Multisensor crop yield estimation with Machine Learning

Laura Martínez-Ferrer, Maria Piles, Gustau Camps-Valls

Machine Learning for Earth System Modelling

EGU General Assembly 2020

VNIVERSITAT DE VALÈNCIA

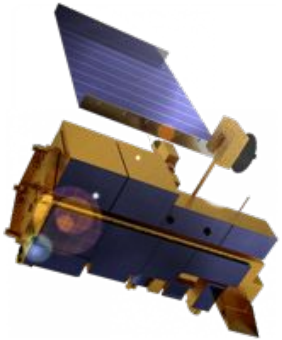
IPL IMAGE PROCESSING LABORATORY

Image Signal Processing - ISP

http://isp.uv.es @isp_uv_es laura.martinez-ferrer@uv.es

Motivation

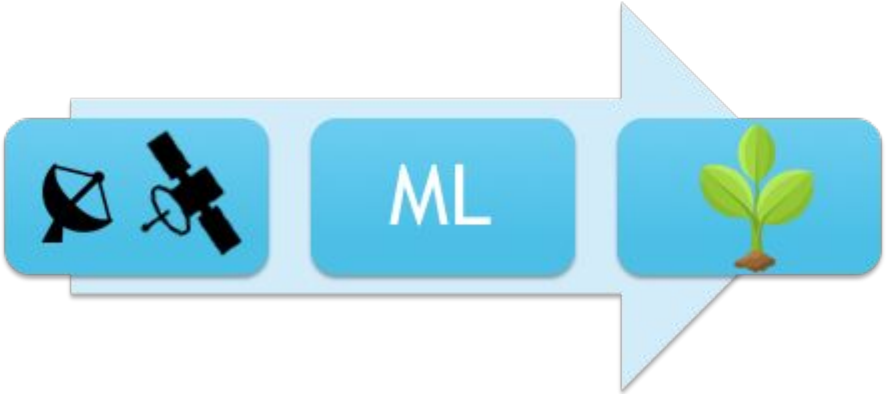
- Population
- Food
- Adverse weather



- Satellite (Optical & Microwave)
- Meteo
- Crop proportion

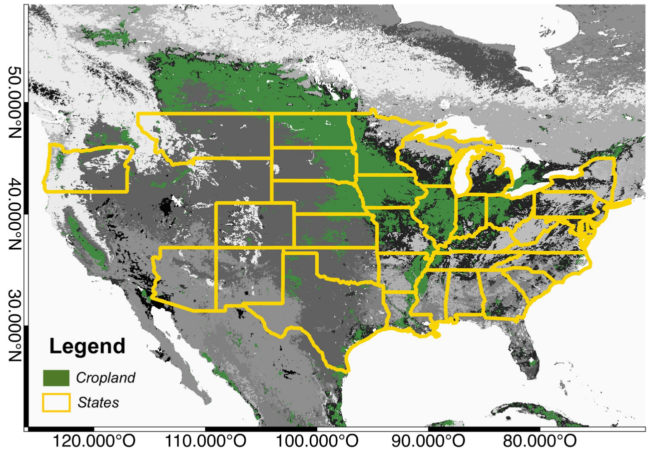
Goals

- Developing an automatic system for crop yield estimation and prediction
- Machine Learning approach to exploit synergies of satellite and meteorological data



Data collection

- Survey data: USDA (county-scale)
  - Crop yield
  - Proportion of each crop planted (CP)
- Products:
  - EVI: MOD13C1, 0.05º, 16 days
  - VOD, SM: SMAP, 9km, 3 days
  - TMAX, P: DAYMET, 1km, monthly
- Study area: CONUS (35 states)
- Years of data: 2015-2018 (growing season Apr-Oct)
- Main Crops:
  - Corn (1744 counties)
  - Soy (2060 counties)
  - Wheat (1036 counties)



