

Supplementary Material

Data-driven lithofacies prediction of unconsolidated sediments from wireline logs

*Sebastian Schaller^{1,2}, Patrico Becerra³, Sarah Beraus^{4,5}, Marius W. Buechi^{1,2}, David Mair¹, Mehrdad Sardar Abadi⁴, Bennet Schuster^{6,1,2}, Flavio S. Anselmetti^{1,2}

¹Institute of Geological Sciences, University of Bern, Baltzerstrasse 1+3, 3012 Bern, Switzerland

²Oeschger Centre for Climate Change Research, University of Bern, Hochschulstrasse 4, 3012 Bern, Switzerland

³Space Research and Planetary Sciences, Physics Institute, University of Bern, Sidlerstrasse 5, 3012 Bern, Switzerland

⁴LIAG Institute for Applied Geophysics, Stilleweg 2, 30655 Hannover, Germany

⁵Section Geology - Institute of Earth System Sciences, Leibniz University Hannover, Callinstrasse 30, 30167 Hannover, Germany

⁶Institute of Earth and Environmental Sciences, University of Freiburg, Tennenbacher Str. 4, 79106 Freiburg, Germany

Author Note

Sebastian Schaller	https://orcid.org/0000-0001-5613-6300
Patrico Becerra	https://orcid.org/0000-0002-2061-4056
Sarah Beraus	https://orcid.org/0009-0000-6984-3098
Marius W. Buechi	https://orcid.org/0000-0001-7638-490X
David Mair	https://orcid.org/0000-0002-7018-6416
Mehrdad Sardar Abadi	https://orcid.org/0000-0003-1751-8642
Bennet Schuster	https://orcid.org/0000-0001-6412-0640
Flavio S. Anselmetti	https://orcid.org/0000-0002-8785-3641

*Corresponding author: Sebastian Schaller (sebastian.schaller@unibe.ch)

