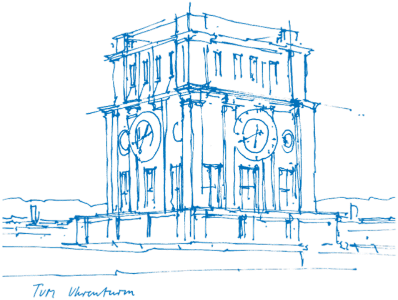
BreizhCrops

A Satellite Time Series Dataset for Crop Type Identification

Marc Rußwurm,[1] Sébastien Lefèvre,[2] Marco Körner[1]

Technical Univ. of Munich, Remote Sensing Technology [1]
IRISA-Obelix, University of South Brittany[2]

June 14th 2019, Time Series Workshop ICML 2019



We challenge the Time Series Community

with a real-world time series dataset $\mathcal{D} = (\boldsymbol{X}, \boldsymbol{y})^i$

with a real-world time series dataset $\mathcal{D} = (\boldsymbol{X}, \boldsymbol{y})^i$

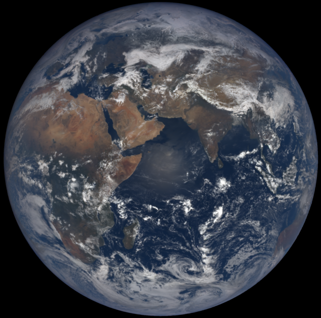
of satellite observations $\boldsymbol{X} = (\boldsymbol{x}_1^T, \boldsymbol{x}_2^T, \ldots, \boldsymbol{x}_N^T)$

with a real-world time series dataset $\mathcal{D} = (\boldsymbol{X}, \boldsymbol{y})^i$

of satellite observations $\boldsymbol{X} = (\boldsymbol{x}_1^T, \boldsymbol{x}_2^T, \ldots, \boldsymbol{x}_N^T)$

vegetation classes $\boldsymbol{y} = (y_{\text{corn}}, y_{\text{meadow}}, \ldots)$

Earth Observation

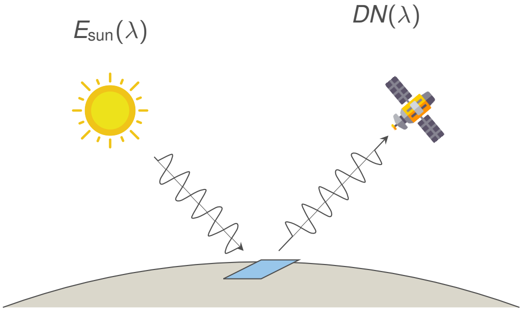


Optical Satellites

Sensor measures **Digital Numbers** DN(λ) for each wavelength λ.

Digital Numbers are normalized to **Radiance** $L(\lambda)$, $\left[\frac{W}{srm^2}\right]$ by gain and offset calibration.

Radiance is normalized to **top-of-atmosphere reflectance** $\rho(\lambda)$

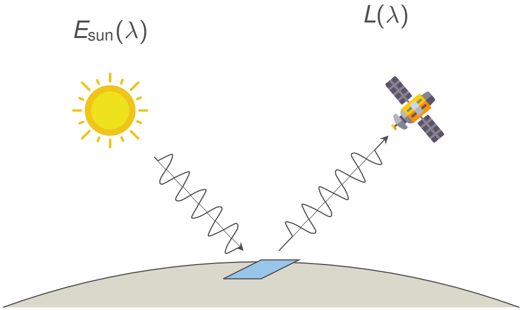


Optical Satellites

Sensor measures **Digital Numbers** DN(λ) for each wavelength λ.

Digital Numbers are normalized to **Radiance** $L(\lambda)$, $\left[\frac{W}{sr m^2}\right]$ by gain and offset calibration.

Radiance is normalized to **top-of-atmosphere reflectance** $\rho(\lambda)$

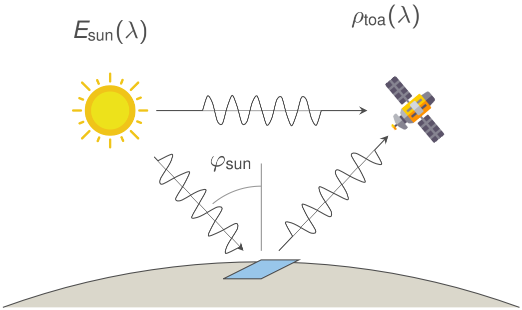


Optical Satellites

Sensor measures **Digital Numbers** DN(λ) for each wavelength λ.

Digital Numbers are normalized to **Radiance** $L(\lambda)$, $\left[\frac{W}{sr m^2}\right]$ by gain and offset calibration.

