Integration of Machine Learning and Remote Sensing Techniques for the Estimation of Grassland Soil Organic Carbon across Temperate Climatic Zone

*S. **Panico**, P. Sciusco, S. Vicario, A. Aiello, M. De Lucia, M. Adamo*
CNR-IIA and Planetek Italia S.p.a. for RIPARTI

ESA Symposium on Earth Observation for Soil Protection and Restoration

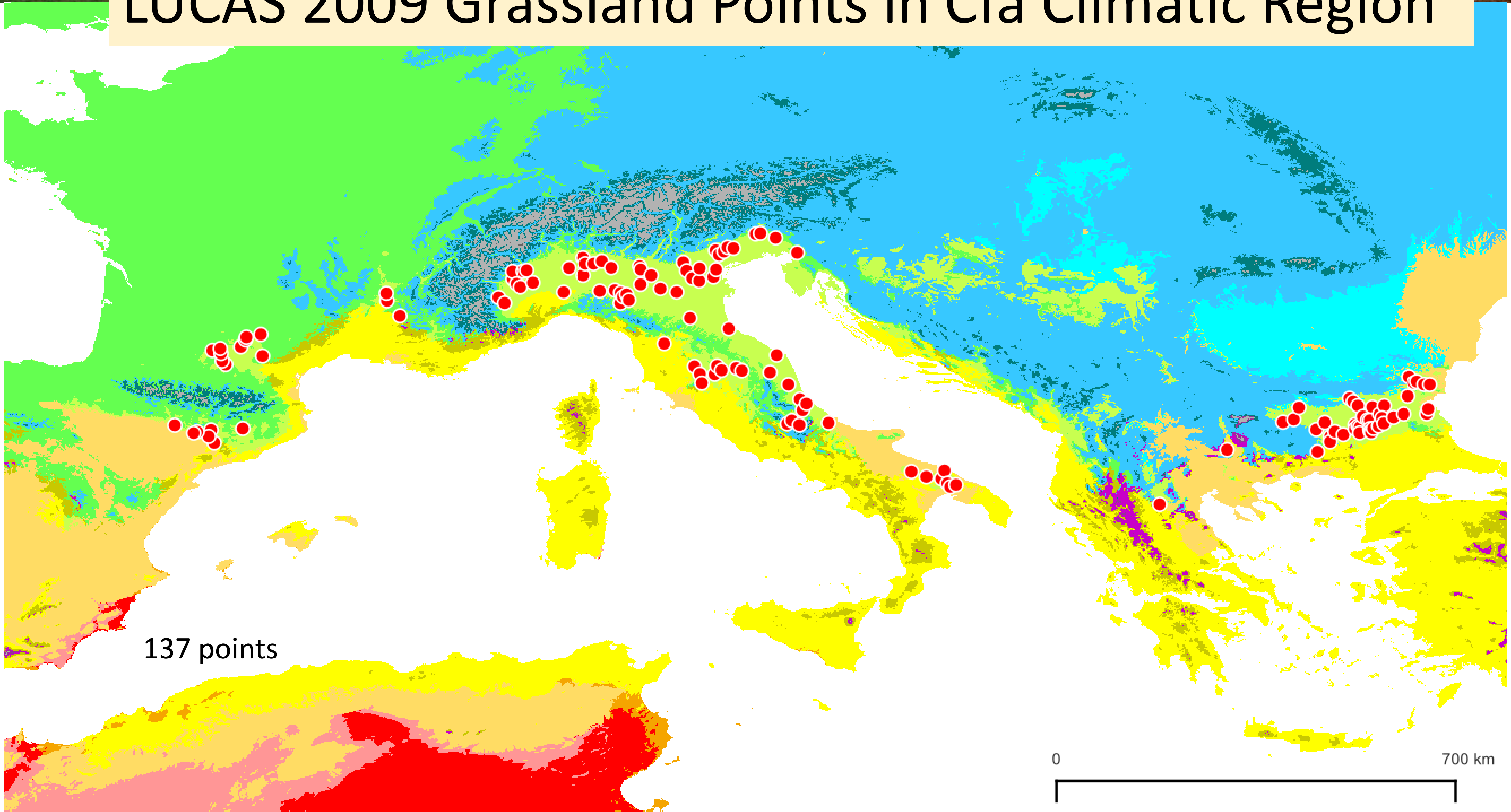
Objectives

- Build a model to infer the SOC concentration for Grassland across the same Climatic Region
- Exploit Vegetation indices time series to consider the biomass production of grassland

