Spatial prediction of the properties of chernozem soils on a field scale using machine learning methods based on data from Landsat 8 OLI and Sentinel 2 images for precision farming

Sahabiev Ilnaz,
Smirnova Elena,
Giniyatullin Kamil

Department of Soil Science, Kazan (Volga) Federal university, 18, Kremlevskaya ave., Kazan, 420008, Russia, E-mail: ilnassoil@yandex.ru

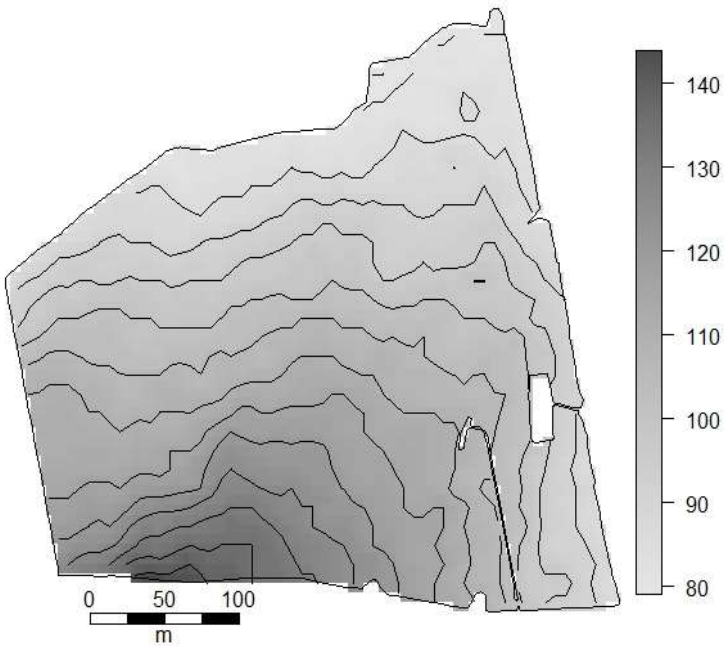
- The purpose of this work is to assess the possibility of using remote sensing data obtained from the Landsat 8 and Sentinel 2 satellites as predictors of the spatial prediction of soil properties using machine learning methods

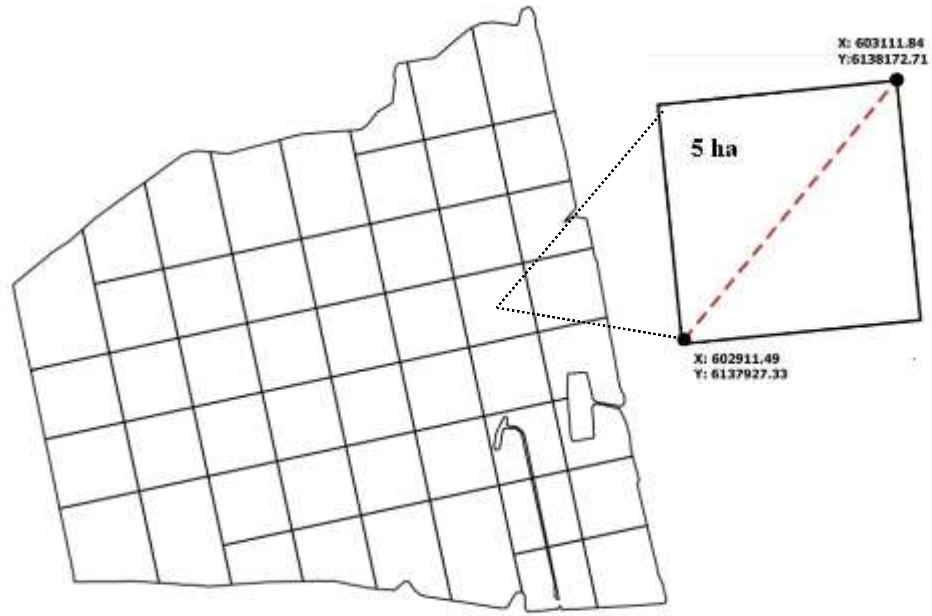
Object of study: Field with an area of 254 ha.