Radiance is normalized to **top-of-atmosphere reflectance** $\rho(\lambda)$



Acquired in regular time intervals

Sentinel 2 Satellite

- polar sun-synchronous orbit
- single orbit circa 100 minutes
- revisit same location after 5 days
- acquisition stripe of 290km width
- 13 spectral bands
- ground resolution 10-60m
- global coverage and free of charge

https://www.esa.int/spaceinvideos/Videos/2016/08/Sentinel-2_global_coverage

Acquired in regular time intervals

Sentinel 2 Satellite

- polar sun-synchronous orbit
- single orbit circa 100 minutes
- revisit same location after 5 days
- acquisition stripe of 290km width
- 13 spectral bands
- ground resolution 10-60m
- global coverage and free of charge

https://www.esa.int/spaceinvideos/Videos/2016/08/Sentinel-2_global_coverage

Acquired in regular time intervals

Sentinel 2 Satellite

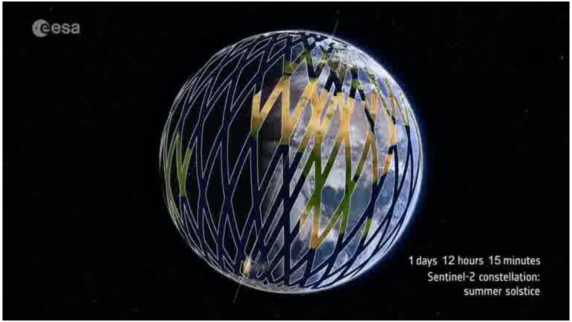
- polar sun-synchronous orbit
- single orbit circa 100 minutes
- revisit same location after 5 days
- acquisition stripe of 290km width
- 13 spectral bands
- ground resolution 10-60m
- global coverage and free of charge

https://www.esa.int/spaceinvideos/Videos/2016/08/Sentinel-2_global_coverage

Acquired in regular time intervals

Sentinel 2 Satellite

- polar sun-synchronous orbit
- single orbit circa 100 minutes
- revisit same location after 5 days
- acquisition stripe of 290km width
- 13 spectral bands
- ground resolution 10-60m
- global coverage and free of charge

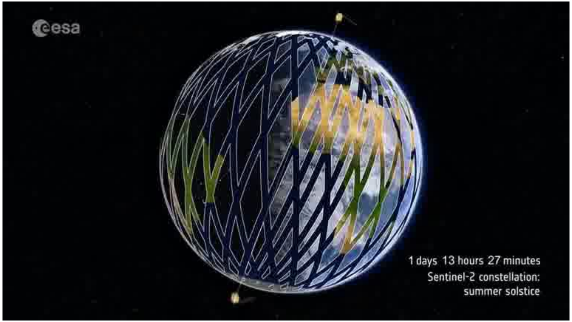


https://www.esa.int/spaceinvideos/Videos/2016/08/Sentinel-2_global_coverage

Acquired in regular time intervals

Sentinel 2 Satellite

- polar sun-synchronous orbit
- single orbit circa 100 minutes
- revisit same location after 5 days
- acquisition stripe of 290km width
- 13 spectral bands
- ground resolution 10-60m
- global coverage and free of charge

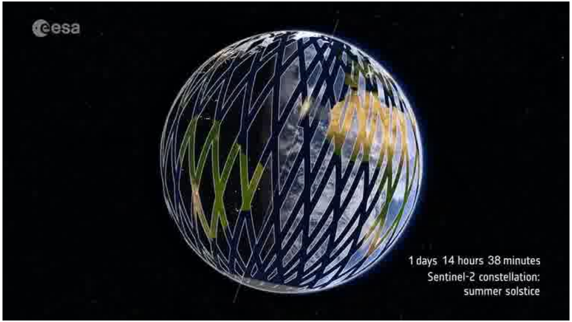


https://www.esa.int/spaceinvideos/Videos/2016/08/Sentinel-2_global_coverage

Acquired in regular time intervals

Sentinel 2 Satellite

- polar sun-synchronous orbit
- single orbit circa 100 minutes
- revisit same location after 5 days
- acquisition stripe of 290km width
- 13 spectral bands
- ground resolution 10-60m
- global coverage and free of charge

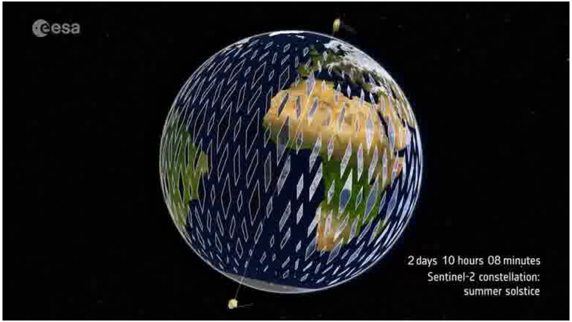


https://www.esa.int/spaceinvideos/Videos/2016/08/Sentinel-2_global_coverage

Acquired in regular time intervals

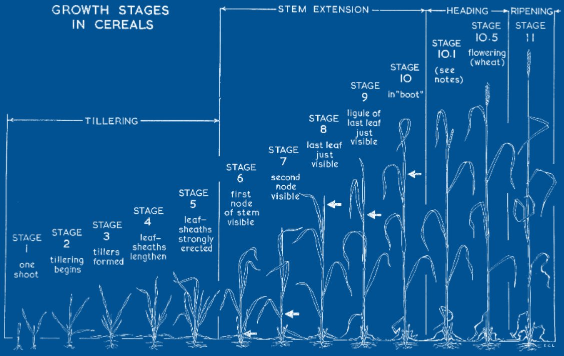
Sentinel 2 Satellite

- polar sun-synchronous orbit
- single orbit circa 100 minutes
- revisit same location after 5 days
- acquisition stripe of 290km width
- 13 spectral bands
- ground resolution 10-60m
- global coverage and free of charge



