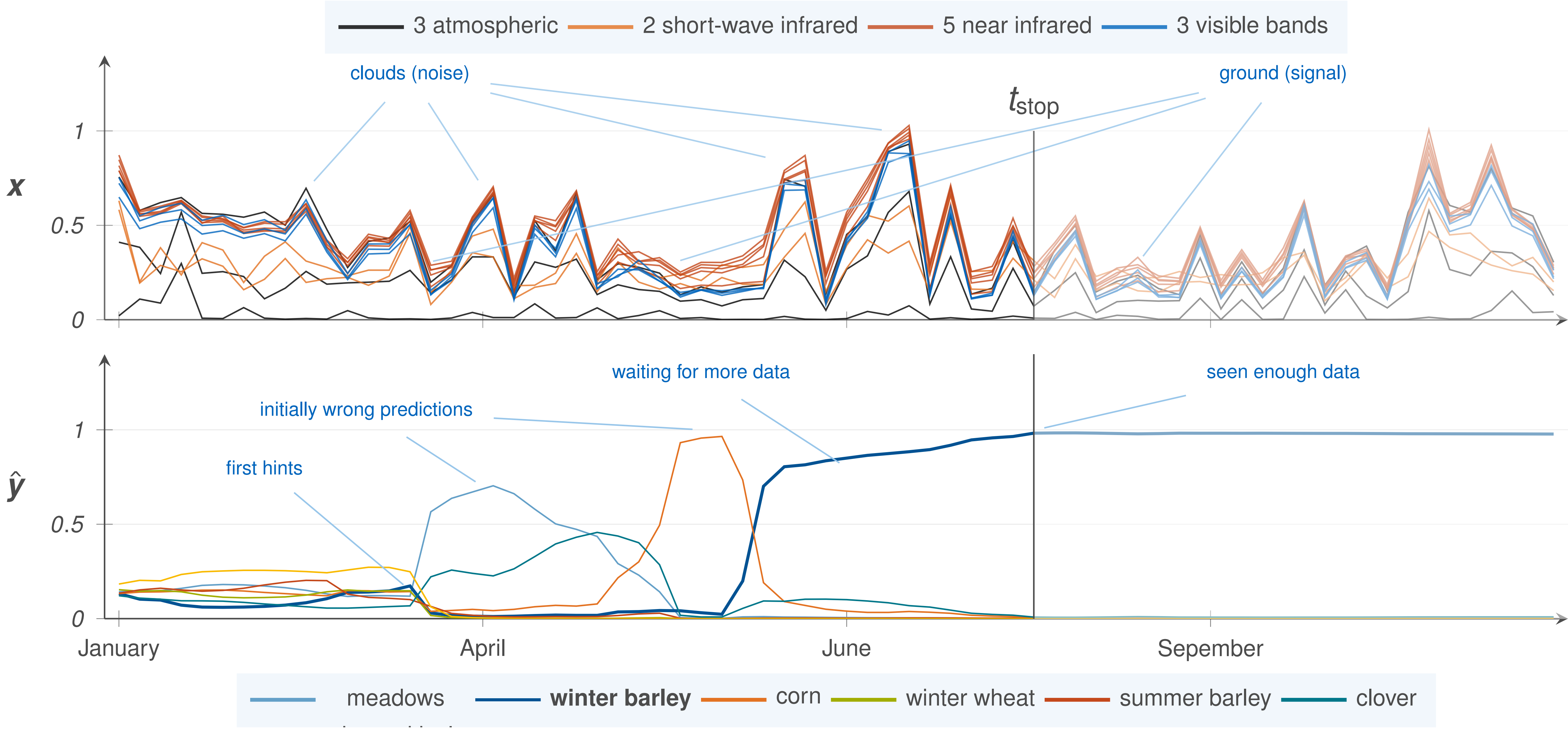
https://doi.org/10.1016/j.isprsjprs.2022.12.016

Get rights and content

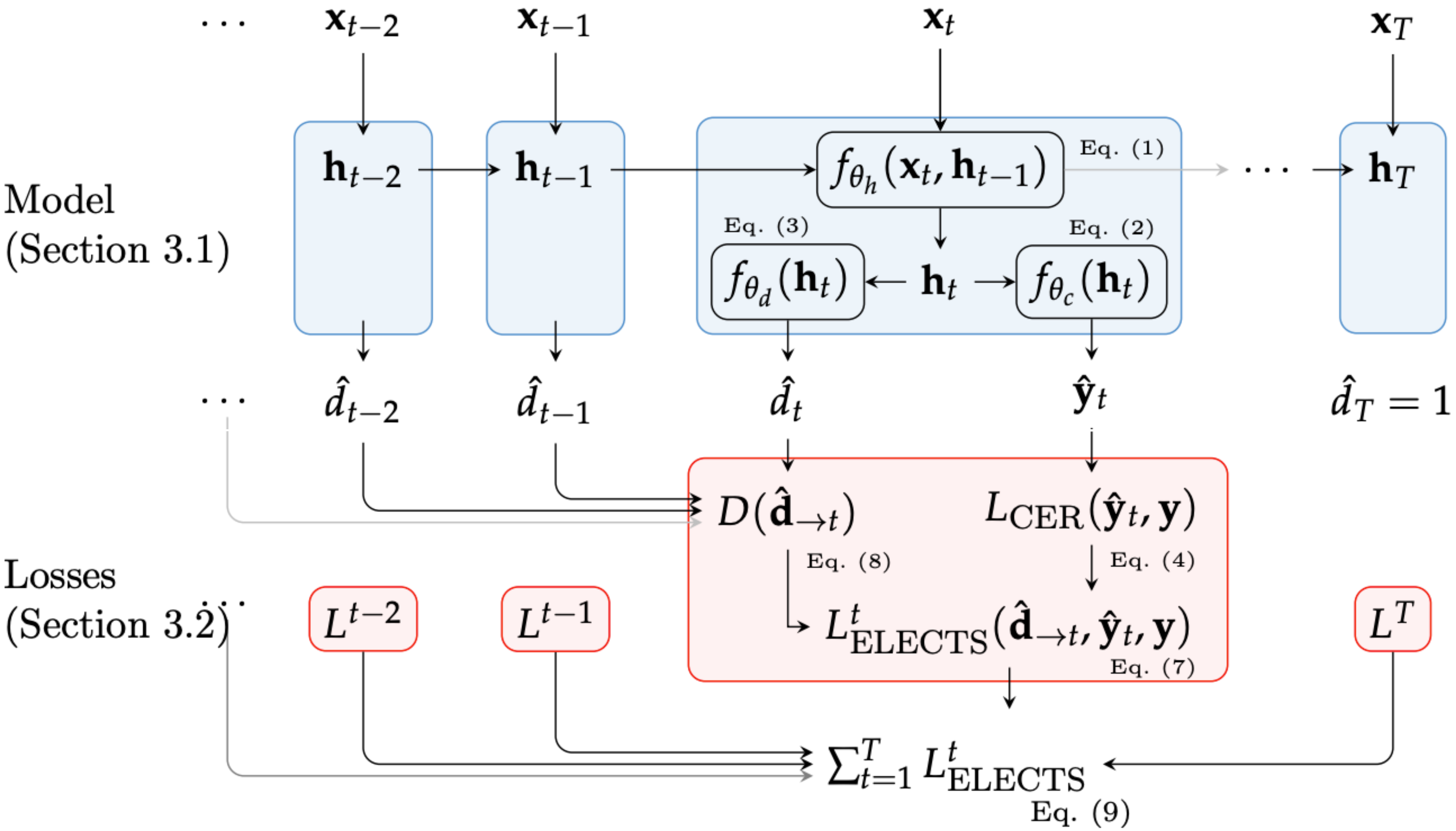
Under a Creative Commons license

open access

Early Classification of Crop Type from Sentinel-2 time series



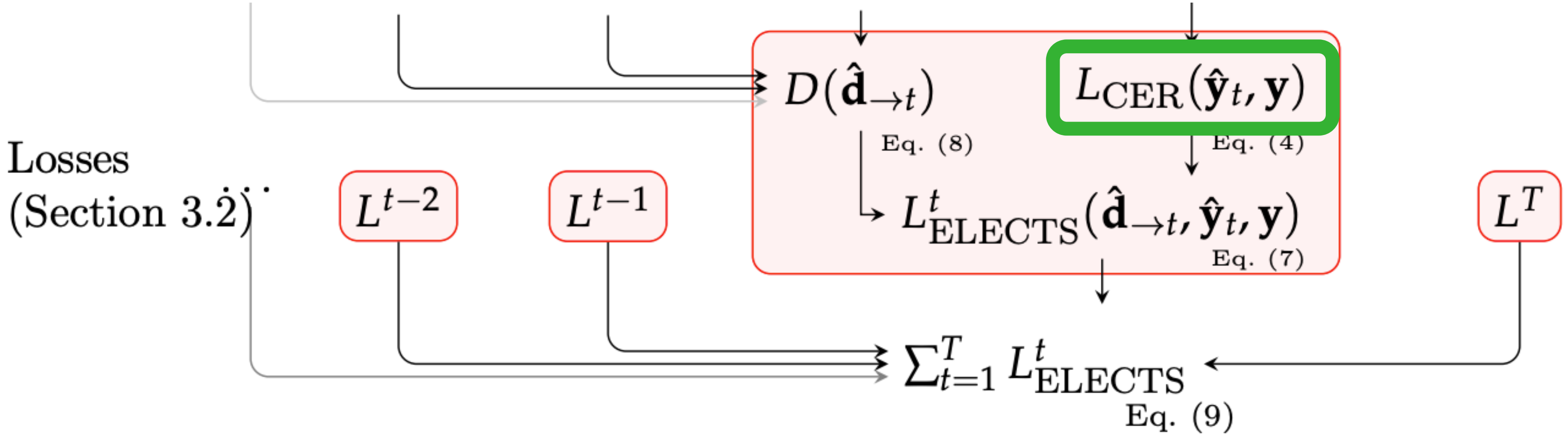
Method: Model and Loss functions



WAGENINGEN UNIVERSITY & RESEARCH

Rußwurm, M., Courty, N., Emonet, R., Lefèvre, S., Tuia, D., & Tavenard, R. (2023). End-to-end learned early classification of time series for in-season crop type mapping. *ISPRS Journal of Photogrammetry and Remote Sensing*, *196*, 445-456.