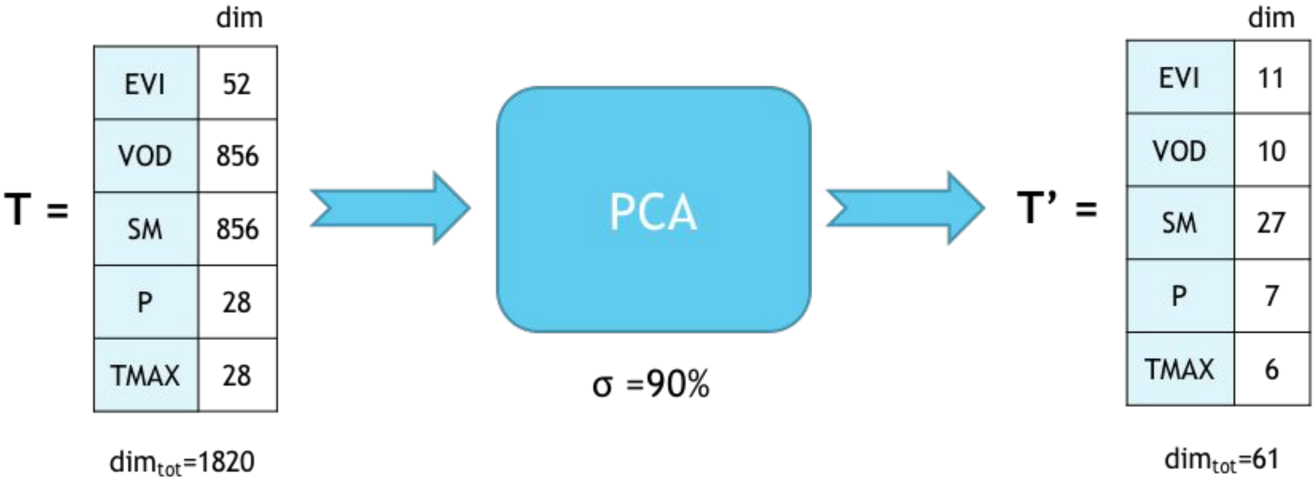
Only satellite data from pure croplands pixels included in the experiment (following MODIS-IGBP land cover)

Methodology

- N...counties (~10^3)
- T...observations (~10^3)
- y...target variable (total yield)

Poorly conditioned matrices

Problem

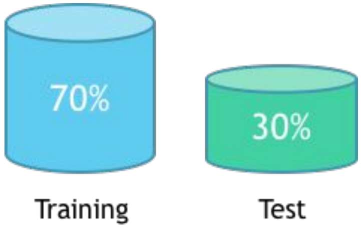


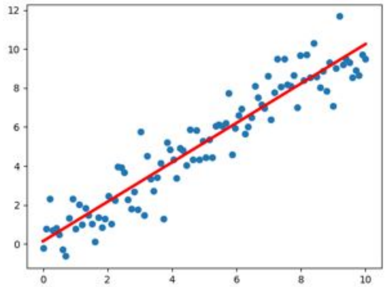
Methodology

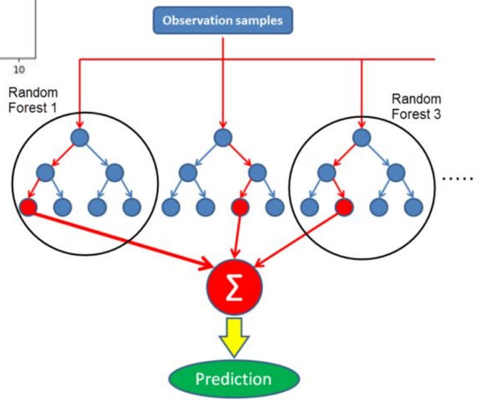
- Machine Learning Methods:
  - Linear Regression
  - Random Forest
  - Gaussian Processes

