Early Classification for Agricultural Monitoring

from Satellite Time Series

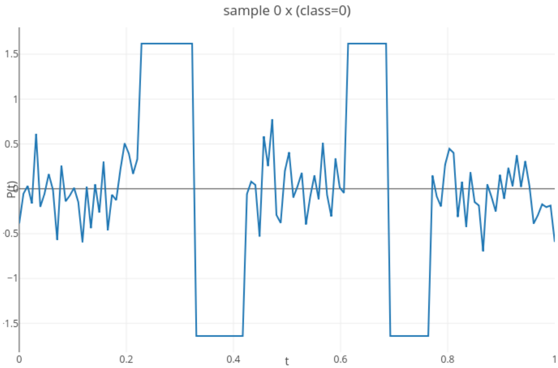
Marc Rußwurm, Romain Tavenard, Sébastien Lefèvre, Marco Körner

July 12, 2019

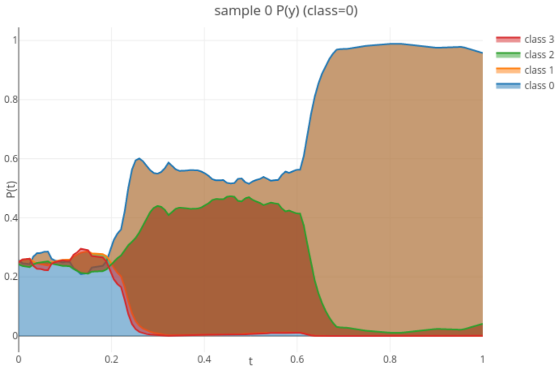
Early Classification

Class Predictions

inputs $\boldsymbol{x}_t$

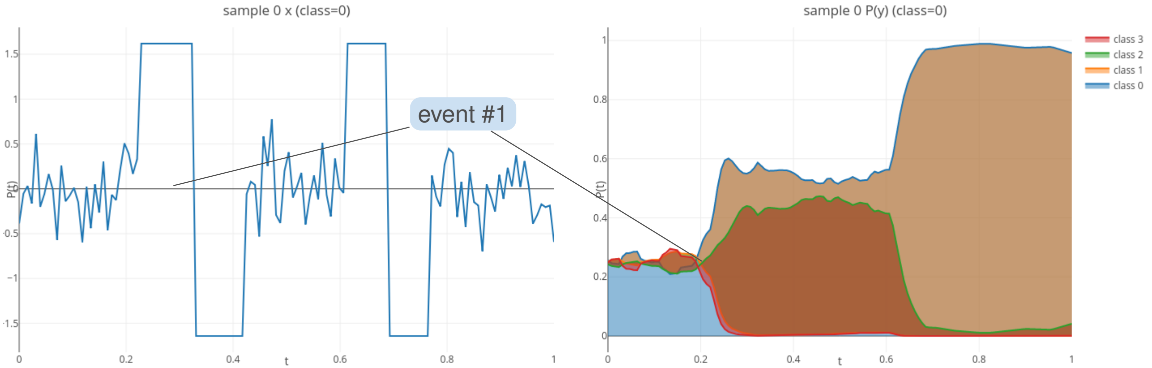


softmaxed class scores $\hat{\boldsymbol{y}}_t$

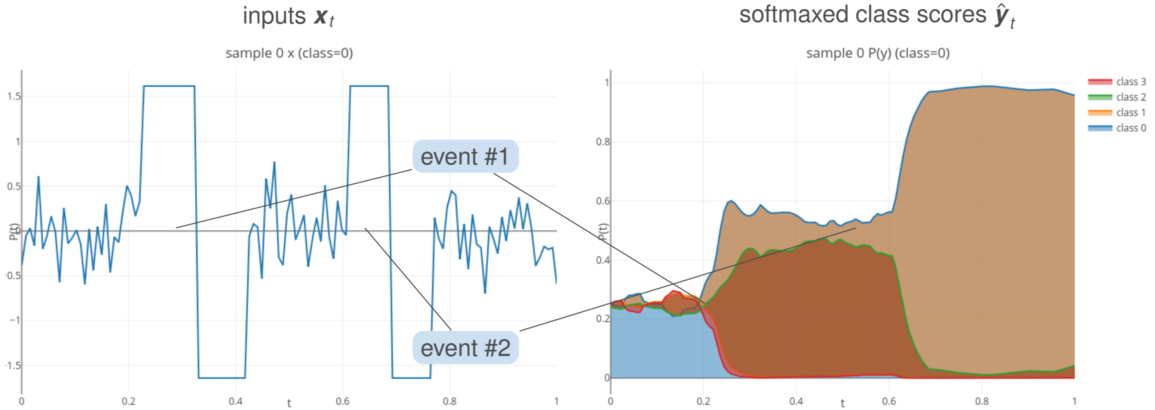


Class Predictions

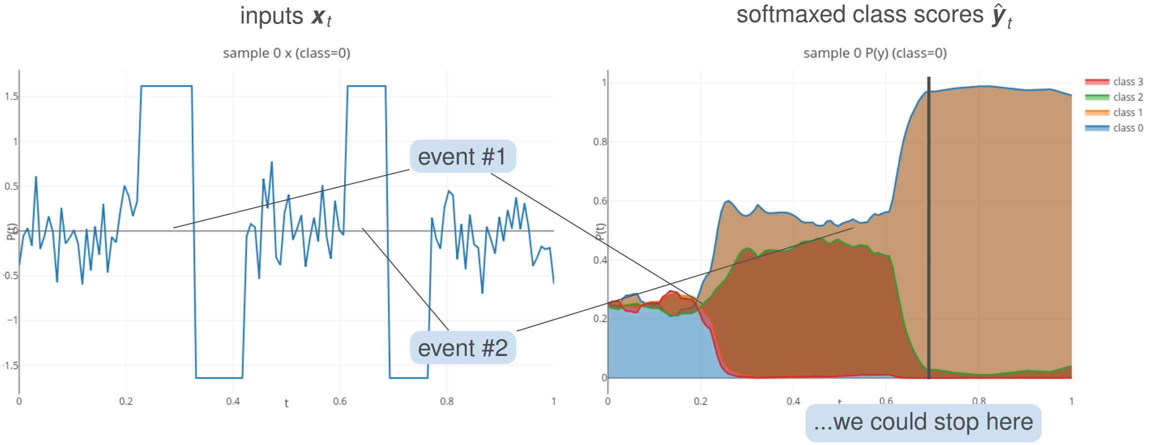
inputs $\boldsymbol{x}_t$ softmaxed class scores $\hat{\boldsymbol{y}}_t$