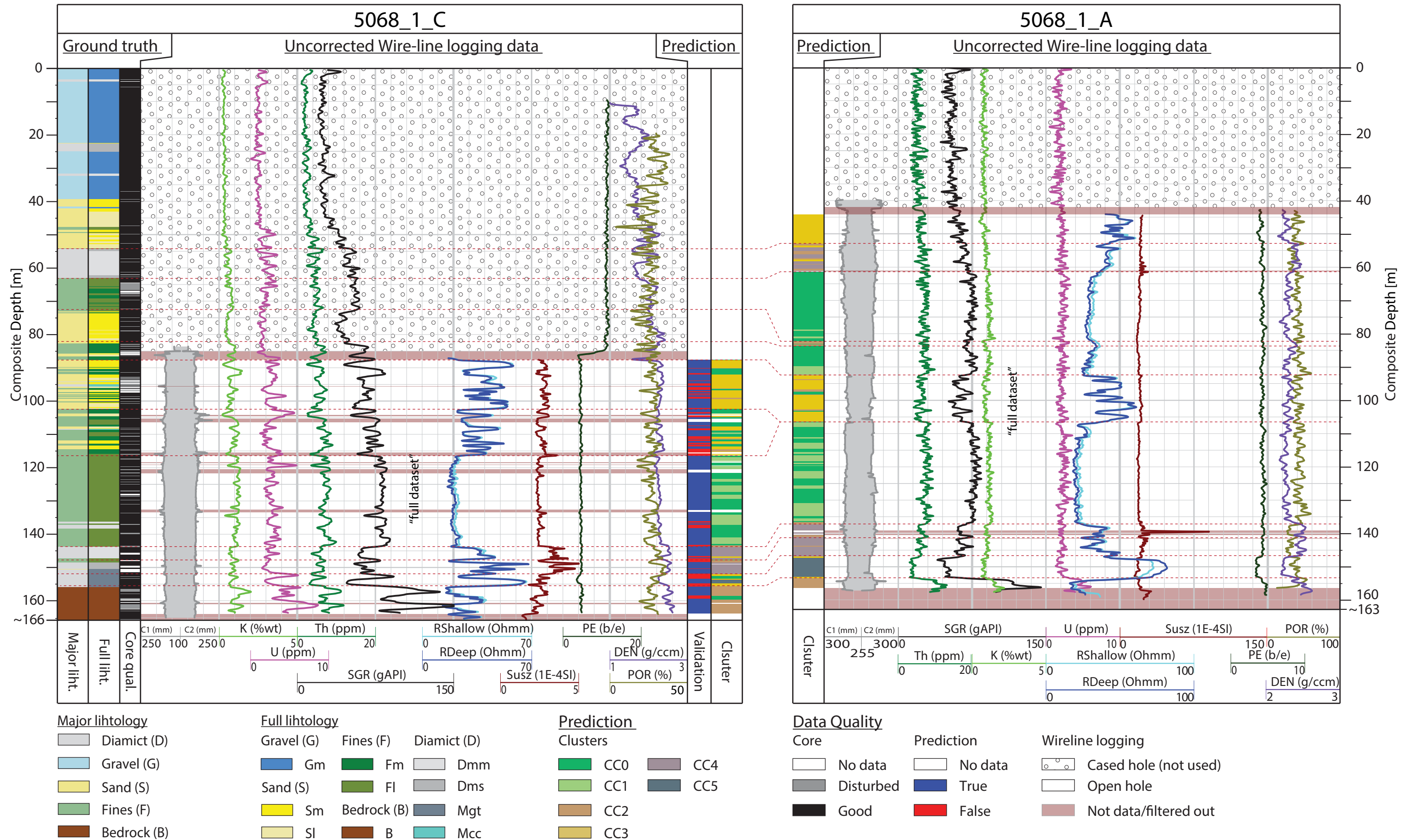


Figure S1: Extended version of Figure 8, including the used sections of the uncorrected wireline logging data of 5068_1_A and C (for used abbreviations of the logs, see Table S5).