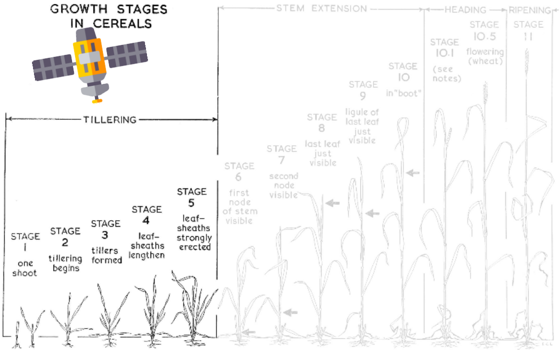
https://www.esa.int/spaceinvideos/Videos/2016/08/Sentinel-2_global_coverage

Vegetation Modeling

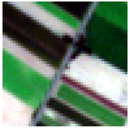


(Large et al., 1954)

Multi-temporal Vegetation Modeling

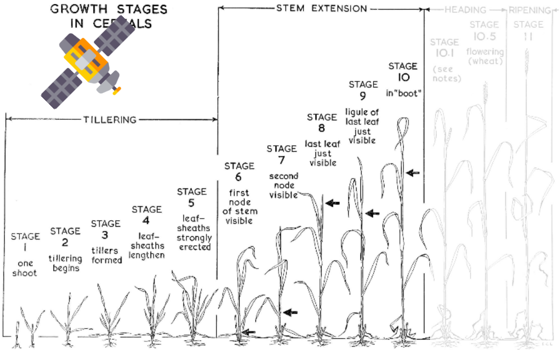


$$f_{\text{vegetation}}(\boldsymbol{x}_t)$$

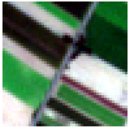


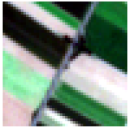
Large, E. C. (1954). Growth stages in cereals illustration of the Feekes scale. Plant pathology, 3(4), 128-129.

Multi-temporal Vegetation Modeling



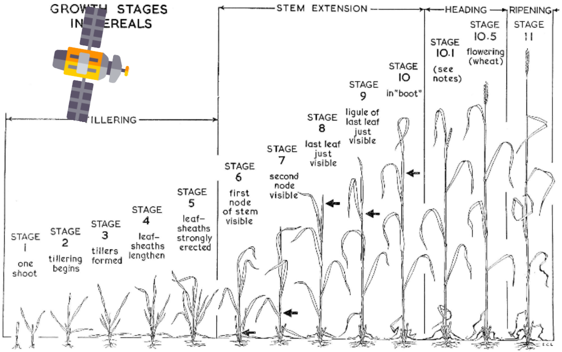
$$f_{\text{vegetation}}(\boldsymbol{x}_t, \boldsymbol{x}_{t+1})$$



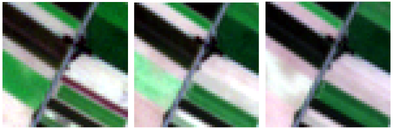


Large, E. C. (1954). Growth stages in cereals illustration of the Feekes scale. Plant pathology, 3(4), 128-129.

Multi-temporal Vegetation Modeling



$$f_{\text{vegetation}}(\boldsymbol{x}_t, \boldsymbol{x}_{t+1}, \boldsymbol{x}_{t+2})$$



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

Problem Definition

$$f_{\text{vegetation}} \quad (\boldsymbol{x}_t, \boldsymbol{x}_{t+1}, \boldsymbol{x}_{t+2})$$

Problem: **unsupervised learning** of a vegetation model **is difficult**

Solution: re-framing as **supervised classification** of crop type labels

Intuition: A **supervised classification model** must **internalize** a learned **discriminative model** for the **vegetation**

Problem Definition

$$\boldsymbol{y} = f_{\text{vegetation}} \left(\boldsymbol{x}_t, \boldsymbol{x}_{t+1}, \boldsymbol{x}_{t+2}\right)$$

Problem: **unsupervised learning** of a vegetation model **is difficult**

Solution: re-framing as **supervised classification** of crop type labels

Intuition: A **supervised classification model** must **internalize** a learned **discriminative model** for the **vegetation**

Problem Definition

$$\boldsymbol{y} = f_{\text{vegetation}} \left(\boldsymbol{x}_t, \boldsymbol{x}_{t+1}, \boldsymbol{x}_{t+2}\right)$$

Problem: **unsupervised learning** of a vegetation model **is difficult**

Solution: re-framing as **supervised classification** of crop type labels

Intuition: A **supervised classification model** must **internalize** a learned **discriminative model** for the **vegetation**

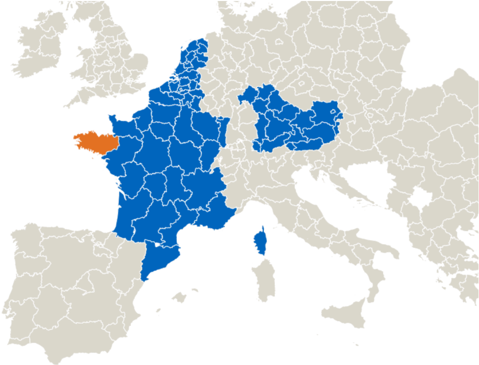
Crop Type Labels in Europe

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

declared by Farmers at **crop subsidy** applications

today slowly made publicly available (on a national basis)

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



The Dataset

Field Parcel Coverage in Brittany

Organization

Nomenclature des unités territoriales statistiques (NUTS) as European standard of administrative Boundaries.