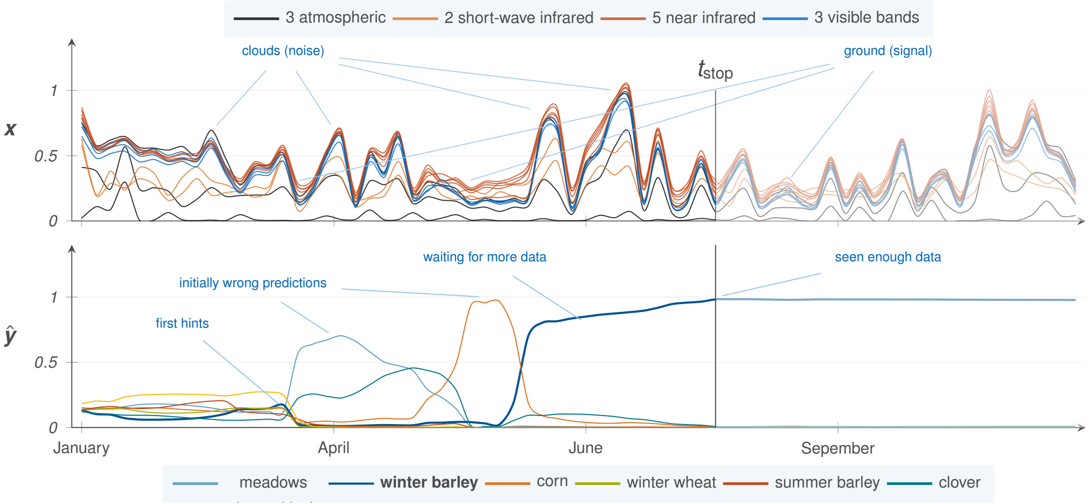
Class Predictions



Class Predictions

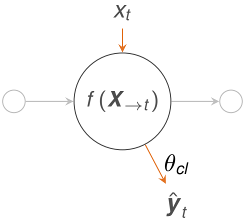


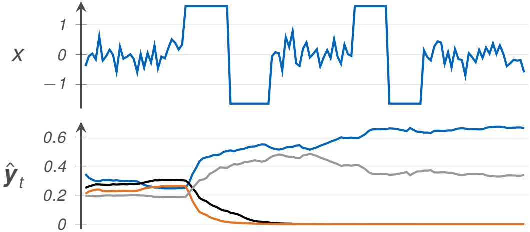
Early Classification on Remote Sensing Data



