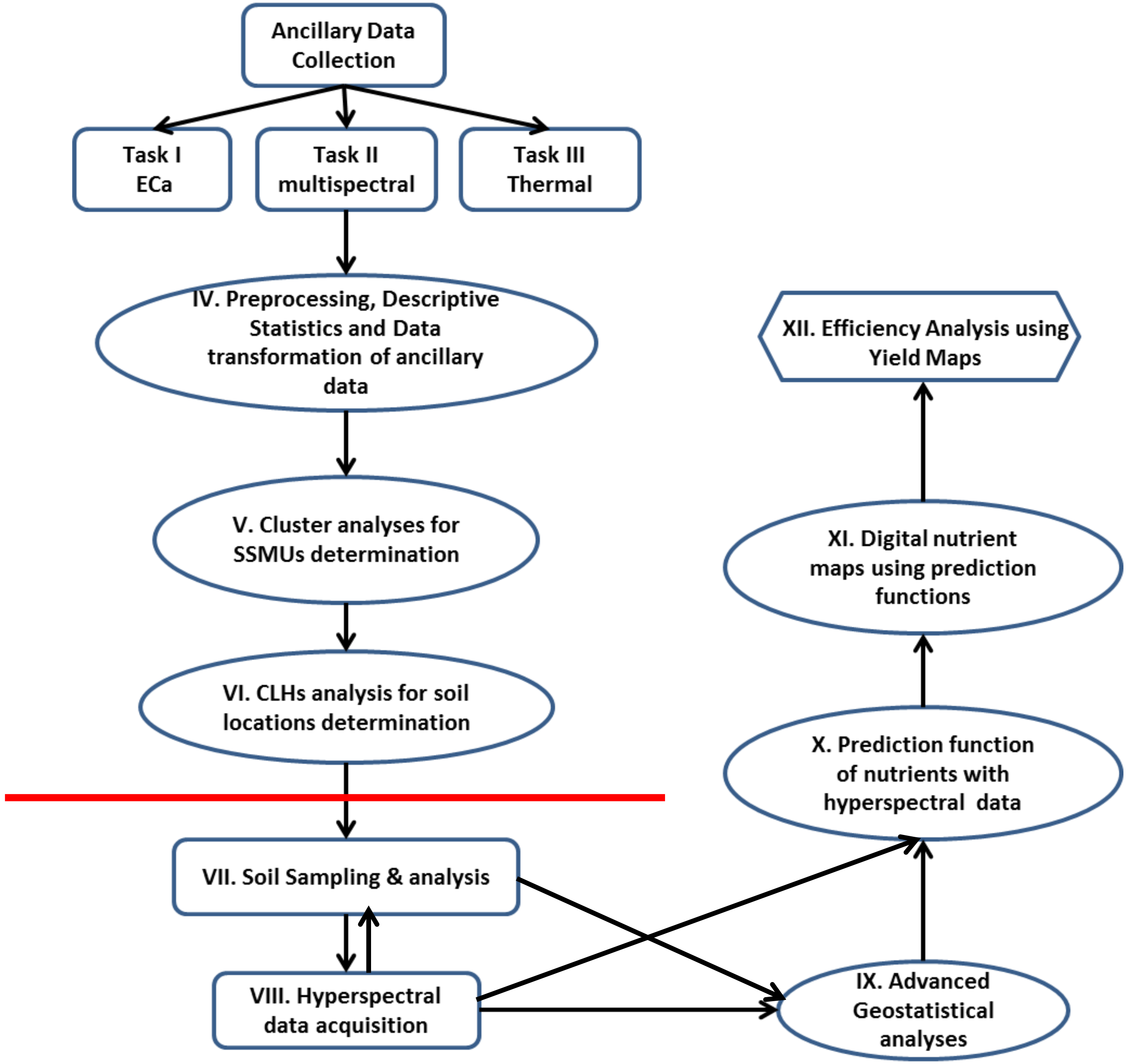
Integrated System for Optimal Soil Sampling using Precision Agriculture Concept