NUTS-0 countries

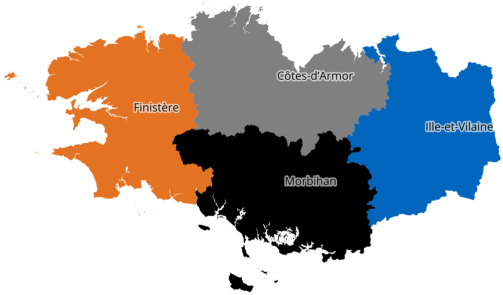
NUTS-1 states

NUTS-2 districts

NUTS-3 municipalities

we suggest a spatially distinct train/test split via these partitions.

Bonus: Eurostat statistics are gathered on NUTS boundaries.



NUTS-3 regions within Brittany, France

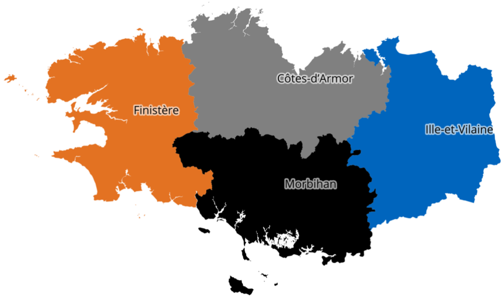
Organization

Field Parcels

Departements	NUTS-3	Parcels	Size
Morbihan	FRH04	158522	4.3 Gb
Côtes-d'Armor	FRH01	221095	6.7 Gb
Finistère	FRH02	180565	6.2 Gb
Ille-et-Vilaine	FRH03	207993	6.8 Gb
Brittany	FRH0	768175	

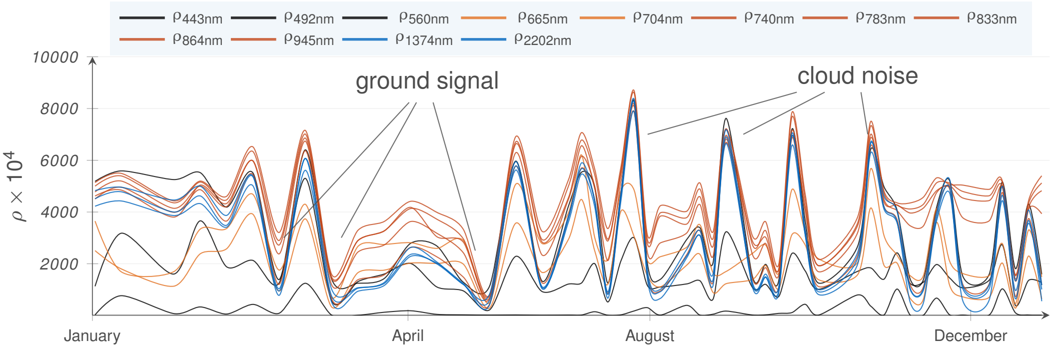
we suggest a spatially distinct train/test split via these partitions.

Bonus: Eurostat statistics are gathered on NUTS boundaries.



NUTS-3 regions within Brittany, France

Corn Example



Baseline Results

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

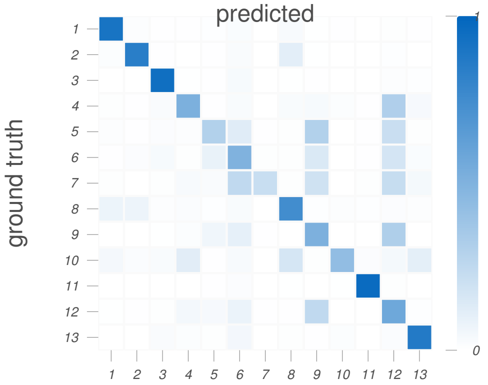
well-defined classes accurately classified

broadly defined classes systematically confused

Results

LSTM Model

#	crop type	prec.	rec.	f_1	#samples
1	barley	90	86	88	4982
2	wheat	83	95	89	13850
3	corn	93	**96**	94	25059
4	fodder	51	34	41	3449
5	fallow	30	2	4	3863
6	misc.	50	49	49	12499
7	orchards	21	7	10	391
8	cereals	74	47	57	4645
9	perm. meadows	51	47	49	20966
10	protein crops	42	61	50	498
11	rapeseed	**96**	94	**95**	2664
12	temp. meadows	56	68	62	29977
13	vegetables	86	69	76	3114
		63	**58**	**59**	125957



Summary

Summary

- we gathered, compiled, harmonized a **large supervised classification dataset** (20 Gb of data) for crop type mapping
- we **prove the feasibility** of classification with two deep neural network classifiers (LSTM and Transformer)
- baslines **leave potential for improvement** by future research