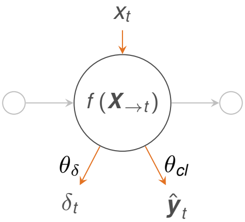
Method

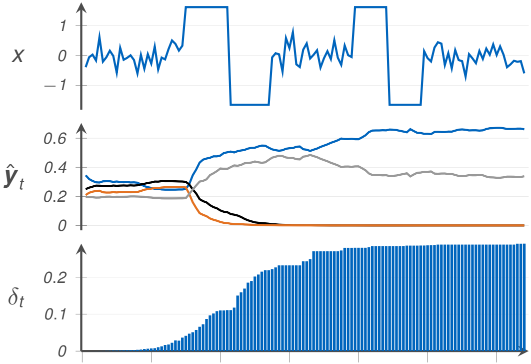
Augmenting Classification Models



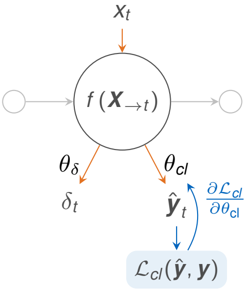


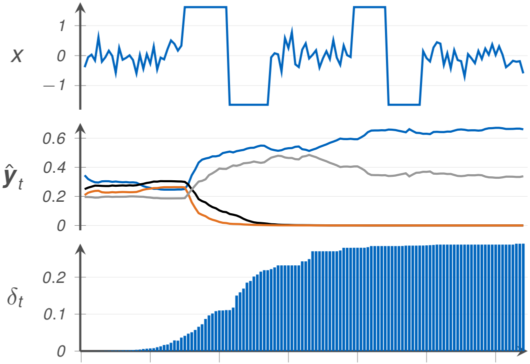
Augmenting Classification Models



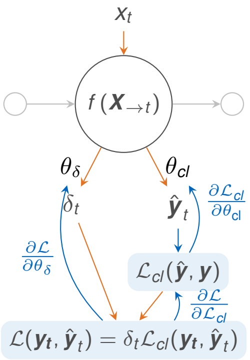


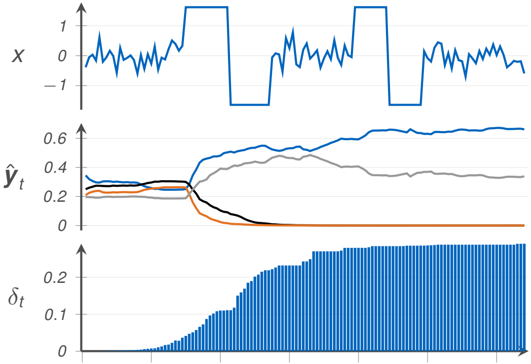
Augmenting Classification Models



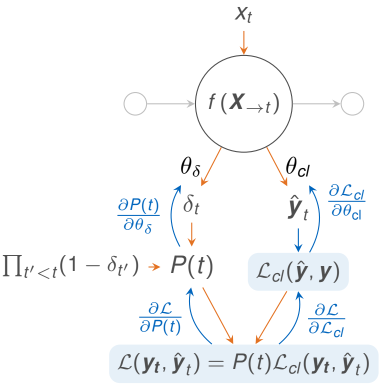


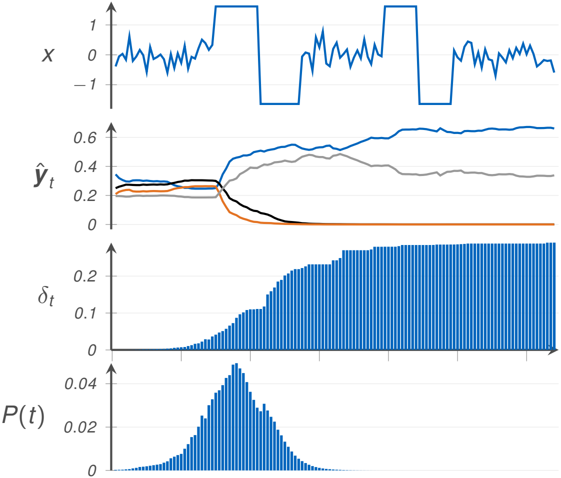
Augmenting Classification Models





Augmenting Classification Models





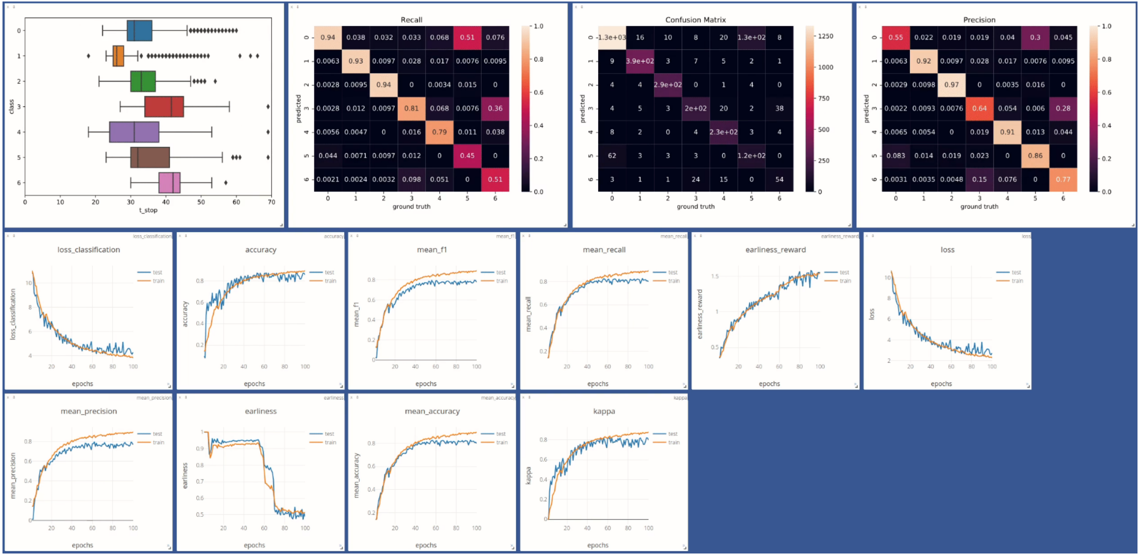
Loss functions

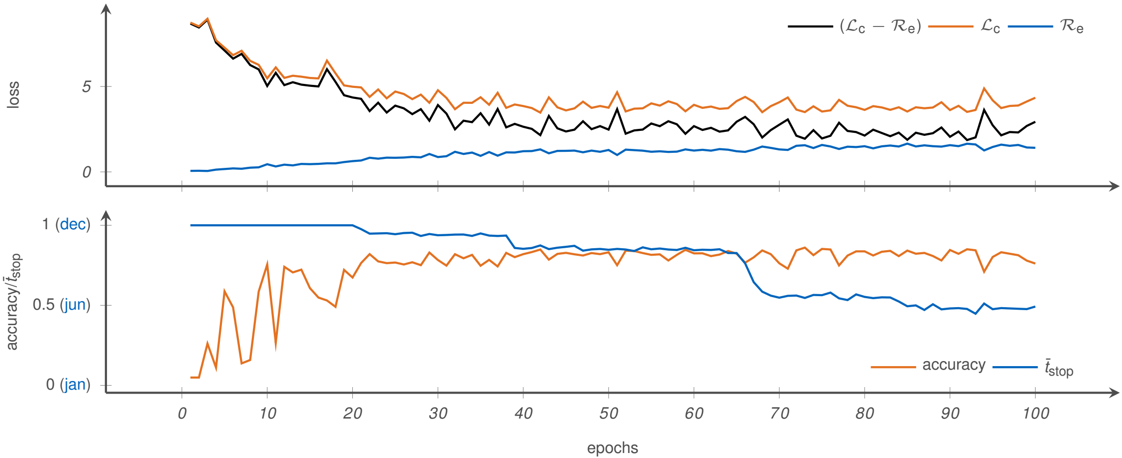
$$\mathcal{L}(\boldsymbol{X}, \boldsymbol{y}) = \sum_{t=1}^{T} P(t) \left(\mathcal{L}_c(\boldsymbol{X}_{\rightarrow t}, \boldsymbol{y}) - \mathcal{R}_e(t, \hat{y}_t^+) \right)$$

earlyness reward: $\mathcal{R}_e(t, \hat{y}_t^+) = \hat{y}_t^+ \left(1 - \frac{t}{T}\right)$ classificaiton loss: $-\log(\hat{y}_t^+)$ (aka. cross entropy)