Not normalized data distribution comparison

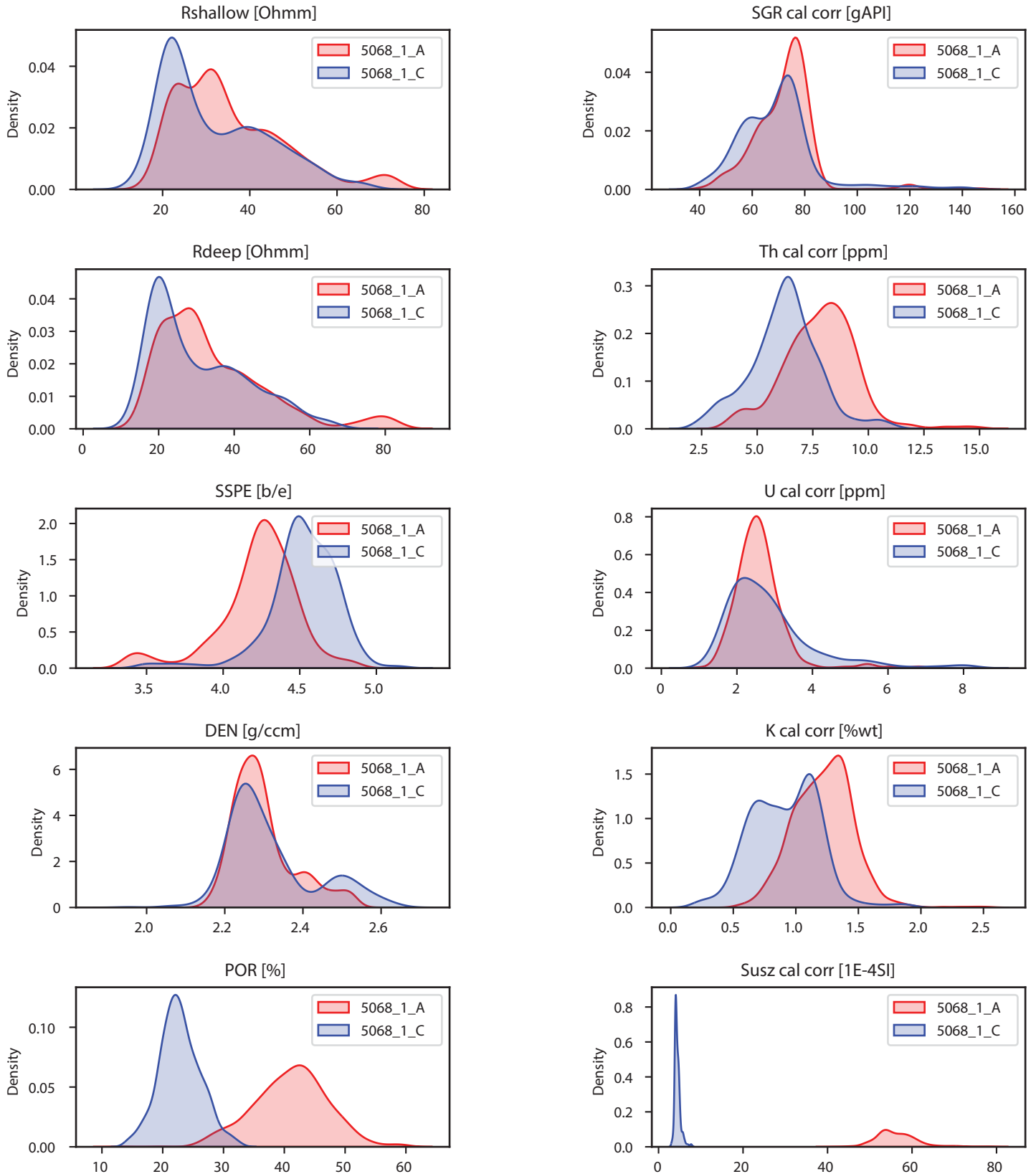


Figure S2: Comparison of the probability density distribution of the non-normalized wireline logging data of 5068_1_A and 5068_1_C.