Challenges

- **Imbalanced** class **labels**
- 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

Summary

Summary

we gathered, compiled, harmonized a **large supervised classification dataset** (20 Gb of data) for crop type mapping

we **prove the feasibility** of classification with two deep neural network classifiers (LSTM and Transformer)

baslines **leave potential for improvement** by future research

Challenges

Imbalanced class **labels**

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

Summary

Summary

- we gathered, compiled, harmonized a **large supervised classification dataset** (20 Gb of data) for crop type mapping
- we **prove the feasibility** of classification with two deep neural network classifiers (LSTM and Transformer)
- baslines **leave potential for improvement** by future research

Challenges

- **Imbalanced** class **labels**
- 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

Summary

Summary

- we gathered, compiled, harmonized a **large supervised classification dataset** (20 Gb of data) for crop type mapping
- we **prove the feasibility** of classification with two deep neural network classifiers (LSTM and Transformer)
- baslines **leave potential for improvement** by future research

Challenges

- **Imbalanced** class **labels**
- 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

Summary

Summary

- we gathered, compiled, harmonized a **large supervised classification dataset** (20 Gb of data) for crop type mapping
- we **prove the feasibility** of classification with two deep neural network classifiers (LSTM and Transformer)
- baslines **leave potential for improvement** by future research

Challenges

- **Imbalanced** class **labels**
- 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

Summary

Summary

- we gathered, compiled, harmonized a **large supervised classification dataset** (20 Gb of data) for crop type mapping
- we **prove the feasibility** of classification with two deep neural network classifiers (LSTM and Transformer)
- baslines **leave potential for improvement** by future research

Challenges

- **Imbalanced** class **labels**
- 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

Summary

Summary

- we gathered, compiled, harmonized a **large supervised classification dataset** (20 Gb of data) for crop type mapping
- we **prove the feasibility** of classification with two deep neural network classifiers (LSTM and Transformer)
- baslines **leave potential for improvement** by future research

Challenges

- **Imbalanced** class **labels**
- 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

Summary

Summary

- we gathered, compiled, harmonized a **large supervised classification dataset** (20 Gb of data) for crop type mapping
- we **prove the feasibility** of classification with two deep neural network classifiers (LSTM and Transformer)
- baslines **leave potential for improvement** by future research

Challenges

- **Imbalanced** class **labels**
- 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

Summary

Summary

- we gathered, compiled, harmonized a **large supervised classification dataset** (20 Gb of data) for crop type mapping
- we **prove the feasibility** of classification with two deep neural network classifiers (LSTM and Transformer)
- baslines **leave potential for improvement** by future research

Challenges

- **Imbalanced** class **labels**
- 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

Summary

Summary

- we gathered, compiled, harmonized a **large supervised classification dataset** (20 Gb of data) for crop type mapping
- we **prove the feasibility** of classification with two deep neural network classifiers (LSTM and Transformer)
- baslines **leave potential for improvement** by future research

Challenges

- **Imbalanced** class **labels**
- 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

Takeaway and Feedback

We hope to have given an **overview over the problem space** of multi-temporal **Earth observation.**

We warmly welcome **feedback and suggestions** as questions or as discussion at the poster

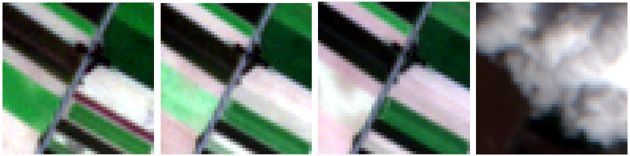
Thank you

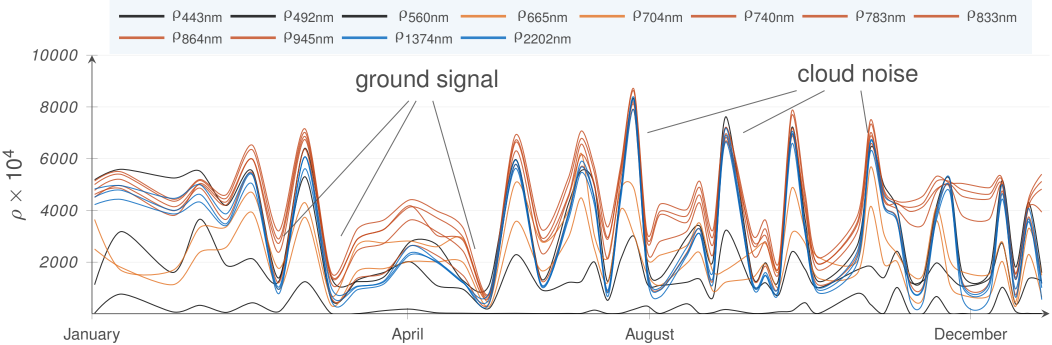
github.com/TUM-LMF/BreizhCrops
lmf.bgu.tum.de/vision
twitter.com/MarcCoru

Backup slides...