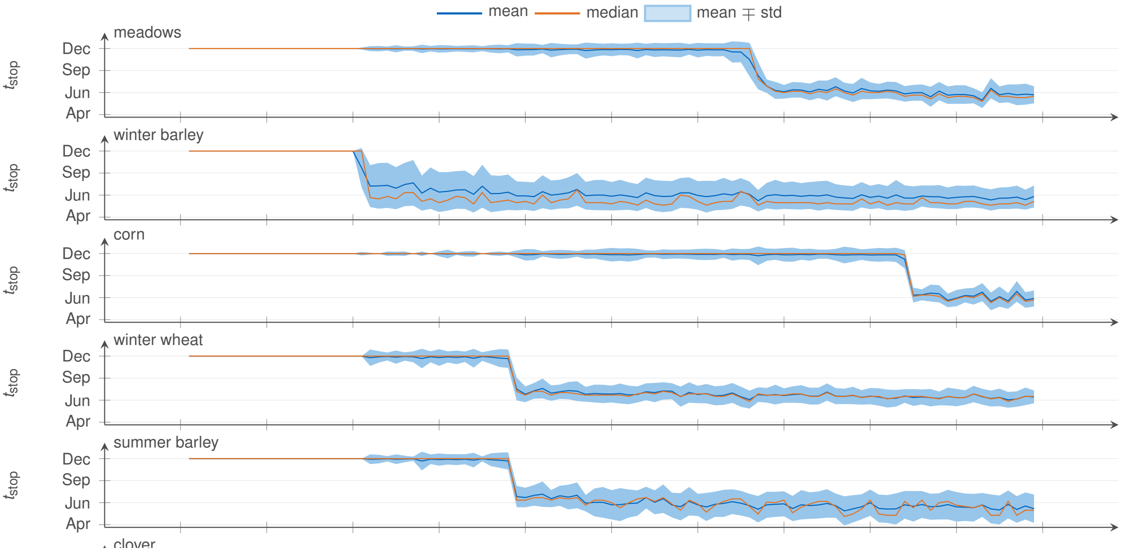
- $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_T)$: entire time series of observations $\boldsymbol{x}$
- $\boldsymbol{X}_{\rightarrow t} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_t)$: time series until time t
- $\boldsymbol{y} \in \{0, 1\}^C$: one-hot vector of the classes
- $\hat{y}^+ \in [0, 1]$: prediction score of the correct class

Results

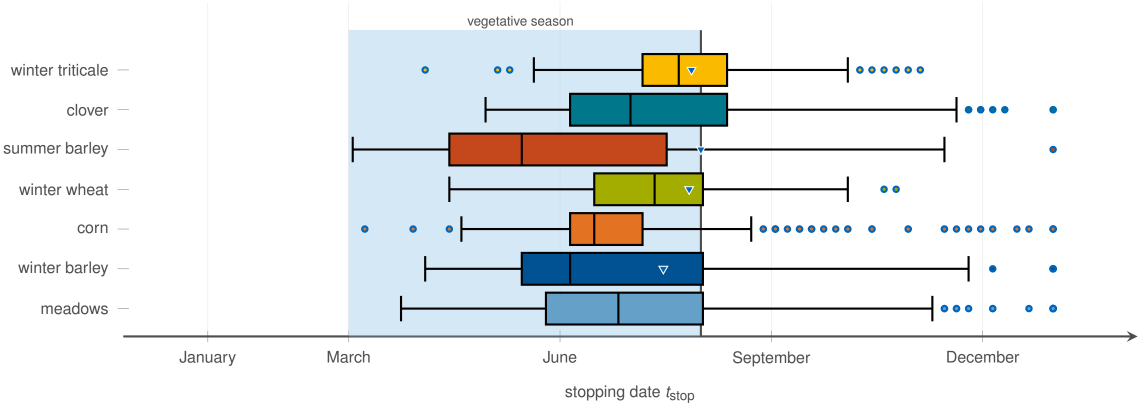

class
t_stop
Recall
predicted
ground truth
0.94 0.038 0.032 0.033 0.068 0.51 0.076
0.0063 0.93 0.0097 0.028 0.017 0.0076 0.0095
0.0028 0.0095 0.94 0 0.0034 0.015 0
0.0028 0.012 0.0097 0.81 0.068 0.0076 0.36
0.0056 0.0047 0 0.016 0.79 0.011 0.038
0.044 0.0071 0.0097 0.012 0 0.45 0
0.0021 0.0024 0.0032 0.098 0.051 0 0.51
Confusion Matrix
predicted
ground truth
1.3e+03 16 10 8 20 1.3e+02 8
9 3.9e+02 3 7 5 2 1
4 4 2.9e+02 0 1 4 0
4 5 3 2e+02 20 2 38
8 2 0 4 2.3e+02 3 4
62 3 3 3 0 1.2e+02 0
3 1 1 24 15 0 54
Precision
predicted
ground truth
0.55 0.022 0.019 0.019 0.04 0.3 0.045
0.0063 0.92 0.0097 0.028 0.017 0.0076 0.0095
0.0029 0.0098 0.97 0 0.0035 0.016 0
0.0022 0.0093 0.0076 0.64 0.054 0.006 0.28
0.0065 0.0054 0 0.019 0.91 0.013 0.044
0.083 0.014 0.019 0.023 0 0.86 0
0.0031 0.0035 0.0048 0.15 0.076 0 0.77
loss_classification
accuracy
mean_f1
mean_recall
earliness_reward
loss
mean_precision
earliness
mean_accuracy
kappa
test
train
epochs