Chernozem zone of the Republic of Tatarstan (RUSSIA).

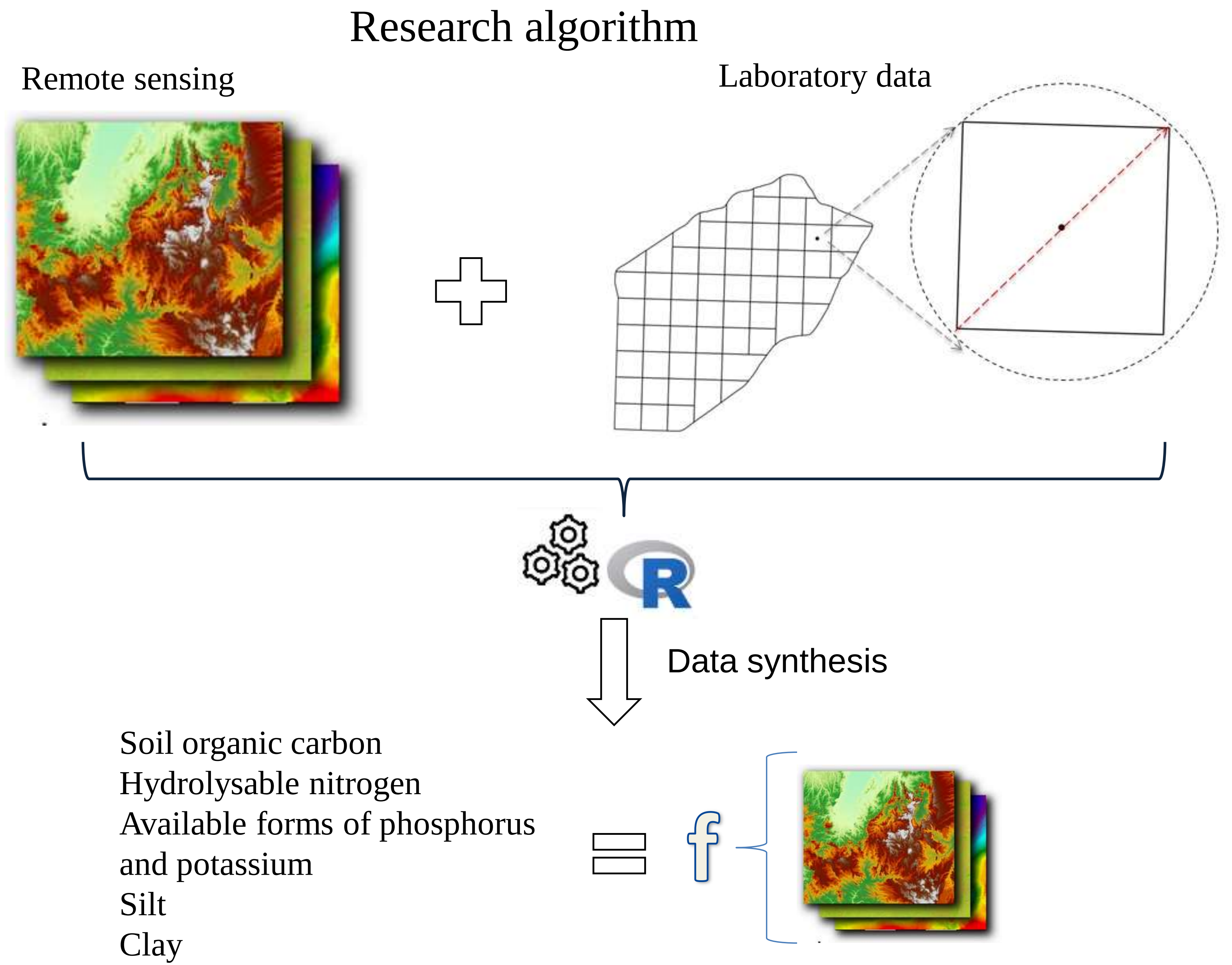
Plot coordinates : 55,182893 N, 51,999358 E

Digital elevation model and soil sampling scheme





The elevation difference of the site is 60 m. For the selection of soil samples, the field was divided into elementary plots (5 ha each), from which diagonally point samples (20-40 pcs.) were selected to compile mixed samples (50 pieces in total).