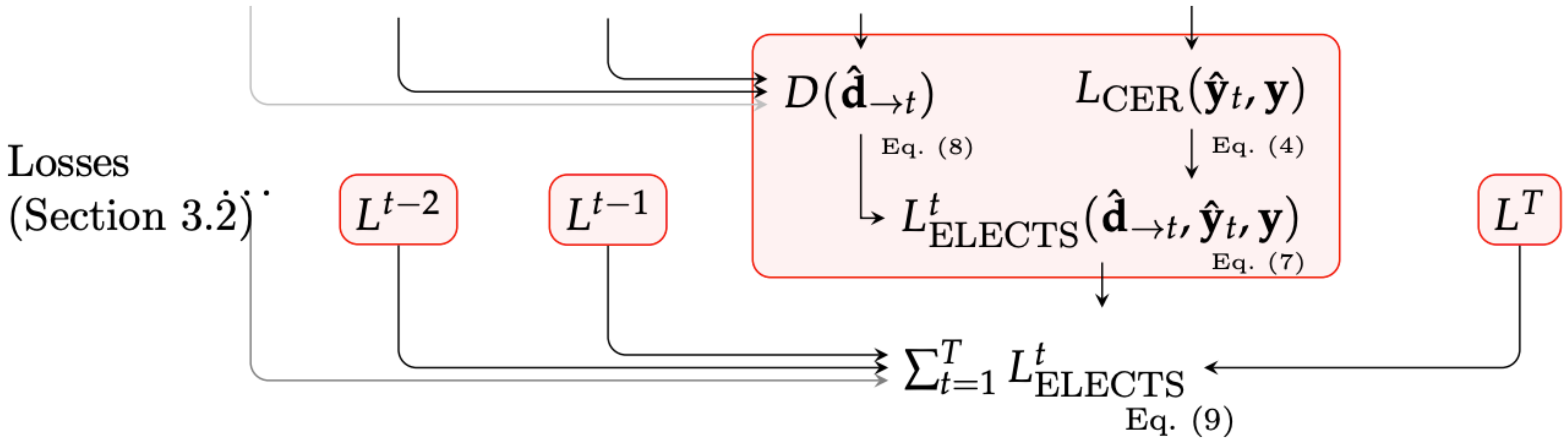
Loss functions



Classification & Early Reward: $L_{\mathrm{CER}}(\hat{\mathbf{y}}_t, \mathbf{y}) = \alpha L_c(\hat{\mathbf{y}}_t, \mathbf{y}) - (1-\alpha) R_e(\hat{\mathbf{y}}_t, \mathbf{y}, t)$

Rußwurm, M., Courty, N., Emonet, R., Lefèvre, S., Tuia, D., & Tavenard, R. (2023). End-to-end learned early classification of time series for in-season crop type mapping. *ISPRS Journal of Photogrammetry and Remote Sensing*, *196*, 445-456.

Loss functions



Classification & Early Reward:

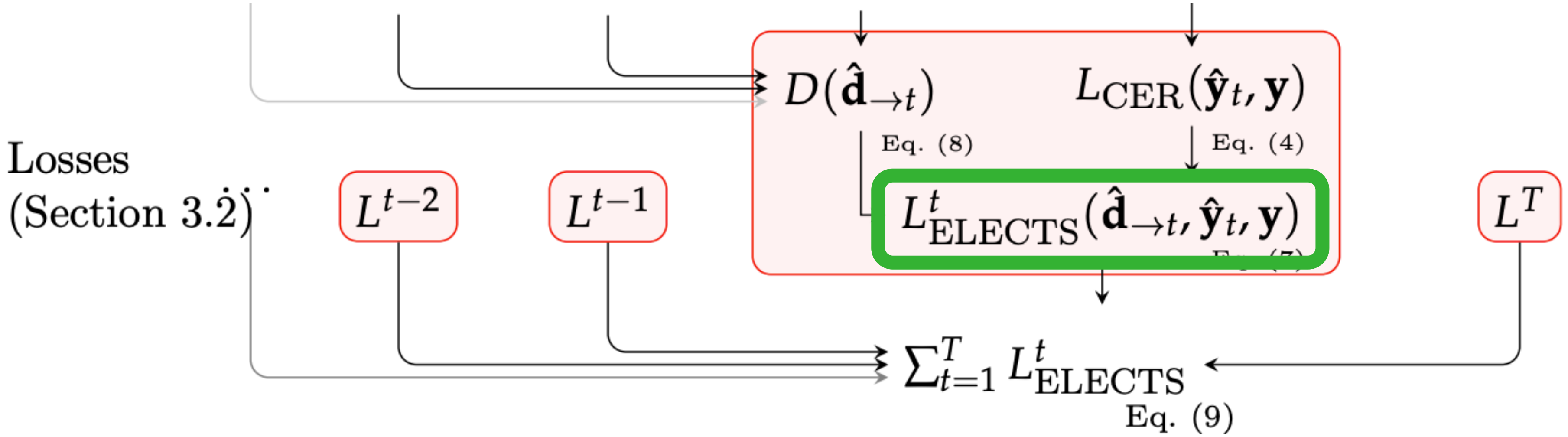
$$L_{\mathrm{CER}}(\hat{\mathbf{y}}_t, \mathbf{y}) = \alpha L_c(\hat{\mathbf{y}}_t, \mathbf{y}) - (1-\alpha) R_e(\hat{\mathbf{y}}_t, \mathbf{y}, t)$$

$$L_c(\hat{\mathbf{y}}_t, \mathbf{y}) = -\sum_{c=1}^{C} y_c \log \hat{y}_{c,t},$$

$$R_e(\hat{\mathbf{y}}_t, \mathbf{y}, t) = \hat{y}_t^{+} \left(\frac{T-t}{T} \right)$$

Rußwurm, M., Courty, N., Emonet, R., Lefèvre, S., Tuia, D., & Tavenard, R. (2023). End-to-end learned early classification of time series for in-season crop type mapping. *ISPRS Journal of Photogrammetry and Remote Sensing*, *196*, 445-456.

Loss functions

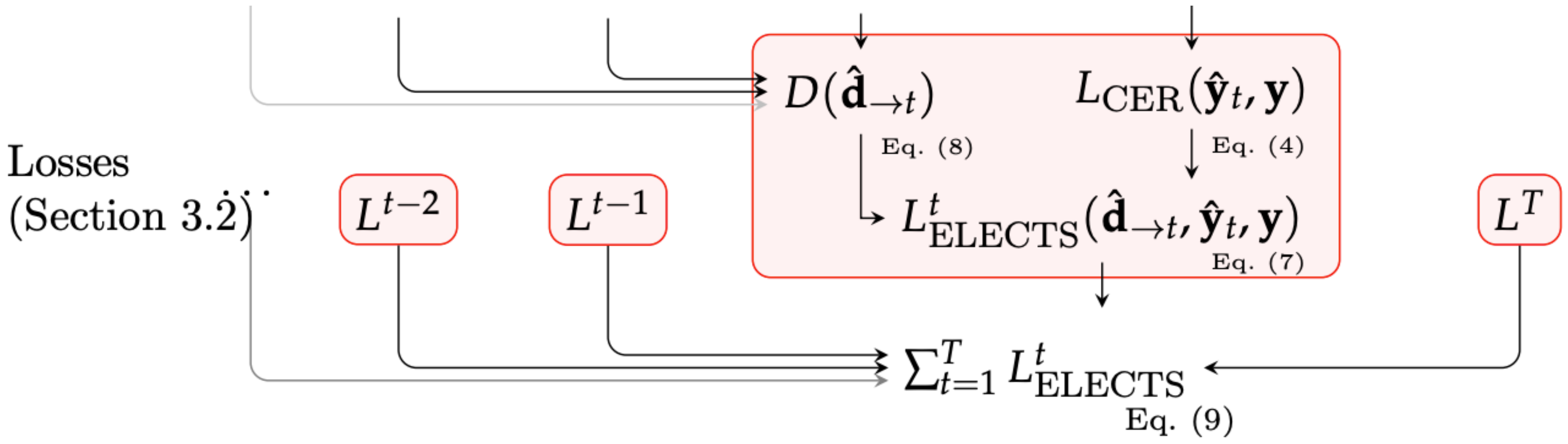


ELECTS Loss: $$L_{\mathrm{ELECTS}}(\hat{\mathbf{d}}_{\rightarrow t}, \hat{\mathbf{y}}_t, \mathbf{y}) = D_t(\hat{\mathbf{d}}_{\rightarrow t}) L_{\mathrm{CER}}(\hat{\mathbf{y}}_t, \mathbf{y})$$

WAGENINGEN UNIVERSITY & RESEARCH

Rußwurm, M., Courty, N., Emonet, R., Lefèvre, S., Tuia, D., & Tavenard, R. (2023). End-to-end learned early classification of time series for in-season crop type mapping. *ISPRS Journal of Photogrammetry and Remote Sensing*, *196*, 445-456.