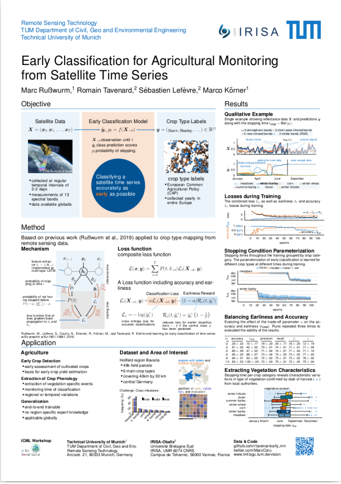
one training phase using the *earliness reward* formulation.


mean
median
mean ∓ std
meadows
winter barley
corn
winter wheat
summer barley
clover
t_{stop}
Dec
Sep
Jun
Apr

Stopping times per crop Class



ICML AI for Social Good Workshop



Early Classification for Agricultural Monitoring from Satellite Time Series

Marc Rußwurm [1] **Romain Tavenard** [2 3] **Sébastien Lefèvre** [4 3] **Marco Körner** [1]

Abstract

In this work, we introduce a recently developed early classification mechanism to satellite-based agricultural monitoring. It augments existing classification models by an additional stopping probability based on the previously seen information. This mechanism is end-to-end trainable and derives its stopping decision solely from the observed satellite data. We show results on field parcels in central Europe where sufficient ground truth data is available for an empiric evaluation of the results with local phenological information obtained from authorities. We observe that the recurrent neural network outfitted with this early classification mechanism was able to distinguish the many of the crop types before the end of the vegetative period. Further, we associated these stopping times with evaluated ground truth information and saw that the times of classification were related to characteristic events of the observed plants' phenology.

1. Introduction

The identification of crop types form space-born imagery forms an important component of agricultural monitoring at a global scale. Information obtained from satellite data is often the only source of information to assess the state national agriculture. This is particularly important in developing countries where agricultural production is rarely monitored or controlled by central authorities. Crucial information on the effect of droughts and shortages are typically available after the end of the vegetative periods which may be too late for effective countermeasures. Obtaining knowledge about the expected crop type as early as possible with space-born crop monitoring can play a significant role in providing food security, preventing famine, and determining policies for sustainable agriculture. Monitoring of vegetation has been an increasing focus of Earth observation programs. NASA's Landsat or ESA's Sentinel satellites observe the surface of the Earth at weekly intervals at spatial resolutions of 10-60m. This allows for regular observations of comparatively small objects, such as field parcels. While enormous quantities of data are collected, methods to extract information, *e.g.*, Torbick et al. (2018); Chen et al. (2018); Rußwurm & Körner (2018), are often applied at later times when the vegetative period has ended to ensure a certain accuracy of the prediction. With the proposed early classification mechanism, we enable a generic multi-temporal classification model to choose the time of classification t_{stop} sampled from a probability p_t. Our results show empirically that the classification stopped when sufficient information has been observed to justify a confident classification, as shown in an example in Fig. 1.

[1]Technical University of Munich [2]LETG-Rennes [3]Institut de Recherche en Informatique et Systèmes Aléatoires (IRISA) [4]Université de Bretagne Sud. Correspondence to: Marc Rußwurm <marc.russwurm@tum.de>.

Appearing at the International Conference on Machine Learning AI for Social Good Workshop, Long Beach, United States, 2019.

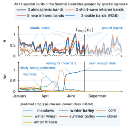


Figure 1: An example of the predictions of the early classification mechanism $\hat{\boldsymbol{y}}, p_t = f(\boldsymbol{X}_{\rightarrow t})$.A binary stopping decision t_{stop} is sampled from the evaluated stopping probability p_t. The upper illustration shows the *Sentinel 2* time series $\boldsymbol{x}$ based on which the class probabilities $\hat{\boldsymbol{y}}$ and the stopping time t_{stop} are produced.

Crop Type Data

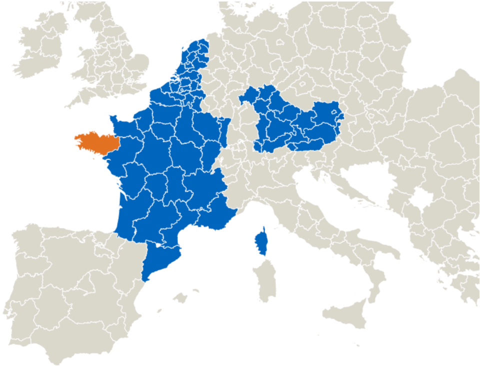
Building Large-Scale Crop Type Mapping Datasets

collected yearly within European **Common Agricultural Policy** (CAP)

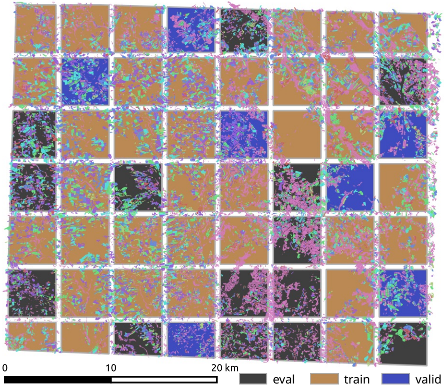
declared by Farmers at **crop subsidy** applications

today slowly made publicly available (on a national basis by French IGN, Bavarian Stmelf, etc.)