Research algorithm
Remote sensing
Laboratory data
Data synthesis
Soil organic carbon
Hydrolysable nitrogen
Available forms of phosphorus
and potassium
Silt
Clay
f

Methods

- Soil organic carbon – dry combustion (organic elemental analyzer Vario Max Cube Elementar);

  Hydrolysable nitrogen, available forms of phosphorus and potassium – chemical definition (by the Kornfield method, by the GOST 26204-91)
- Content of silt fractions, clay fractions – laser sedimentography (particle size analyzer Microtrac SDC)

Remote sensing data collection

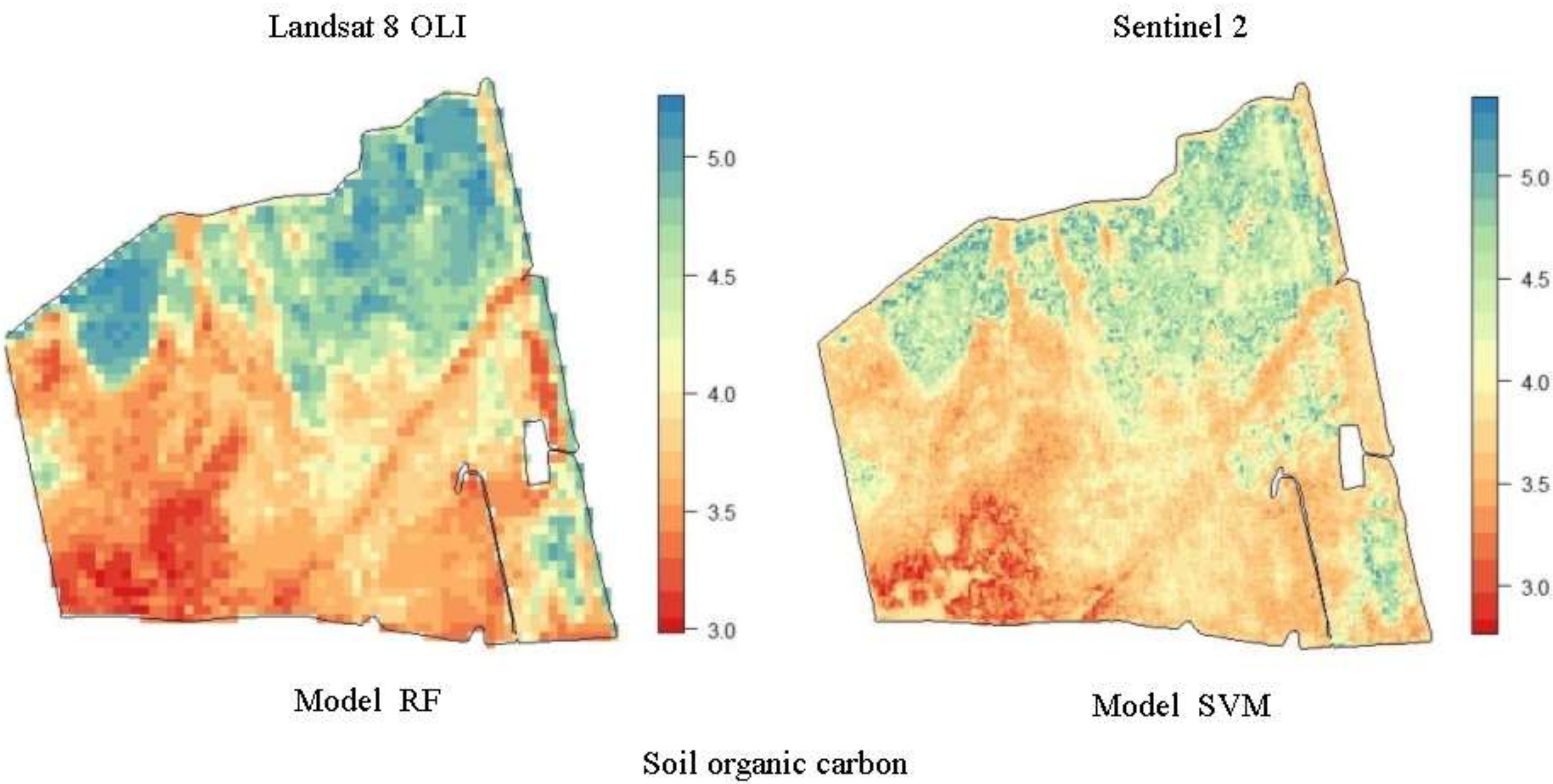
Used materials available in open sources - satellites (Landsat 8 OLI and Sentinel 2). To obtain reflectance data, images with open soil with minimal vegetation development were used. For the investigated area, these are Landsat 8 OLI images from 05/31/2019, Sentinel 2 from 05/12/2019. Data from individual bands and spectral indices were extracted and averaged over elementary sampling sites (5 ha).