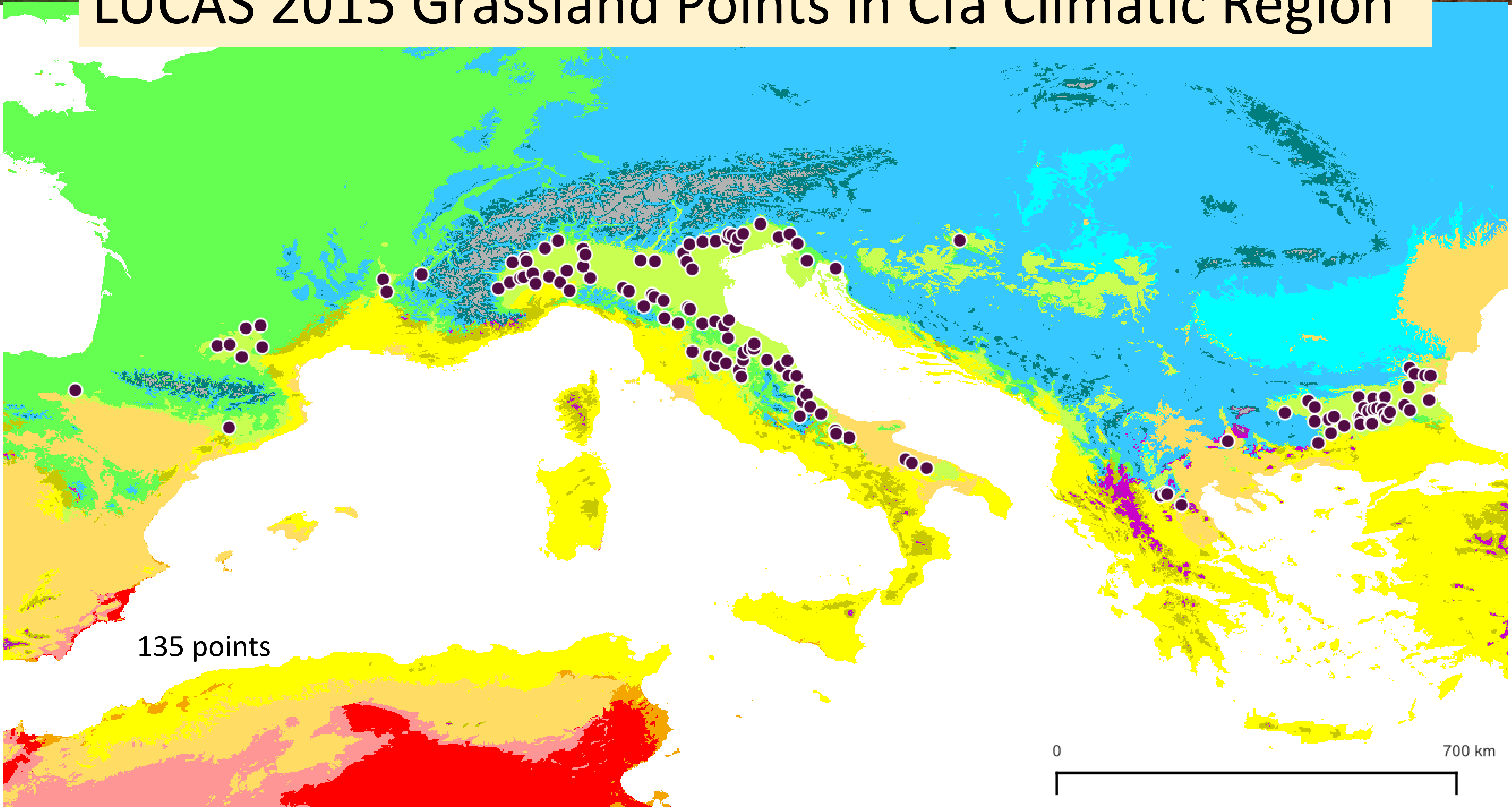
Input Data and Processing

Ground Truth Data	**LUCAS** (Land Use and Coverage Area frame Survey) Points with Topsoils infos from the field campaigns of 2009,2015,2018
	Filtered by Land Cover (LC0) **Grassland** (E10, E20, E30)
Climatic Zones Map	Selection of the LUCAS POINTS falling in the **Cfa** Climatic Zone (Temperate - Without dry season - Hot summer)
	Köppen-Geiger climate classification maps at 1-km resolution (at 95% of confidence)