Loss functions



ELECTS Loss:

$$L_{\mathrm{ELECTS}}(\hat{\mathbf{d}}_{\rightarrow t}, \hat{\mathbf{y}}_t, \mathbf{y}) = D_t(\hat{\mathbf{d}}_{\rightarrow t}) L_{\mathrm{CER}}(\hat{\mathbf{y}}_t, \mathbf{y})$$

$$\boxed{D(\hat{\mathbf{d}}_{\rightarrow t}) = \hat{d}_t \prod_{i=1}^{t-1} (1 - \hat{d}_i) + \frac{\varepsilon}{T}}$$

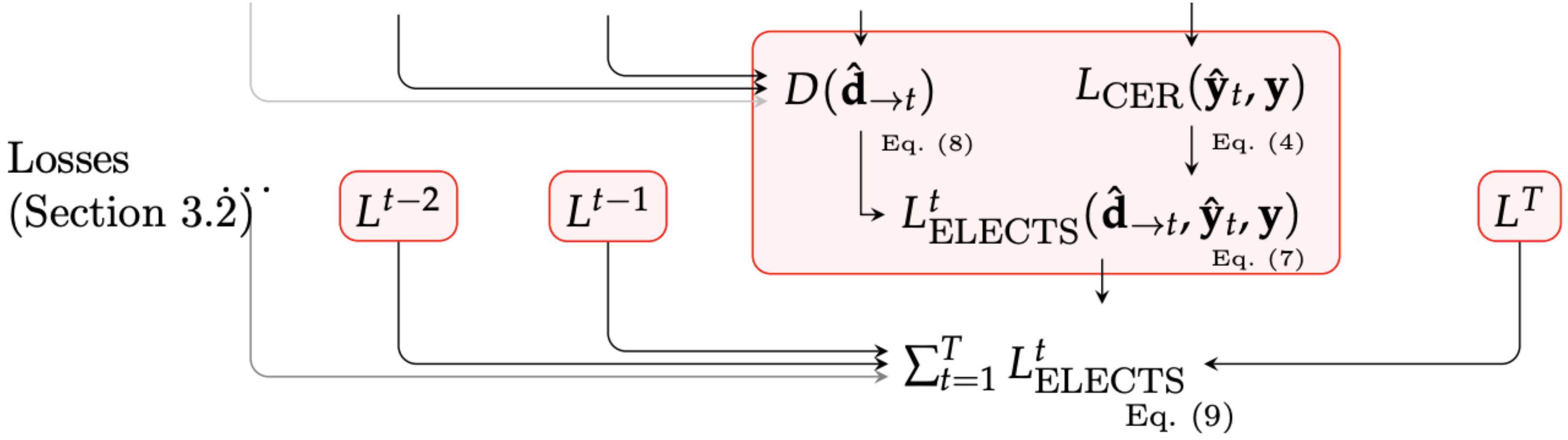
Probability of stopping

and not having stopped before

WAGENINGEN UNIVERSITY & RESEARCH

Rußwurm, M., Courty, N., Emonet, R., Lefèvre, S., Tuia, D., & Tavenard, R. (2023). End-to-end learned early classification of time series for in-season crop type mapping. *ISPRS Journal of Photogrammetry and Remote Sensing*, *196*, 445-456.

Loss functions



ELECTS Loss:

$$L_{\mathrm{ELECTS}}(\hat{\mathbf{d}}_{\rightarrow t}, \hat{\mathbf{y}}_t, \mathbf{y}) = D_t(\hat{\mathbf{d}}_{\rightarrow t}) L_{\mathrm{CER}}(\hat{\mathbf{y}}_t, \mathbf{y})$$

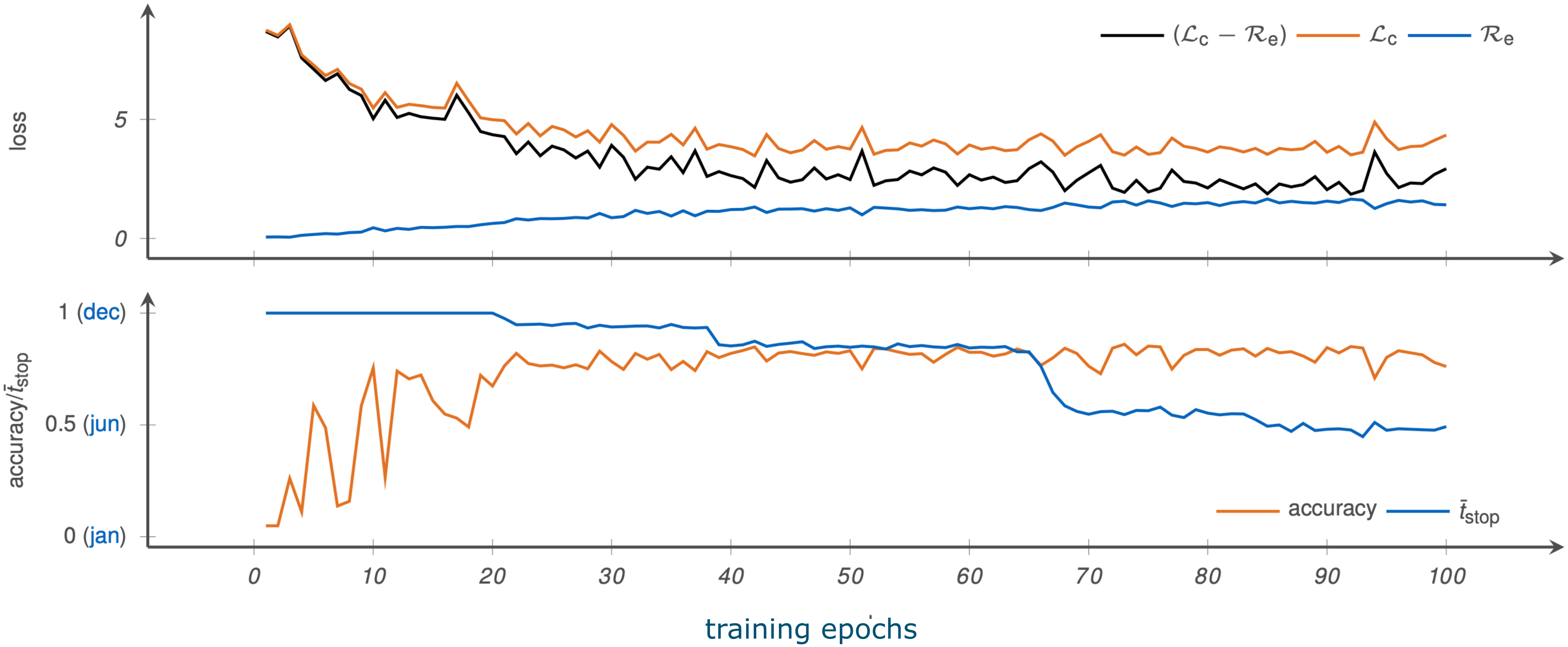
$$\boxed{D(\hat{\mathbf{d}}_{\rightarrow t}) = \hat{d}_t \prod_{i=1}^{t-1} (1 - \hat{d}_i) + \frac{\varepsilon}{T}}$$

constant allows the model to look-ahead for more stable training

WAGENINGEN
UNIVERSITY & RESEARCH

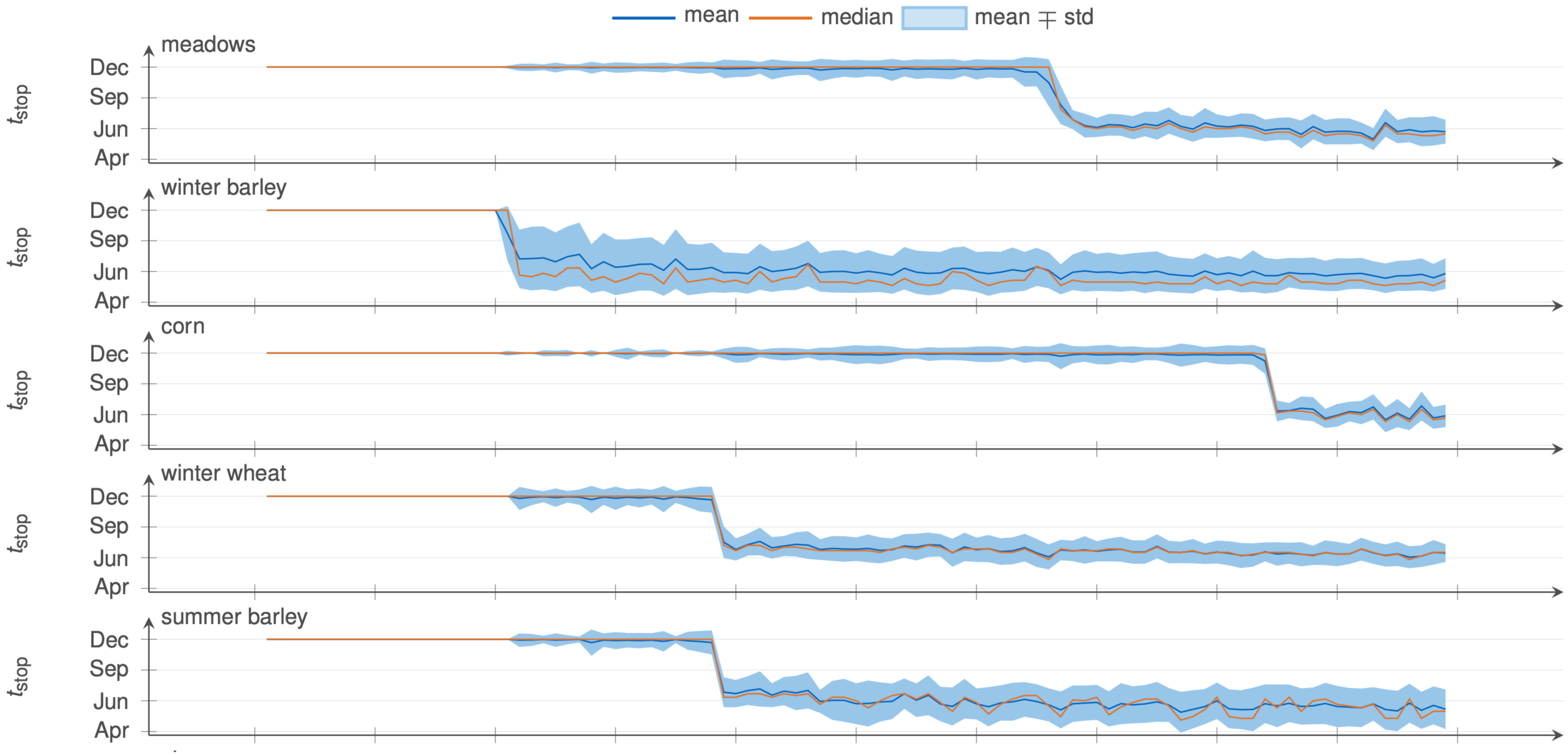
Rußwurm, M., Courty, N., Emonet, R., Lefèvre, S., Tuia, D., & Tavenard, R. (2023). End-to-end learned early classification of time series for in-season crop type mapping. *ISPRS Journal of Photogrammetry and Remote Sensing*, *196*, 445-456.

