SHORT TERM SCIENTIFIC MISSION (STSM) – SCIENTIFIC REPORT

Action number: **CA15140**
STSM title: **Nature-Inspired Multi-Objective Optimization in Agricultural Geo-Statistics**
STSM dates: **14/08/2018 to 24/08/2018**
Grantee name: **Assaf Israeli**

PURPOSE OF THE STSM

The aim of this short term scientific mission was to advance my capabilities of Multi-Objective Optimization (MOO) methodologies implemented by Evolution Strategies (ES), applied to our current geo-statistical research problem of efficient soil nutrients content prediction using a minimal set of soil samples.

Ancillary data is collected to assess the spatial variability of a given field, thus assisting in formulating a soil sampling scheme, and serving as a trend for the Universal Kriging interpolator. Such data includes soil Apparent Electrical Conductivity (ECa) and aerial imagery from which Normalized Difference Vegetation Index (NDVI) is derived, among other available sources such as crop yield, digital elevation maps or other relevant spatial data.

Prior to this STSM, field experiments were conducted, in which a sampling scheme was devised by optimization of the Conditional Latin Hypercube Sampling (cLHS) objective function that stratifies the marginal distribution of the ancillary data, using an existing R package with the renowned heuristic of Simulated Annealing (SA). This approach has some inherent drawbacks in terms of computational efficiency and in being indifferent to spatial distribution of the sampling points, therefore it was followed by the space-filling algorithm Spatial Coverage Sampling Design to compensate for the spatial dispersion.

One main concern of this activity, which was not previously addressed is the number of sampling points - a parameter that is directly related to the total operational cost. It was suggested to optimize a bi-objective function that would minimize the cost (number of points) while maximizing the information in the sample (an information criterion). The theoretical background and tools required to carry out such an approach were lacking, so we turned to Prof. Emmerich for bridging the gap to utilizing state-of-the-art multi-objective optimization methods.

DESCRIPTION OF WORK CARRIED OUT DURING THE STSM

During the visit we have revised the soil sampling methodology and considered a number of paths to explore.

Work carried out and tasks included:

- Introduction to MOO concepts.

- Some theory on Kriging, random processes and multi-dimensional search spaces.

- Code implementation of ES-based Multi-Objective algorithms (using ecr package in R).

- Integration of the cLHS objective function in the R procedure.

- Study and realization of several candidate objective functions to represent the information criteria of the sample. Some functions make use of ancillary data (ECa) as a "ground truth" reference, among them are Kullback–Leibler divergence (KL divergence), Kolmogorov–Smirnov test (KS test), chi-squared test (χ^2 test), , and Residuals Root Mean Squared Error (RMSE). These were coupled with an independent objective function of Mean Ordinary Kriging Variance (MOKV).

- Consideration of a spatial dispersion measures such as max-min distance or Hausdorff distance, either as a constraint or as an objective to maximize.

- Presentation and analysis of the unfolding efficiency fronts, assessing the feasibility of these objective functions as targets for optimization.

DESCRIPTION OF THE MAIN RESULTS OBTAINED

Initial results suggest there is no discrepancy (conflict) between the cLHS and several information-based objective functions, therefore they are suitable to be used as simultaneous targets for optimization. However, these results were obtained by short runs with a minimal number of iterations, thus a systematic evaluation on a high-performance machine should follow.