Normalized data distribution comparison

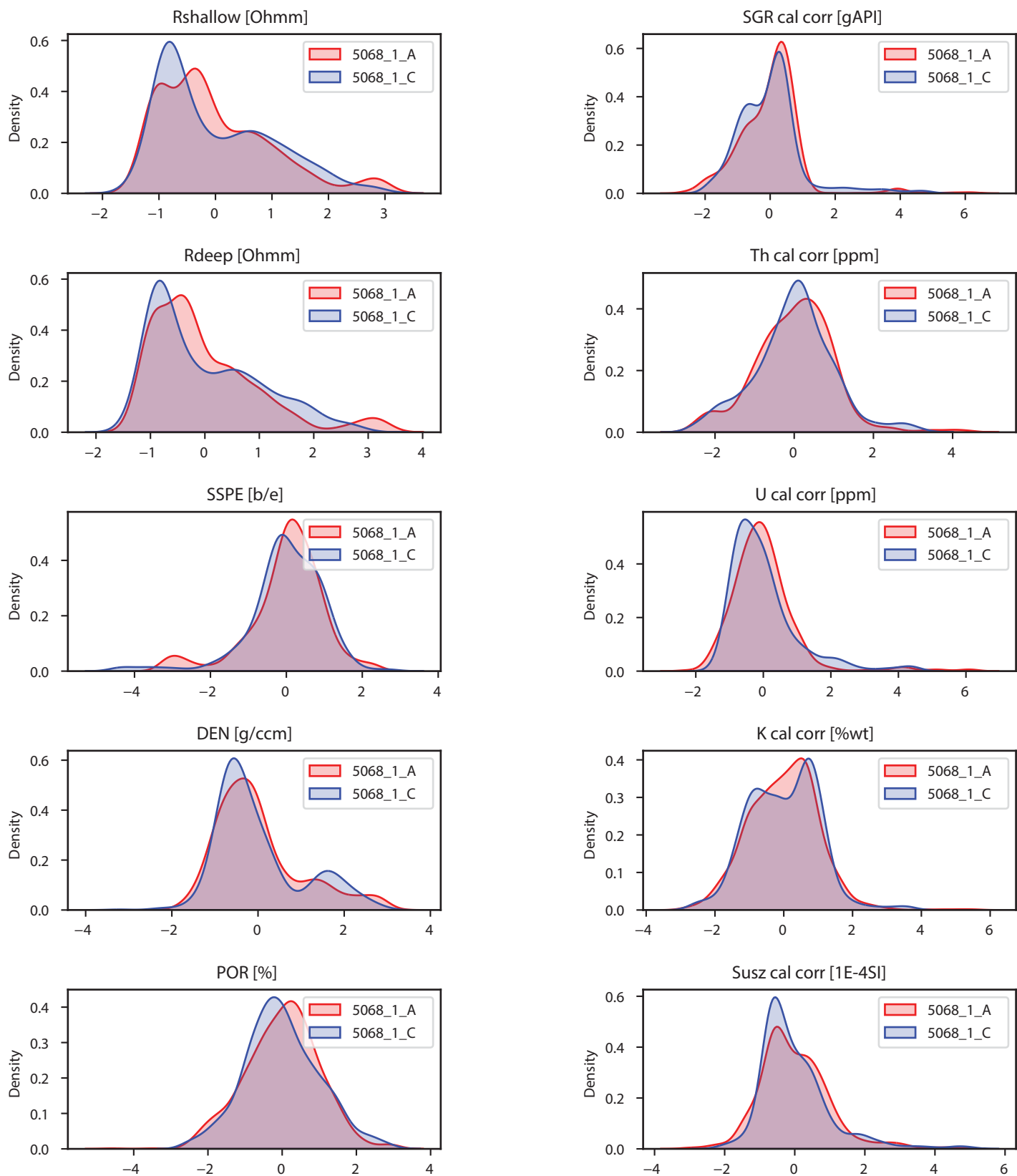


Figure S3: Comparison of the probability density distribution of the normalized (normal distribution around mean = 0 and std = 1) wireline logging data of 5068_1_A and 5068_1_C, used for combined clustering.

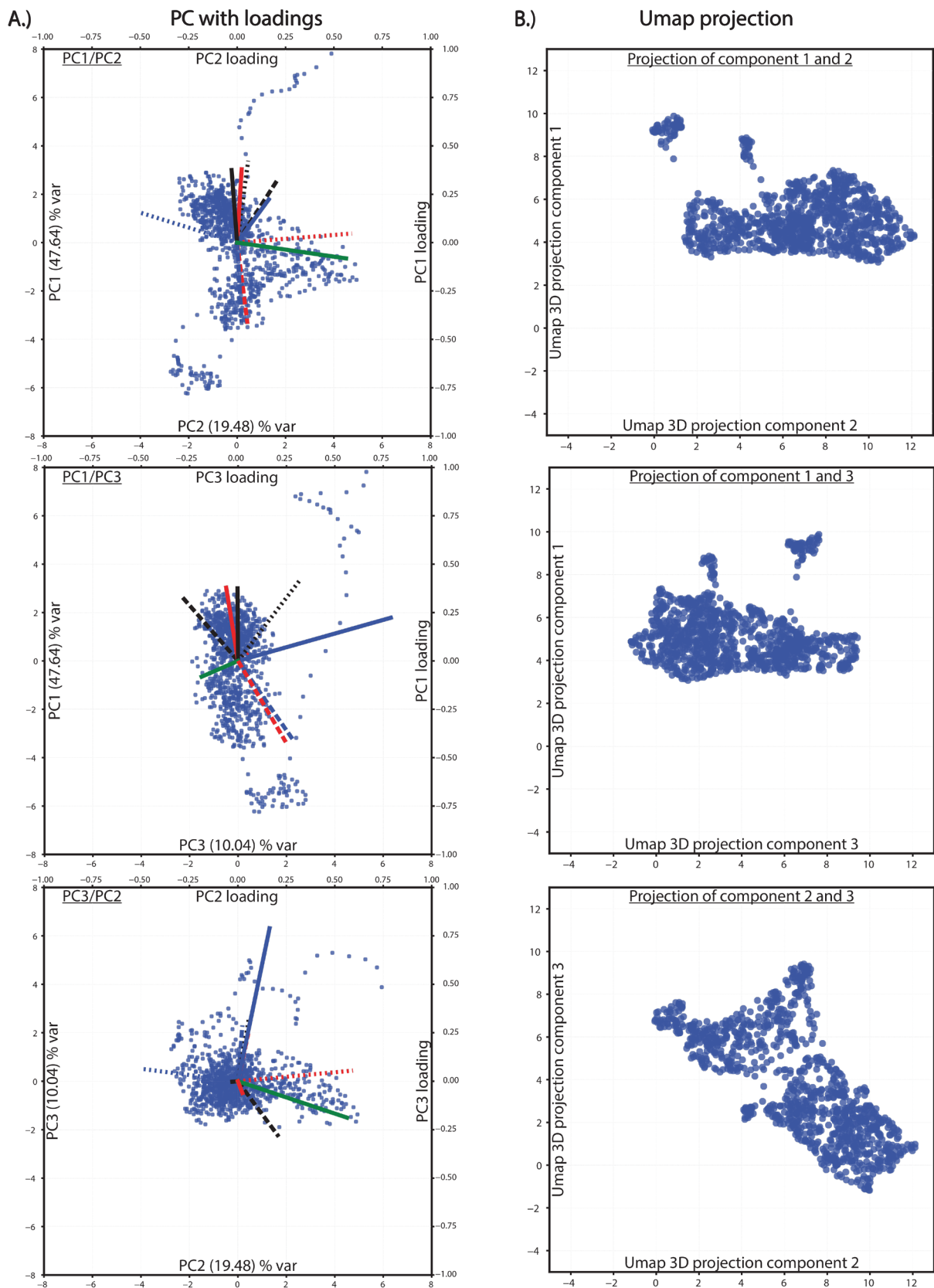


Figure S4: Biplot projections of the dimension reduction to 3d by PCA (A) and Umap (B) of 5068_1_A. The corresponding loading plots (secondary x and y-axis between 1 and -1), indicating the impact of each log, accompany the PCA biplots. Color codes for scatter data (assigned major lithology) of biplots and loading plots are included in Figure 1. The data for the loading plots are provided in Appendix Tables 4 (vector coordinates = direction starting from 0/0) and 5 (vector value = length).