M. Iggy Litaor, Ofer Shir, Assaf Israeli, Oren Reichman, Nitzan Malachy and Shalev Malul

Flow Chart Diagram of the Integrated System: Optimal Soil Sampling

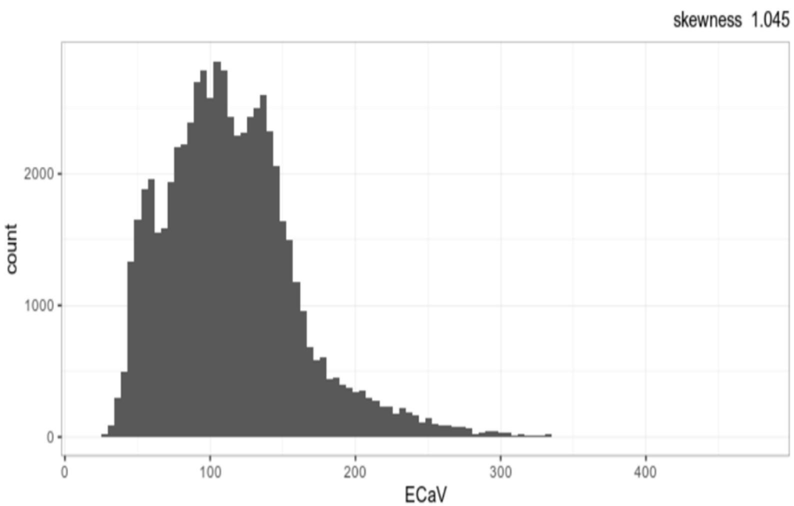


Pre-processing

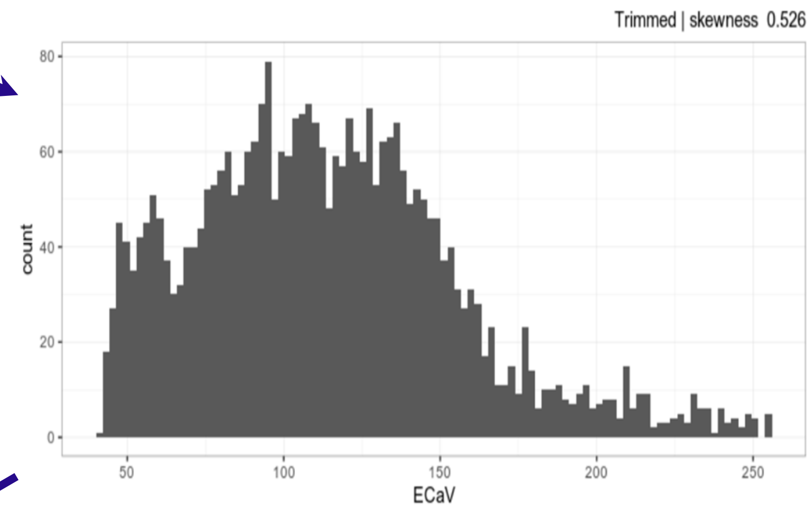
EM38 MK2 ECa Vertical

Cleaning
Merging
Compaction

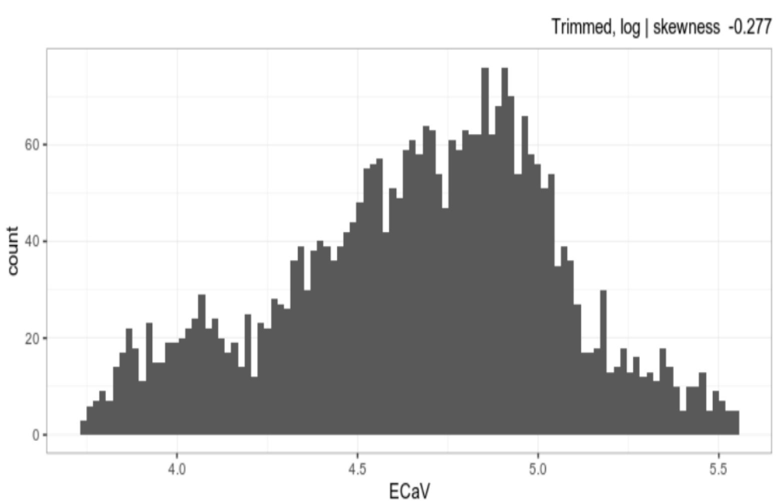
//UTM zone 36,	Datum: WGS84							
Easting[m]	Northing[m]	CV-1.0m	CV-0.5m	Quality	Sat	HDOP	Time	Date
704640.284	3621267.751	69.57	50.977	2	8	1.2	44:44.1	12/03/2018
704640.284	3621267.751	69.766	50.82	2	8	1.2	44:44.4	12/03/2018
704640.305	3621267.744	69.805	51.289	2	8	1.2	44:54.7	12/03/2018
704640.305	3621267.744	69.648	51.211	2	8	1.2	44:54.9	12/03/2018
704640.305	3621267.744	70.078	51.484	2	8	1.2	44:55.1	12/03/2018
704640.306	3621267.744	69.609	51.289	2	8	1.2	44:55.3	12/03/2018
704640.306	3621267.745	70.117	51.133	2	8	1.2	44:55.5	12/03/2018
704640.306	3621267.745	70.156	50.898	2	8	1.2	44:55.7	12/03/2018
704640.307	3621267.746	70.117	51.094	2	8	1.2	44:55.9	12/03/2018
704640.307	3621267.745	70.039	50.82	2	8	1.2	44:56.1	12/03/2018
704640.307	3621267.743	70.508	50.742	2	8	1.2	44:56.3	12/03/2018
704640.307	3621267.741	70	50.781	2	8	1.2	44:56.5	12/03/2018
704640.308	3621267.739	70.312	50.859	2	8	1.2	44:56.7	12/03/2018



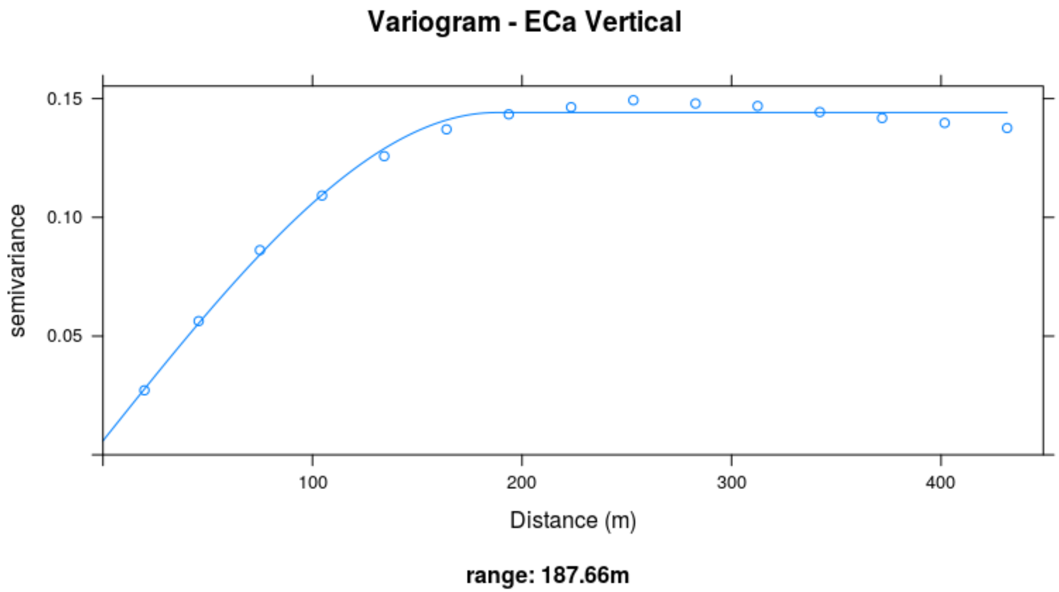
Trim to 99% percentile (~3 S.D.)



