Compressing soil structural information into parameterized correlation functions

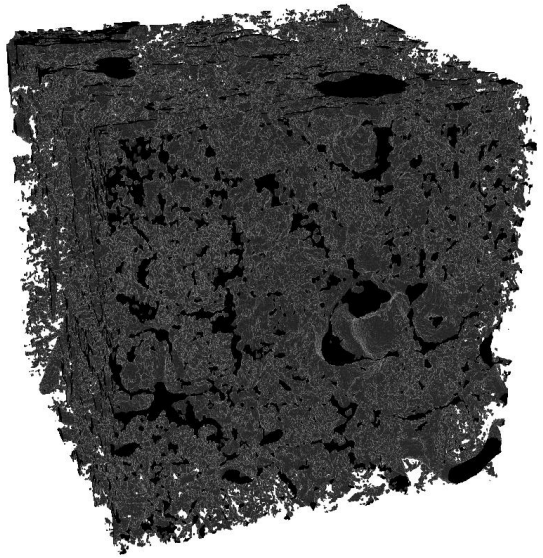
[1]Marina V. Karsanina, [1,2]Efim V. Lavrukhin, [3]Dmitry S. Fomin, [3]Anna V. Yudina, [3]Konstantin N. Abrosimov, [1,*]Kirill M. Gerke

[1]*Schmidt Institute of Physics of the Earth of the Russian Academy of Sciences (IPE RAS), Moscow, Russia*

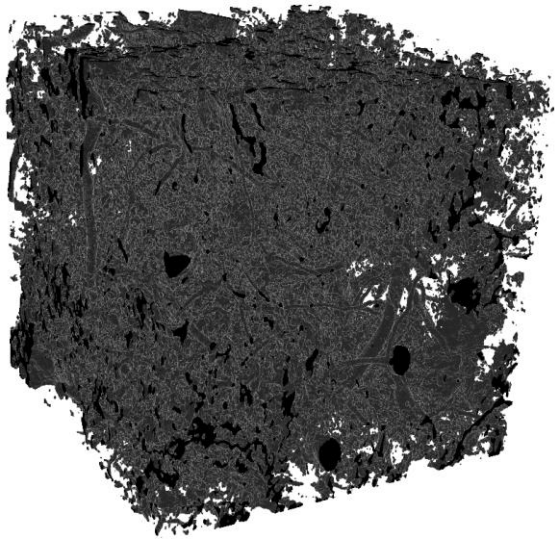
[2]*Computational Mathematics and Cybernetics, Lomonosov Moscow State University, Moscow, Russia*

[3]*Dokuchaev Soil Science Institute of Russian Academy of Sciences, Moscow, Russia*

EGU2020 – for live chat



In total 16 3D images with sizes of 900-1300^3 voxels with porosities of 0.009-0.202. They were classified into two and three groups according to soil use.