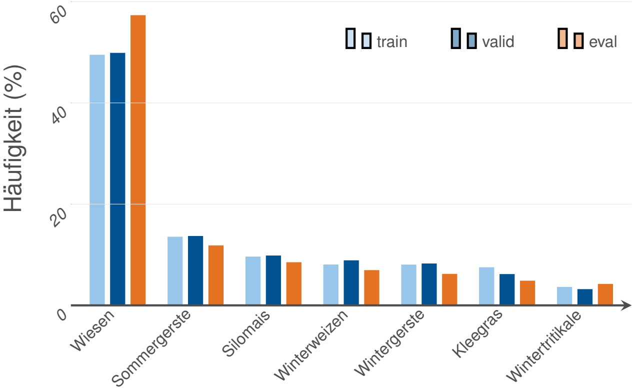
in future further harmonized within **Europe's INSPIRE** directive



Area of Interest for Early Classification



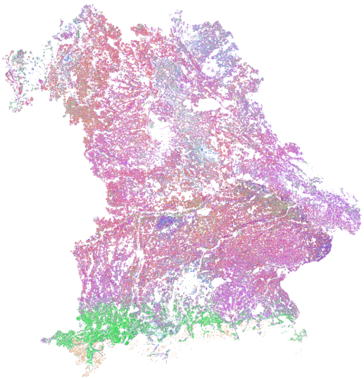
The area of interest and partitioning in blocks of 4.5 km for training validation and evaluation

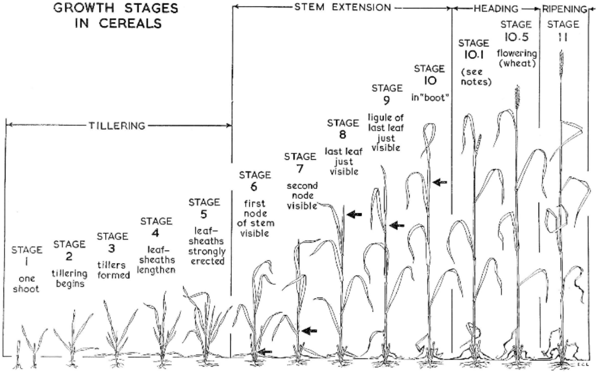


Class distribution in the dataset with block-wise separation of train, valid and evaluation partitions.

Here, we hand-selected 7 classes from a small region of interest...

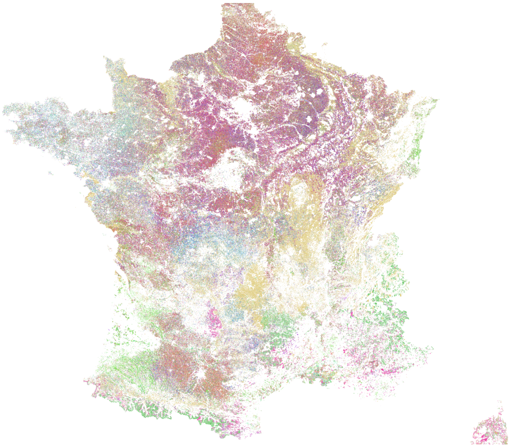
Large Scale Regional Variations





Large, E. C. (1954). Growth stages in cereals illustration of the Feekes scale. Plant pathology, 3(4), 128-129.

Large Scale Regional Variations



Questions:

- how do we learn models on these inter-regional scales?
- the same class label will correspond to different representations in the data.

Outlook

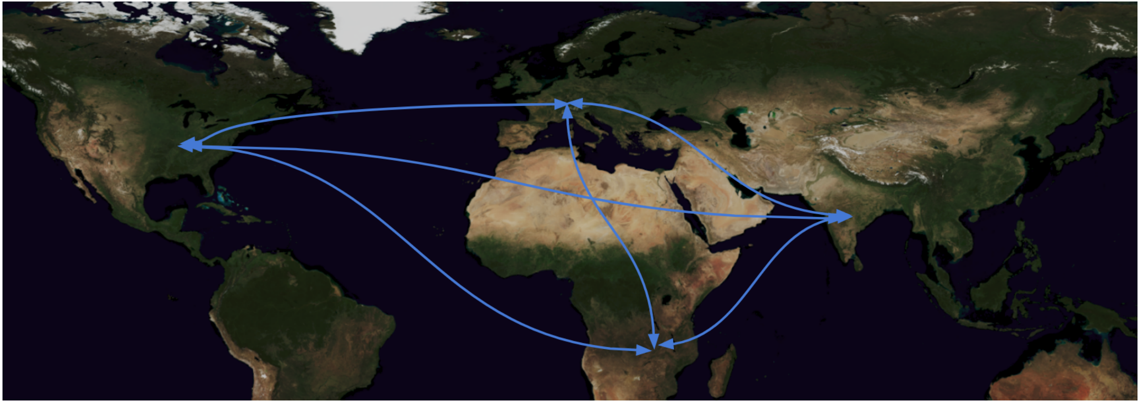
Goals:

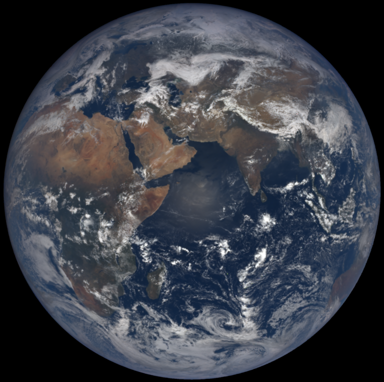
- large-scale domain adaptation between regions of multi-temporal vegetation data
- addressing long-tailed class distributions of $>$300 distinct (overlapping) categories with 90% of data in $<$20 classes

Short-Term Objective:

- compile a large-scale inter-regional crop type mapping dataset to be able to evaluate these questions

The Goal

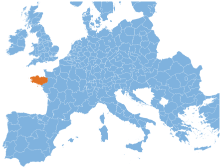


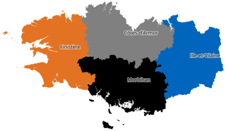


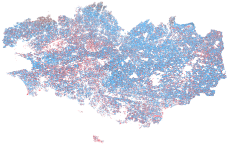
Thank you!

