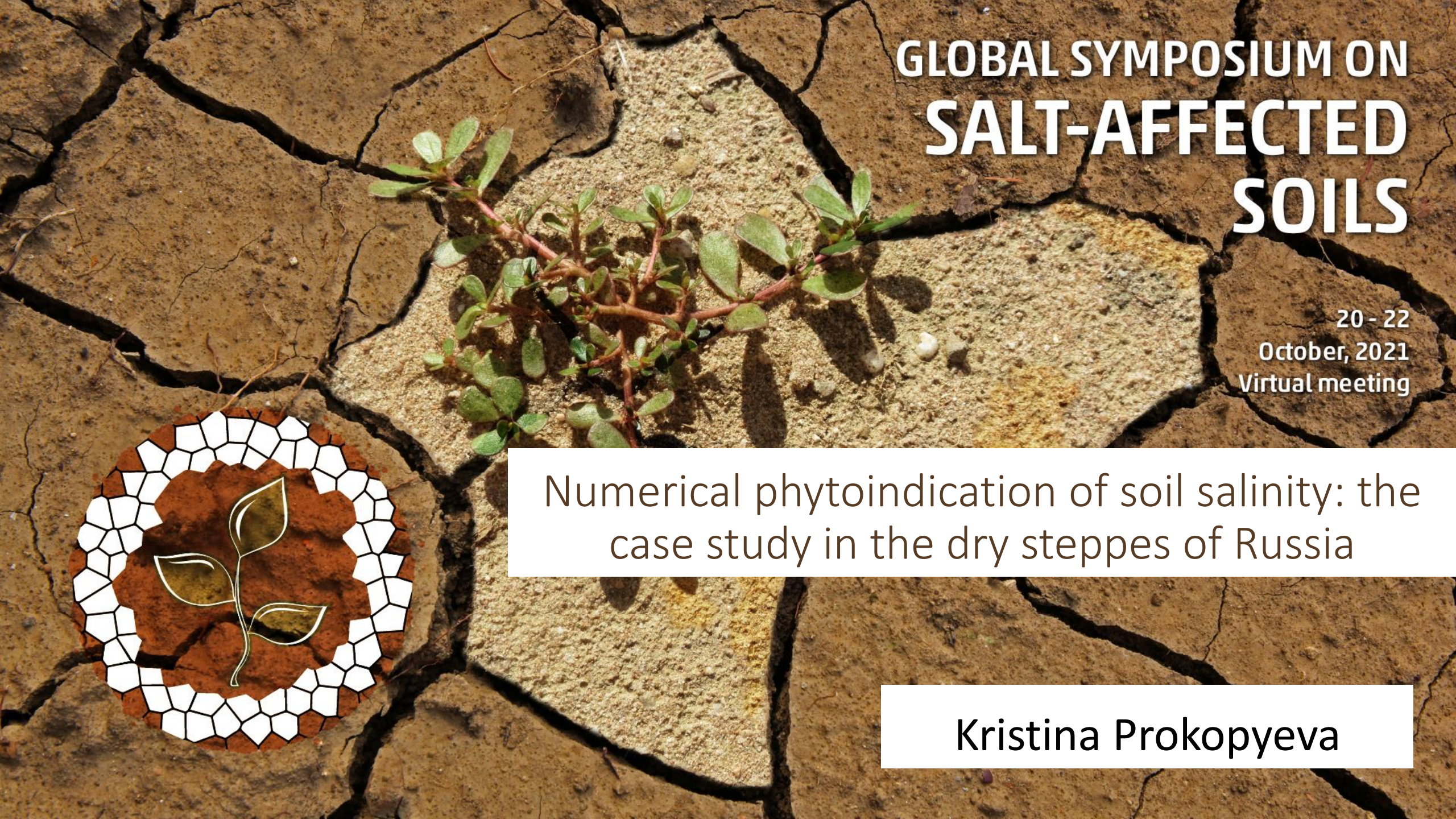


Numerical phytoindication of soil salinity: the case study in the dry steppes of Russia

Kristina Prokopyeva

Introduction

ML techniques such as Support Vector Machines, K-Means, Artificial Neural Networks, Decision Trees are currently being used in soil science for classification and mapping

- Phytoindication uses information about vegetation as an indicator of the state of the studied environmental components.
- Phytoindication is very powerful method, however:
  - **The published data** on the confinement of species and communities to quantitative salinity values and their validity as indicators **is very sparse**.
  - **No** commonly **accepted** methodical **approaches** to the acquisition of such data.
- Purpose of our study:
  - To identify the strength of the relationship between plant species and communities, on the one hand, and the salinity depth and degree in soils of the solonetzic complex, on the other.
  - To explore the possible production of quantitative predictions of the soil salinity based on the presence of certain plant species typical of the Sarpa lowland plain.