Temporal Series Extraction of Vegetation Indexes (VIs) and Soil indexes for each LUCAS selected POINT	**Google Earth Engine** Platform
	LUCAS 2009: Landsat 5/7/8 startDate: 2004-01-01 endDate: 2012-12-31 LUCAS 2015: Landsat 5/7/8/9 startDate: 2007-01-01 endDate: 2023-09-30 LUCAS 2018: Sentinel 2 startDate: 2017-03-30 endDate: 2023-09-30
	Buffer size : 10 m radius for Sentinel, 15 m radius for Landsat
Temporal Series Analysis	**Python**
DEM	Elevation, Aspect and Slope extracted by **Google Earth Engine**

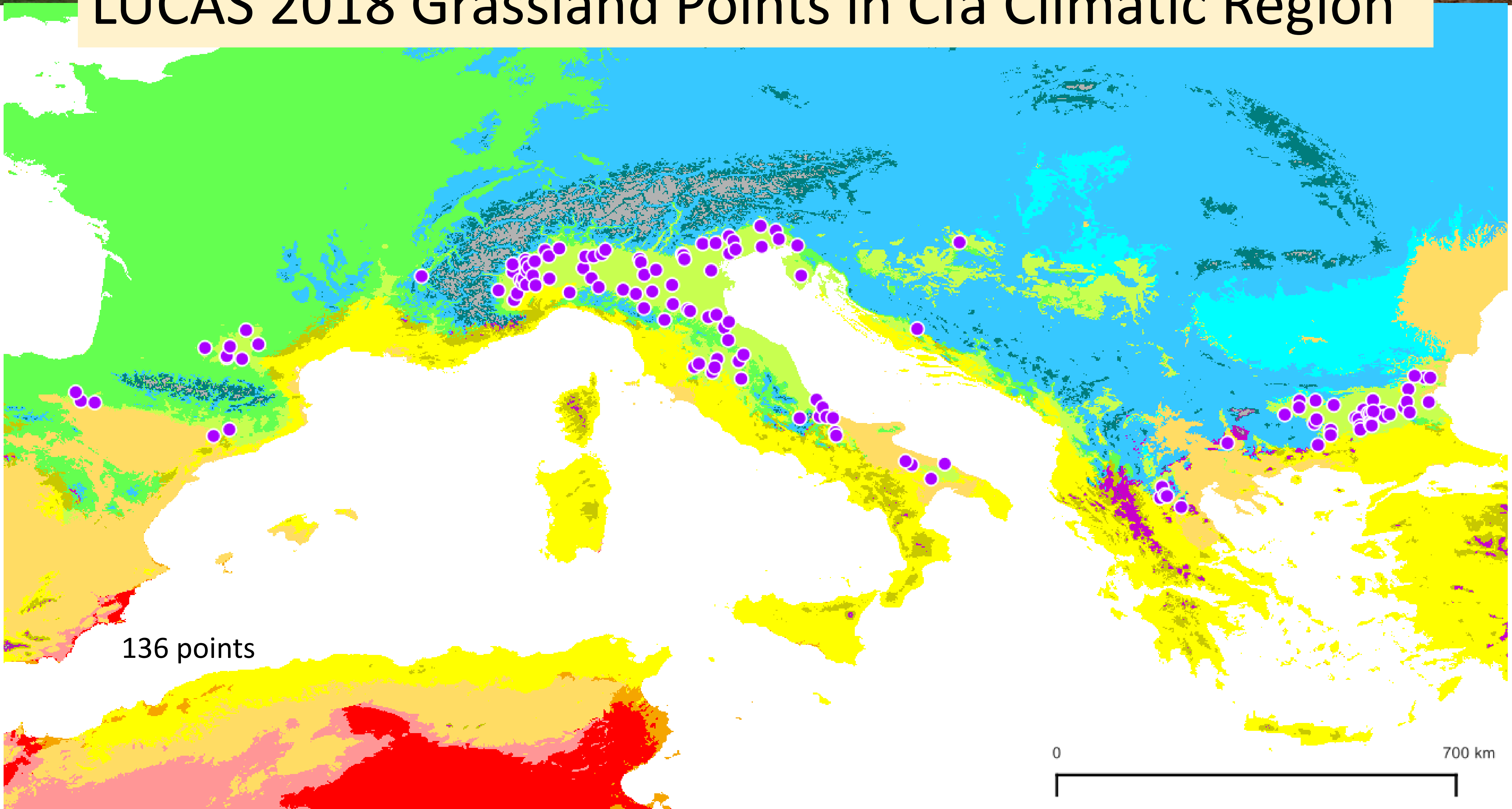
LUCAS 2009 Grassland Points in Cfa Climatic Region



