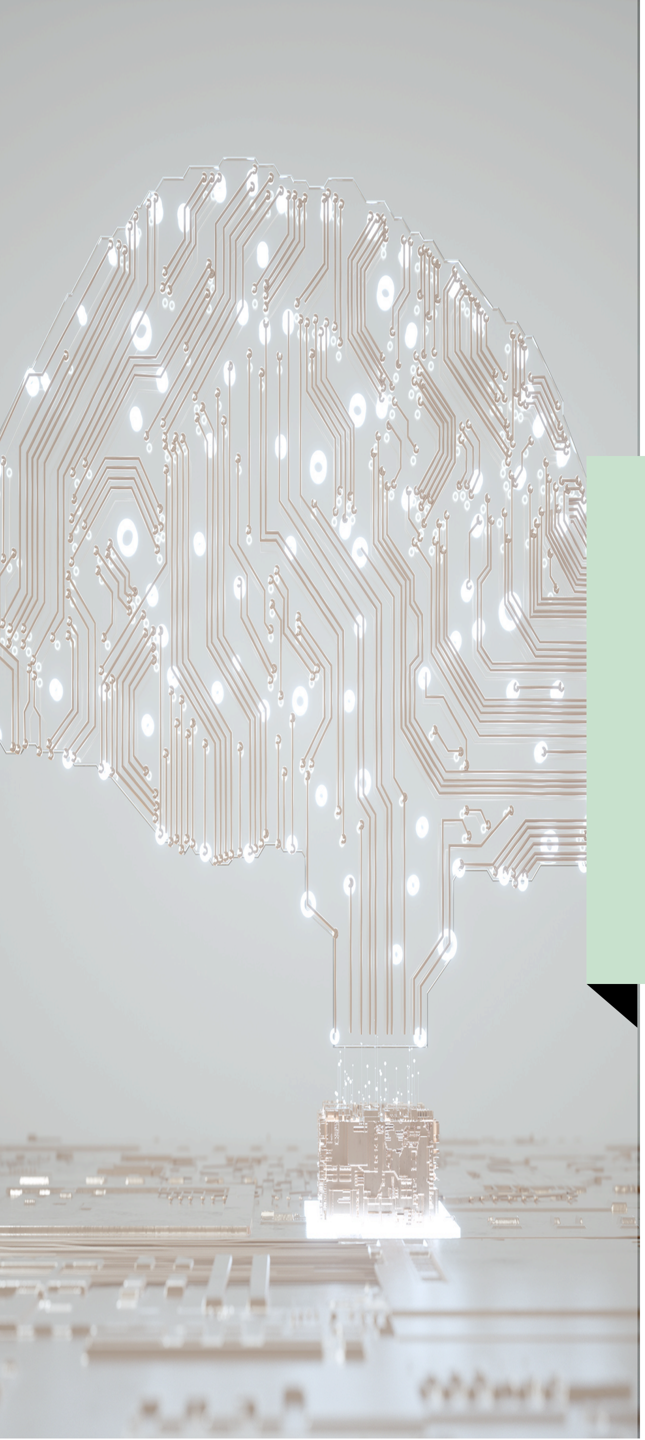
Soil Moisture Prediction based on Satellite Data using a novel Deep Learning model

Amina Habiboullah and Mohamed-Abdellahi Louly

GTI



Outline

1. GTI – AI Unit
2. Why Soil Moisture ?
3. Main contributions
4. Data
5. Architecture
6. Entreprise-grade deployment
7. Experiments
8. Results

1. WHO are we ?

Amina Habiboullah

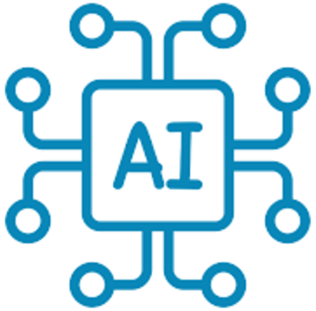
GTI-AI Lead
a.habiboullah@gti-intl.com

Mohamed Abdellahi Louly

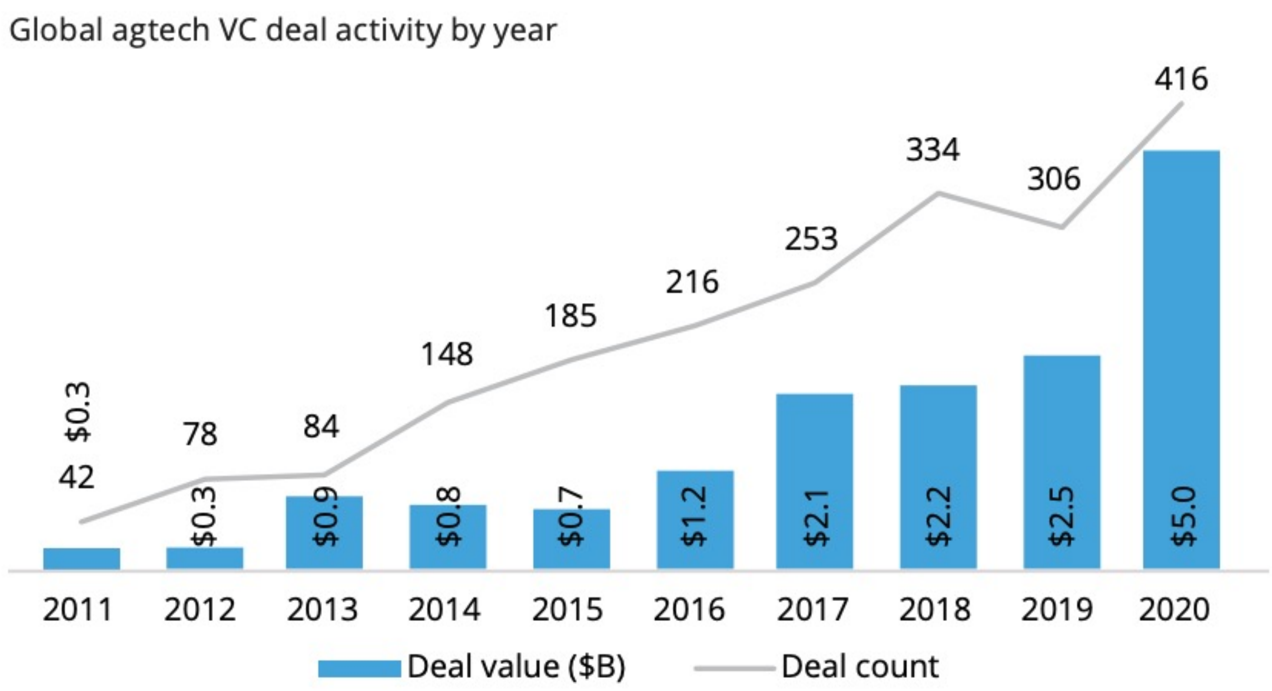
GTI international CEO
ma.louly@gti-intl.com