log

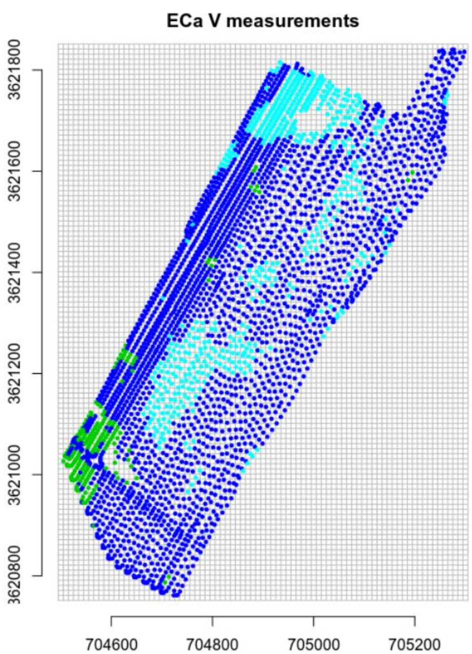


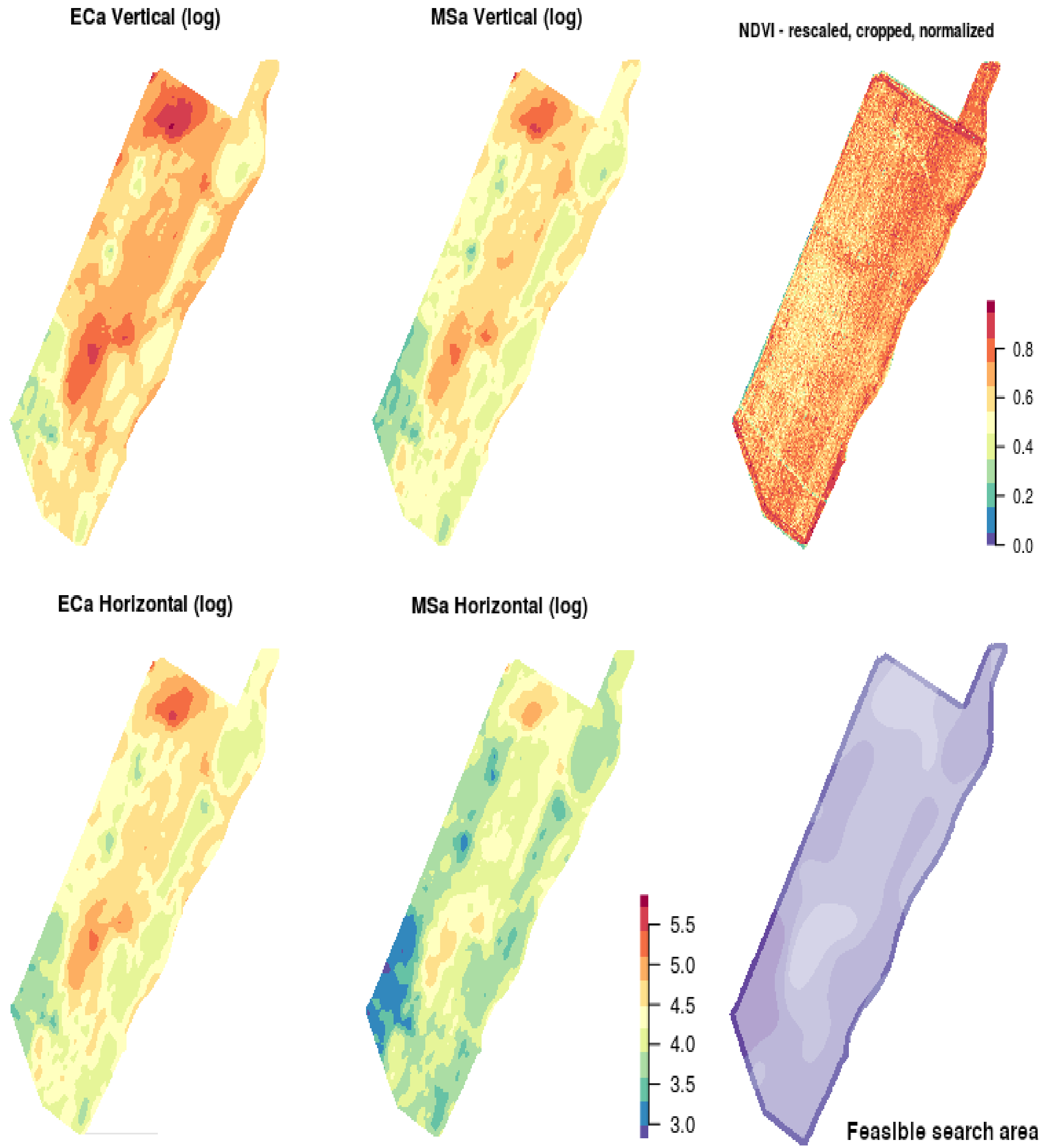
Spatial Model



Ancillary Data

Crop / Clip by perimeter mask
Point data ⟶ Variogram
Ordinary Kriging (1 x 1 m)

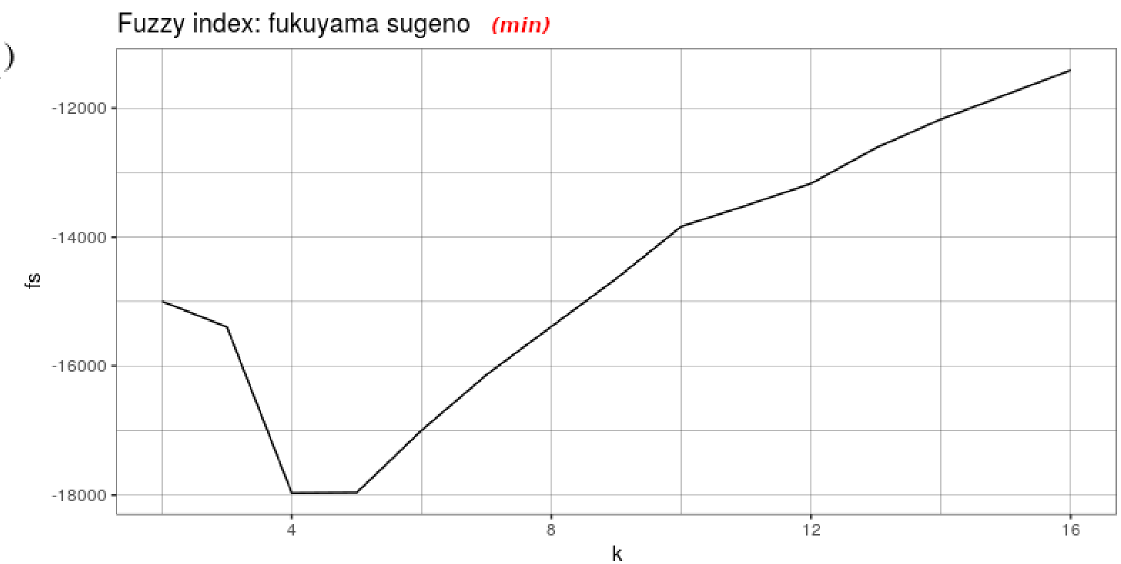


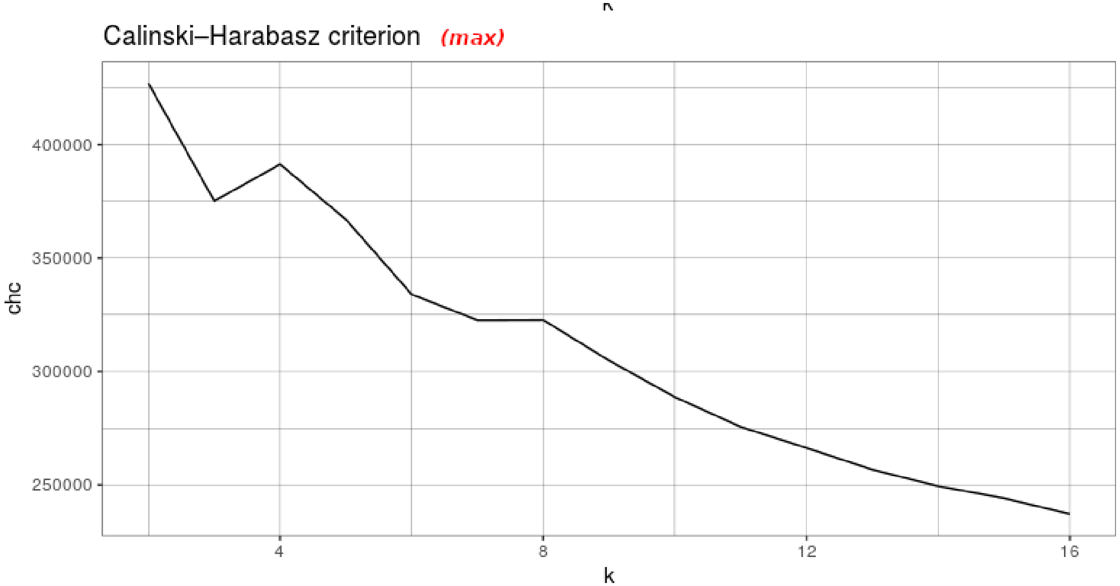


Determination of MZ using Validation Index for fuzzy c-means

$$V_{\mathrm{FS},m}(\boldsymbol{U},\boldsymbol{V};X) = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^{m}\left(\|x_k - v_i\|^2 - \|v_i - \bar{v}\|_A^2\right)$$

x_k – kth data point
v_i – cluster center
c_i – number of clusters
$\bar{U}$ – grand mean of all data x_k
u_{ik} – membership of data x_k of class ci