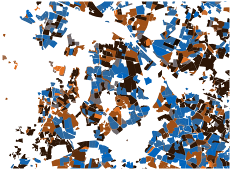
backup slides...

BreizhCrops Dataset (ICML Workshop Submission)

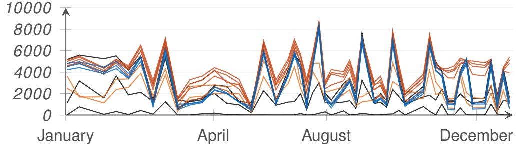




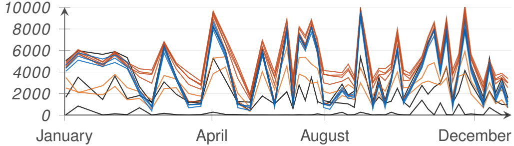




corn grain and silage



temporary meadows



580k samples of Time Series ***X*** and labels ***y***.

`https://github.com/TUM-LMF/BreizhCrops`

Rußwurm M., Lefèvre S., and Körner M (2019). ***BreizhCrops: A Satellite Time Series Dataset for Crop Type Identification****. ICML 2019 Time Series Workshop. arXiv:1905.11893*

Two Baseline Models

Inspired by Models used in NLP, we implemented a **multi-layer LSTM** and a **(minified) Transformer encoder**.

baseline	accuracy	κ	mean f1	mean precision	mean recall
Transformer (Vaswani et al., 2017)	**0.69**	**0.63**	0.57	0.60	0.56
LSTM (Hochreiter and Schmidhuber, 1997)	0.68	0.62	**0.59**	**0.63**	**0.58**

Takeaway:

models perform quite similar

potential for improvement

well-defined classes accurately classified

broadly defined classes systematically confused

Challenges and Impact

Impact

- large scale **real-world dataset**
- effectively **unlimited data** (spatially and temporally)
- **assessing generalization** over large regions
- extraction for further **vegetation characteristics** in future work (drought indicator, early classification, crop yield)

Challenges and Impact

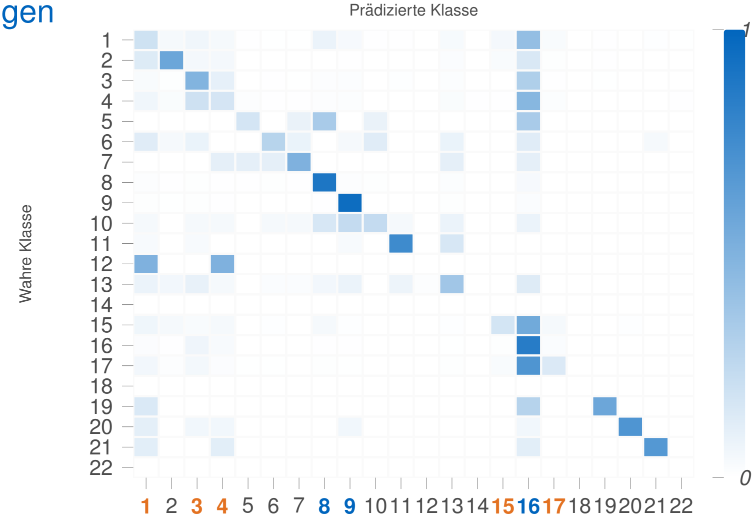
Impact

- large scale **real-world dataset**
- effectively **unlimited data** (spatially and temporally)
- **assessing generalization** over large regions
- extraction for further **vegetation characteristics** in future work (drought indicator, early classification, crop yield)

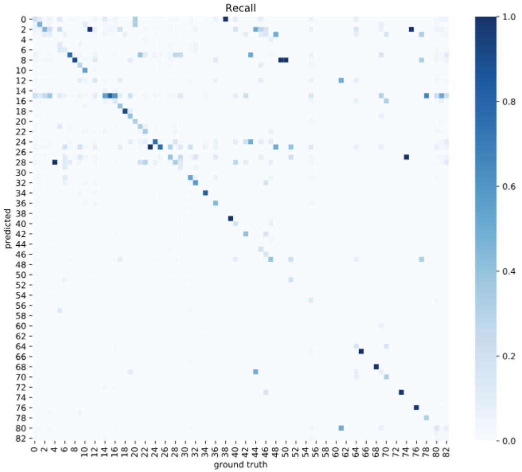
Challenges

- **imbalanced** class **labels** (¿300 raw classes)
- classes with **similar characteristics**
- non-Gaussian noise induced by **clouds**
- **regional variations** in the class distributions
- **spatial autocorrelation**
- **irregular** temporal **sampling** distance
- **variable sequence** length

Verwechslungen



83 Finale Kategorien



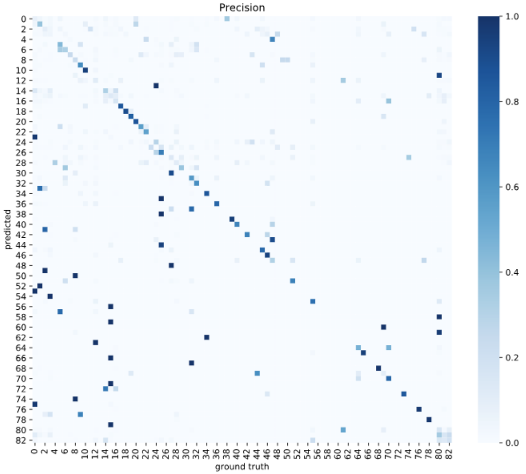


Table: Varying the weighting factor α for the two loss formulations *early reward* (**??**) and *lateness penalty* (**??**).