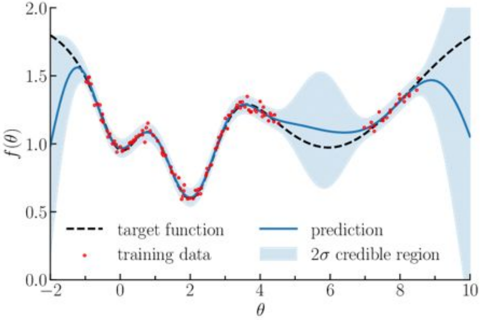
- Approaches:
  - Individual (each variable)
  - Global (combination of variables)

- Cross validation:

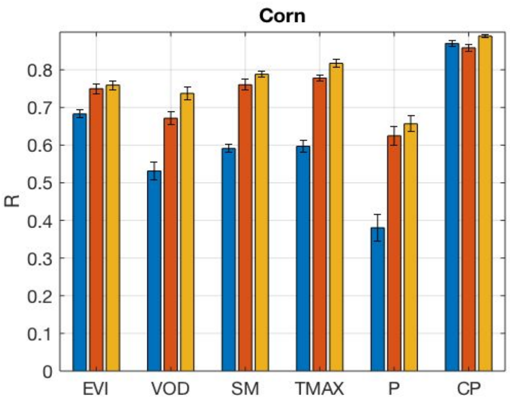


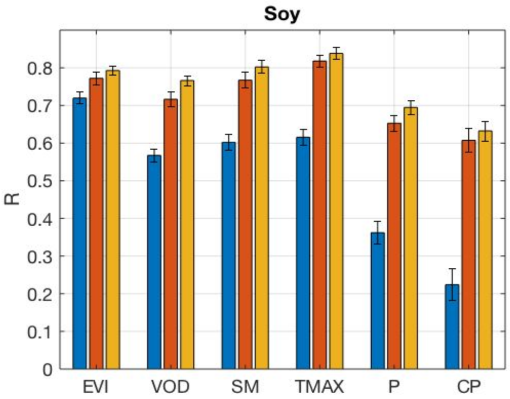


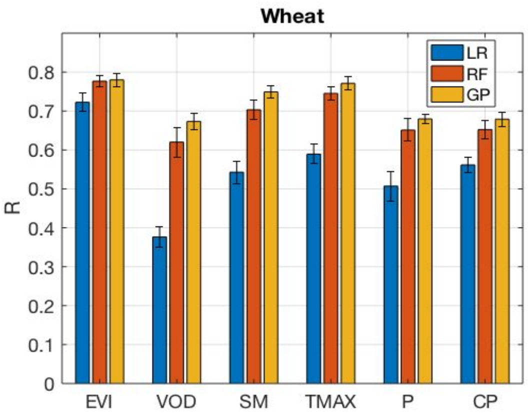


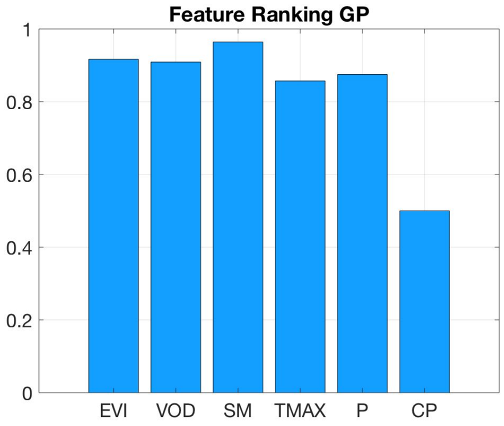


Results from individual approach



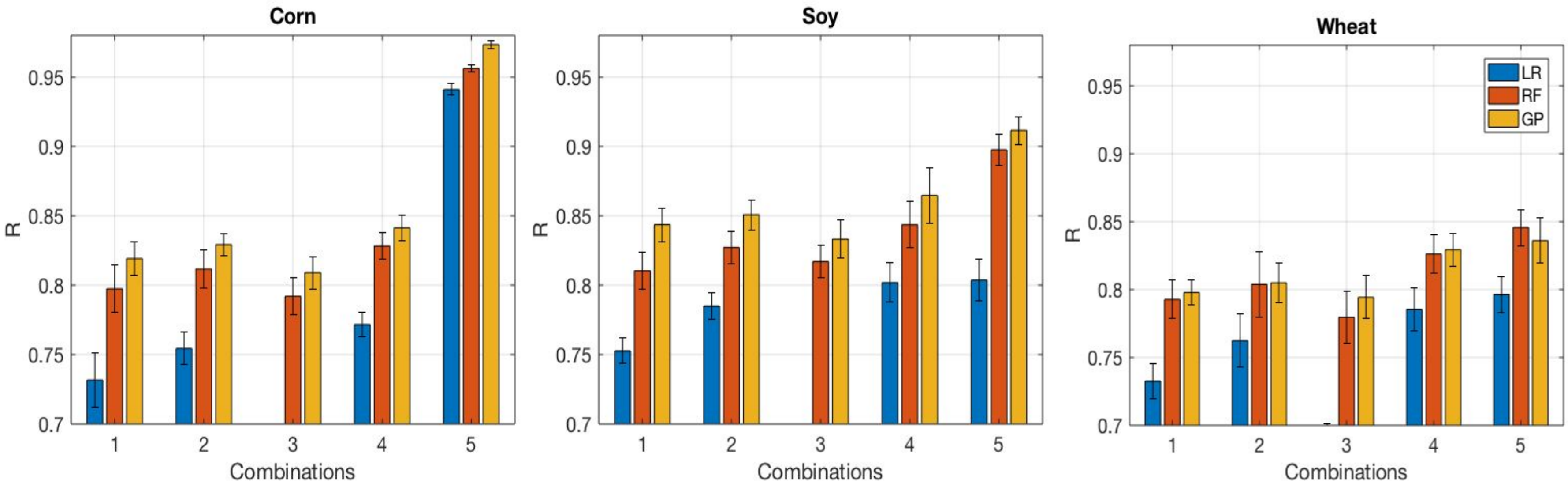






- ❑ Nonlinear methods (GP, RF) perform better: GP leads to the highest scores
- ❑ EVI, VOD, and SM are all capturing relevant information to estimate crop yield
- ❑ TMAX high relation to crop growth in corn/soy is confirmed (highest scores)
- ❑ SM leads the feature ranking while CP seems to have less relevance

Global experiments



1	2	3	4	5
EVI + VOD	EVI + VOD + SM (SAT)	TMAX + P (METEO)	SAT + METEO	SAT + METEO + CP

Results

- Best solution for all crops obtained with non-linear methods & all input features:

  R (0.84-0.97) and RMSE (0.51-1.2 t/ha)