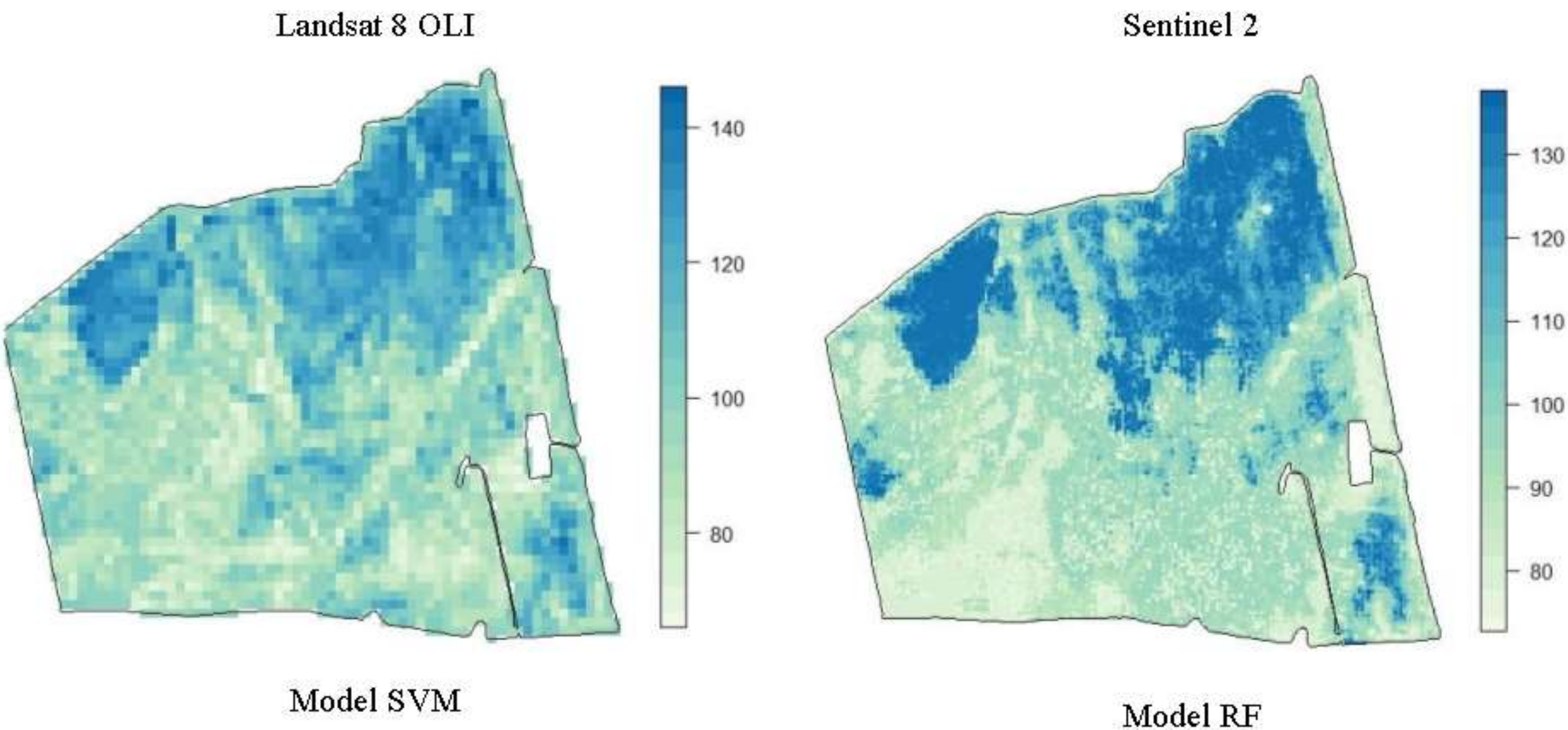
Models

- Linear models (MLR), support vector regression (SVM) and random forest (RF) were used as models. The models were validated using the bootstrap procedure, taking into account performance optimism [4].
- Model Evaluation Criteria - standard error (RMSE), mean absolute error (MAE) и coefficient of determination (R2).