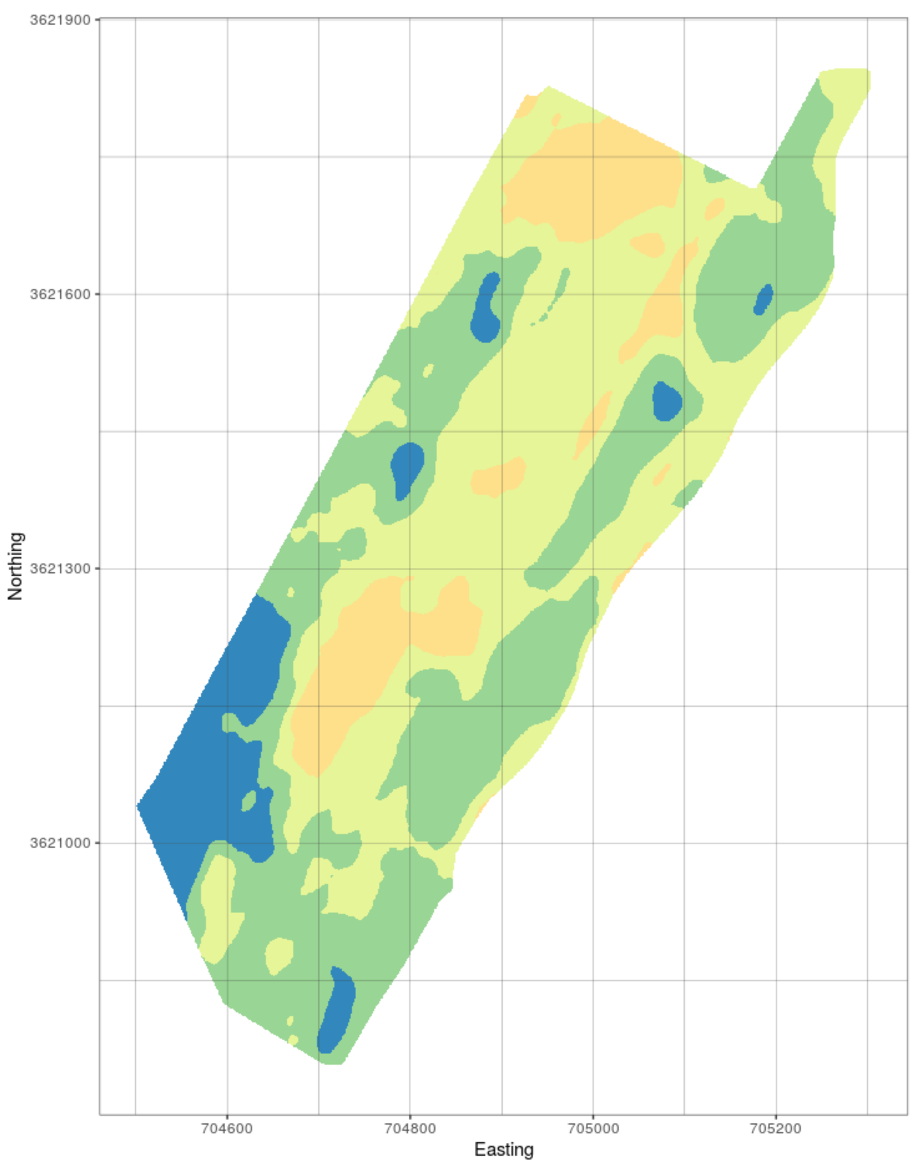




$$CHC = \frac{BMZSS/(MZn-1)}{WMZSS/(N-MZn)}$$

N – number of data points
MZ – management zones
BMZSS – between MZ sum of squares
WMZSS – within group sum of squares

Management Zones Determination using Fuzzy c-means cluster Analysis



SSMU	ECa V	MSa V	ECa H	MSa H	count
Zone 1	0.221	0.189	0.165	0.2	33226
Zone 2	0.425	0.382	0.366	0.419	140964
Zone 3	0.572	0.525	0.539	0.573	152468
Zone 4	0.736	0.675	0.728	0.737	46275

Optimal Sampling Design

- The main objective is to design sampling plan where soil sampling sites ***n*** **<<** ***N*** ancillary data, that best represent the soil domain and also spatially dispersed in geographical space.
- **Conditional Latin Hypercube Sampling (*cLHS*) solves a single objective optimization problem by maximizing the stratification of the multivariate distribution of *N* by forming a Latin hypercube of their quantiles.**
- *cLHS* results in perfect stratification of the ancillary data in the soil domain but fails to account to distribution of important/interesting locations within the soil domain.
- We devised a new procedure by augmenting the cLHS single objective function with spatial dispersion objective function (Israeli et al., *in review*).

cLHS - conditioned Latin Hypercube Sampling

Stratification of the feature space

Candidate samples p are evaluated by:

1. Number of occurrences in each quantile - 1. (optimal=0)

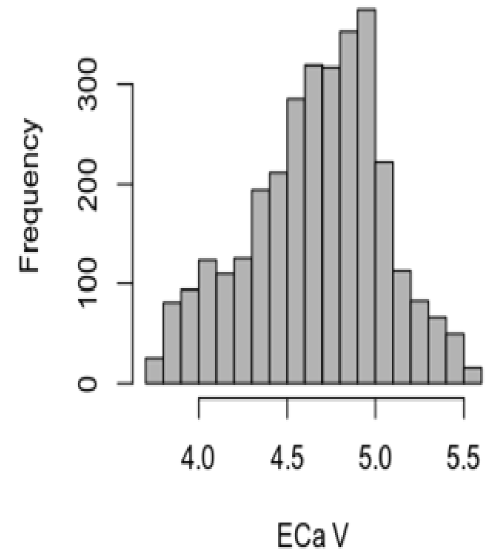
$$\psi_1\left(\mathbf{A}^{(p)}\right) = \sum_{i=1}^{n}\sum_{j=1}^{k}\left|\eta\left[q_j^{(i)} \leq \alpha_{\pi(i),j} \leq q_j^{(i+1)}\right] - 1\right|$$

2. Correlation matrices difference. (similar=0)

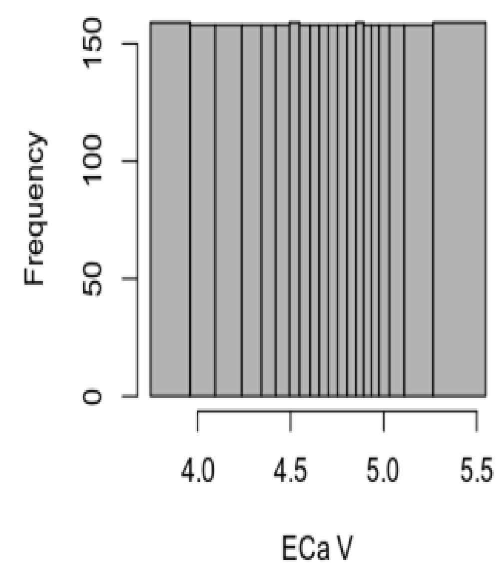
$$\psi_2\left(\mathbf{A}^{(p)}\right) = \sum_{i=1}^{k}\sum_{j=1}^{k}\left|\mathbf{C}_{i,j}^{(\mathcal{A})} - \mathbf{C}_{i,j}^{(\mathbf{A}^{(p)})}\right|$$

$$f_{\mathrm{cLHS}}(p) = \omega_1 \cdot \psi_1\left(\mathbf{A}^{(p)}\right) + \omega_2 \cdot \psi_2\left(\mathbf{A}^{(p)}\right) \longrightarrow \min$$

Distribution



Distribution by quantiles



Multi-Objective Optimization

$$\min(f_1(x), f_2(x), \ldots, f_k(x))$$
$$\text{s.t. } x \in X,$$

Where k is the number of objectives and X defines the feasible search space.