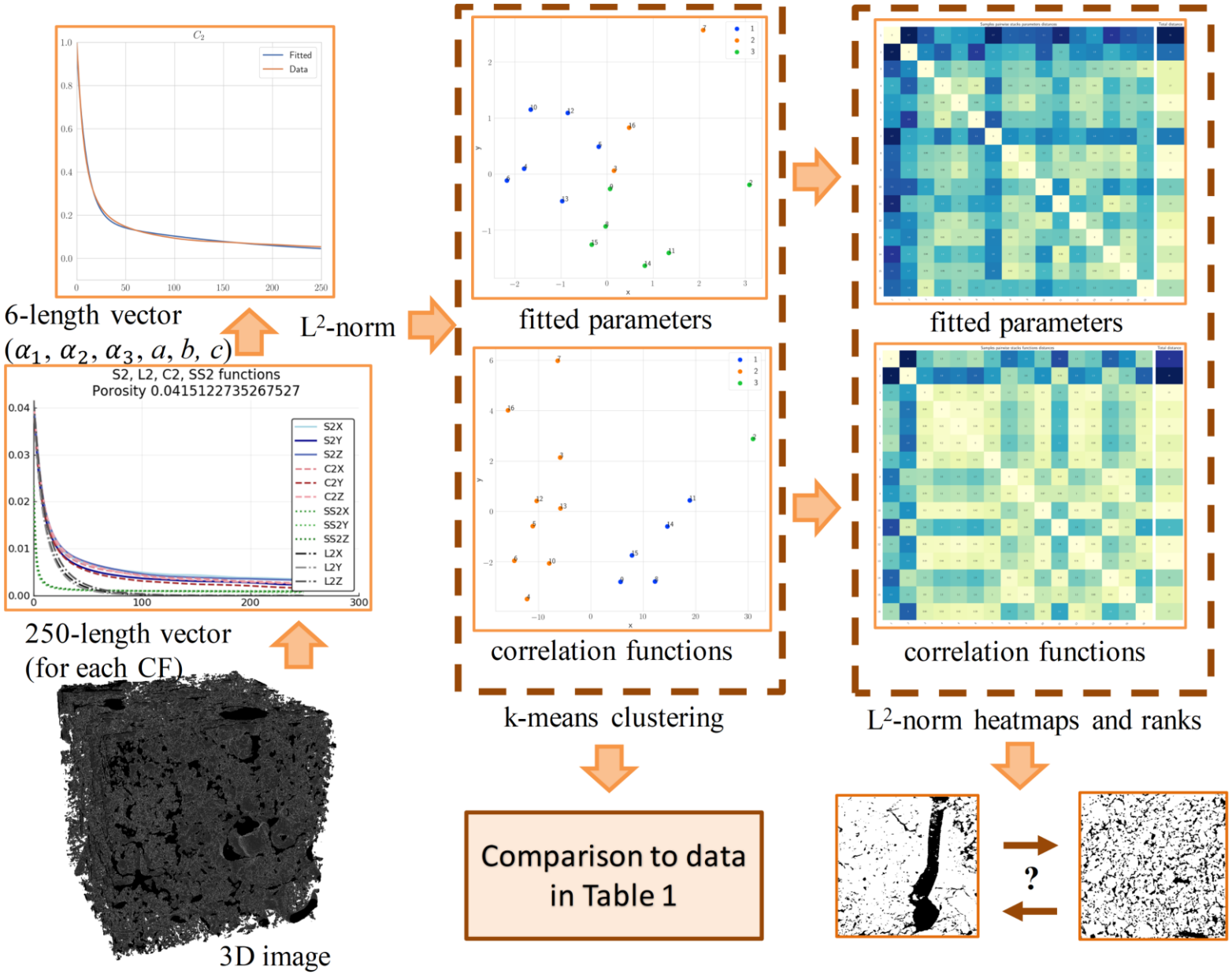


Fig.1. The general scheme of all structure data processing starting from 3D pore images: CFs computation (250 parameters for each CF) → CFs fitting to 6 parameters → cluster analysis → L^2-norm analysis → comparison against land use information → detailed comparison of all results to pore images.

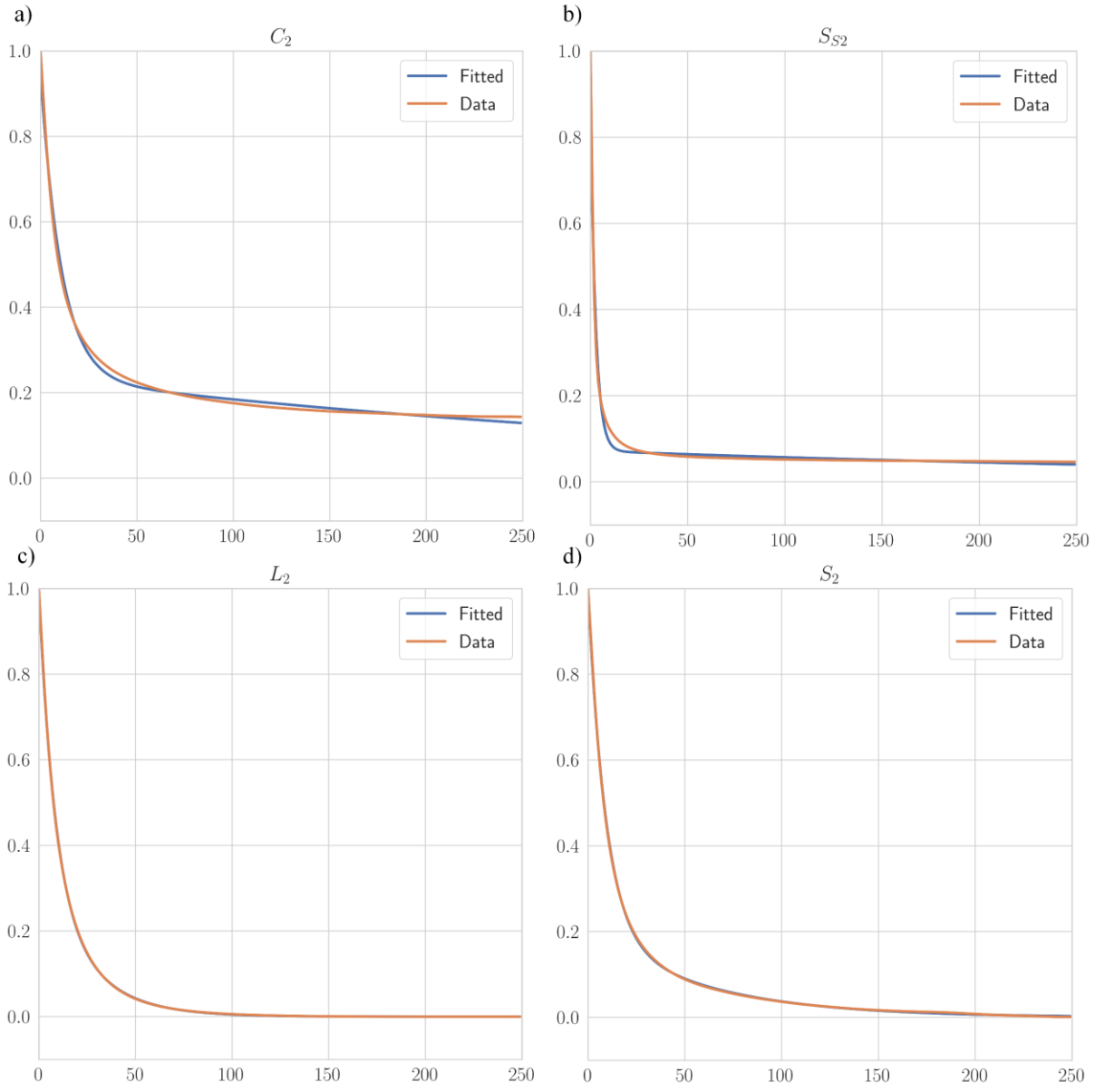
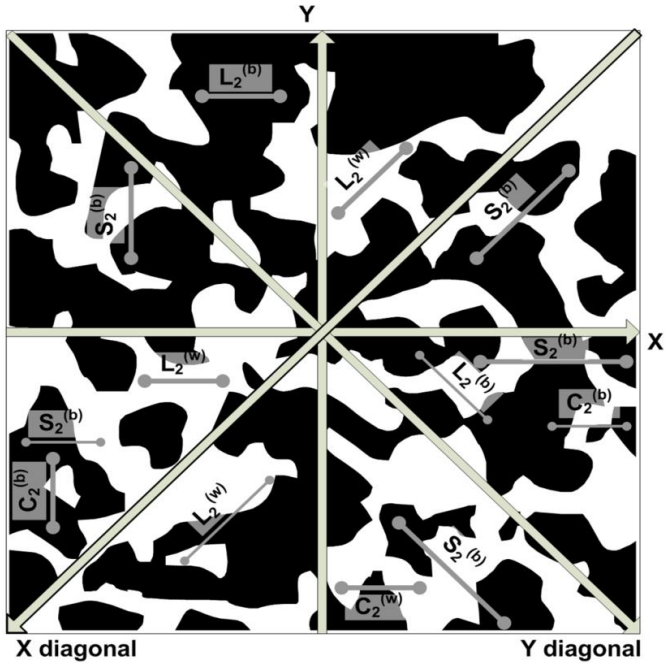