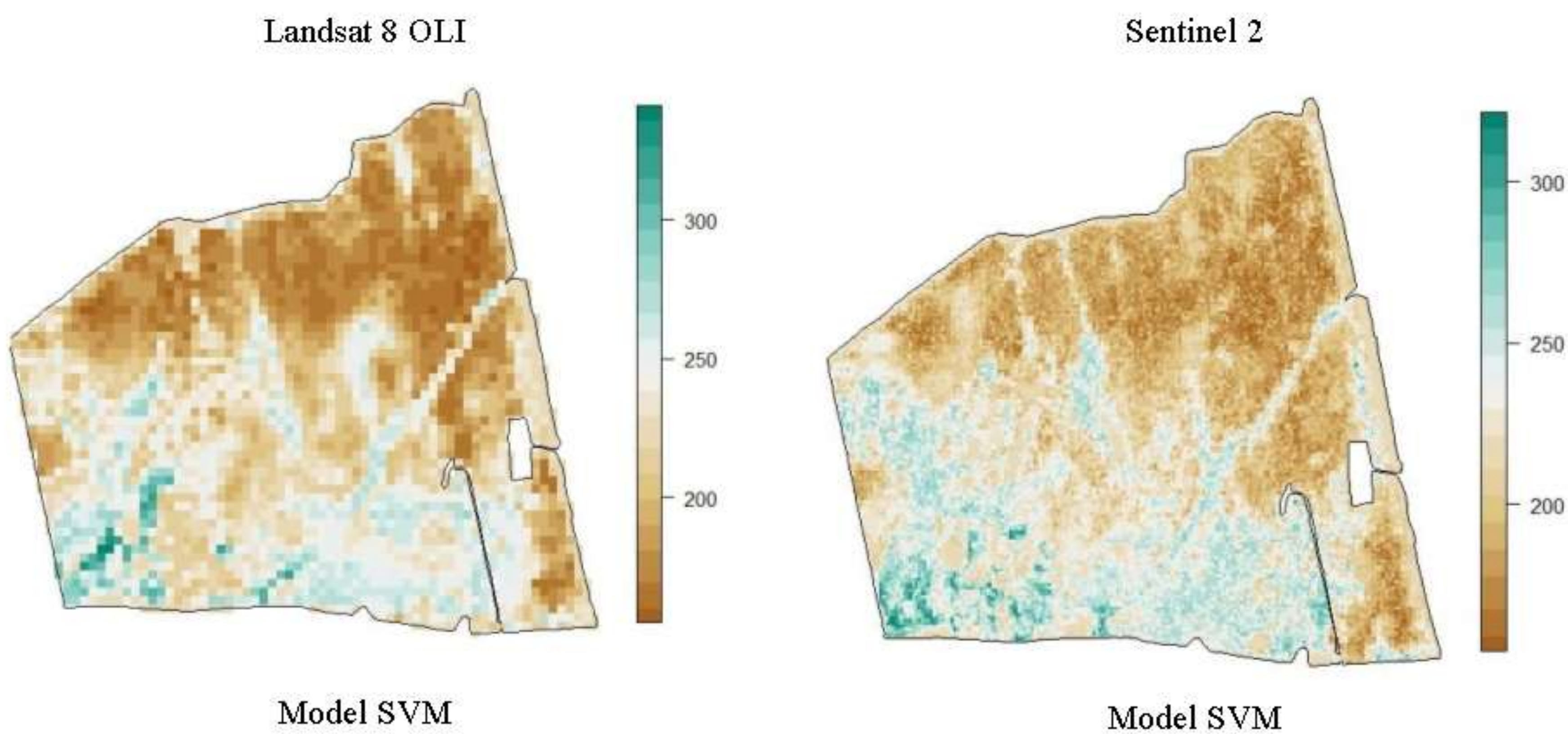
Table. Model performance estimates

Property	Model	RMSE	MAE	R^2	RMSE	MAE	R^2
		Landsat 8 OLI			Sentinel 2		
Hydrolysable nitrogen	MLR	11.00	0.86	0.69	12.38	0.92	0.60
	RF	9.21	0.28	0.83	8.51	0.47	0.85
	SVM_R	4.51	0.64	0.95	8.91	0.23	0.79
Available phosphorus	MLR	47.62	4.19	0.12	48.96	4.29	0.07
	RF	33.71	1.39	0.66	34.62	3.10	0.65
	SVM_R	33.39	6.80	0.57	35.26	7.83	0.52
Available potassium	MLR	28.80	2.27	0.57	31.59	2.56	0.48
	RF	23.89	1.72	0.77	25.34	1.48	0.74
	SVM_R	19.23	2.31	0.81	22.01	4.28	0.75
Soil organic carbon	MLR	0.53	0.04	0.51	0.49	0.04	0.58
	RF	0.35	0.02	0.83	0.35	0.02	0.84
	SVM_R	0.32	0.04	0.82	0.37	0.07	0.77
Silt	MLR	6.86	0.55	0.18	6.95	0.56	0.16
	RF	4.23	0.17	0.75	5.55	0.21	0.57
	SVM_R	2.64	0.34	0.87	2.84	0.46	0.76
Clay	MLR	3.02	0.26	0.09	2.99	0.26	0.11
	RF	2.24	0.15	0.61	2.23	0.21	0.61
	SVM_R	2.13	0.24	0.54	1.93	0.25	0.62


Landsat 8 OLI
Sentinel 2
5.0
4.5
4.0
3.5
3.0
Model RF
Model SVM
Soil organic carbon



Hydrolysable nitrogen



Available potassium

Conclusions

- The use of spectral reflectance data to refine digital maps of SOC content, available forms of nitrogen, phosphorus, potassium, PSD fractions on a single field scale provides a significant improvement in prediction when using machine learning methods (RF and SVMr) compared to traditional linear models (MLR).
- When using the RF and SVMr models, the SOC, available nitrogen, and available potassium is well predicted; the content of available phosphorus and PSD fractions is somewhat worse.
- Refined digital maps based on Sentinel 2 data are characterized by a greater degree of detail in the spatial variability of soil parameters; at the same time, the use of Landsat data can using machine learning methods, make it possible to obtain maps that are sufficient in terms of spatial prediction accuracy for the requirements of digital farming.

Thanks for attention !

This work was supported in part by the Russian Foundation for Basic Research, research project № 19-29-05061-mk.