A feasible solution $x^1 \in X$ is said to (Pareto) dominate another solution $x^2 \in X$, if

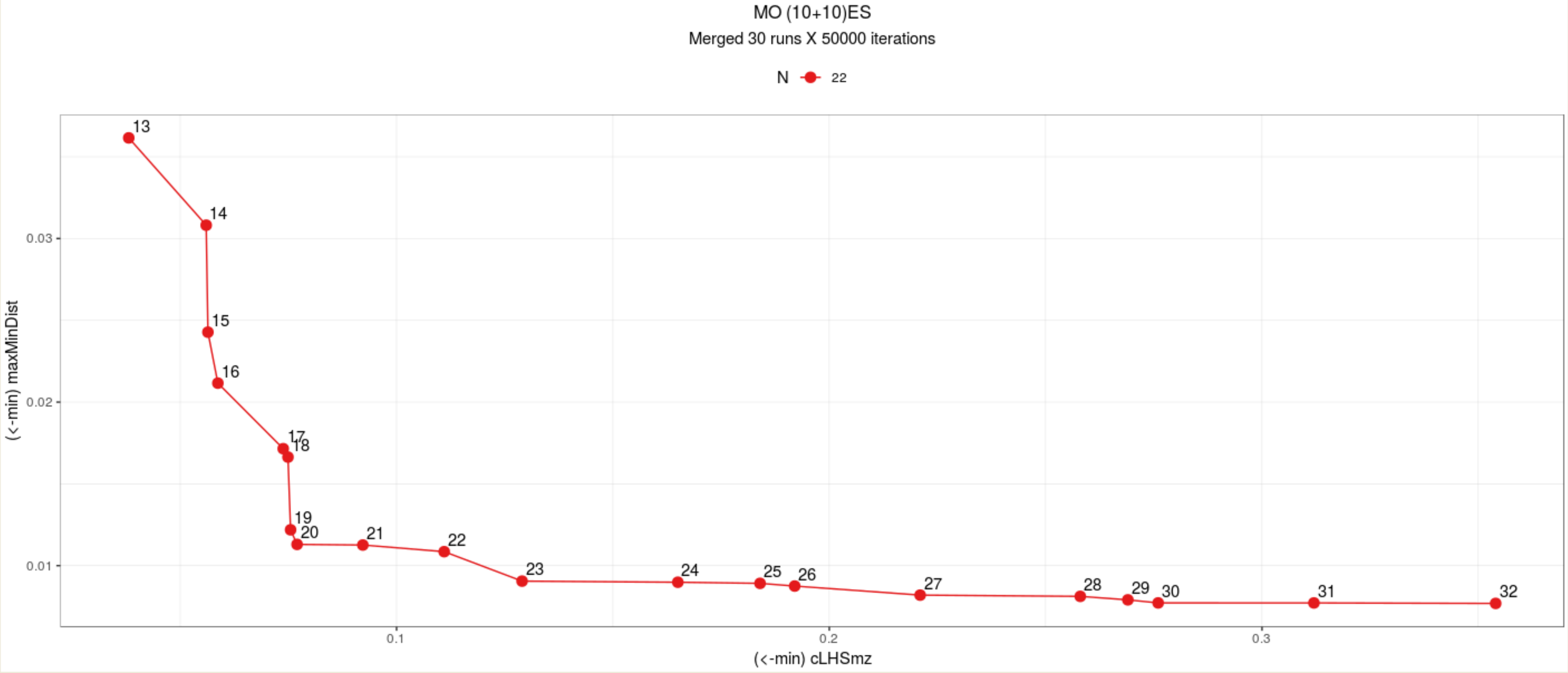
1. $f_i(x^1) \leq f_i(x^2)$ for all indices $i \in \{1, 2, \ldots, k\}$ and
2. $f_j(x^1) < f_j(x^2)$ for at least one index $j \in \{1, 2, \ldots, k\}$.

A solution $x^* \in X$ (and the corresponding outcome $f(x^*)$) is called Pareto optimal, if there does not exist another solution that dominates it.

The goal is to obtain the non-dominated set for $F = f(X)$, entitled the *Efficient Frontier*, and its pre-image in X, the *Pareto optimal* set.

Multi-objective Optimization to Determine n Soil Samples

$$\text{minimize } f_1(\mathbf{x}), \ldots, \text{ minimize } f_m(\mathbf{x}), \mathbf{x} \in \mathcal{X}$$



Efficiency fronts obtained using optimization of cLHS and Max-min distance per sample size N
The Max-min function aims at maximize the minimal pairwise distances among all sampling points:

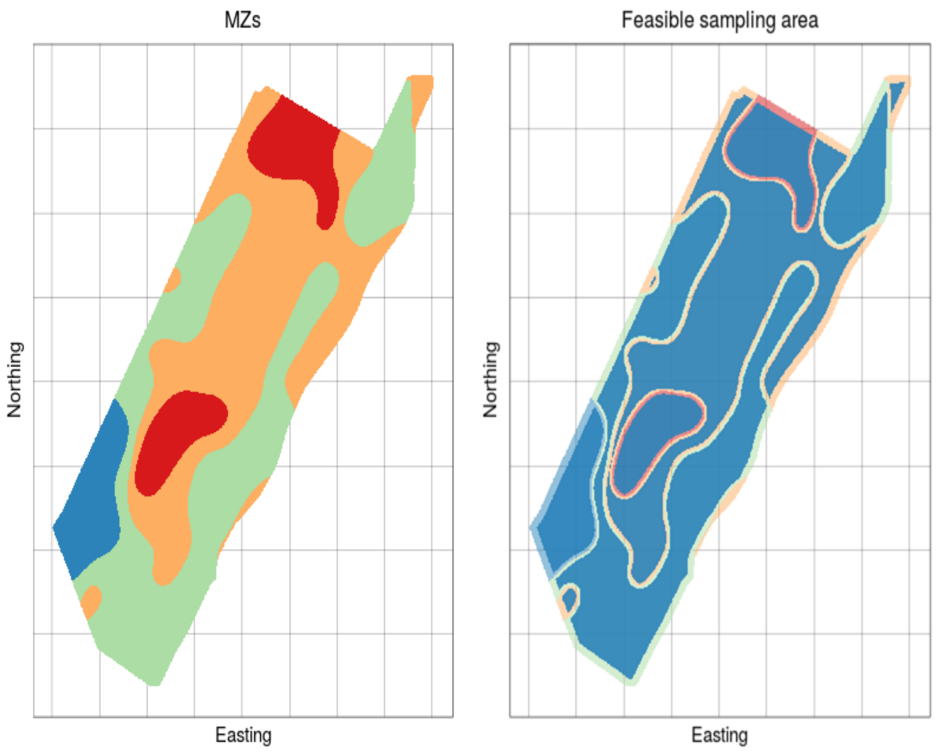
$$f_{d_{\min}^{(G)}}(p) = \min_{\pi(i),\pi(j)} \left\{ d_{\pi(i),\pi(j)}^{(G)} \right\} \longrightarrow \max \qquad i, j \in 1, \ldots, n, \; i \neq j.$$

Sample Design as Bi-Objective optimization problem

$$f_1 := 1 / f_{d_{\min}}(p) \longrightarrow \min$$

$$f_2 := f_{\mathrm{cLHS}}(p) \longrightarrow \min$$

Feasible sampling area defined with edge-detection filter: gap of 7*m* from MZs boundaries