Experimental setup

Geographical location:

- Right bank of the Volga River
- Altitude: 8m above sea level
- Coordinates: 47.965° N, 45.551° E
- Soil types: light chestnut soils form complexes with solonetz soils (25–50% or more).

The field studies were carried out in September 2010 and April and September 2011.

Details of the *in situ* measurements:

- A 64-m transect was laid from the center of one microdepression to the center of another.
- Boreholes 1–2 m deep were drilled at 1-m intervals
- samples were collected from the following depths: 0–2, 2–10, 10–20, 20–30, 30–40, 40–50, 50–70, 80–100, 100–120, 140–160, and 180–200 cm.

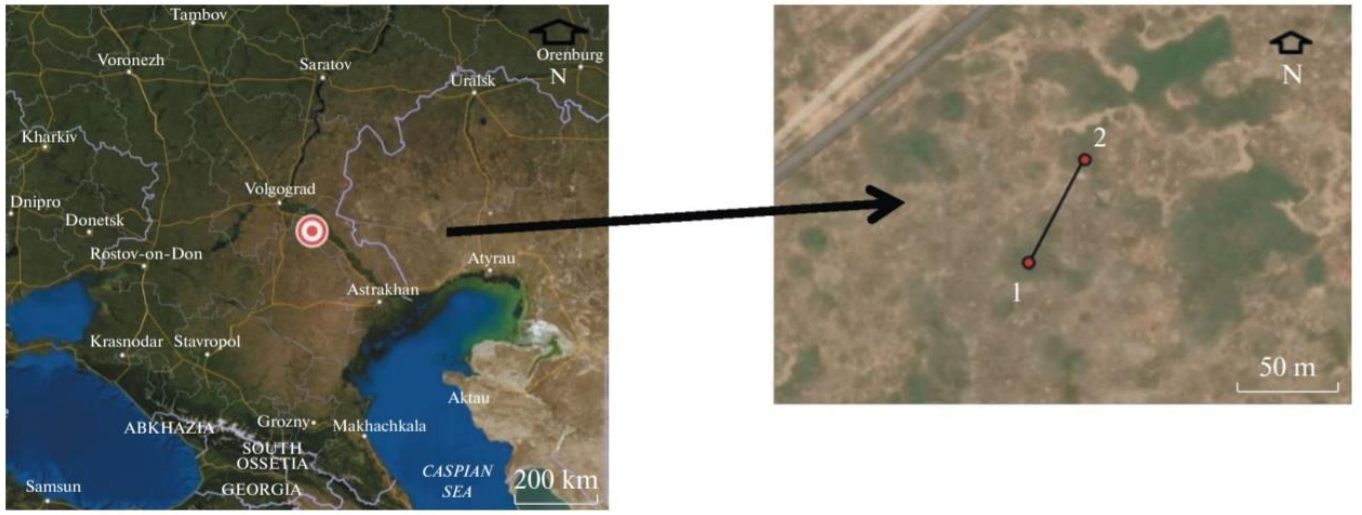
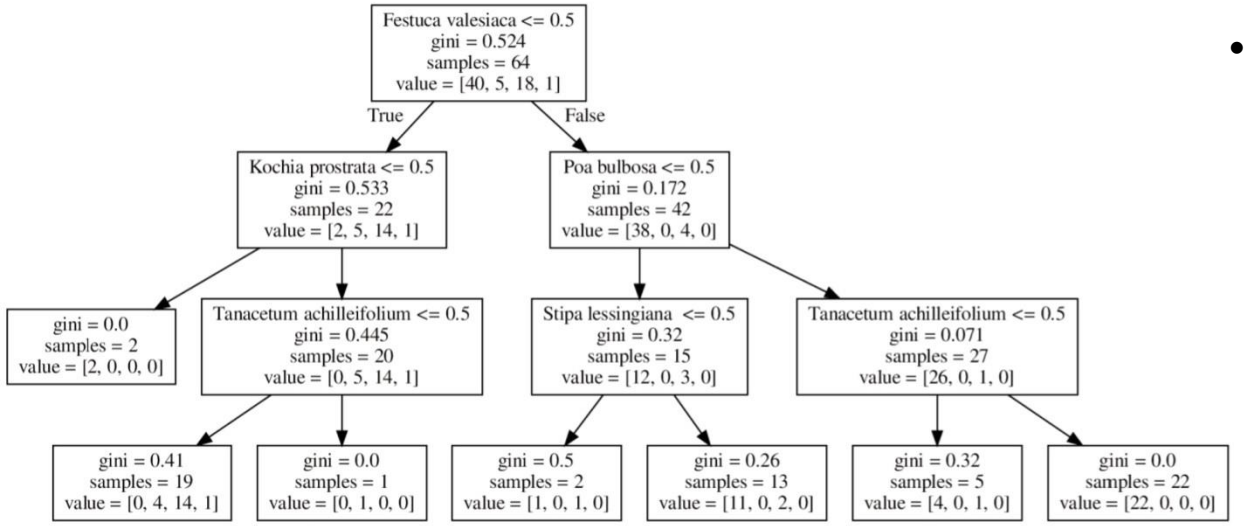