Earliness decreases step-wise during training

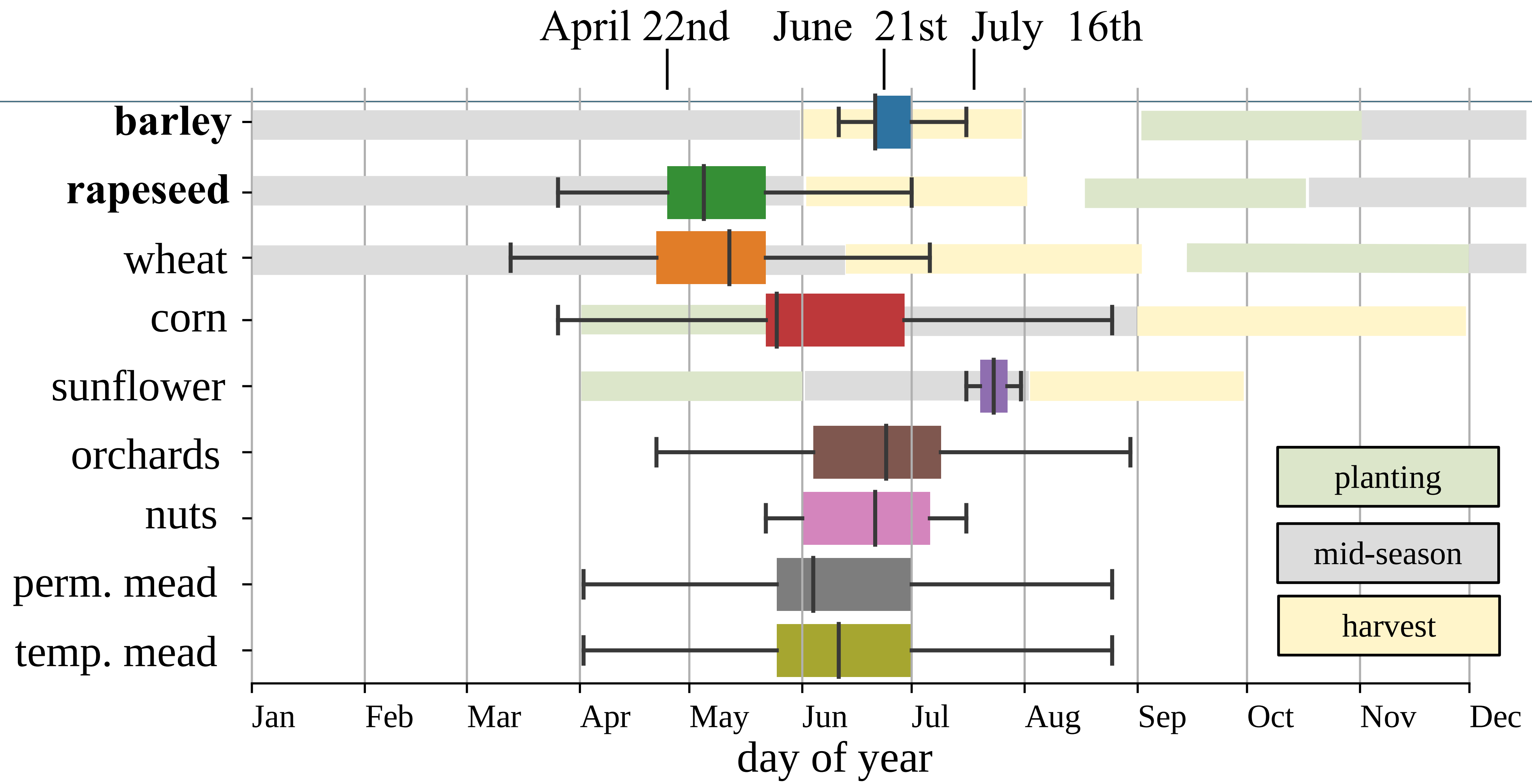


WAGENINGEN UNIVERSITY & RESEARCH

Step-wise learning of stopping times per class



Results: Different Crop Types are Stopped at different times



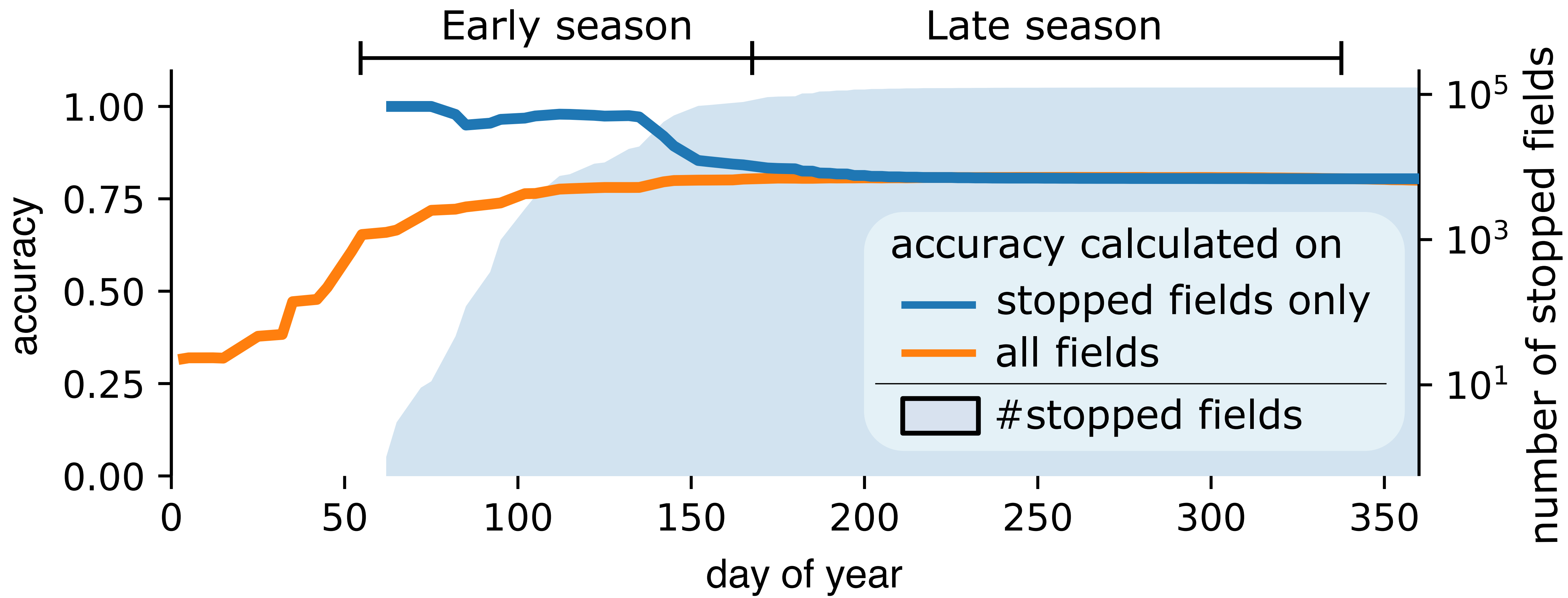
WAGENINGEN
UNIVERSITY & RESEARCH

Rapeseed fields are stopped after blossoming

Barley fields are stopped with harvesting

WAGENINGEN
UNIVERSITY & RESEARCH

Results: Early Stopped field classifications are more accurate



Motivational property of selective prediction models