Fig.2. The worst (upper row) and best (lower row) fits (Fitted) using Eq.3 to describe computed correlation functions (Data) for: a) ensemble C_2 function for Sample 13, b) ensemble SS_2 function for Sample 1, c) ensemble L_2 function for Sample 8, and d) ensemble S_2 function for Sample 9.



Calculation of directional correlation functions: S_2-two-point probability, L_2 – linear, C_2 – cluster functions.

$$S_{2norm}(\boldsymbol{r}) = \frac{S_2(\boldsymbol{r}) - \varphi}{\varphi(1-\varphi)}$$

Normalization by porosity for all utilized correlation functions: two-point probability, lineal, cluster and surface-surface.

Each correlation function was fitted with the superposition of the three most flexible basis functions, i.e., exponential, dumped oscillating and polynomial functions as proposed by Jiao et al. (2008):

$$\begin{cases} f(r) = \alpha_1\, exp(-r/a) + \alpha_2\, exp(-r/b) + \alpha_3 \begin{cases} (1-r/c)^2, 0 \le r \le c \\ 0, r > c \end{cases} \\ \alpha_1 + \alpha_2 + \alpha_3 = 1 \\ \alpha_1 \ge 0, \alpha_2 \ge 0, \alpha_3 \ge 0 \\ a \ge b \end{cases}$$

where α_1, α_2 and α_3 the weights for each basis function, a, b and c are the scaling coefficients for the basis functions, q is oscillating frequency and ψ is the phase angle.