GTI-AI targets industrial application of Deep Learning technologies in **agriculture** and **environment** fields using satellite data.

GTI-AI's value proposition is based on products and services in its sector through engineering combining **professional quality** and **solid scientific foundation**.



2. WHY soil moisture ?



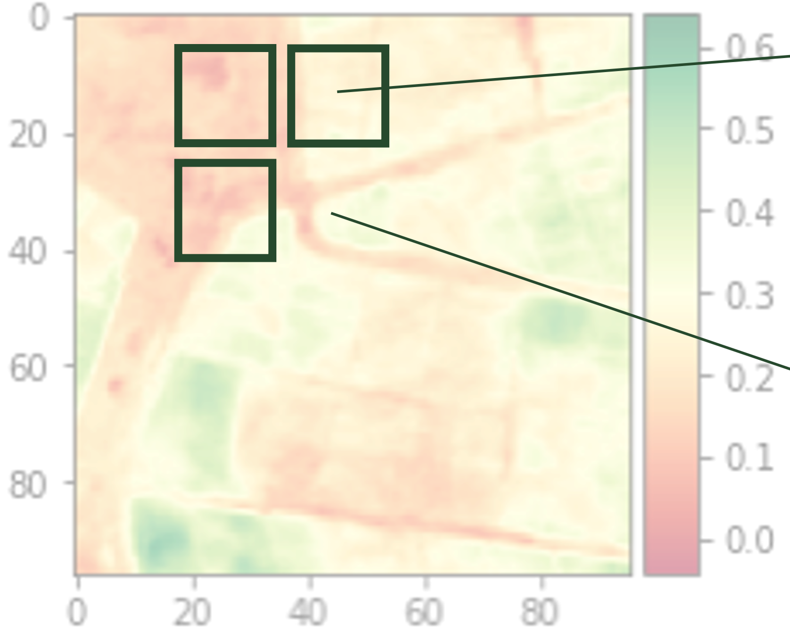
Source : 2020 Agrifood Tech Investment Review.
Finistere Ventures, LLC.

soil moisture is KEY for proper plant formation and high crop yields

2. HOW : the intuition

A whole literature on soil moisture content (SMC) retrieval from satellite data using machine learning (Random Forest, SVM, …) as well as deep learning (mainly CNN based).

But no prior works, **to the best of our knowledge**, targeting SMC as a "semantic segmentation" of the satellite image AND taking into account the "spatial correlation"



Semantic segmentation →U-Net architecture

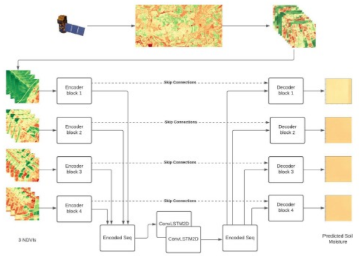
Spatial correlation …. NLP inspired architecture

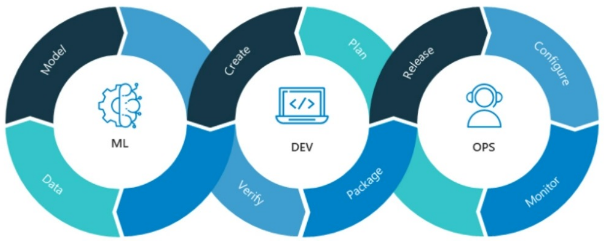
3. Main novelties/contributions

Novel DL architecture integrating both **semantic segmentation** nature as well as **sequences context**