- CP makes a difference in the global experiments

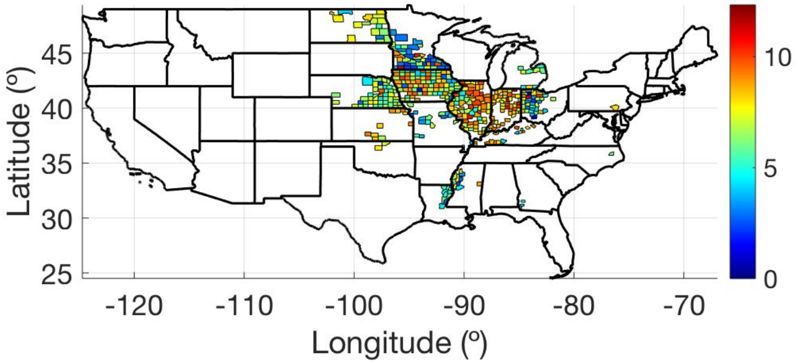
Model	T	LR			RF			GP		
		ME	RMSE	R	ME	RMSE	R	ME	RMSE	R
EVI+VOD	21									
Corn		0.05	1.50	0.73	0.04	1.35	0.80	0.04	1.27	0.82
Soy		0.03	1.44	0.75	0.03	1.30	0.81	0.02	1.17	0.84
Wheat		0.06	1.50	0.73	0.07	1.36	0.79	0.07	1.33	0.80
EVI+VOD+SM	48									
Corn		-0.01	1.43	0.75	0.01	1.29	0.81	0.01	1.21	0.83
Soy		-0.02	1.35	0.78	-0.02	1.25	0.83	-0.03	1.15	0.85
Wheat		-0.01	1.41	0.76	-0.05	1.30	0.80	-0.02	1.29	0.80
TMAX+P	13									
Corn		0.01	1.62	0.68	-0.01	1.37	0.79	0.00	1.30	0.81
Soy		-0.06	1.62	0.67	-0.04	1.27	0.82	-0.04	1.20	0.83
Wheat		0.01	1.61	0.67	-0.04	1.38	0.78	-0.04	1.32	0.79
SAT+METEO	61									
Corn		-0.03	1.41	0.77	-0.03	1.27	0.83	-0.01	1.20	0.84
Soy		0.01	1.30	0.80	-0.01	1.19	0.84	0.01	1.09	0.86
Wheat		0.04	1.36	0.78	0.01	1.26	0.83	0.03	1.23	0.83
SAT+METEO+CP	62									
Corn		-0.01	0.75	0.94	-0.01	0.68	0.96	**-0.01**	**0.51**	**0.97**
Soy		-0.05	1.31	0.80	-0.05	1.00	0.90	**-0.06**	**0.90**	**0.91**
Wheat		-0.03	1.32	0.80	-0.06	1.18	0.84	**-0.04**	**1.20**	**0.84**

MODIS-based corn grain yield estimation model incorporating crop phenology information.
Toshihiro Sakamoto, Anatoly A.Gitelson y Timothy J.Arkebauerc. (2013) **RMSE=0.81 y 2.09 t/ha**

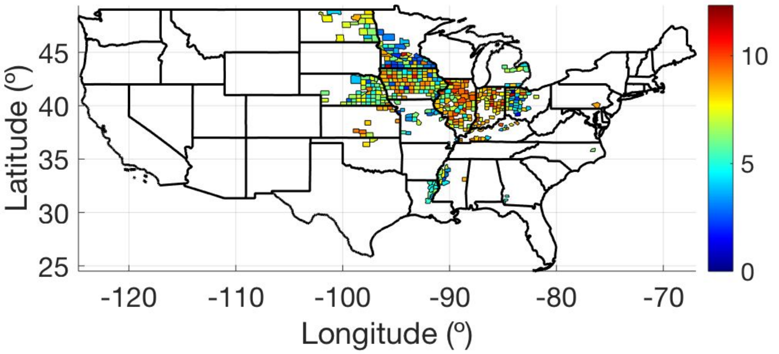
Estimation maps (Corn)

2017

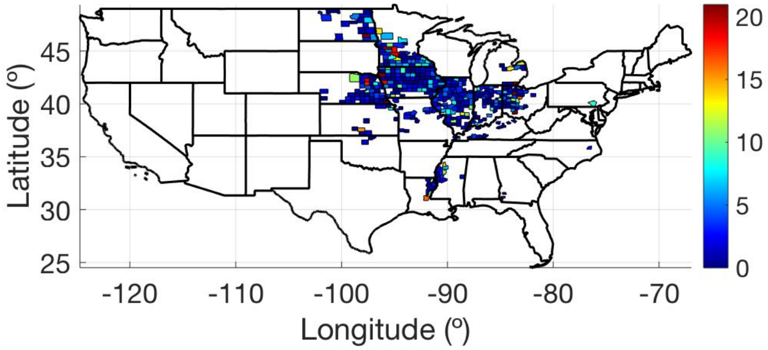
Corn yield (t/ha)



Corn yield estimation (t/ha)