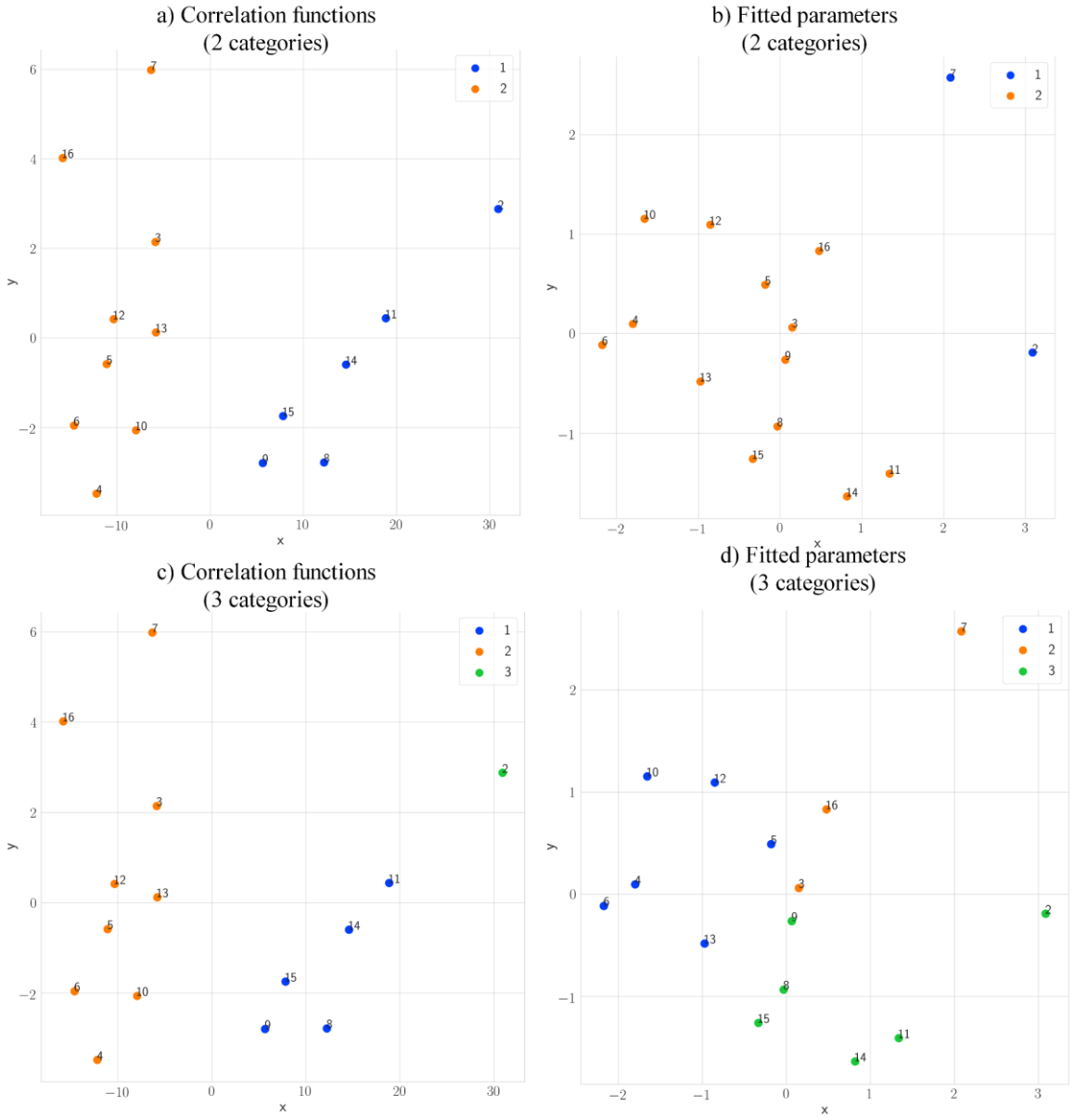


Fig.3. The results of clustering of: a) based L^2-norm for 250-length vector of ensemble average for all CFs into 2 categories, b) for 6-length vector of fitted parameters into 2 categories, c) for 250-length vector of ensemble average for all CFs into 3 categories, and d) for 6-length vector of fitted parameters into 3 categories. Note that sample 1 is removed from the cluster analysis.

The clustering is completely different for CFs and fitted parameters. It turned out that the parameters may be very sensitive to the smallest changes in the functions. In other words, two correlation functions than differ only slightly may have parameters that vary significantly. On the other hand, if parameters are very close for two CFs, this means that they are indeed matching. See next slide.

The clustering based on either CFs or parameters was unable to catch the trends in soil use (2-category classification: arable and post arable, 3-category classification: bare fallow, afforestation, cereals)

One can immediately pose a question if we can find such basis functions that fitted parameters always behave similar to CFs. This is a non-trivial question from the mathematical point of view, and our current vision suggests it is unlikely that any potential set of basis functions may result in such fits.

Regardless of future direct applicability of fitted parameters one can always “unpack” correlation functions from such parameters (Fig.2) to utilize for analysis or machine learning. In this sense fitting using basis functions is still a very useful procedure to compress data.