Production-ready industrial AI cloud service

Remote-sensed data engineering



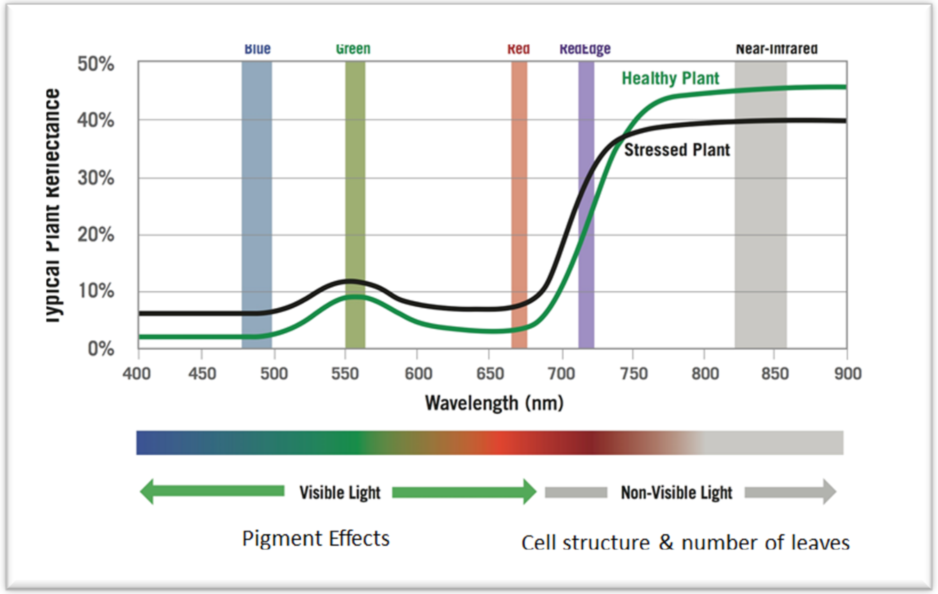


4. Data

NDVI Data

Sentinel 2 : MultiSpectral Instrument – Spatial resolution of 10 m – June 2015

European Space Agency

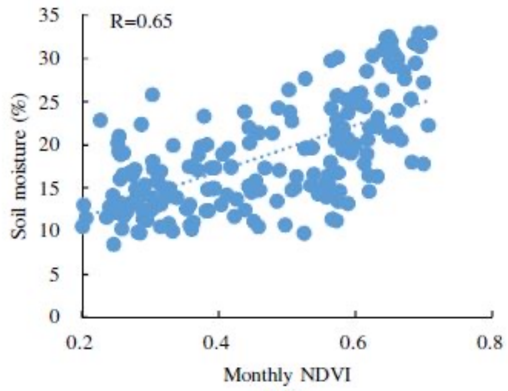


Source : micasense.com

$$NDVI_{red} = \frac{NIR - Red}{NIR + Red}$$

$$NDVI_{green} = \frac{NIR - Green}{NIR + Green}$$

$$NDVI_{blue} = \frac{NIR - Blue}{NIR + Blue}$$



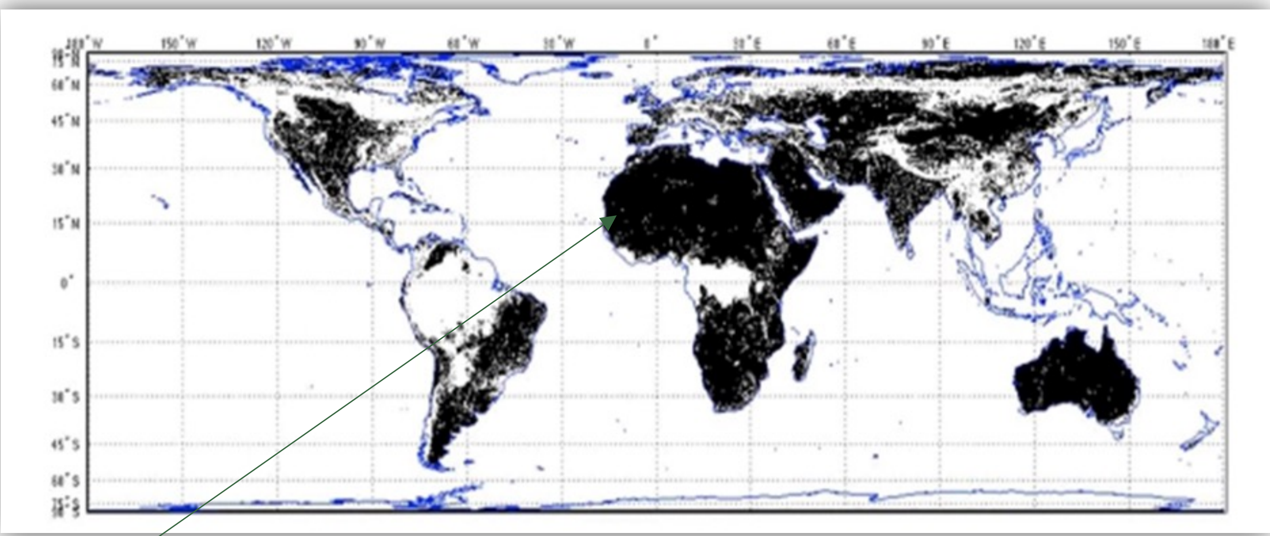
(d)

Source : NDVI dynamic changes and their relationship with meteorological factors and soil moisture, Zhang et al., 2018

4. Data

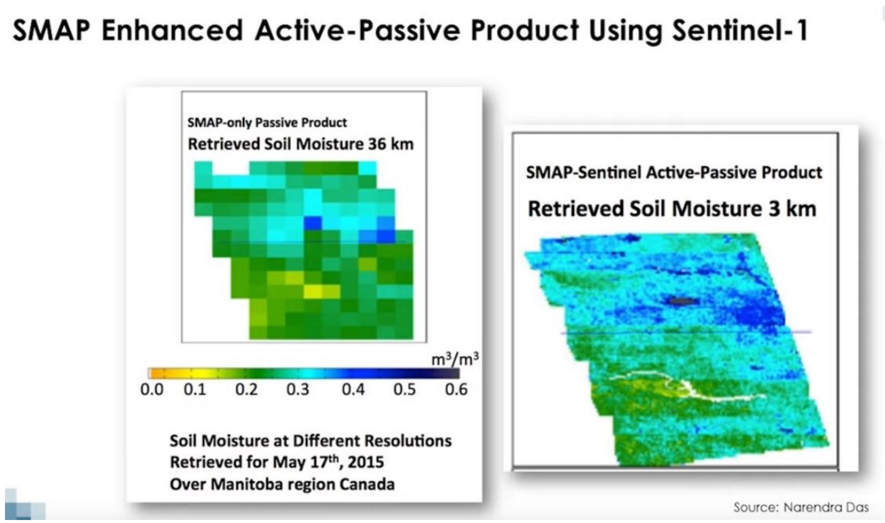
Soil Moisture Data

SMAP (NASA) : Microwave sensors (1.26 GHz) – Spatial resolution of 3 km – January 2015



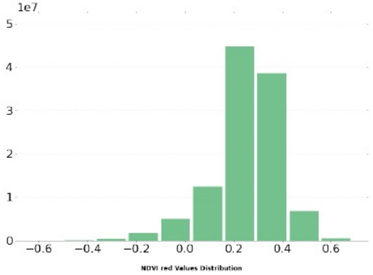
Source : NASA ARSET : Soil Moisture for Agricultural Applications

SMAP data is quite accurate in our target region

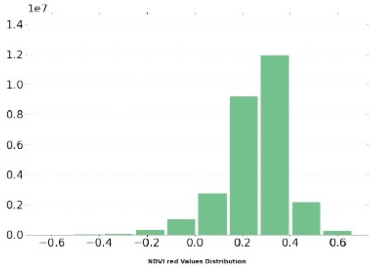


4. Data statistics

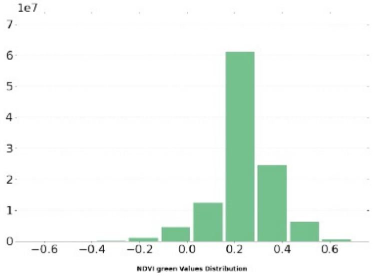
Only data with less than 10% cloud coverage has been used for training and validation purposes.