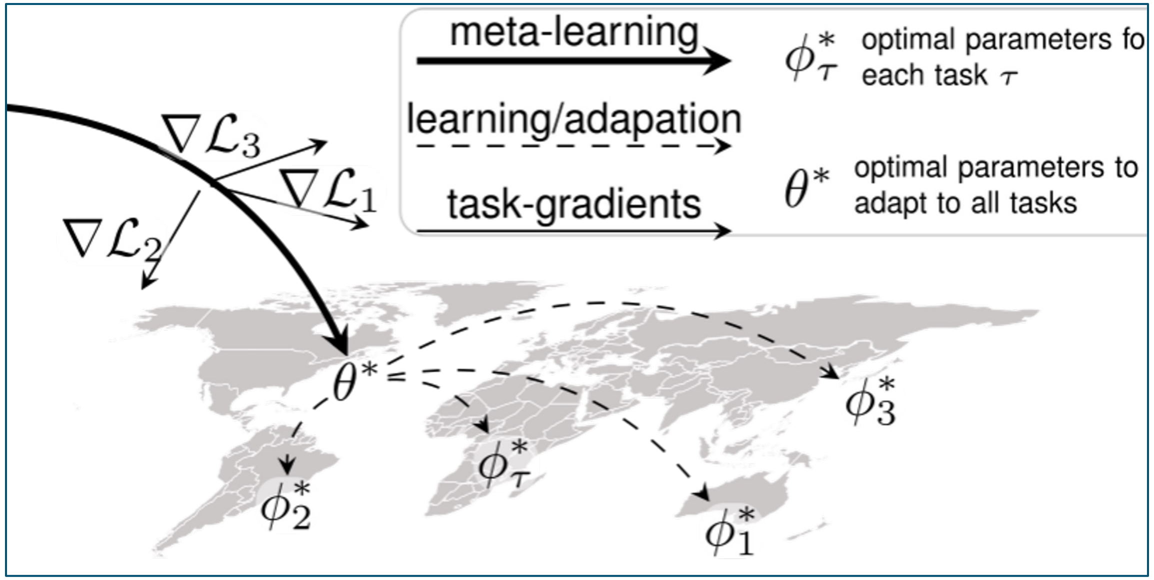
Takeaways Early Classification

Method from the time series community the runs really well on crop type data

niche of early classification

niche of crop type mapping

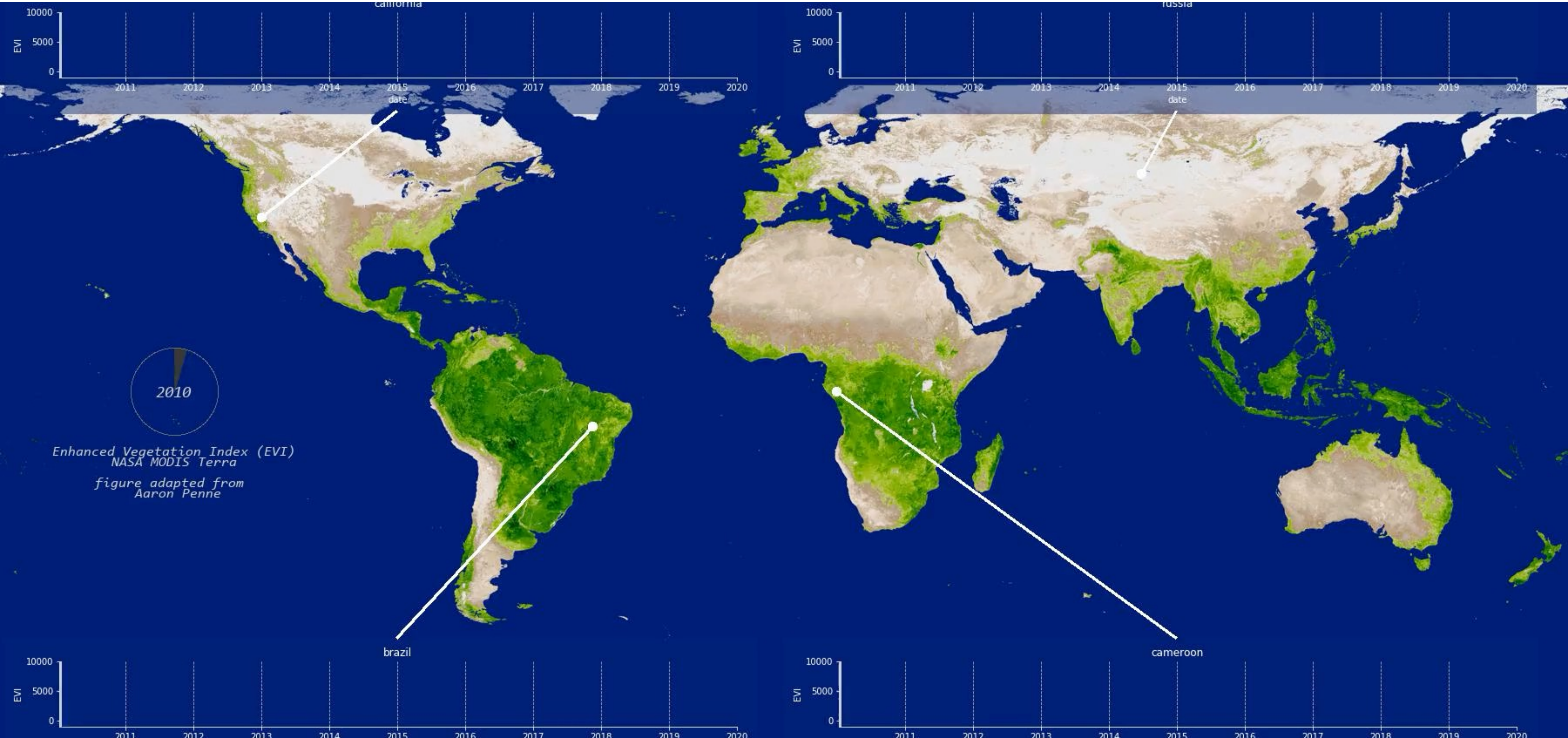
very rewarding,
but difficult area to publish



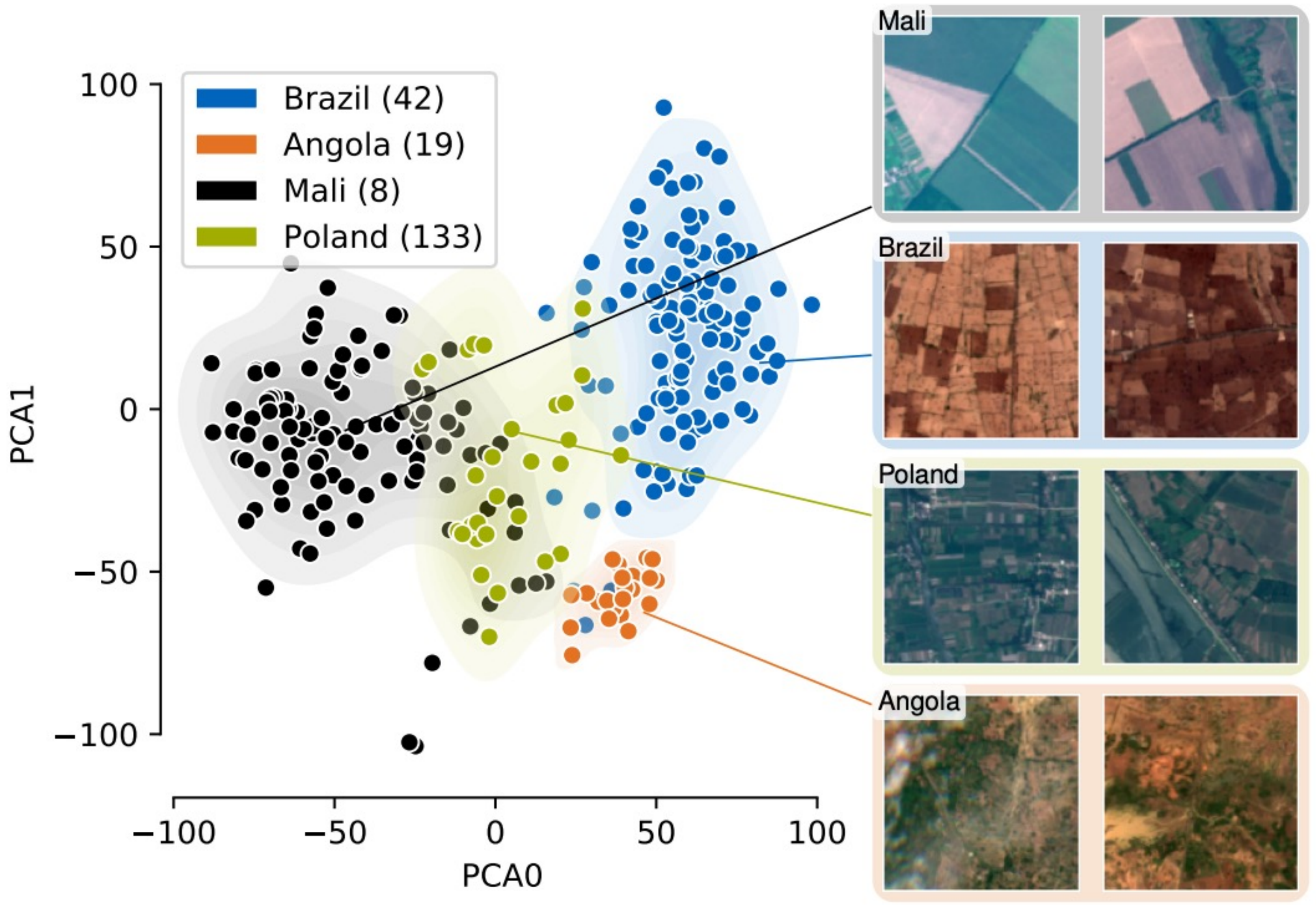
2 Transfer Learning across geographies

WAGENINGEN UNIVERSITY & RESEARCH

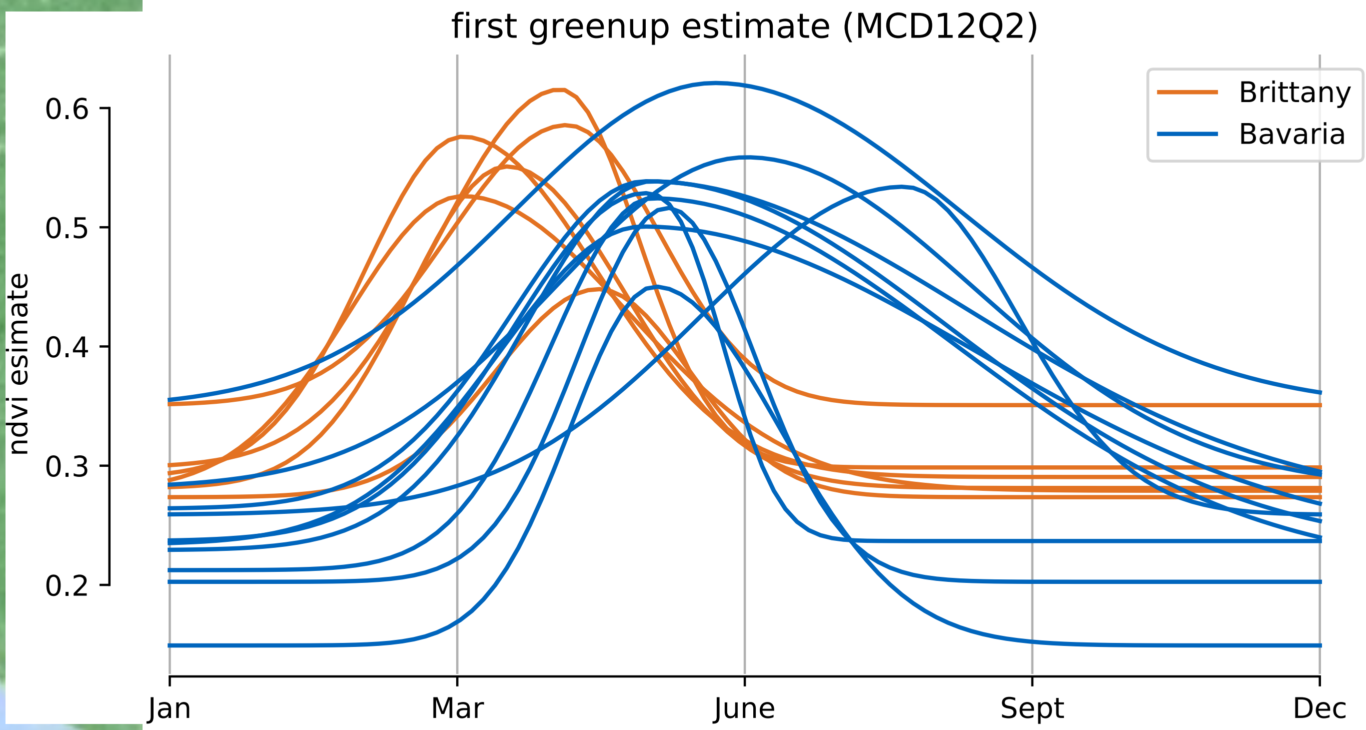
Regional-shifts of Vegetation on the Planet