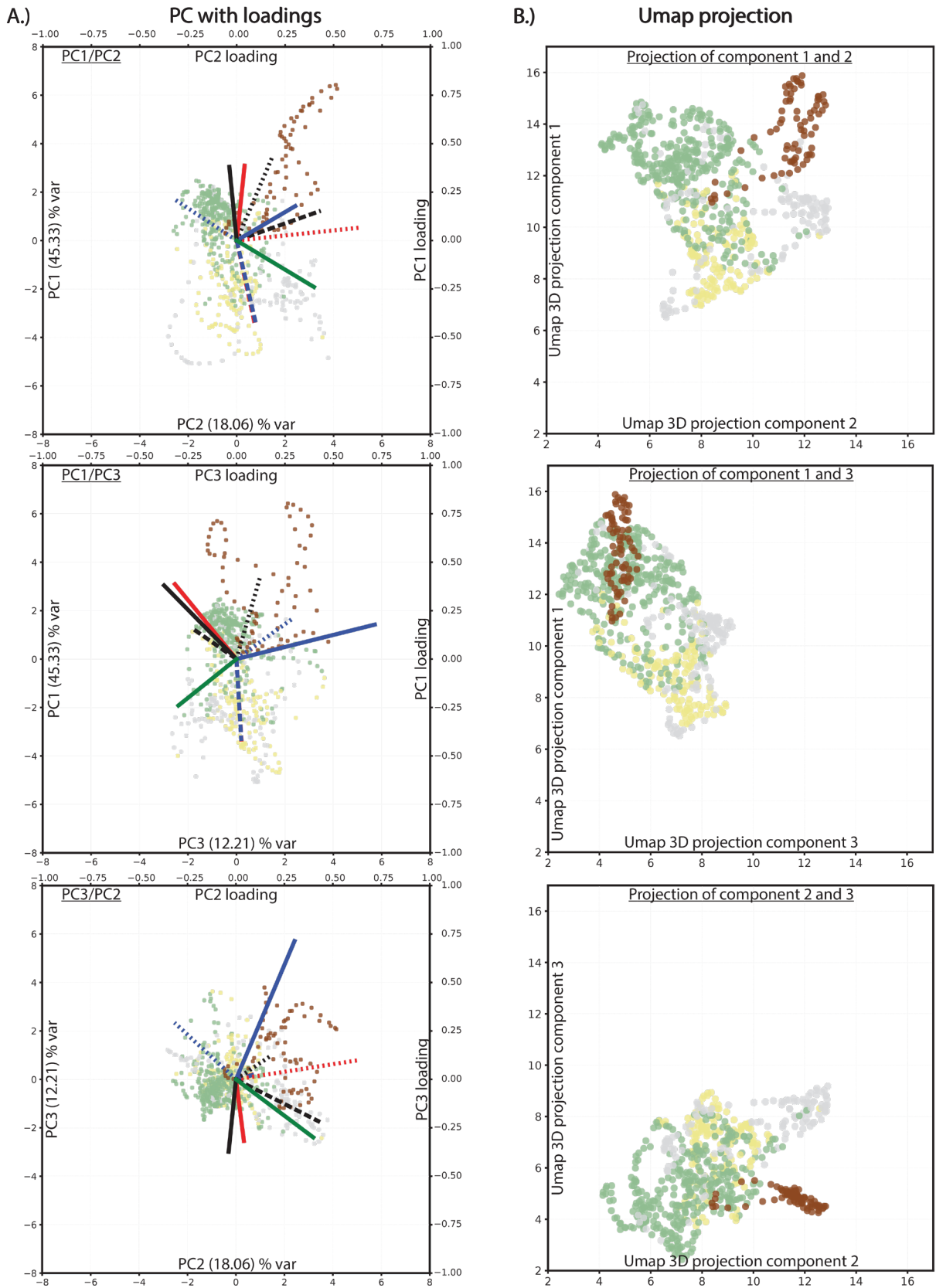


Figure S5: Biplot projections of the dimension reduction to 3d by PCA (A) and Umap (B) of 5068_1_C. The corresponding loading plots (secondary x and y-axis between 1 and -1), indicating the impact of each log, accompany the PCA biplots. Color codes for scatter data (assigned major lithology) of biplots and loading plots are included in Figure 1. The data for the loading plots are provided in Appendix Tables 4 (vector coordinates = direction starting from 0/0) and 5 (vector value = length).