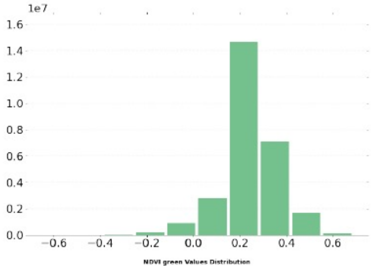
(a) NDVI red Training set



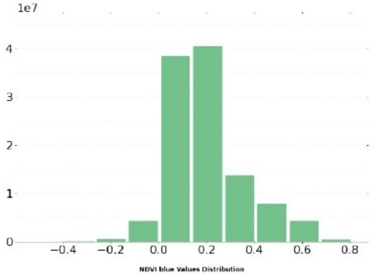
(b) NDVI red Validation set



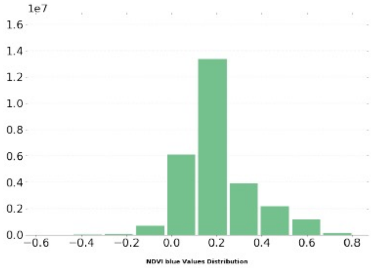
(c) NDVI green Training set



(d) NDVI green Validation set

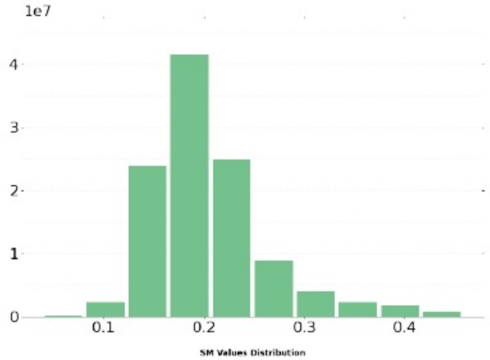


(e) NDVI blue Training set

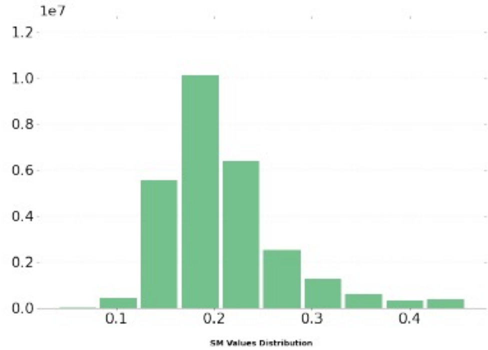


(f) NDVI blue Validation set

NDVI Data Distribution



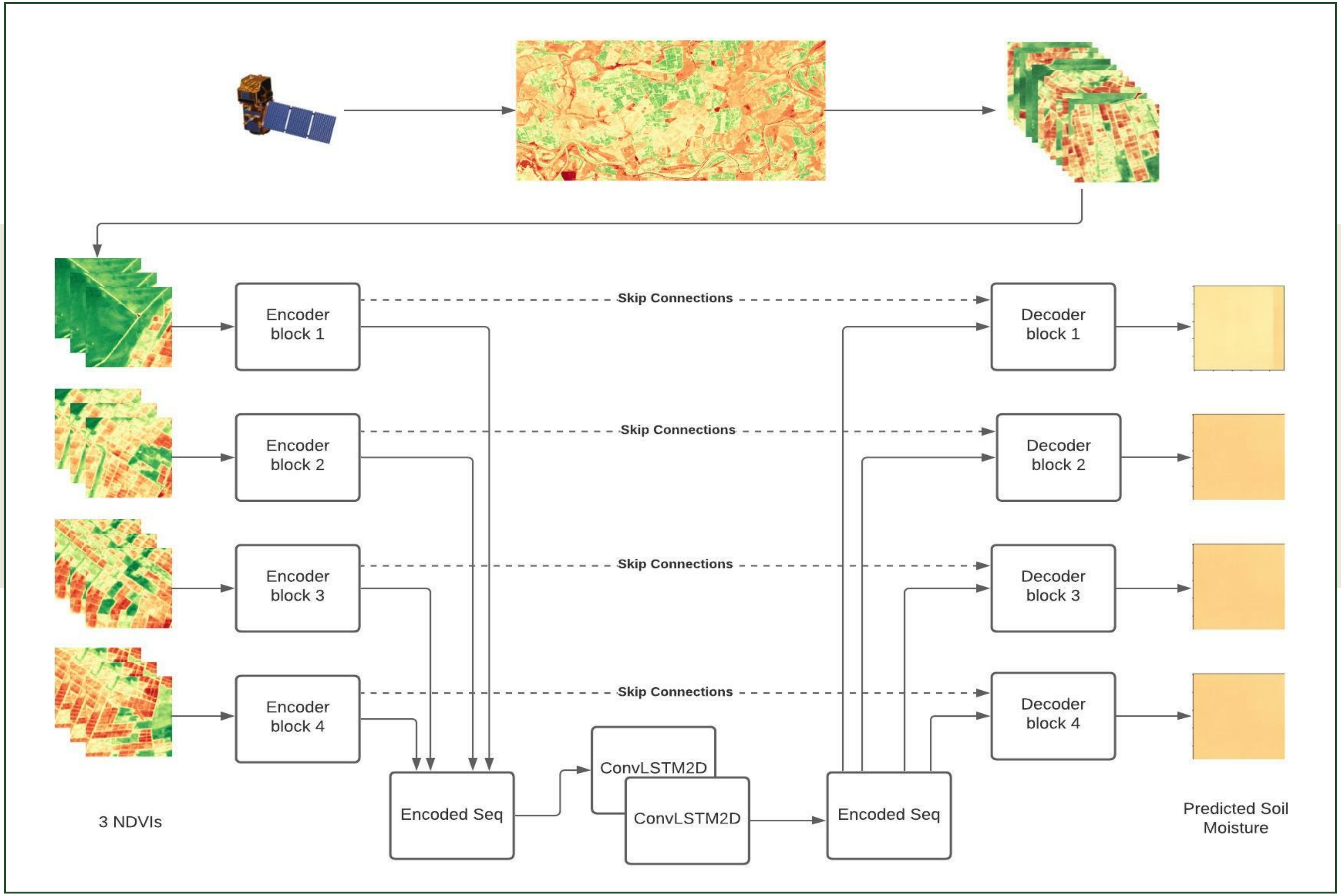
(a) Training set



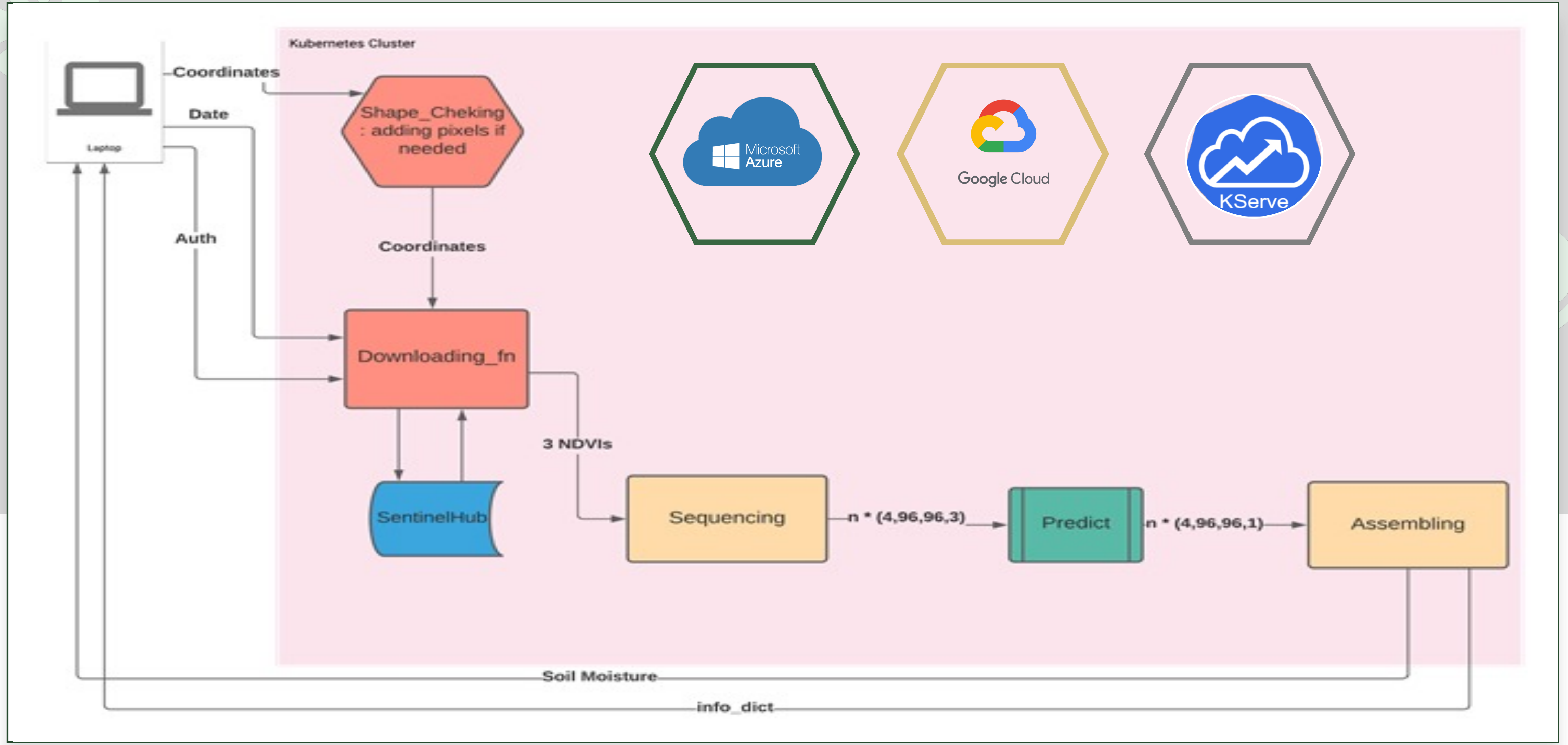
(b) Validation set

Soil Moisture Data Distribution

5. Architecture



6. Enterprise-grade deployment



7. Experiments : Evaluation Metrics

- Mean Squared Logarithmic Error : Allows to split each image into independent and equally weighted regression problems.

$$L(y,\hat{y}) = \frac{1}{N}\sum_{i=0}^{N}(\log(\hat{y}_i+1)-\log(y_i+1))^2 = \frac{1}{N}\sum_{i=0}^{N}(\log(\frac{\hat{y}_i+1}{y_i+1}))^2$$

- Evaluation Metrics :
  - RMSE and ubRMSE are quite convenient to the SMP task according to Entekhabi et al., 2009