Images of Cropland are very different between regions



Environment drives the changes in plant life cycles



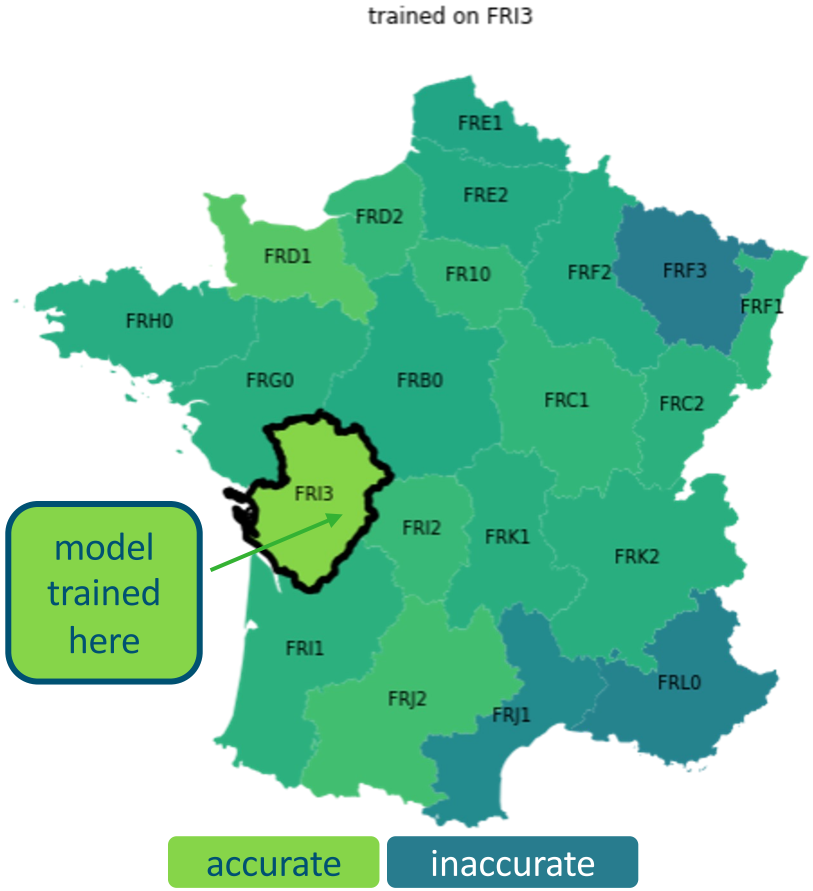
Environment drives farming choices

- Different environmental conditions (temperature, precipitation, soil) lead to different farming practices

- very region-specific dominant crop types

Example Domain Shifts in Crop-Type Mapping

Regional Domain Shift



Temporal Domain Shift

Accuracy crop type mapping in Germany (Kondmann et al., 2021) with attention-based time series classifier (PSE-TAE; Garnot et al., 2020)

trained on 2018 data

tested on:	2018	2019
accuracy	78.77%	67.25%

Kondmann, L., Toker, A., Rußwurm, et al., (2021, August). DENETHOR. In *NeurIPSDatasets and Benchmarks Track*

Garnot, V. S. F., et al., (2020). Satellite image time series classification with pixel-set encoders and temporal self-attention.

Problem:
testing models in areas with training data is not useful!

Real-world data is not I.I.D.

development time

collected once:

$$\mathcal{D} = \{X_i, y_i\}_{i=1}^{N} \sim p(X, Y)$$

then partitioned randomly:

random sampling

$$\mathcal{D}_{\text{train}} \longleftrightarrow \mathcal{D}_{\text{test}} \sim p(X, Y)$$

in-distribution generalization

deployment time

🤷

out-of-distribution generalization

$$\mathcal{D}_{\text{test}} \sim q(X, Y)$$

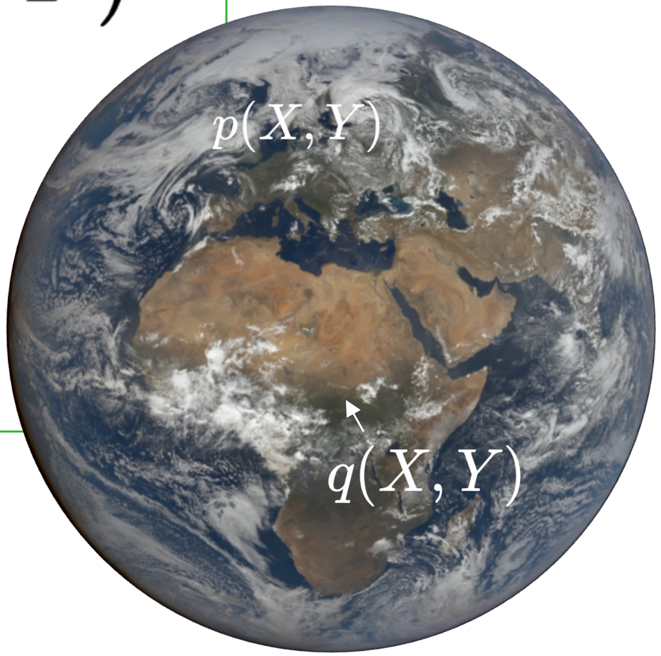
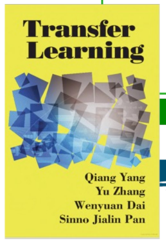
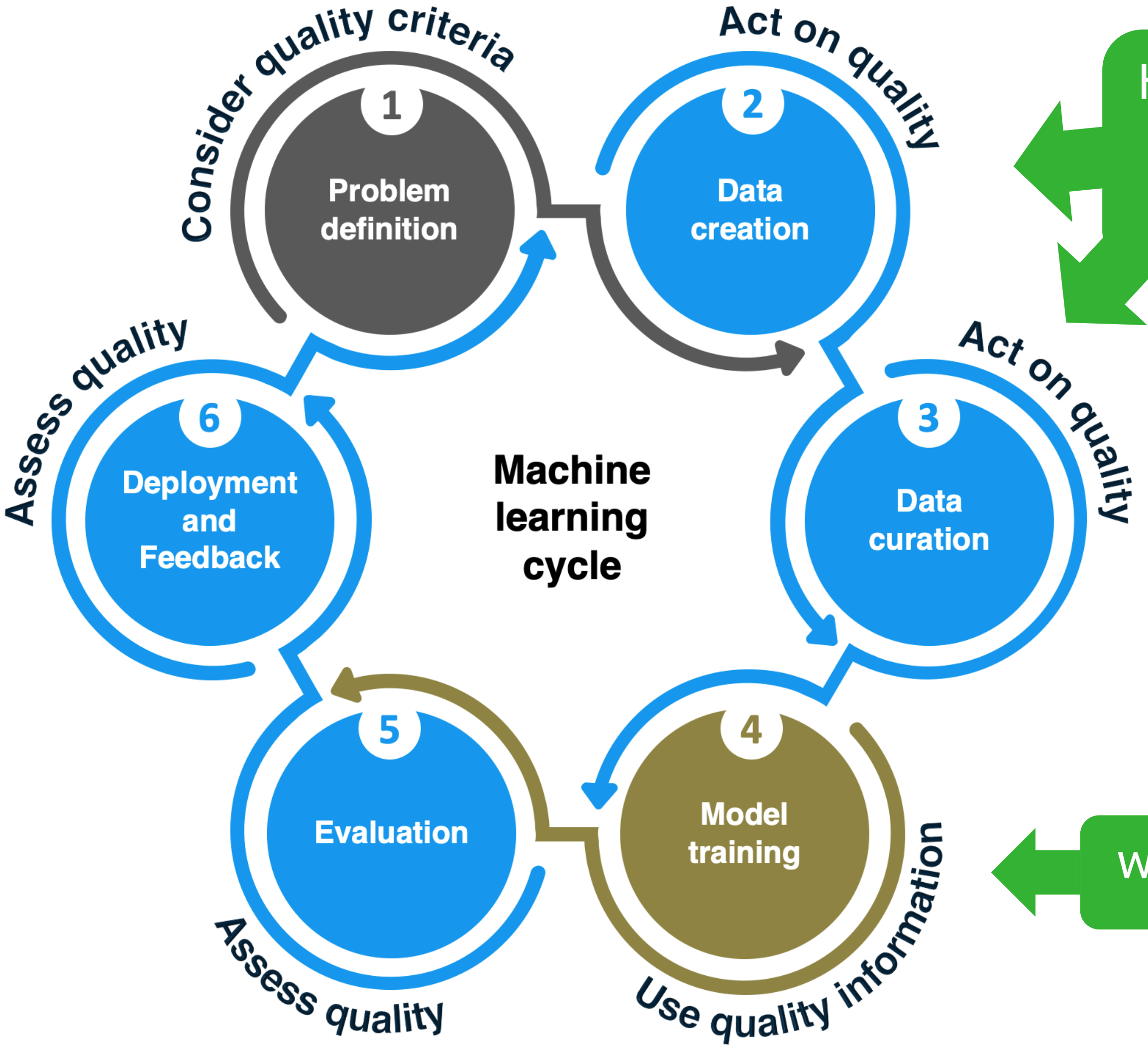