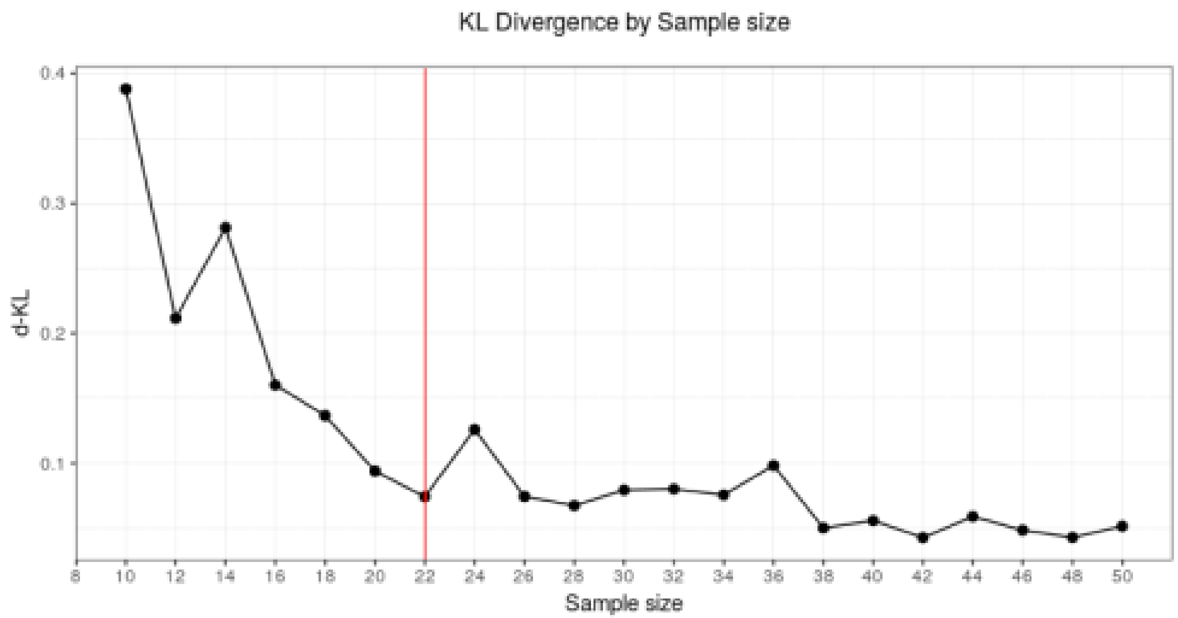


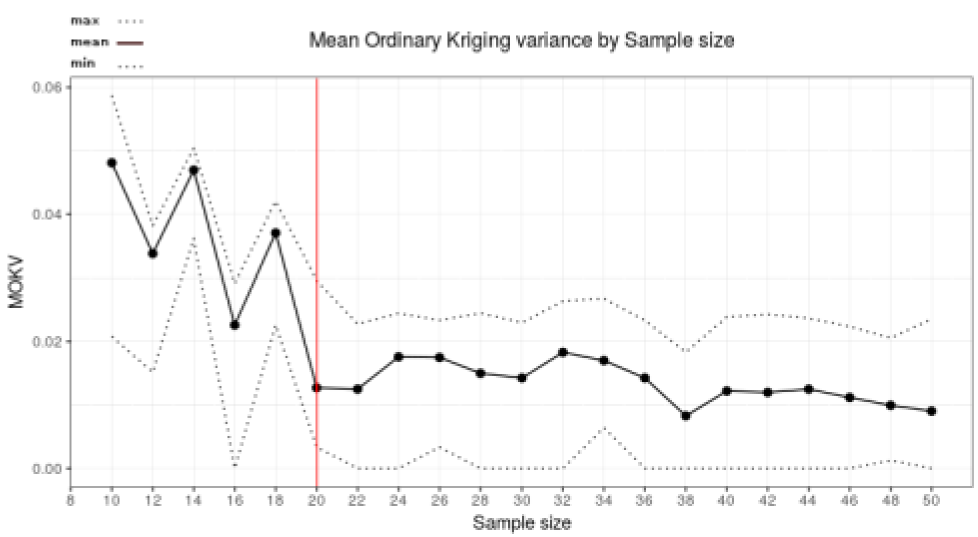
Additional constraints:
minimum 3 points per each MZs

Optimization Test for number of sampling

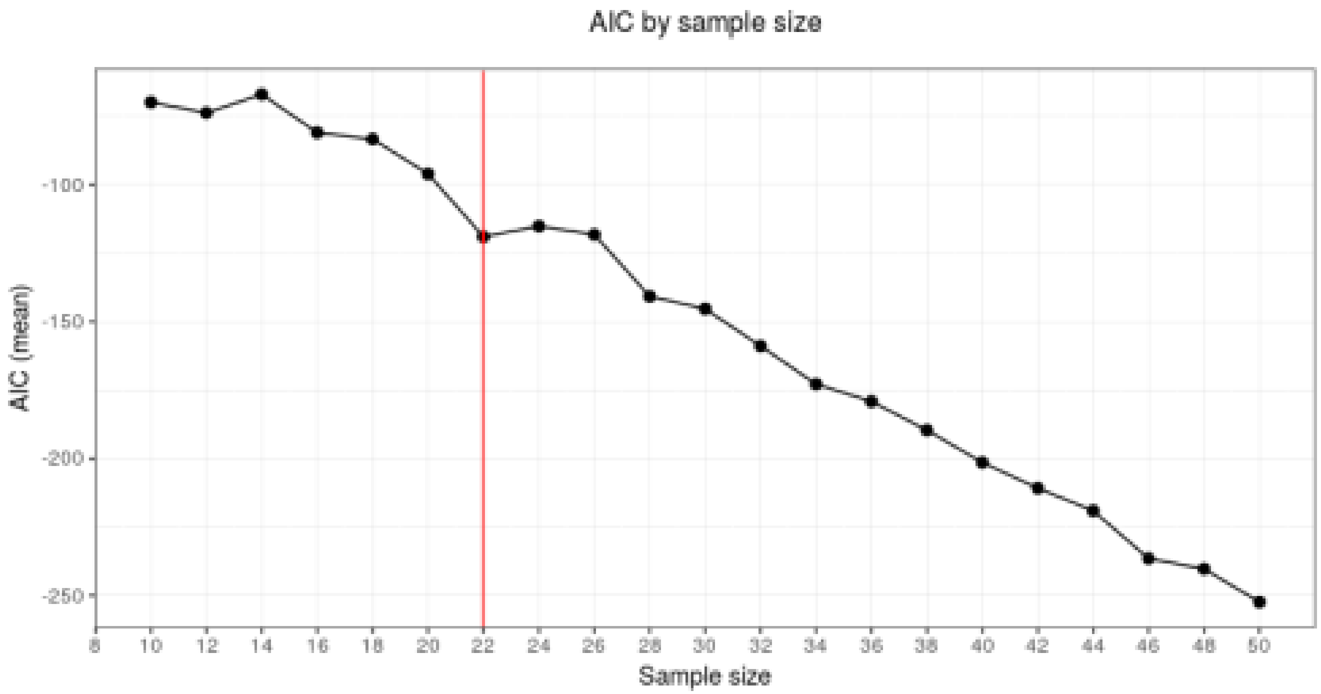
KL- Divergence (d-KL) of the ancillary data probabilities distribution compared with full field to produce indices by sample size. The index decrease as the model's goodness of fit improves.