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=0}^{N}(\hat{y}_i - y_i)^2}$$

$$ubRMSE = \sqrt{RMSE^2 - MBE^2}$$

  - Mean Absolute Error

$$MAE = \frac{1}{N}\sum_{i=0}^{N}|\hat{y}_i - y_i|$$

7. Experiments : Optimization

- Adam's update Rule :

$$\alpha_t = \alpha \times \frac{\sqrt{1-\beta_2^t}}{1-\beta_1^t}$$

$$\theta_t = \theta_{t-1} - \alpha_t \times \frac{m_t}{\sqrt{v_t} + \hat{\epsilon}}$$

with

$$\beta_1 = 0.9$$

$$\beta_2 = 0.999$$

- Since soil moisture values on our training and validation set are between 0 cm^3/cm^3 and 0.45 cm^3/cm^3, the MSLE loss is quite low and decreases quickly → vanishing gradients problem

Our proposed solution is to schedule ϵ parameter for the 30 first epochs of training in order to find the best ϵ parameter value, which will scale up gradients

7. Experiments : Optimization

- Adam's update Rule :

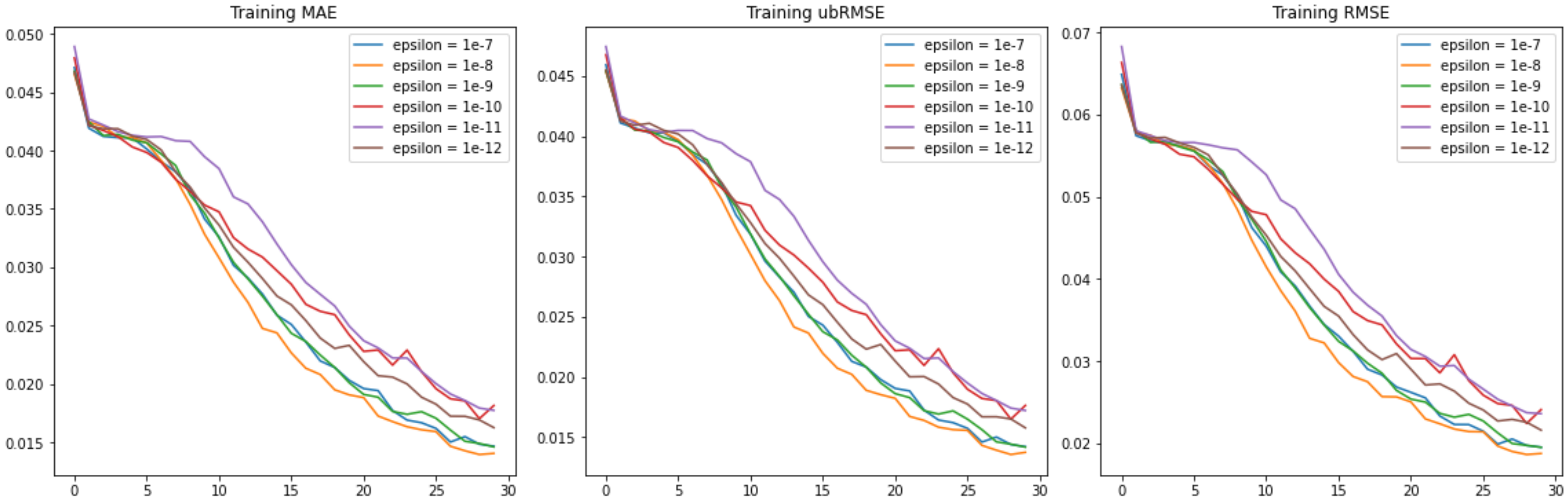
$$\alpha_t = \alpha \times \frac{\sqrt{1-\beta_2^t}}{1-\beta_1^t}$$

$$\theta_t = \theta_{t-1} - \alpha_t \times \frac{m_t}{\sqrt{v_t} + \hat{\epsilon}}$$

with

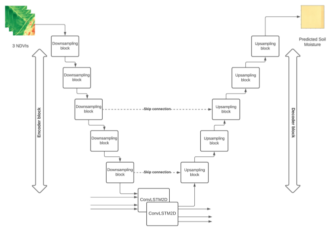
$$\beta_1 = 0.9$$

$$\beta_2 = 0.999$$



7. Experiments : Training

Configuration	Search space	Loss($\times 10^{-3}$)	MAE	RMSE	ubRMSE	Epoch
Skip connections	L0	1.618	0.0371	0.0486	0.0465	29
	L2/4	1.818	0.0379	0.0513	0.0501	13
	L3/5	**1.535**	**0.0347**	**0.0468**	**0.0455**	**47**
	L1/3/5	1.587	0.0353	0.0483	0.0472	22
	L2/3/5	1.808	0.0374	0.0511	0.0498	18
	L1/2/3/4/5	1.738	0.0366	0.0503	0.0493	28
Batch size	4	1.897	0.0400	0.0497	0.0436	43
	8	1.671	0.0363	0.0482	0.0456	47
	16	**1.535**	**0.0347**	**0.0468**	**0.0455**	**47**
	32	1.799	0.038	0.0522	0.0513	13
Activation function	**ReLU**	**1.358**	**0.0325**	**0.0447**	**0.0435**	**35**
	ELU	1.535	0.0347	0.0468	0.0455	47