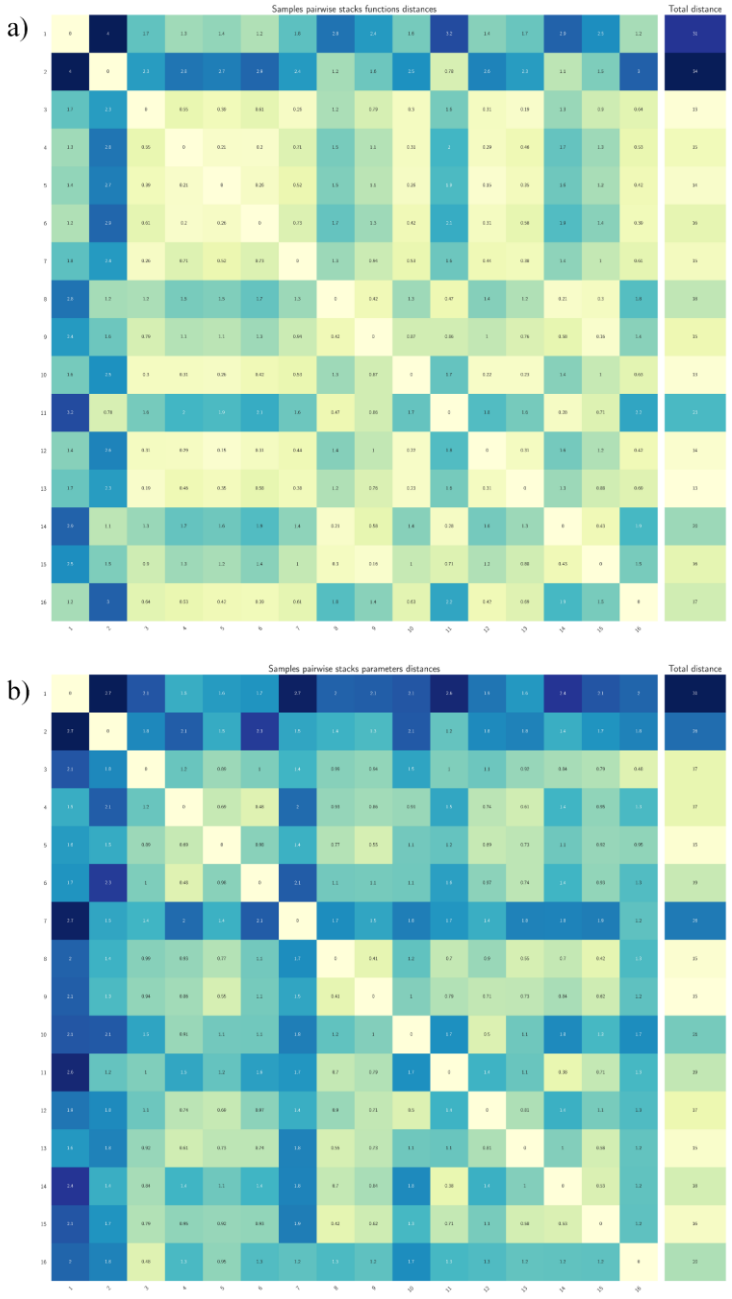


Fig.4. Heatmaps of differences between soil samples in terms of: a) ensemble averaged correlation functions and b) fitted parameters for these correlation functions.

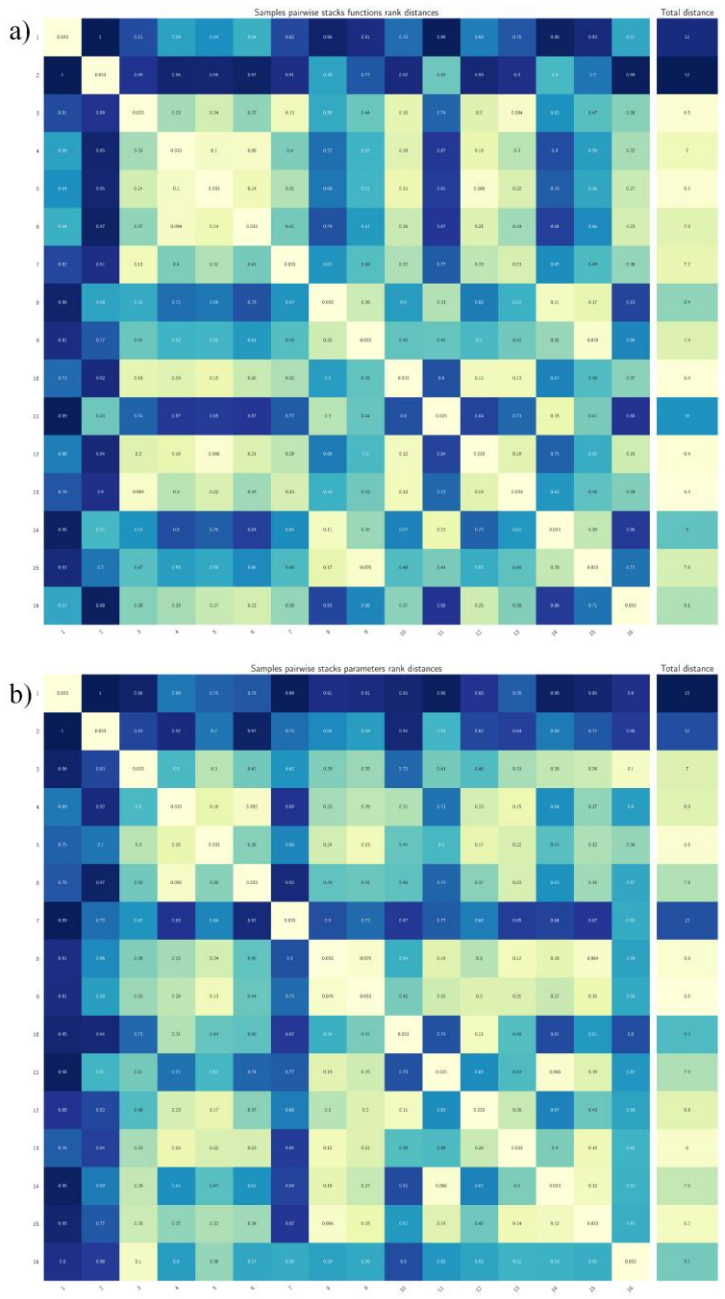


Fig.5. Heatmap rankings of differences between soil samples in terms of: a) ensemble averaged correlation functions and b) fitted parameters for these correlation functions.

Again, the differences between samples in terms of CFs and fitted parameters do not agree completely. Yet, we do observe all major outliers such as Sample 1.
Thus, the usage of fitted parameters for statistical analysis may be compromised.

careful analysis of soil 3D images (and their cross-sections) and correlation functions clearly indicated that there were visually very similar pore structures within different soil horizons as developed under different land use conditions (Fig.6). On the other hand, soil taken from similar horizon or land use may have a very different structure (Fig.7).