Error (%)

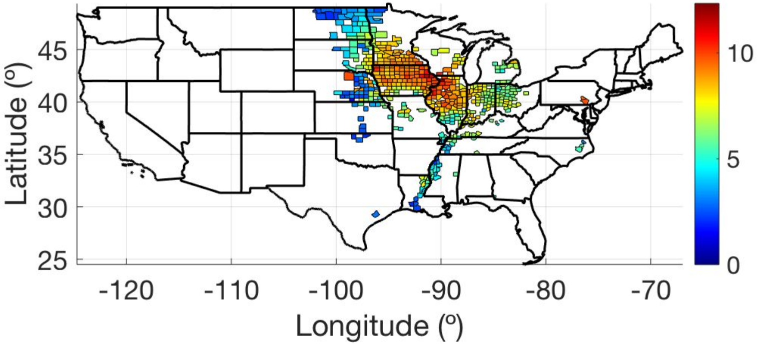


- ❑ Most error is concentrated in only a few counties
- ❑ Overall, good results in terms of R, RMSE and relative error

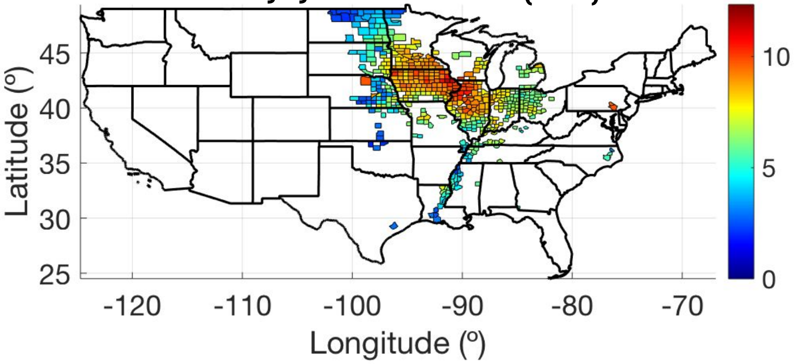
Estimation maps (Soy)

2017

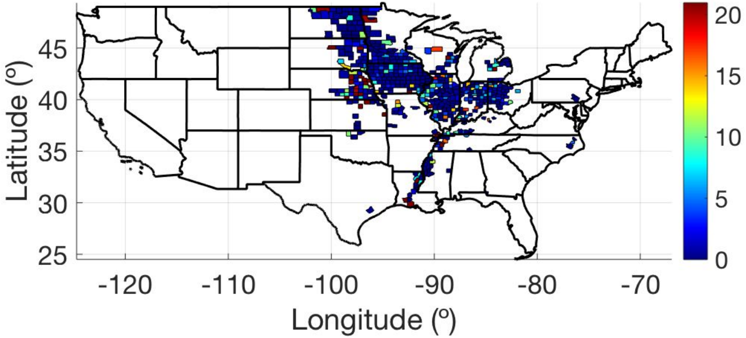
Soy yield (t/ha)



Soy yield estimation (t/ha)



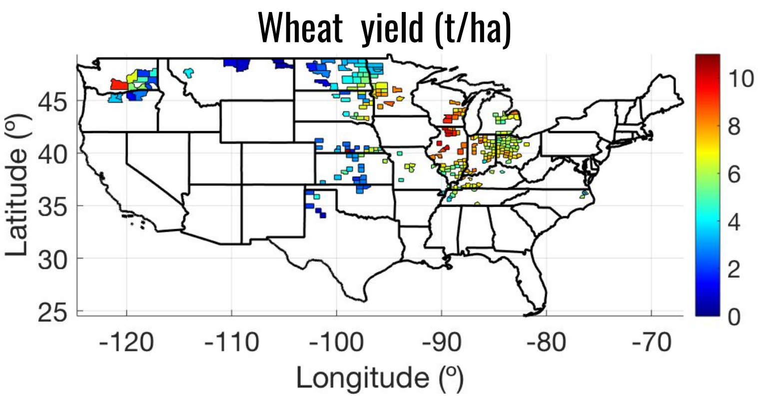
Error (%)