Image from
DSCOVR:EPIC 2022-06-16



WAGENINGEN
UNIVERSITY & RESEARCH

Domain Shift from a Data-Centric Perspective



Here, we define

$p(X,Y)$

Here, we have

$q(X,Y)$

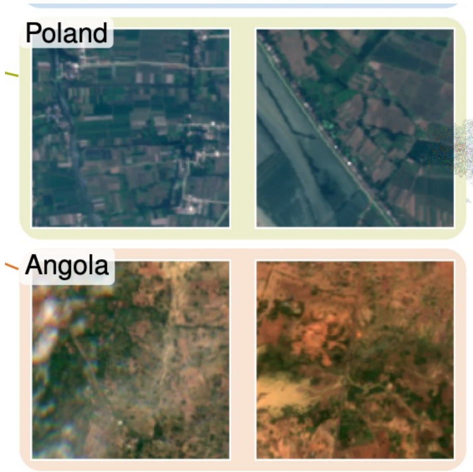
we focus a lot on models

WAGENINGEN UNIVERSITY & RESEARCH

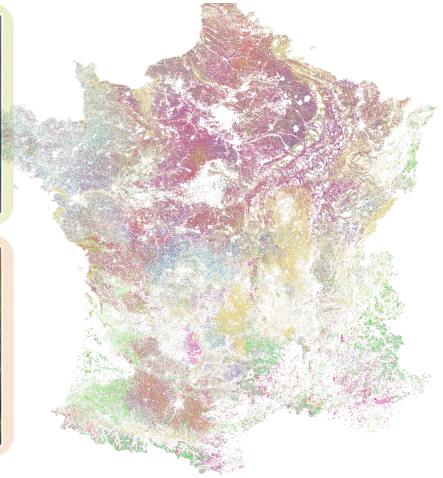
Roscher et al., 2023 (to appear) Data-Centric Machine Learning for Geospatial Data

Four canonical types of domain shift

manifestation shift



prior shift



Distribution Shift

$$p(X,Y) \neq q(X,Y)$$

discriminative perspective:

1. covariate shift

$$p(X)p(Y|X) \neq q(X)p(Y|X)$$

2. label shift

$$p(X)p(Y|X) \neq p(X)q(Y|X)$$

generative perspective:

3. prior shift

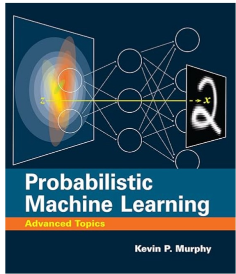
$$p(Y)p(X|Y) \neq q(Y)p(X|Y)$$

4. manifestation shift

$$p(Y)p(X|Y) \neq p(Y)q(X|Y)$$

WAGENINGEN
UNIVERSITY & RESEARCH

Kevin P. Murphy. Probabilistic Machine Learning: Advanced Topics



Learn-to-Learn Regional Models

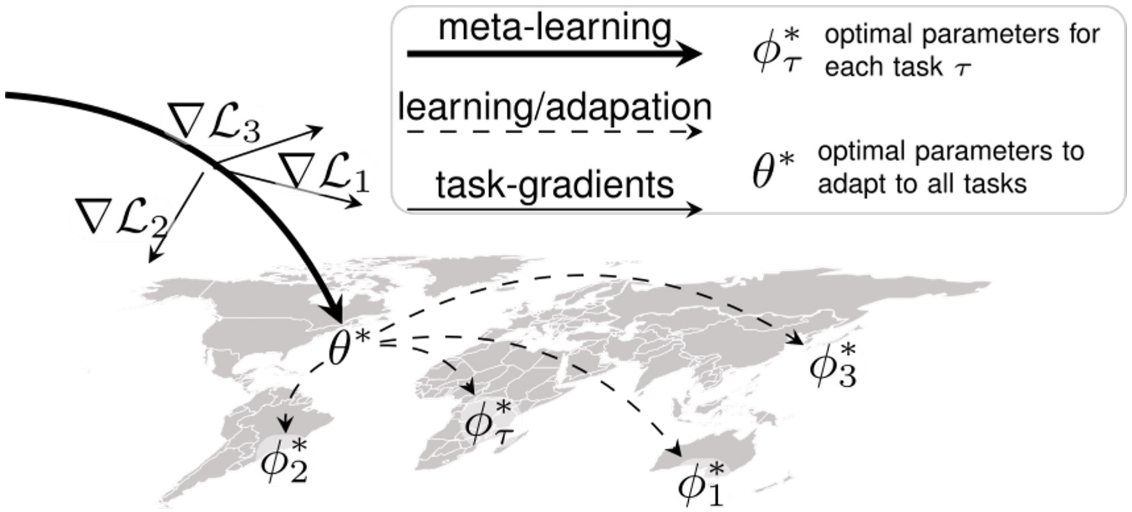
concept

meta-learning
a global adaptive model

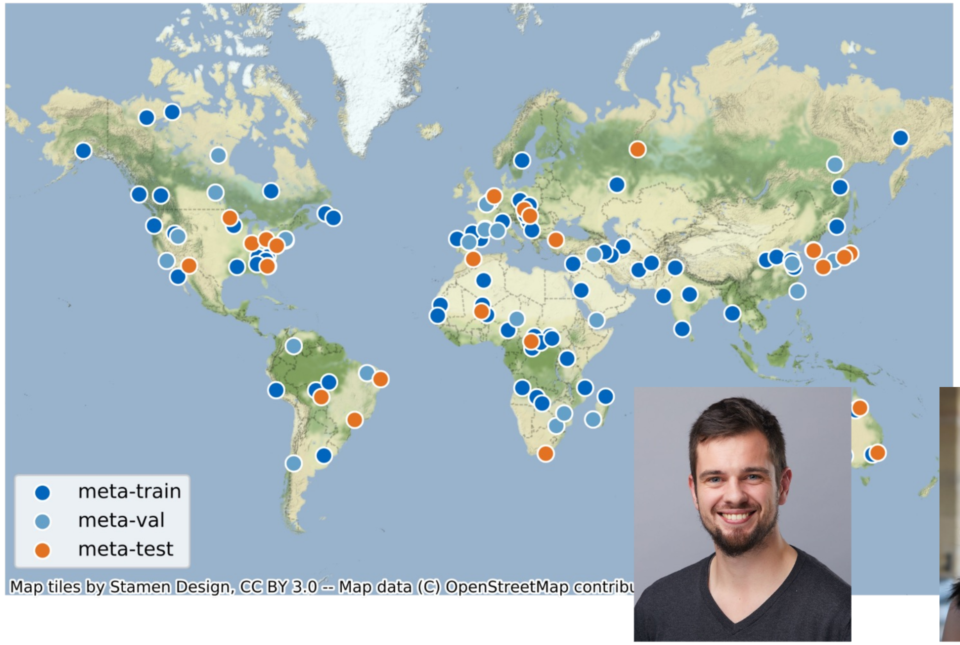
fine-tune

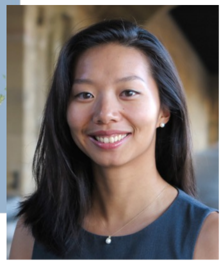
regional model
with few annotated samples

Model-agnostic meta-learning (Finn et al., 2017) **naturally allows for different data distributions** between tasks



globally distributed land cover



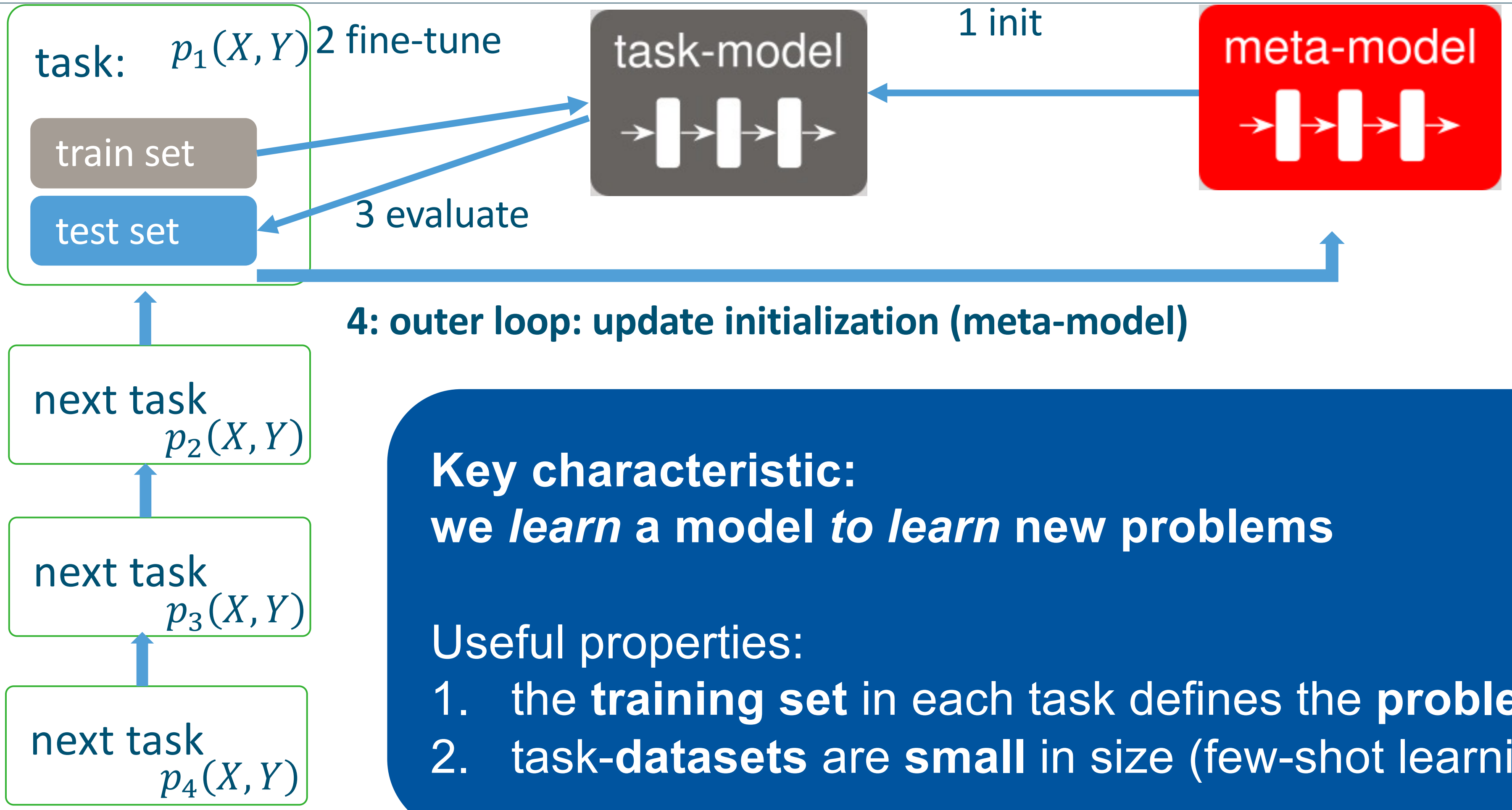


Marc Rußwurm* Sherrie Wang*
* equal contribution

WAGENINGEN
UNIVERSITY & RESEARCH

Rußwurm & Wang et al., (2020); CVPR Earthvision Workshop (best paper award)