The choice of three basis functions as employed in this work provides a reasonable trade-off between the number of parameters and goodness of fit, but a potentially better set of correlation functions could be found in future studies.

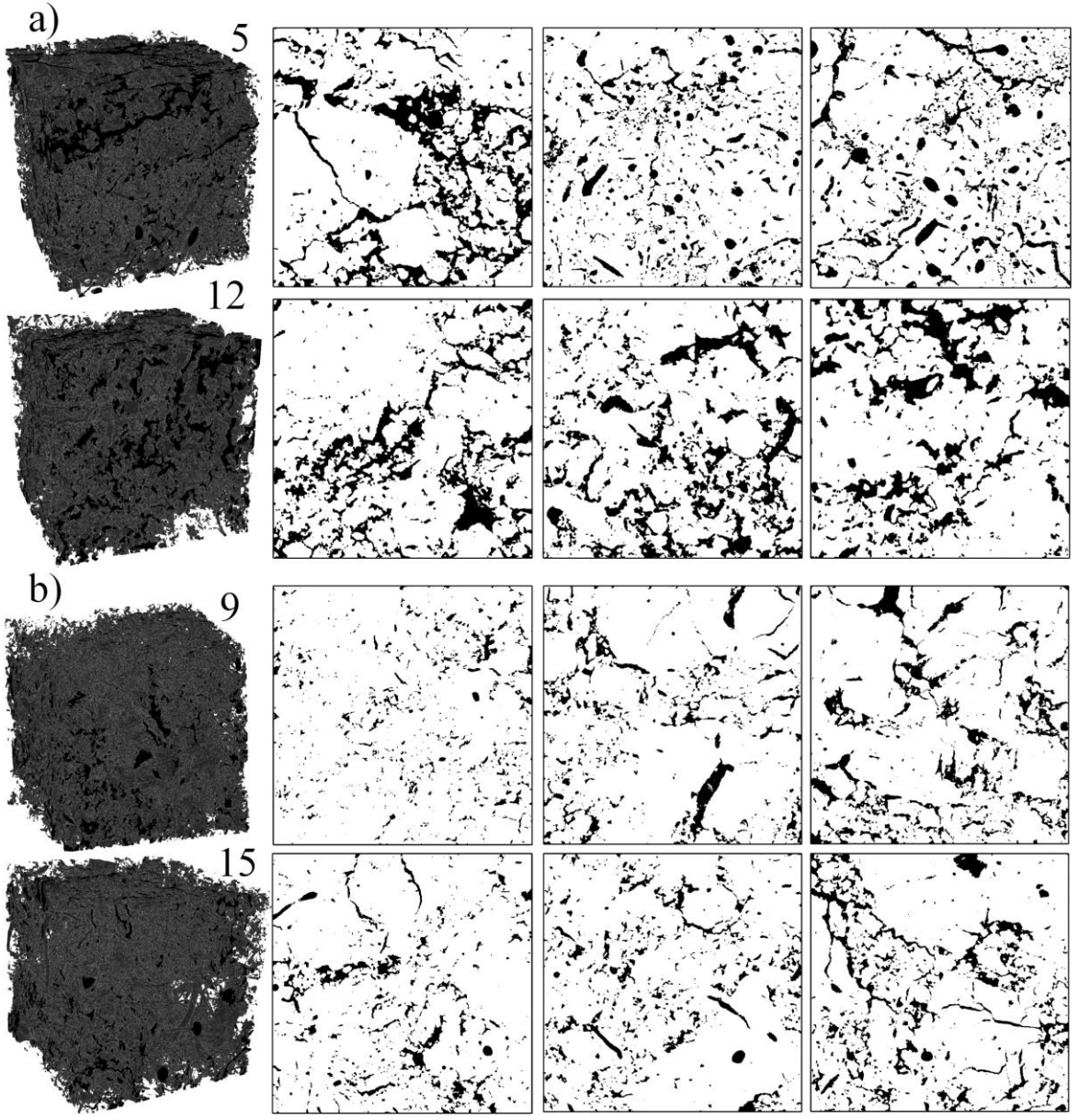


Fig.6. The 3D pore structure visualizations and three cross-sections close to top, middle and bottom along the z-axis for samples with minimum difference in terms of L^2-norm for two pairs of samples: a) Samples 5-12 (arable/bare fallow and arable/cereals land use categories, respectively) and b) Samples 9-15 (post arable/afforestation and arable/cereals).