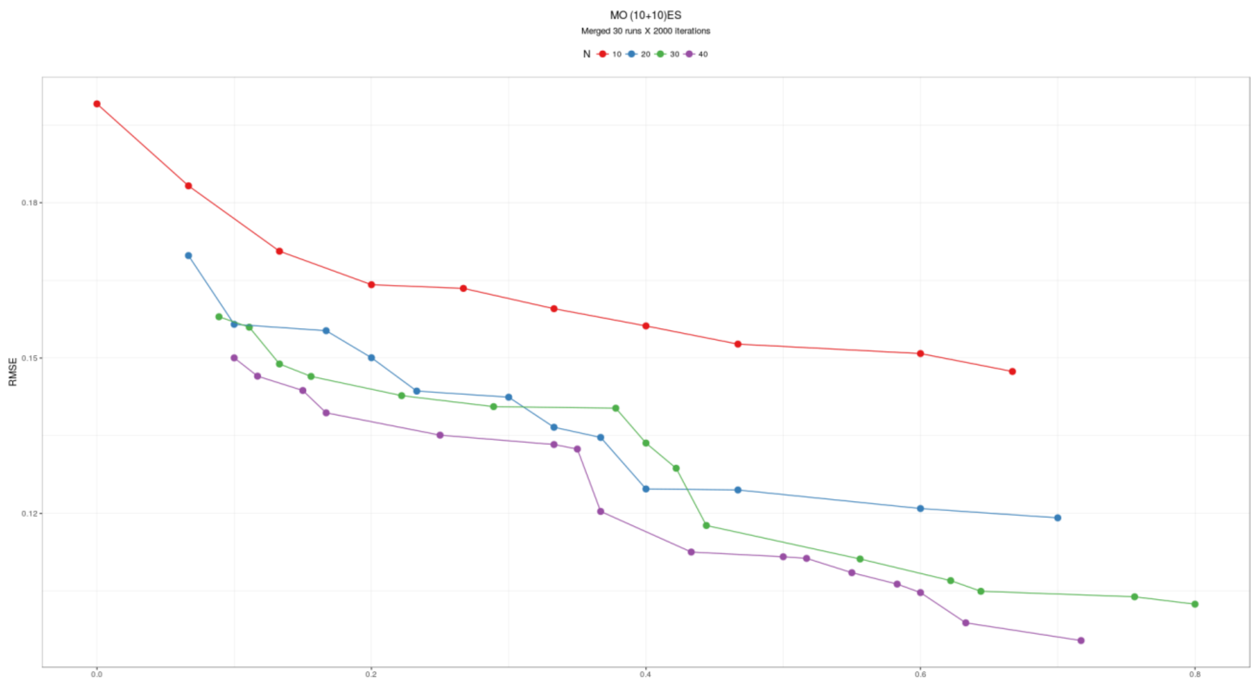
New tools were created to realize bi-objective optimization of the sampling problem at stake, with cLHS as an objective function along another information index, with the number of sampling points as another parameter that is gradually increased in steps (10, 20, 30, 40). Each run results in an efficiency (Pareto) frontier for several sample sizes, presenting the attainable trade-off of the underlying characteristics of each sampling scheme - that is cLHS and information ranks, and the corresponding operational cost (reflected by the number of samples).

This information can support the decision process of a farmer considering the scope of a soil sampling operation, while simultaneously providing an efficient sampling scheme for each given cost step.

An interesting idea that was raised suggests solving the cLHS problem by initializing the search with a state of hyper-dimensional identity (diagonal), which forms a Latin Hypercube, then mix the search by hyper-planes' swaps that preserve this property.

Another promising idea, yet to be verified for applicability, is dividing the sampling operation into phases, in each a few soil samples are extracted and analyzed, then the data is used to compute the next sampling locations by relying on the Kriging variance, until a satisfactory level of confidence is attained.

Exemplary results obtained with the tools developed during the STSM are shown below, depicting the Pareto frontiers of the bi-objective optimization task to minimize cLHS and RMSE.



FUTURE COLLABORATIONS

This multi-disciplinary STSM has strengthened the collaboration between MIGAL/Tel-Hai and LIACS/Leiden University within the COST action of Improving Applicability of Nature-Inspired Optimization by Joining Theory and Practice (ImAppNIO).

The tools and methods established during the mission will be refined and tested on a field scale research, to be carried out in the Upper Galilee, Israel. The application of such methods in agricultural geo-statistics constitutes a novel approach to the best of our knowledge, and should accordingly be reported in the scientific literature. We intend to maintain this collaboration in order to assess our scientific hypothesis, and once validated, we plan to report it as a joint publication on the subject. Furthermore, we expect that additional research questions will be raised in this future work, and we hope to keep this collaboration alive for addressing them.