Fig. 1. Geographical location of the key site and transect. Coordinates of the transect beginning (1) and end (2) are 47.96472° N, 45.55139° E and 47.96524° N, 45.55180° E, respectively.

CART and Random Forest

- CART = The **C**lassification **A**nd **R**egression **T**ree (Breiman et al. (1984))
- A data-classification algorithm based on the sequential division of data based on decision rules (i.e., whether a certain plant species is present in the soil sample?).
- The algorithm selects the classification path with the maximum Gini coefficient value.
- In order to avoid overfitting, the number of levels is limited to three (see Figure)



- An *ensemble* of decision trees.
- Each of these trees is trained independently of others with a random data sample that constitutes 70% of the initial training sample.
- The number of levels in Random Forest is not limited, and this does not result in overfitting due to the large number of algorithms (decision trees) in the ensemble.
- After learning, new data are processed with each of the algorithms (decision trees), and the final prediction is produced “by majority vote”.
- In this study, the ensemble size is set to 100.

Results. Salinity profile

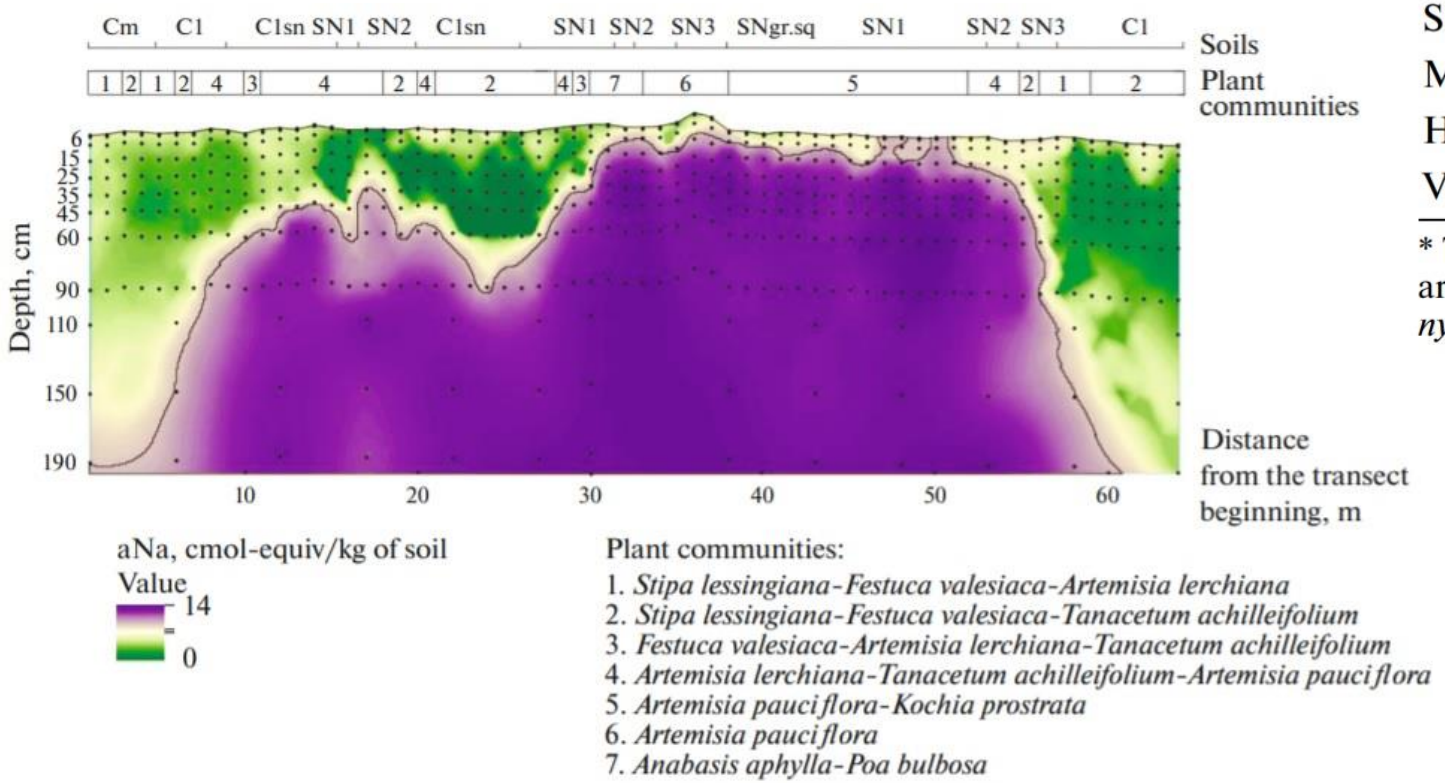


Fig. 2. Soil-salinity profile (aNa, cmol-equiv/kg of soil) along the transect and distribution of plant communities. Dots mark the location of soil samples and average sampling depths. The line divides the saline and nonsaline layers ($aNa_{1:5}$ = 1 cmol-equiv/kg of soil). Soil codes: (Cm) chestnut meadow, (C1) light chestnut, (C1sn) light-chestnut solonetzic, (SN0) crusted solonetz (0- to 5-cm horizon above the solonetz horizon), (SN1) shallow solonetz (5–10 cm), (SN2) medium solonetz (10–15 cm), (SN3) deep solonetz (>15 cm), and (SNgr.sq) solonetz under earth thrown from a ground squirrel's burrow.

Table 1. Assessment of the soil-salinity degree based on the content of Na ions and water-soluble salts per 100 g of soil

Salinity degree	Na^+, mmol-equiv/100 g of soil*
Nonsaline	1
Slightly saline	1–2
Medium saline	2–4
Highly saline	4–8
Very highly saline	>8

* The criteria for the chloride and sulfate–chloride salinity types are provided in accordance with *Saline Soils of Russia* (*Zasolennye...*, 2006).

The following soils are distinguished based on the upper boundary of the saline horizon:

- solonchak (the boundary is within the upper 30-cm layer),
- solonchak-like (30–100 cm),
- deep saline (150–200 cm)

Results. Prediction of CART/Random Forest

CART method:

- Accuracy of prediction: 51 of the 64 observations (80%) in the 30-cm layer, 52 of the 64 observations (81%) in the 50-cm layer, and 41 of the 64 observations (64%) in the 100-cm layer
- The plants important for classification: *Kochia prostrata, Tanacetum achilleifolium, Artemisia austriaca*, and *Festuca valesiaca* for the 30-cm layer*; F. valesiaca, T. achilleifolium*, and *K. prostrata* for the 50-cm layer; and *F. valesiaca* and *K. prostrata* for the 100-cm layer.

The Random Forest:

- The prediction accuracy for the test sample is 68% (0-30 cm), 84% (0-50 cm), and 72% (0-100 cm).
- The plants important for classification: *Festuca valesiaca* and *Tanacetum achilleifolium* in the 30-cm layer; *Festuca valesiaca* in the 50-cm layer; and *Artemisia pauciflora*, *Kochia prostrata*, *Tanacetum achilleifolium*, and *Festuca valesiaca* in the 100-cm layer.

Conclusions

- Statistical parameters of quantitative soil salinity values at the depths of 0–30, 0–50, and 0–100 cm were computed for each of the 12 plant species typical for solonetz soils.
- Three plant groups were identified based on their tolerance to soil salinity:
  - species confined to nonsaline soils (*Stipa lessingiana, Tanacetum achilleifolium, Festuca valesiaca, and Limonium caspium*)
  - species that tolerate a broad range of salinity values and prefer nonsaline soils (*Anabasis aphylla, Artemisia lerchiana, and Poa bulbosa*)
  - five salt loving species (halophytes) confined to saline and highly saline soils (*Bassia sedoides, Artemisia austriaca, Leymus ramosum, Kochia prostrata, and Artemisia pauciflora*)
- The application of the CART method made it possible to identify the indicatory importance of individual species for various soil-salinity degrees.
- The produced models predict the salinity degree with the accuracy of 80% for the 0- to 30-cm layer, 81% for the 0- to 50-cm layer, and 64% for the 0- to 100-cm layer.
- Based on the Random Forest method, the accuracy of the soil salinity indication is 79%
- Further plans: Add data from Sentinel-2B satellite into analysis in order to improve accuracy of soil salinity prediction

GLOBAL SYMPOSIUM ON SALT-AFFECTED SOILS

20 - 22
October, 2021
Virtual meeting