Empiric evaluation for n clusters for combined 5068_1 data

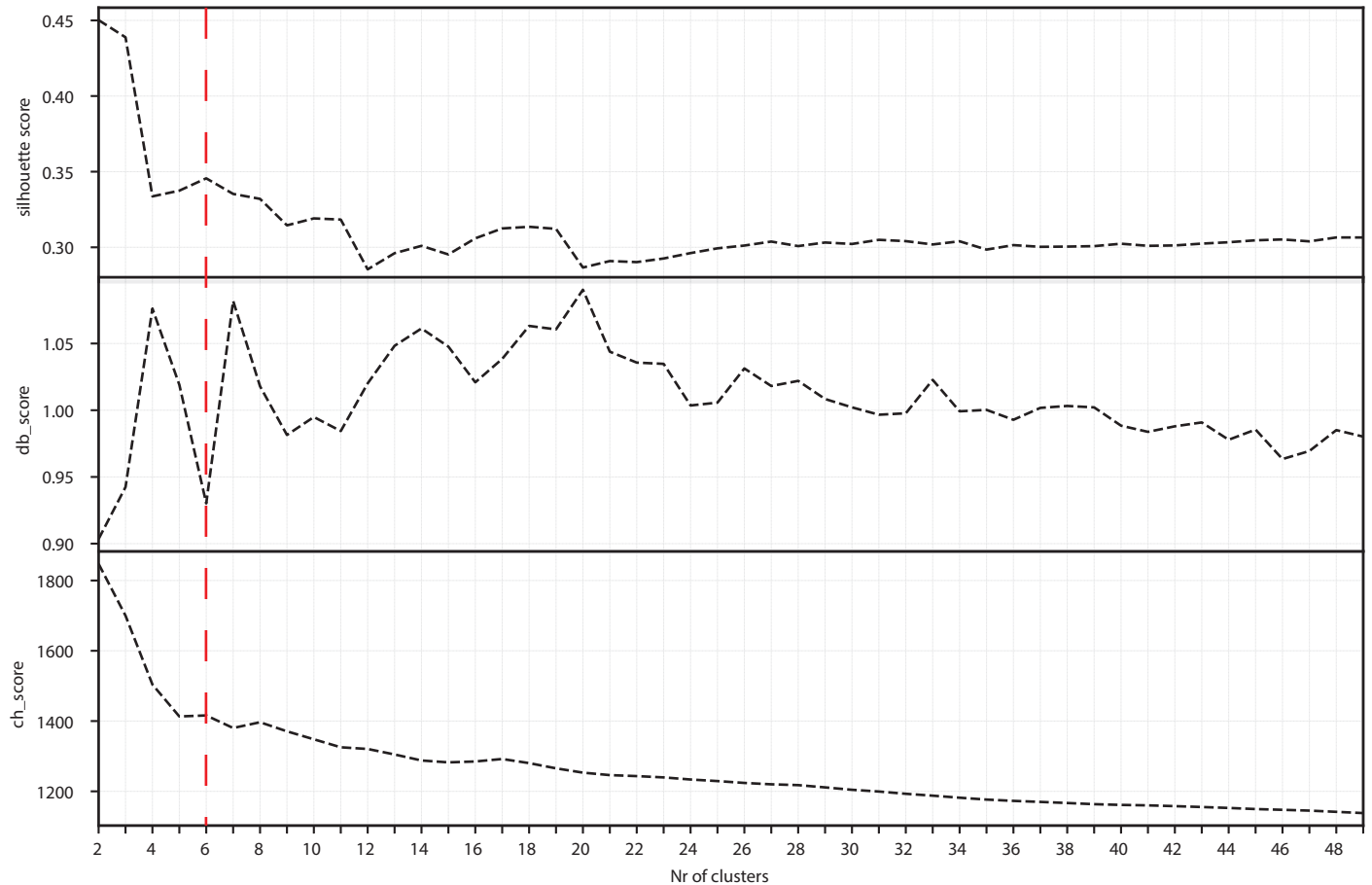


Figure S5: Cluster metrics for the combined dataset of 5068_1_A and C, displaying on the y-axis the calculated Silhouette-, Davies-Bouldin- (db), and Calinski-Harabasz (ch) Scores for (x-axis) a series of numbers of clusters ($n=2-49$). The red vertical line marks the selected numbers of clusters, leading to an acceptable solution