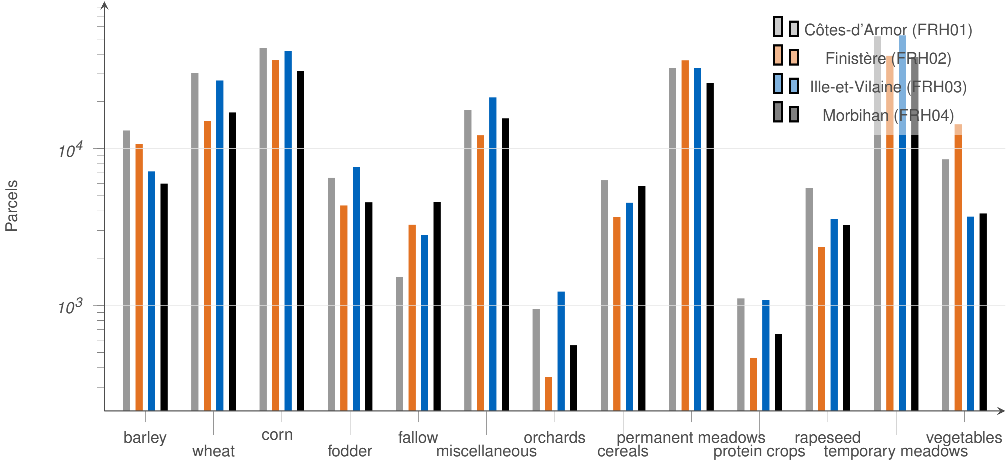
Challenges

Challenge 1: Clouds adding a positive bias





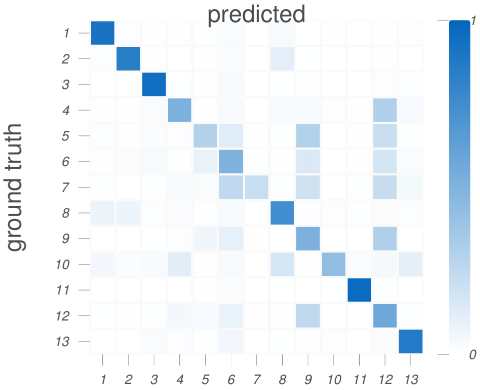
Challenge : Imbalanced class labels



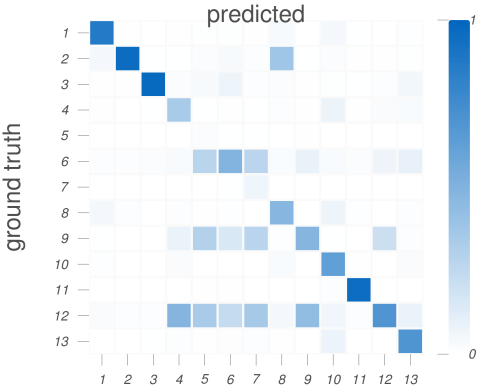
Challenge: Classes with similar characteristics