Model-agostic Meta-learning Algorithm (Finn et al., 2017)



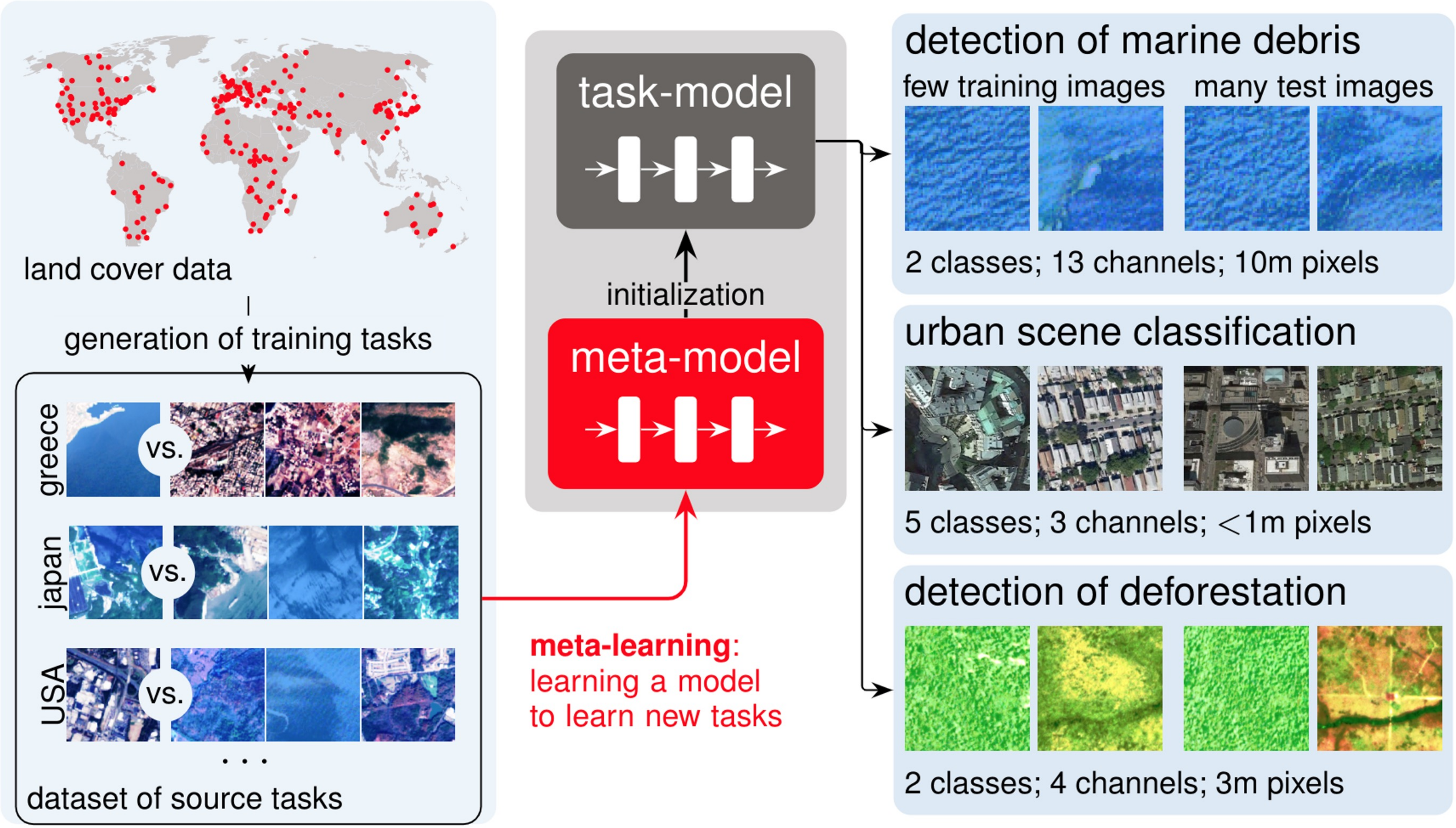
Key characteristic:
we *learn* a model *to learn* new problems

Useful properties:

1. the **training set** in each task defines the **problem**
2. task-**datasets** are **small** in size (few-shot learning)

WAGENINGEN UNIVERSITY & RESEARCH

Meta-learning across heterogeneous remote sensing problems

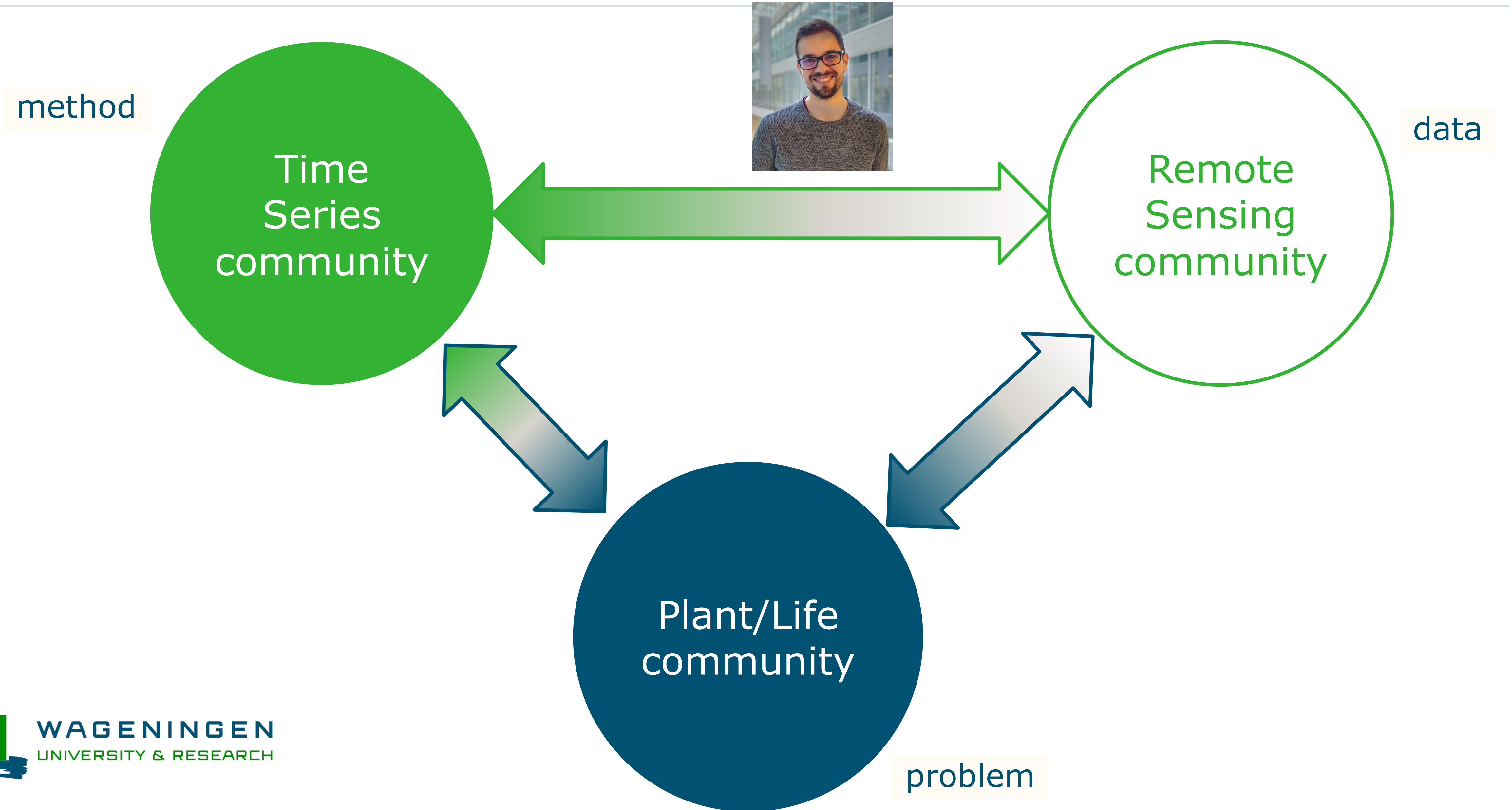


WAGENINGEN UNIVERSITY & RESEARCH

M. Rußwurm, R. Roscher, B. Kellenberger, S. Wang, D. Tuia, (2023) Meta-learning to address diverse Earth observation problems across resolutions. To appear in Nature Communications Earth & Environment

Outlook and future work

Interdisciplinary Approaches



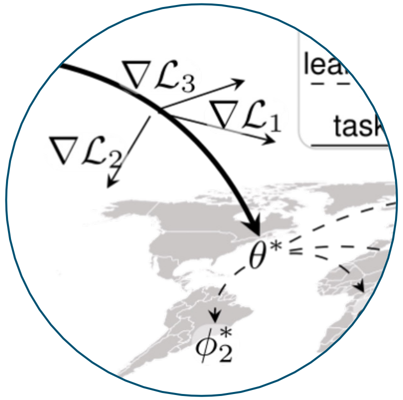
Researching towards these **main challenges** in future

1: timely in-season classification

selective prediction
early time series classification

2: transfer learning across regions

e.g., though
model-agnostic meta-learning



Explore these directions systematically in future

Thank you!

My Mission:

to **envision** and **build intelligent systems** that help address **relevant** and **meaningful** Earth observation **problems**

Rußwurm, M., Courty, N., Emonet, R., Lefèvre, S., Tuia, D., & Tavenard, R. (2023). End-to-end learned early classification of time series for in-season crop type mapping. *ISPRS Journal of Photogrammetry and Remote Sensing*, *196*, 445-456.
https://www.sciencedirect.com/science/article/pii/S092427162200332X

Source code:
http://github.com/marccoru/elects

Contact: marc.russwurm@wur.nl