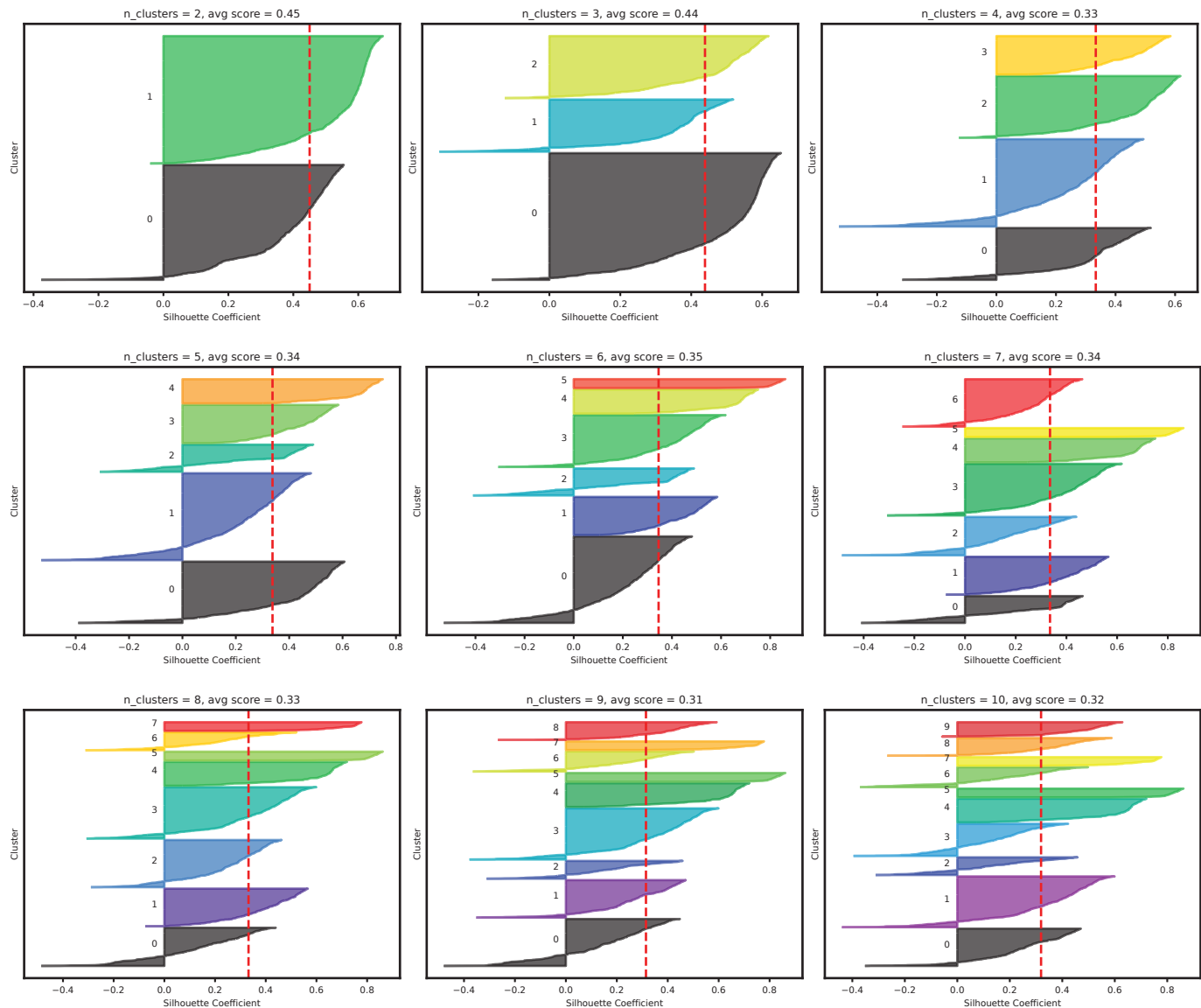


Figure S7: Silhouette plots for $n=2-10$ for the combined dataset of 5068_1_A and C, including the silhouette score for the corresponding number of clusters; colors of the clusters are randomly assigned and have no further meaning.