Multi-Layer RNN baseline

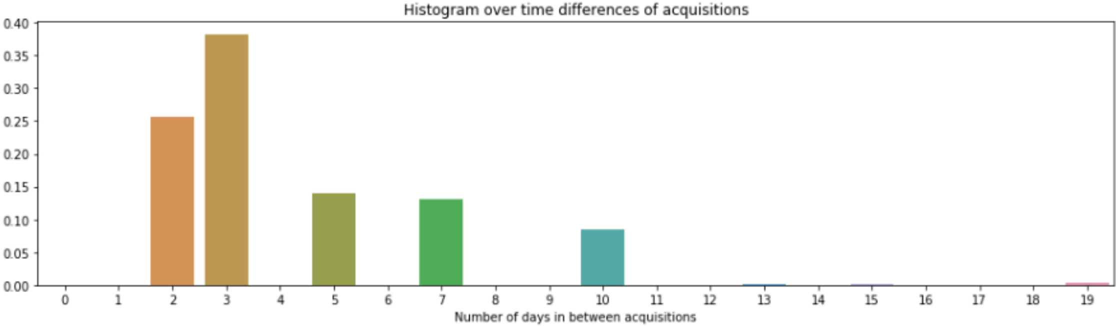
Precision



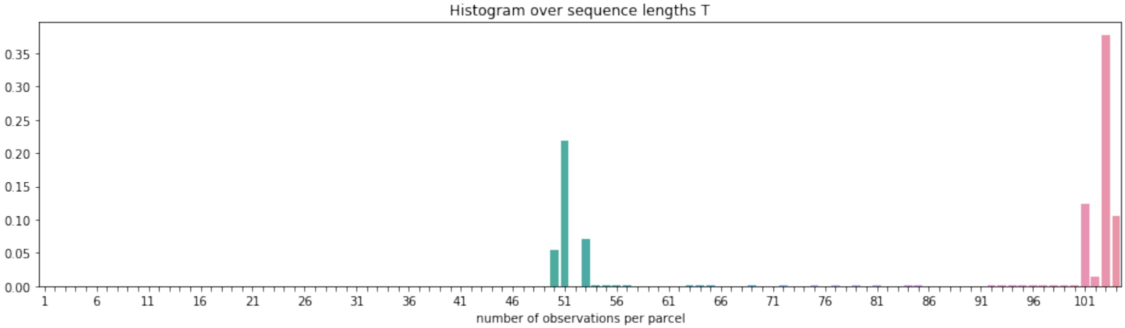
Recall



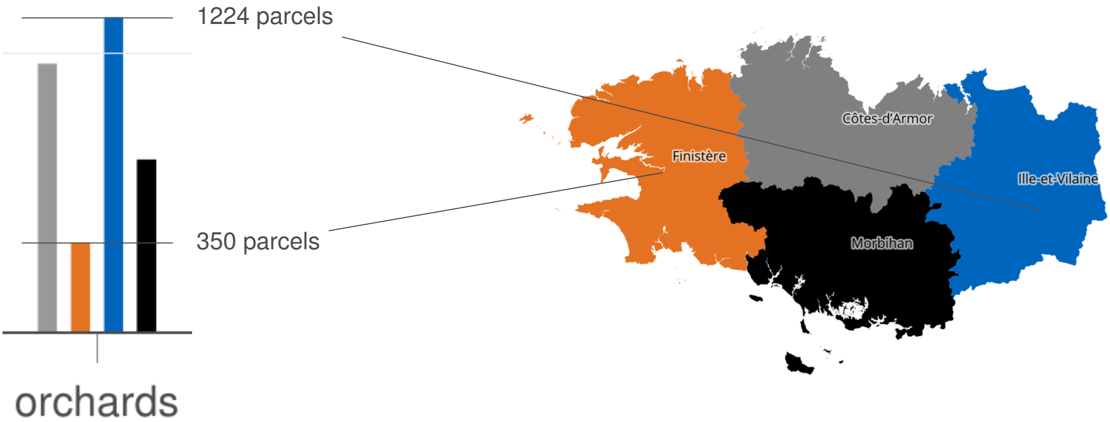
Challenge: Irregular sampling distance.



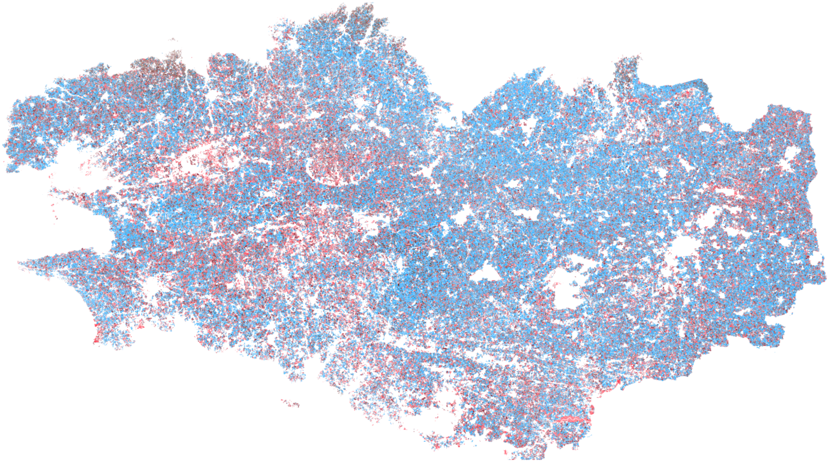
Challenge: Variable sequence length



Challenge: Regional variations in the class distributions

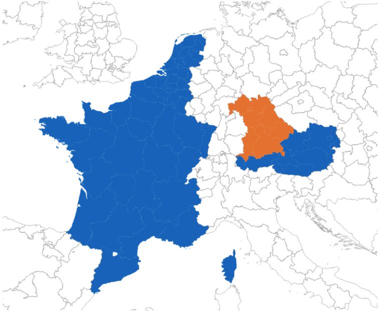


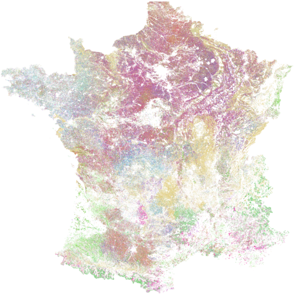
Spatial autocorrelation

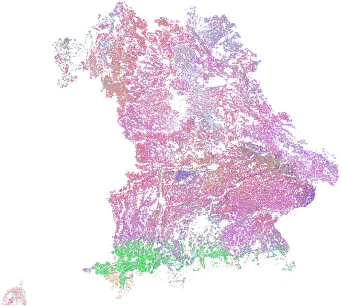


Summary and Outlook

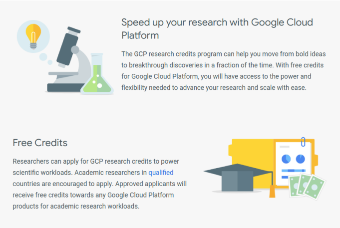
Scaling up







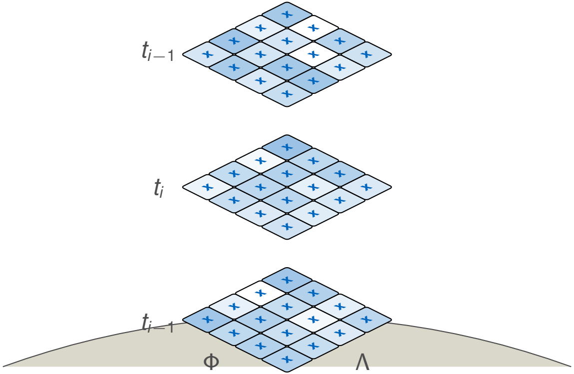
Supported by Google Research Credits



Spatial and Temporal Discretization

$$\mathbf{x}_t = \begin{pmatrix} \rho_{\lambda_1} \\ \rho_{\lambda_2} \\ \dots \\ \rho_{\lambda_n} \end{pmatrix}$$