$$D_{KL} = -\sum P(A) \log \left(\frac{P(A^{(p)})}{P(A)} \right)$$



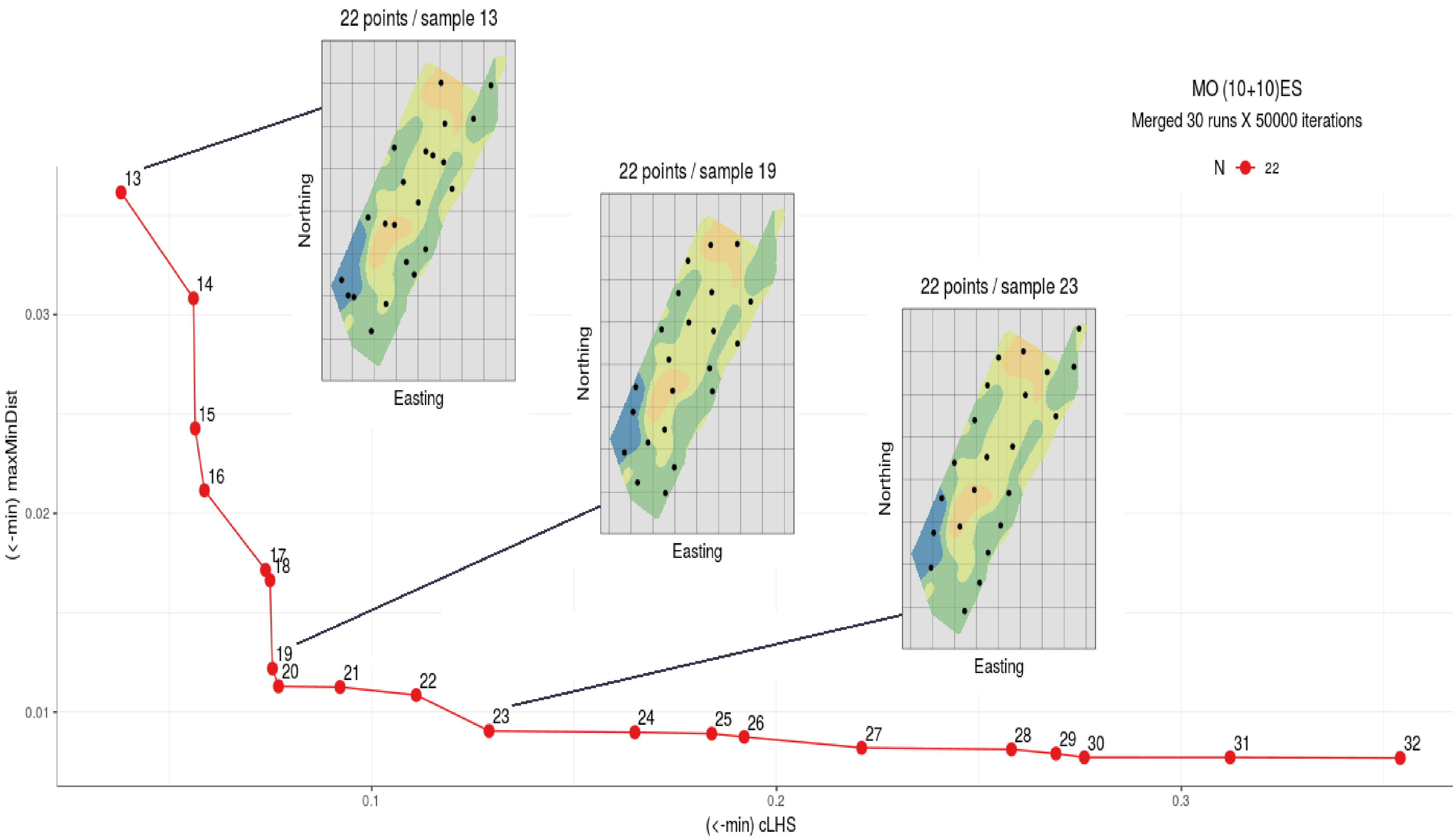


Optimization test for number of sampling



AIC – Akaike Information Criterion is an estimator of the relative quality of statistical models for a given set of data. Given collection of models AIC estimates the quality of each model relative to each of the other models, so AIC provides means for model selection.

Multi-objective evolutionary optimization algorithm



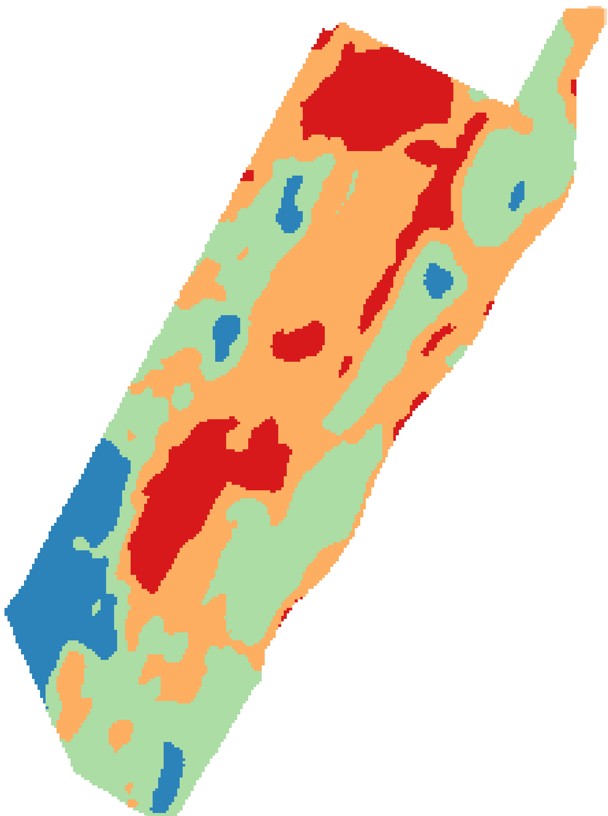
Spatial Smoothing of the Management Zones

original

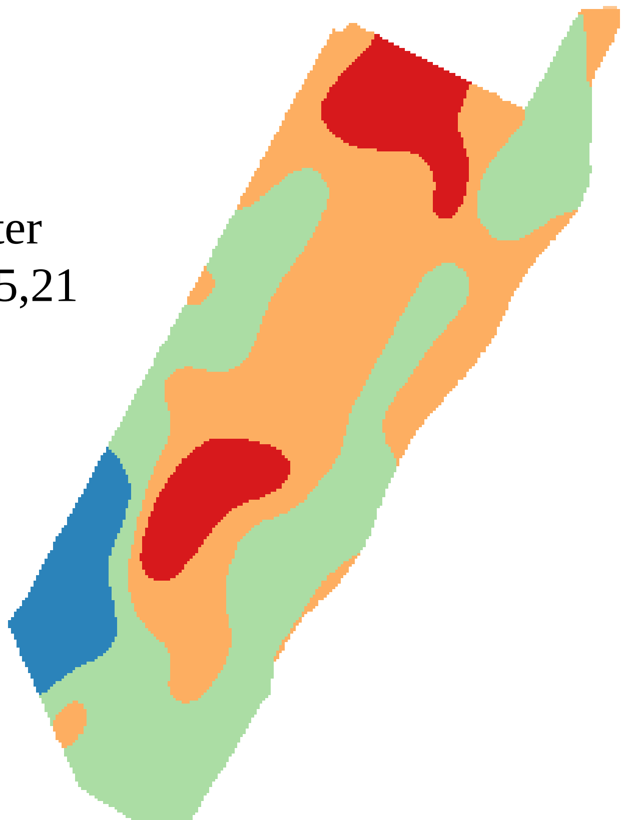
The *median filter* is an effective method that can, to some extent, distinguish out-of-range isolated noise from legitmate image features such as edges and lines. Specifically, the median filter replaces a pixel by the median, instead of the average, of all pixels in a neighborhood w

$$y[m,n] = median\{x[i,j], (i,j) \in w\}$$

where w represents a neighborhood defined by the user, centered around location $[m,n]$ in the image.