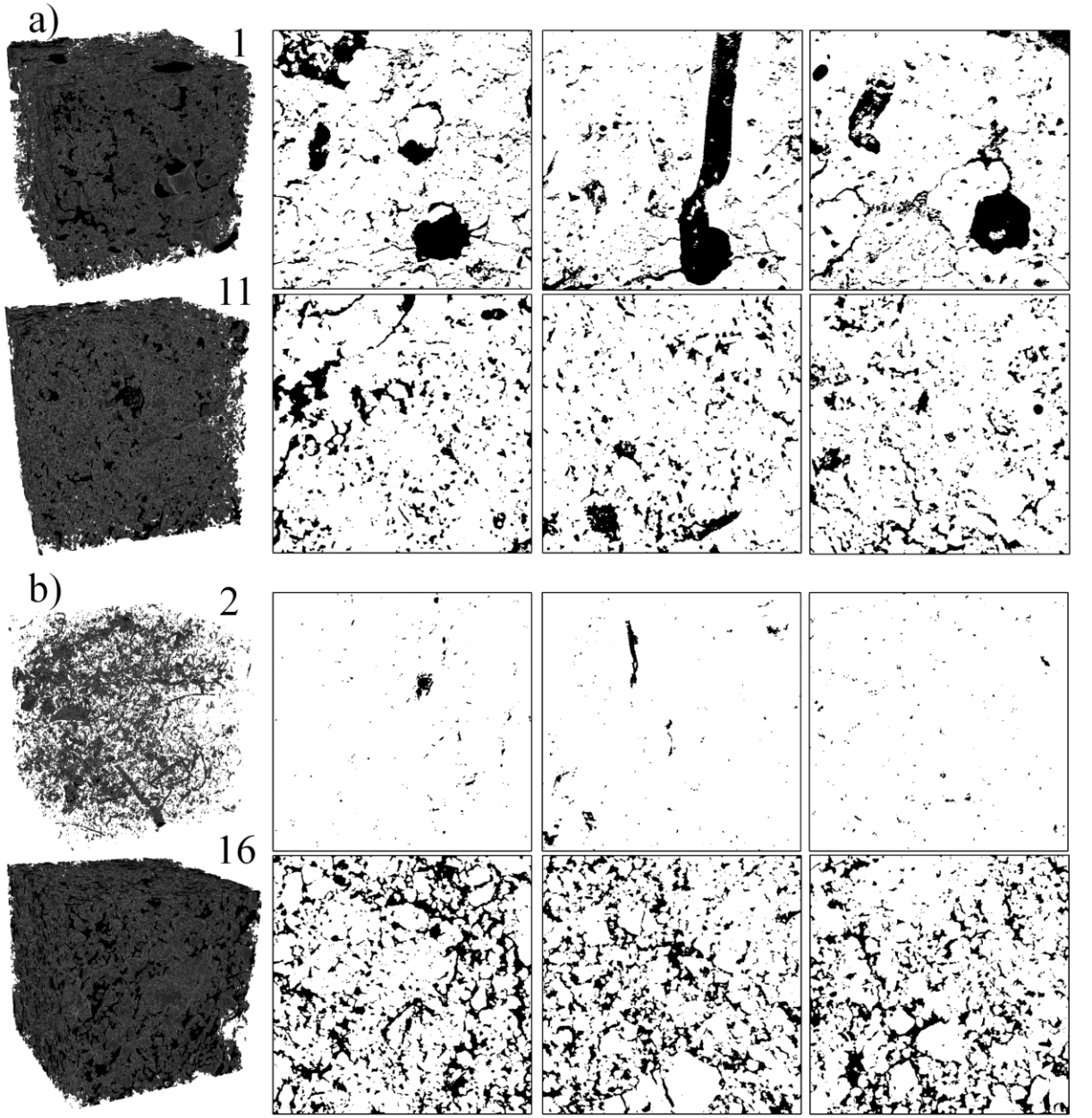
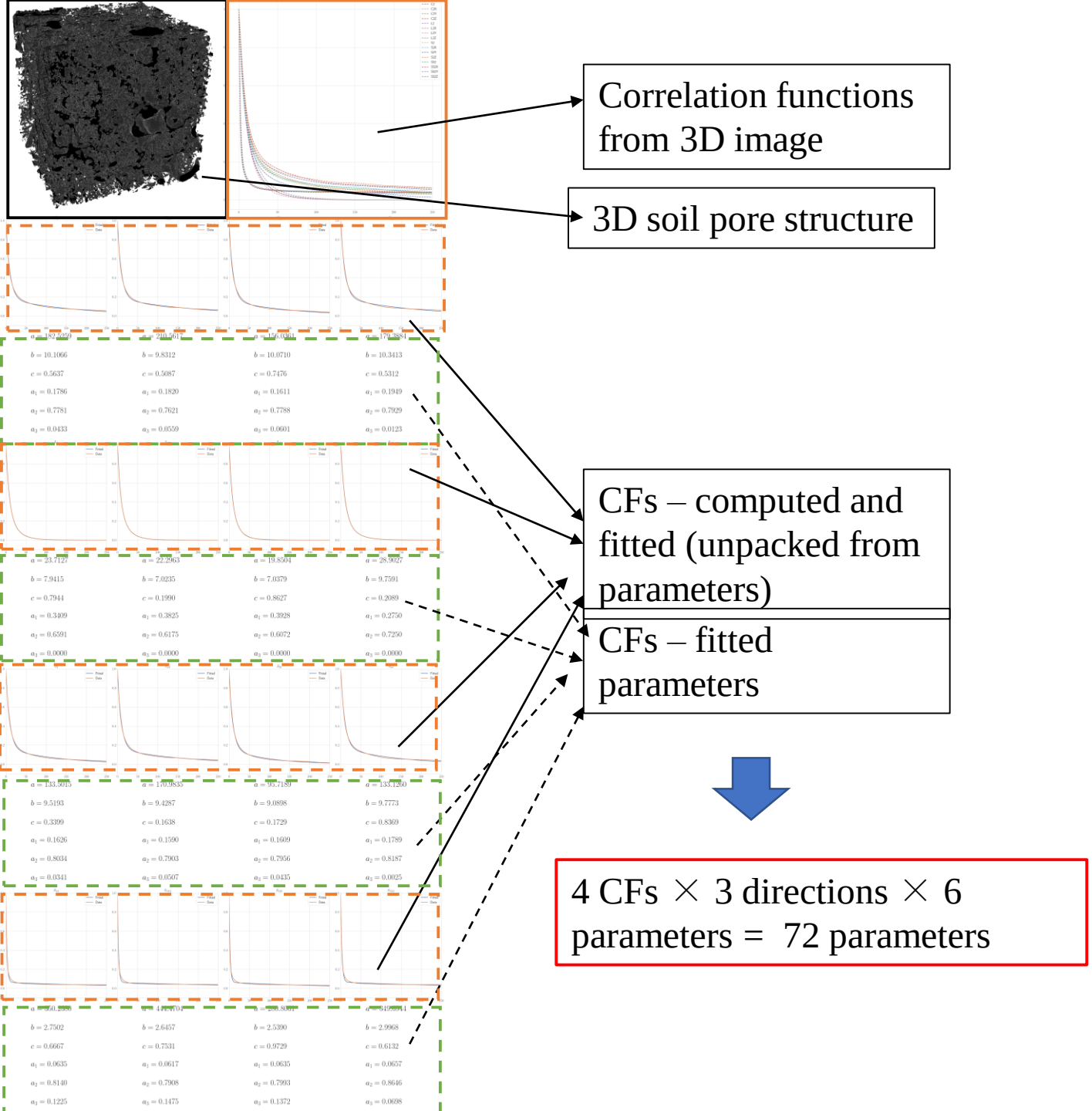