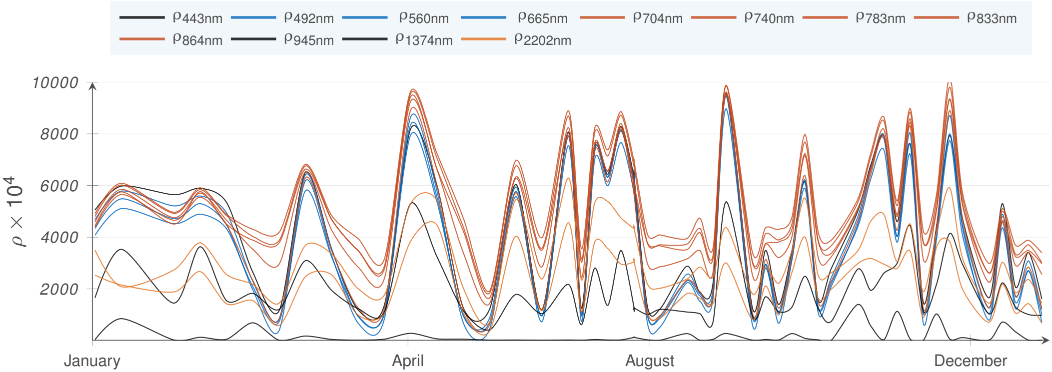
Spectral reflectance of **spectral bands** disctretized on a **spatial grid**. Each grid cell is georeferenced by its Longitude Λ and Latitude Φ. Acquisitions in regular **temporal intervals**.



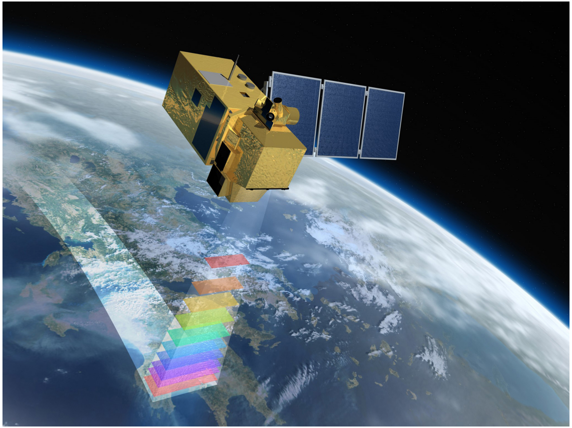
Photosynthesis

$$6CO_2 + 6H_2O \xrightarrow{\text{light absorbtion } \Delta \boldsymbol{x}} C_6H_{12}O_6 + 6O_2$$

Meadow Example



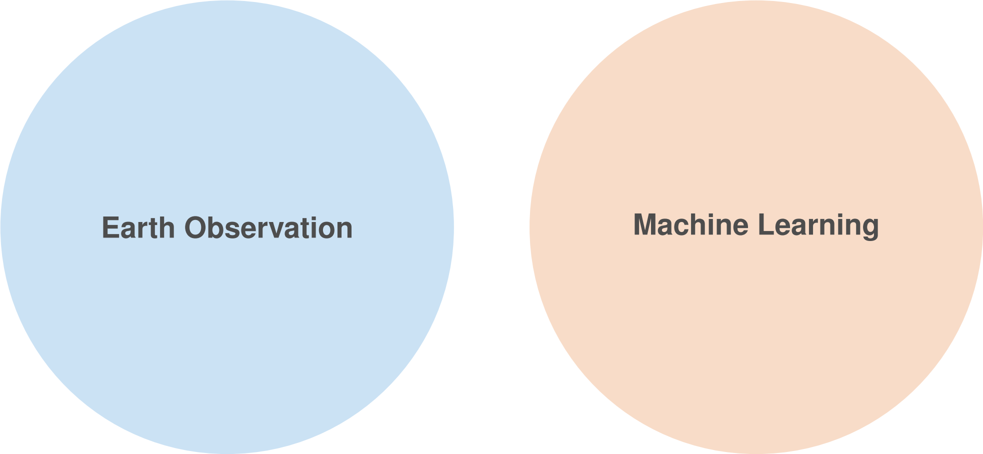
Sentinel 2: Satellite Time Series $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_N)$



$$\boldsymbol{x}_t = \begin{pmatrix} \rho_{UV} \\ \rho_{Blue} \\ \rho_{Green} \\ \rho_{Red} \\ \rho_{NIR} \\ \rho_{NIR} \\ \rho_{NIR} \\ \rho_{NIR} \\ \rho_{NIR} \\ \rho_{SWIR} \\ \rho_{SWIR} \\ \rho_{SWIR} \\ \rho_{SWIR} \end{pmatrix}$$

UV=ultra-violet, NIR=near infra-red, SWIR=short-wave infra-red

Two Research Fields



Common Datasets