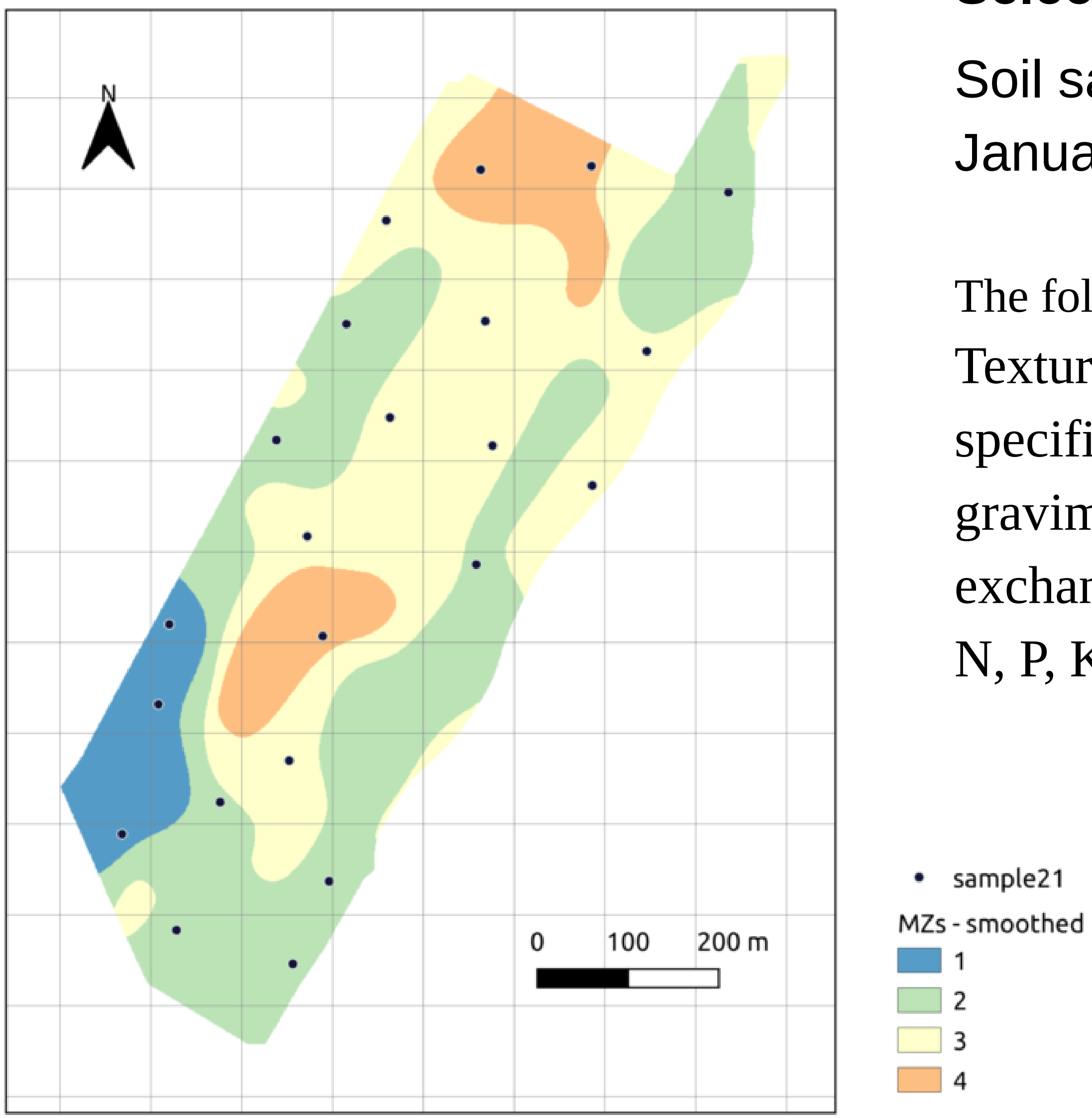
Median filter
57,57,35,35,21



The numbers represent the size and times of the moving windows runs

Soil Sampling map at Neve Yaar Research Station

Neve Yaar - Selected Soil Sampling Scheme (#21)

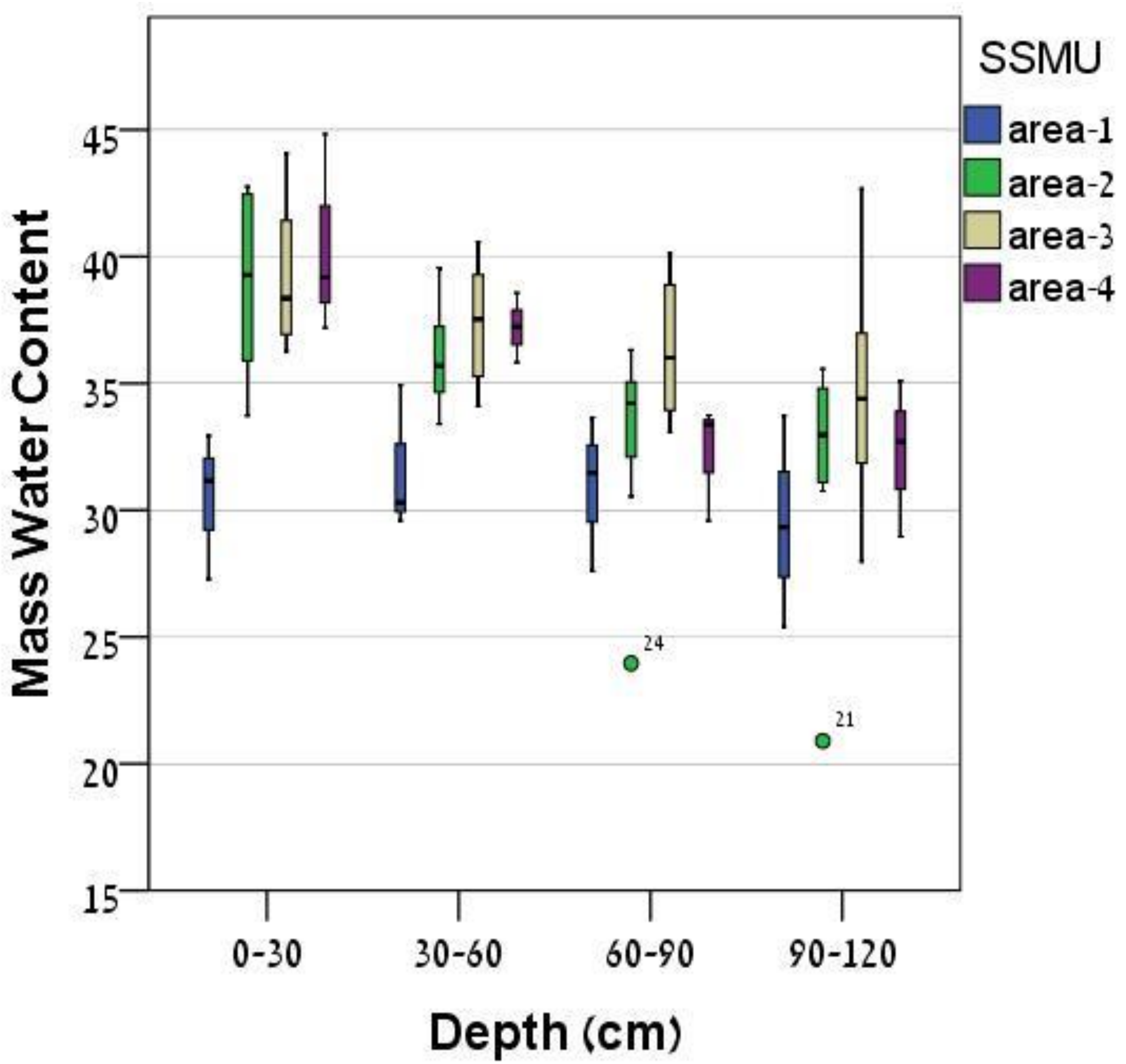


Selected Scheme of 22 points.

Soil sampling conducted on January 27th, 2019.

The following analyses are underway: Texture, hydraulic conductivity, specific surface, bulk density, gravimetric soil moisture, CEC, exchangeable ions, SOM, available N, P, K, pH, ECe, $CaCO_3$ content.

Preliminary Soil Results



Conclusions

- We devised a sampling plan by locating $\boldsymbol{n} \ll \boldsymbol{N}$ sites whose ancillary data (ECa) vectors best represent the soil attribute space's distribution and concurrently are optimally dispersed in the geographical space.
- We wrote a multi-objective evolutionary optimization algorithms (MOEAs) combined with problem-specific search operator (Israeli et al., *in review*).