Fig.7. The 3D pore structure visualizations and three cross-sections close to top, middle and bottom along the z-axis for samples with maximum difference in terms of L^2-norm for two pairs of samples: a) Samples 1-11 (arable/bare fallow and arable/cereals land use categories, respectively) and b) Samples 2-16 (post arable/afforestation and post arable/afforestation).



Summary

In this study we successfully compressed the soil structural information in the form of 3D binary images into a set of correlation functions each of which is described using fitted 6 parameters. We utilized four different correlation functions (two-point probability, lineal, cluster and surface-surface functions) computed in three orthogonal directions for the pores. The methodology was applied to 16 different soil 3D images obtained using X-ray microtomography and segmented into pores and solids. All computed CFs were fitted using a superposition of three basis functions. In other words, we reduced 900-1300^3 voxels images into sets of 72 parameters.

Fitting of computed correlation functions and reducing them to a number of parameters is a powerful way of compressing soil structural information. However, the analysis based on parameters alone is different from the one where correlation functions are used, as was observed from clustering statistics. This problem can be negated by uncompressing the correlation functions back from these parameters before any application. This way correlation functions are not only a way to compress the soil structural information with minimal loss, but also may be utilized to solve a number of additional problems: comparison and differentiation of soil samples, location of elementary volumes, effective physical property prediction using machine learning, fusion of hierarchical soil structures.

The paper with detailed description of all results is currently under review with *EJSS* journal.

Most significant references for the methods used here:

- Karsanina, M. V., Gerke, K. M., Skvortsova, E. B., Ivanov, A. L., & Mallants, D. (2018). Enhancing image resolution of soils by stochastic multiscale image fusion. Geoderma, 314, 138-145.
- Jiao, Y., Stillinger, F. H., & Torquato, S. (2008). Modeling heterogeneous materials via two-point correlation functions. II. Algorithmic details and applications. Physical Review E, 77(3), 031135.
- Gerke, K. M., Karsanina, M. V., & Mallants, D. (2015). Universal stochastic multiscale image fusion: an example application for shale rock. Scientific reports, 5, 15880.
- Karsanina, M. V., & Gerke, K. M. (2018). Hierarchical Optimization: Fast and Robust Multiscale Stochastic Reconstructions with Rescaled Correlation Functions. Physical Review Letters, 121(26), 265501.
- Gerke, K. M., & Karsanina, M. V. (2015). Improving stochastic reconstructions by weighting correlation functions in an objective function. EPL (Europhysics Letters), 111(5), 56002.
- Gerke, K. M., Karsanina, M. V., Vasilyev, R. V., & Mallants, D. (2014). Improving pattern reconstruction using directional correlation functions. EPL (Europhysics Letters), 106(6), 66002.
- Karsanina, M. V., Gerke, K. M., Skvortsova, E. B., & Mallants, D. (2015). Universal spatial correlation functions for describing and reconstructing soil microstructure. PLoS ONE, 10(5), e0126515.

Acknowledgements

- This work was supported by Russian Foundation for Basic Research grant 18-34-20131 мол_а_вед. We thank Pavel Chuyko for administrating our HPC resources.