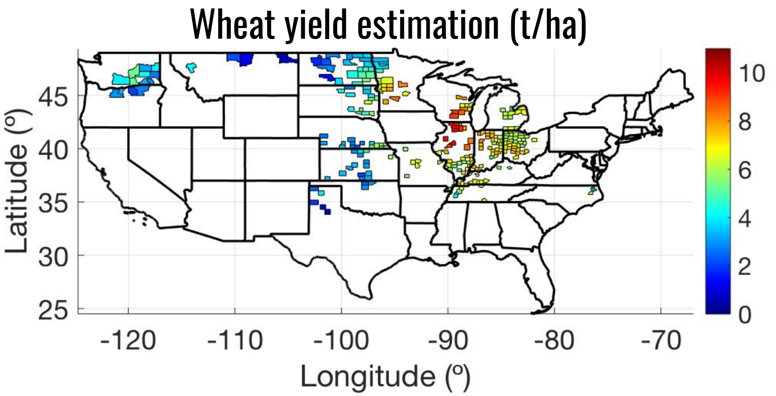


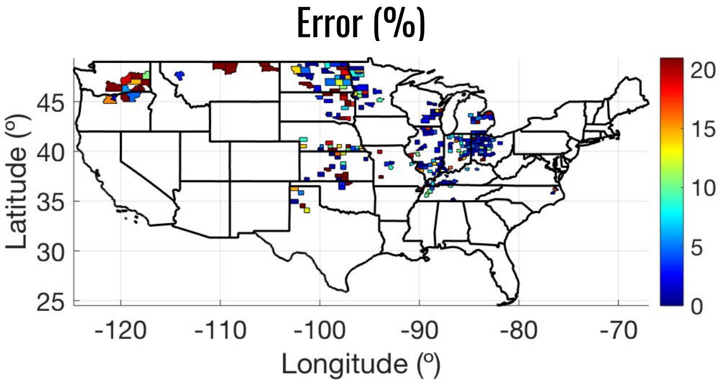
- Most error is concentrated in only a few counties
- Overall, good results in terms of R, RMSE and relative error

Estimation maps (Wheat)

2017







- Higher geographic variability
- Overall, good results in terms of R, RMSE and relative error

Conclusions

- The proposed ML framework for crop yield estimation works:
  - method: non-linear regression using GPs or RF
  - input features: time series of data summarized using PCA & proportion of planted crop CP
- Corn, soy and wheat estimated over CONUS R (0.84-0.97) and RMSE (0.51-1.2 t/ha)
- GPs allow ranking the input features:
  - Satellite data (optical, microwaves) lead the ranking, followed by meteorological
  - CP makes a difference in global experiments, while ranking is lowest in individual experiments
- Best results for Corn & Soy: higher number of observations available, concentrated in US Corn Belt

References

- A. Mateo-Sanchis et al., Synergistic integration of optical and microwave satellite data for crop yield estimation, RSE (2019)
- D. Chaparro et al., L-band vegetation optical depth seasonal metrics for crop yield assessment, RSE (2018)
- Konings, A., Piles, M., et al., Vegetation optical depth and scattering albedo retrieval using time series of dualpolarized L-band radiometer observations, RSE (2016)
- G. Camps-Valls et al., A survey on gaussian processes for earth observation data analysis, IEEE GRS Magazine (2016)

Thanks!

Laura Martínez-Ferrer, Maria Piles, Gustau Camps-Valls

VNIVERSITAT DE VALÈNCIA

Machine Learning for Earth System Modelling

EGU General Assembly 2020

IPL IMAGE PROCESSING LABORATORY

http://isp.uv.es @isp_uv_es laura.martinez-ferrer@uv.es

Image Signal Processing - ISP

erc