LUCAS 2015 Grassland Points in Cfa Climatic Region



LUCAS 2018 Grassland Points in Cfa Climatic Region

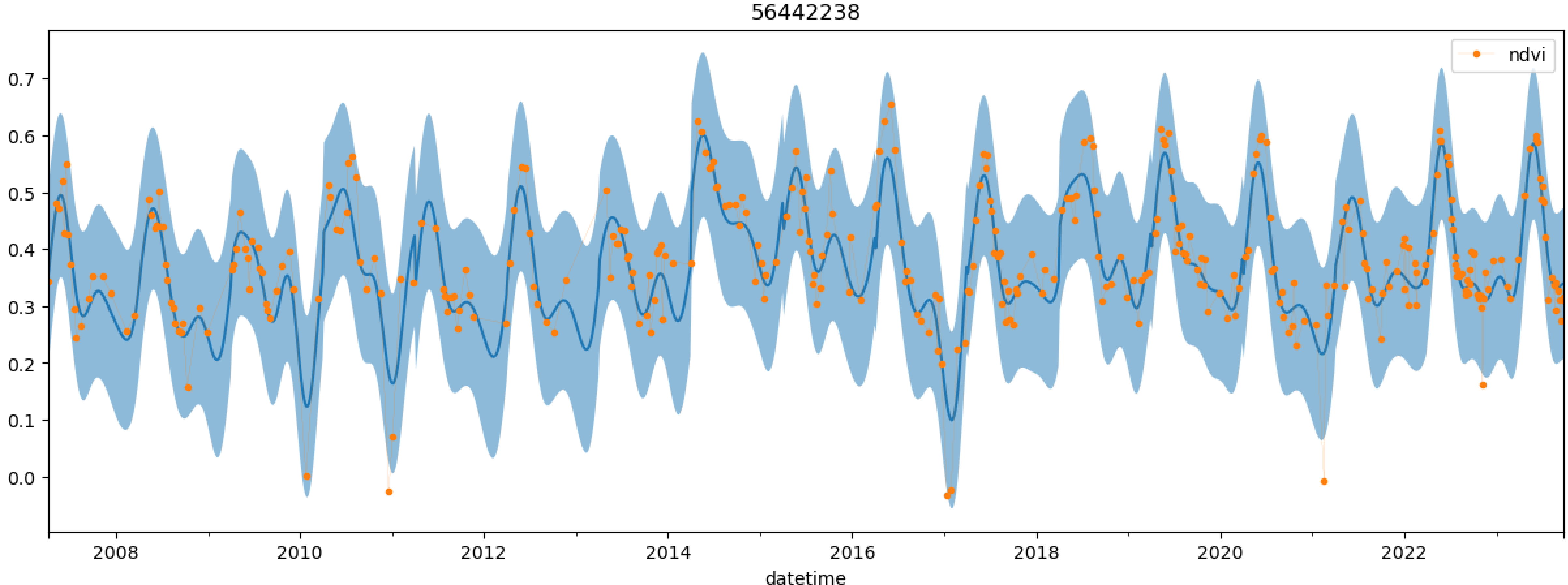


Interpolation of data through an Harmonic Model with 3 components

The Satellite's Bands, the VIs and Soil indexes' temporal series extracted from Google Earth Engine, for each point of the three LUCAS datasets, was then interpolated with Harmonic Model with 3 components that allow annual change implemented within a hierarchical Bayesian framework, obtaining:

Expected values every 5 days

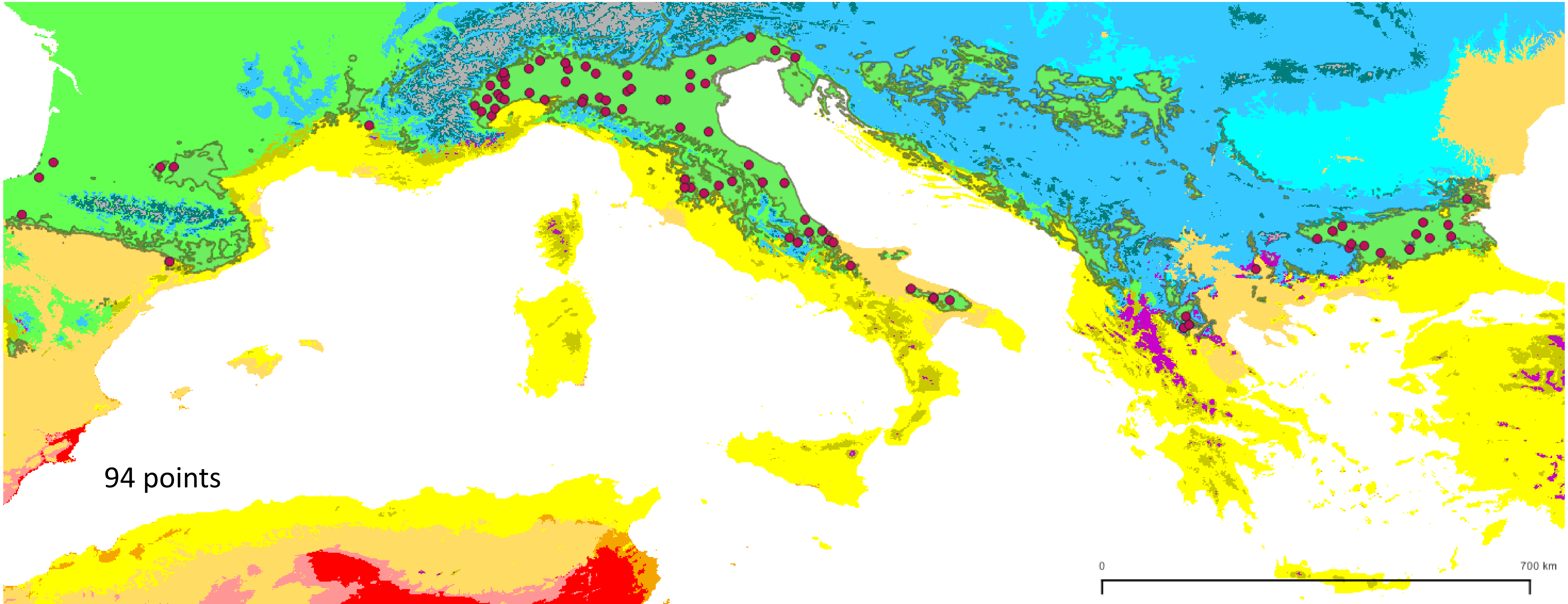
The relative **Standard Deviation** for each expected value



Extraction of Bare Soil Data and Dataset construction

The Expected values was then filtered to Extract only Bare Soil Data for each LUCAS point

- 0 < NDVI_Expected < 0.24
- temporal filter:
  - Start of the LUCAS dataset year < Date of Expected value < End of the LUCAS dataset year



Extraction of Bare Soil Data and Dataset construction

For each remaining LUCAS Point we made a composite of the Bare Soil data, producing:

- Bands and Soil indexes **mean**
- Bands and Soil indexes **standard deviation mean**
- Bands and Soil indexes **sigma** (sample standard deviation)

The VIs instead were used to produce:

- Mean value of VIs for the relative LUCAS dataset year point (**1 year**)
- Mean value of VIs for the relative LUCAS dataset year point and of the two years before (**3 years**)
- Mean value of VIs from the beginning of the relative LUCAS dataset year point and the survey date on field (**1 year to survey date**)

Machine Learning Model Application

We applied a **Random Forest** Model with a **Recursive feature Elimination** (RFE) to the dataset produced (30 features)

We obtained the highest R^2 (0.45) with 18 features, the RMSE values is 8.65 g/Kg

Among the most important features selected there are

- NDMI mean values 1 year to survey date (VI features)
- Latitudine
- Swir and NIR mean value (Bare Soil related feature)
- Elevation
- NDVI mean values 1 year to survey date (VI features)

Future works

- To develop specific predictive models based on ML for each climatic region
- Use of soil texture and type as predictors
- Apply the model by using Copernicus Grassland High Resolution Layer

Thank you

*S. **Panico**, P. Sciusco, S. Vicario, A. Aiello, M. De Lucia, M. Adamo*
CNR-IIA and Planetek Italia S.p.a. for RIPARTI

ESA Symposium on Earth Observation for Soil Protection and Restoration