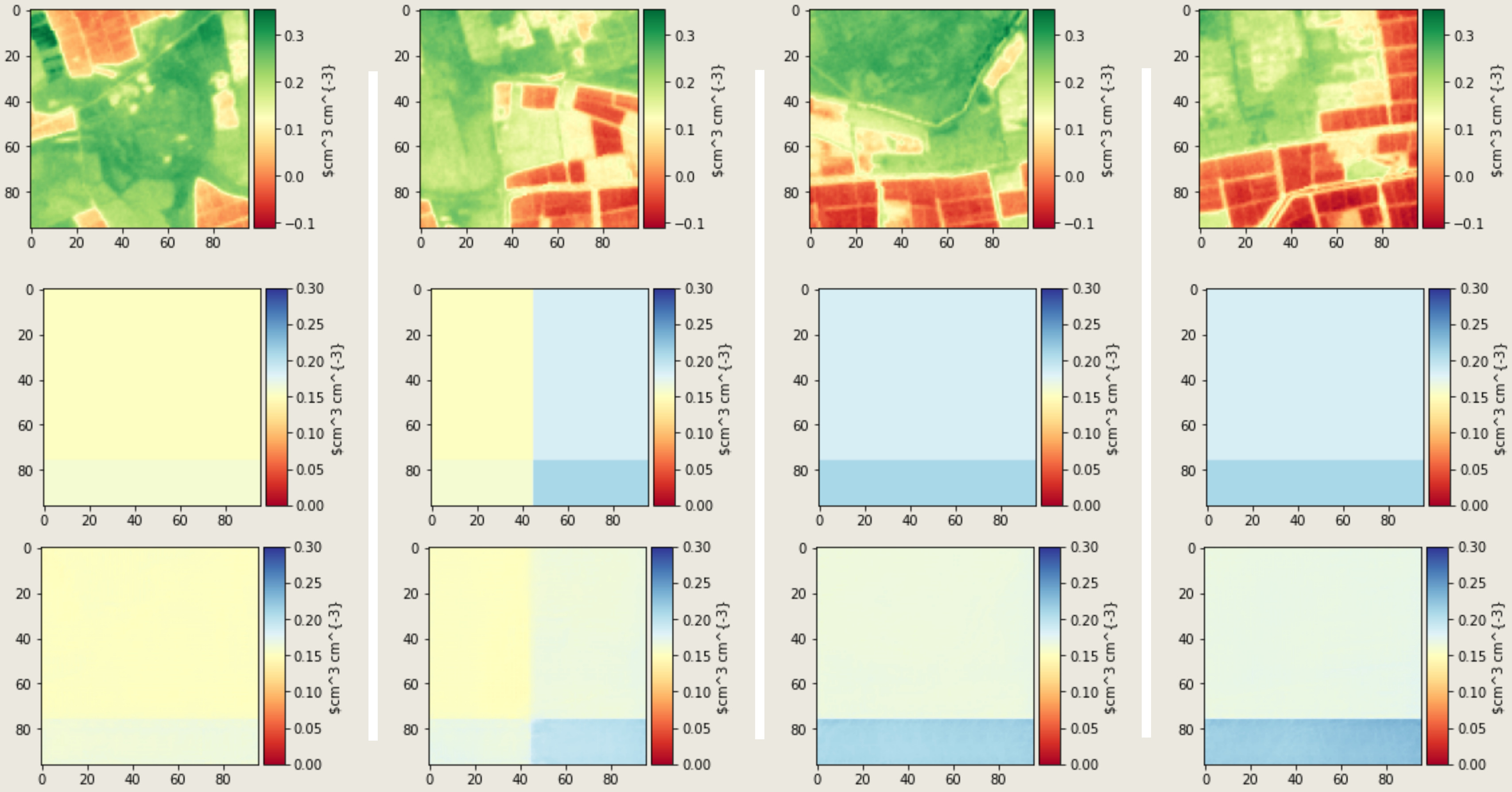
8. Results : Example

MAE = 0.0097
RMSE = 0.012

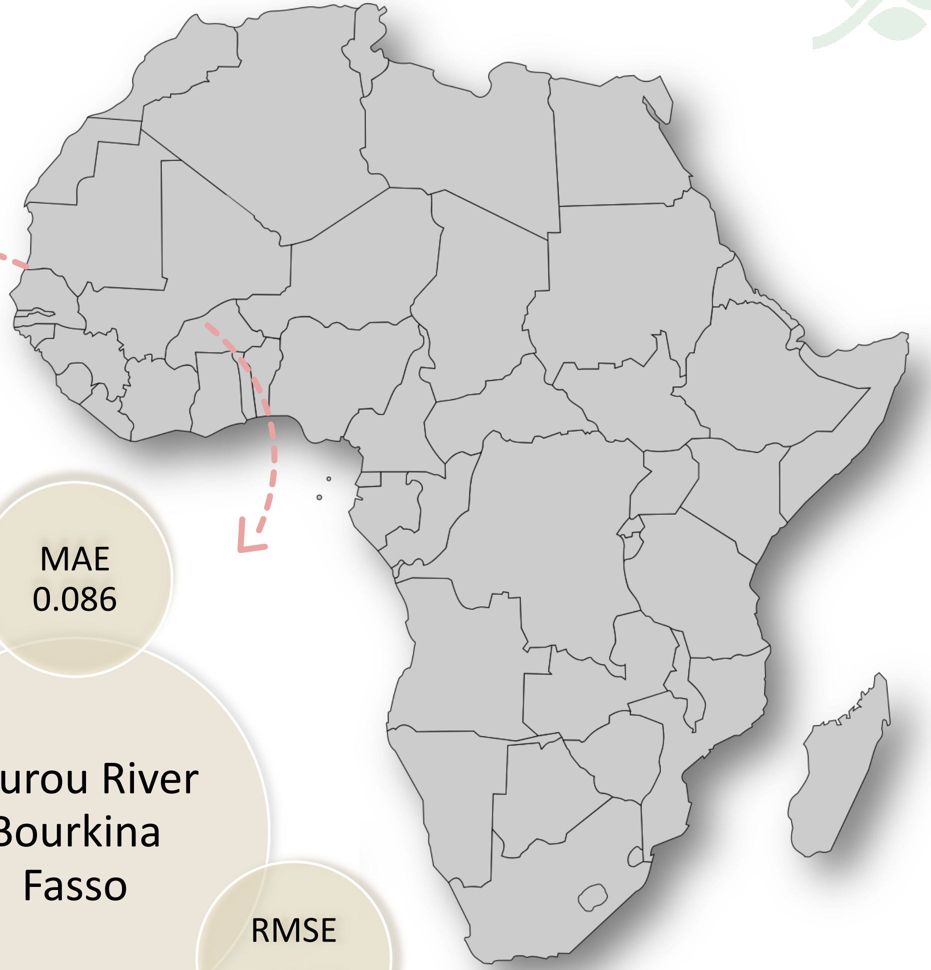
NDVI

Ground Truth

Predicted Soil Moisture



8. Results : Performance



Senegalese-Mauritanian river

MAE
0.0347

ubRMSE
0.0435

RMSE
0.0447

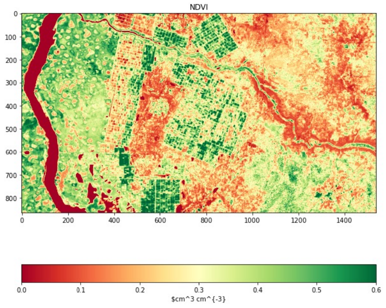
Sourou River Bourkina Fasso

MAE
0.086

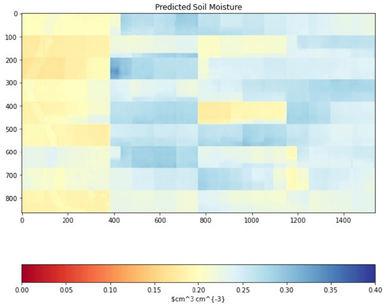
ubRMSE
0.053

RMSE
0.097

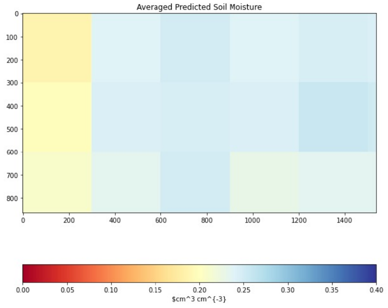
8. Results: Block Averaging



NDVI

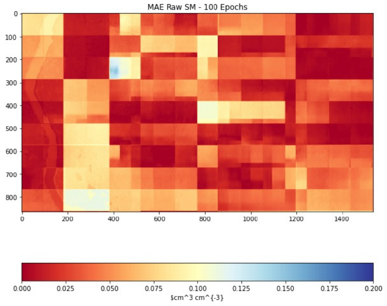


Raw Predicted SM

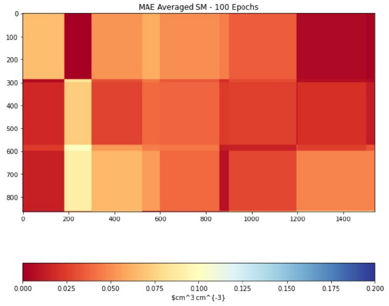


Avg Predicted SM

Error



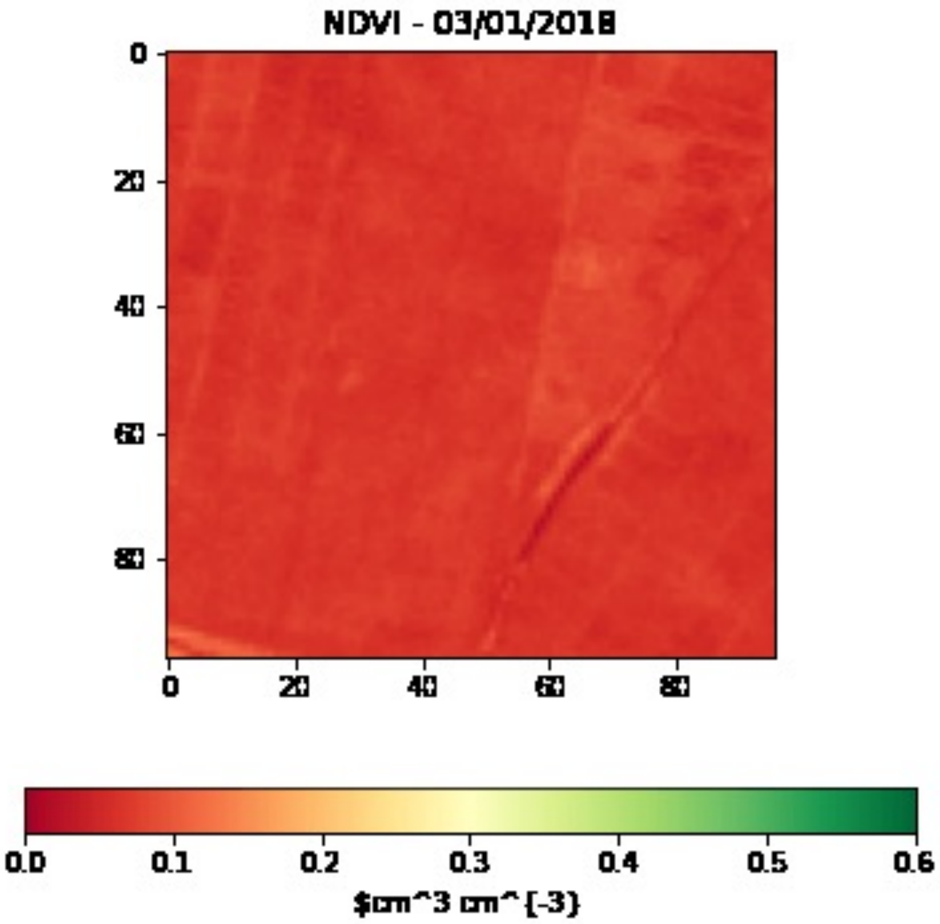
MAE = 0.036
RMSE = 0.046
ubRMSE = 0.043



MAE = 0.036
RMSE = 0.042
ubRMSE = 0.039

Inference Pipeline	MAE	RMSE	ubRMSE	Epoch
Without block averaging	0.089	0.100	0.055	35
	0.087	0.099	0.056	100
With block averaging	0.088	0.099	0.053	35
	0.086	0.097	0.053	100

8. Further work and applications



- Filling the gap on NDVI data
- IoT-based system for automated irrigation
- Wildfires prevention

Acknowledgments

- R, G, B and NIR bands are collected from Sentinelhub. The authors wish to thank the European Space Agency for sponsoring the access to Sentinelhub.

- We wish also to thank the National Snow and Ice Data Center Distributed Active Archive Center (NSIDC DAAC) and NASA Earthdata Search for providing free access to SMAP data.

Thanks !