α	accuracy	$\bar{t}_{stop}$	precision	recall	f_1	κ
.0	.24 ± .29	.00 ± .00	.04 ± .04	.15 ± .00	.05 ± .05	.00 ± .00
.2	.15 ± .08	.00 ± .00	.10 ± .05	.15 ± .01	.08 ± .03	.01 ± .01
.4	.79 ± .05	.39 ± .03	.67 ± .05	.71 ± .04	.68 ± .04	.69 ± .06
.6	.83 ± .03	.59 ± .17	.71 ± .03	.74 ± .02	.72 ± .02	.74 ± .04
.8	.86 ± .01	.86 ± .12	.74 ± .01	.76 ± .02	.75 ± .02	.79 ± .01
1.0	.86 ± .01	1.00 ± .00	.75 ± .02	.76 ± .02	.75 ± .02	.79 ± .02

(a) two-phase *lateness penalty* loss formulation

α	accuracy	$\bar{t}_{stop}$	precision	recall	f_1	κ
.0	.25 ± .22	.10 ± .17	.19 ± .20	.25 ± .17	.16 ± .20	.12 ± .19
.2	.81 ± .03	.40 ± .02	.70 ± .01	.74 ± .01	.71 ± .01	.71 ± .04
.4	.80 ± .09	.47 ± .03	.71 ± .02	.74 ± .01	.71 ± .02	.71 ± .10
.6	.85 ± .02	.88 ± .07	.73 ± .04	.74 ± .03	.73 ± .03	.77 ± .03
.8	.84 ± .01	.93 ± .05	.72 ± .02	.75 ± .01	.73 ± .02	.76 ± .02
1.0	.83 ± .03	1.00 ± .00	.72 ± .03	.75 ± .01	.72 ± .03	.75 ± .04

(b) one phase *early-reward* formulation

Table: Quantitative analysis of the ε parameter on different trade-off factors between α earliness and accuracy. The illustrated figures show the mean and standard deviation of three runs with same parameters, but different initial random initialization.

ε	accuracy	$\bar{t}_{stop}$	f_1	precision	recall	κ
0	$.10 \pm .02$	$.02 \pm .00$	$.07 \pm .01$	$.13 \pm .06$	$.17 \pm .00$	$.02 \pm .00$
1	$.75 \pm .09$	$.44 \pm .06$	$.65 \pm .05$	$.64 \pm .03$	$.69 \pm .03$	$.64 \pm .10$
10	$.81 \pm .03$	$.40 \pm .02$	$.71 \pm .01$	$.70 \pm .01$	$.74 \pm .01$	$.71 \pm .04$

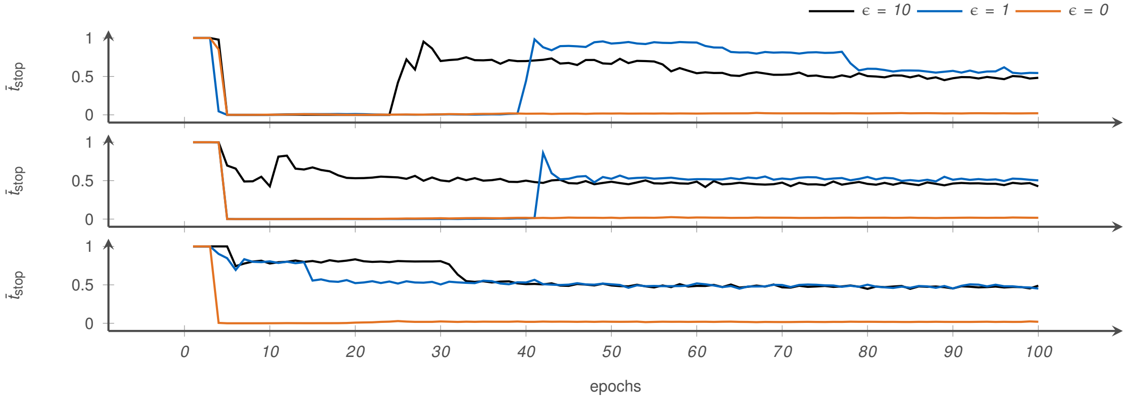
(a) $\alpha = 0.2$

ε	accuracy	$\bar{t}_{stop}$	f_1	precision	recall	κ
0	$.21 \pm .20$	$.02 \pm .00$	$.09 \pm .04$	$.18 \pm .03$	$.16 \pm .01$	$.04 \pm .03$
1	$.80 \pm .02$	$.50 \pm .05$	$.68 \pm .05$	$.67 \pm .05$	$.70 \pm .07$	$.70 \pm .03$
10	$.80 \pm .09$	$.47 \pm .03$	$.71 \pm .02$	$.71 \pm .02$	$.74 \pm .01$	$.71 \pm .10$

(b) $\alpha = 0.4$

ε	accuracy	$\bar{t}_{stop}$	f_1	precision	recall	κ
0	$.13 \pm .04$	$.02 \pm .00$	$.08 \pm .01$	$.16 \pm .01$	$.16 \pm .01$	$.02 \pm .01$
1	$.80 \pm .05$	$.85 \pm .14$	$.71 \pm .02$	$.70 \pm .02$	$.74 \pm .01$	$.70 \pm .06$
10	$.85 \pm .02$	$.88 \pm .07$	$.73 \pm .03$	$.73 \pm .04$	$.74 \pm .03$	$.77 \pm .03$

(c) $\alpha = 0.6$



The three runs with different random initialization per ε offset and $\alpha = 0.4$ of Table 4b. The $\varepsilon > 0$ offset factor allows the models to recover from a too early classification, as is visible in the top two plots.