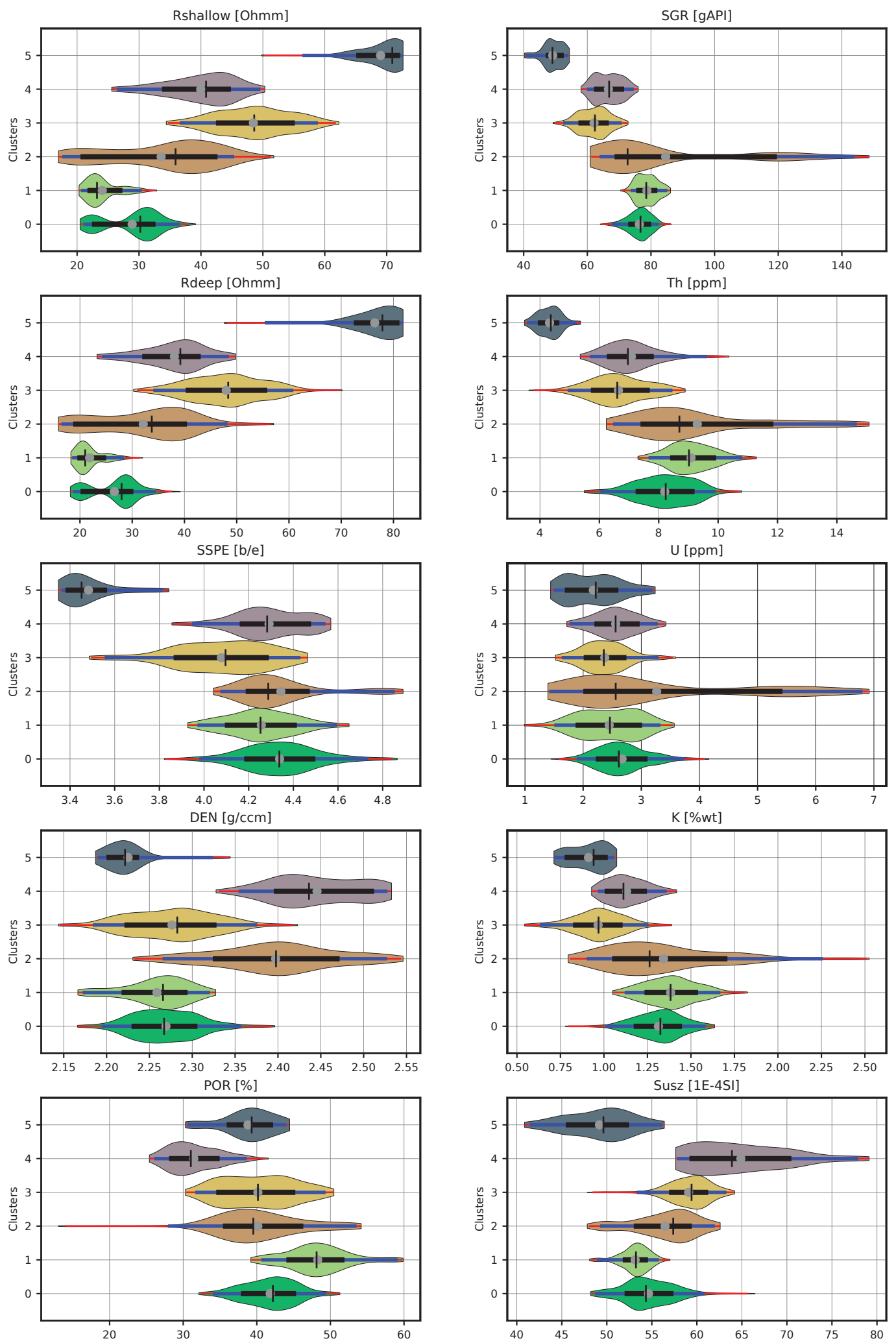


Figure A8: Violin plot of the individual log data distribution of the clusters of 5068_1_A, log with the corresponding unit is displayed as the title of each subplot, clusters are labeled on the y-axis form 0-5, and color-coded. Each violin plot includes the position of the mean (grey dot), the median (centered vertical black line), and the 1-, 2-, and 3- σ -ranges are displayed on the horizontal (1 σ : black, 2 σ : black + red; 3 σ : black + red + blue)

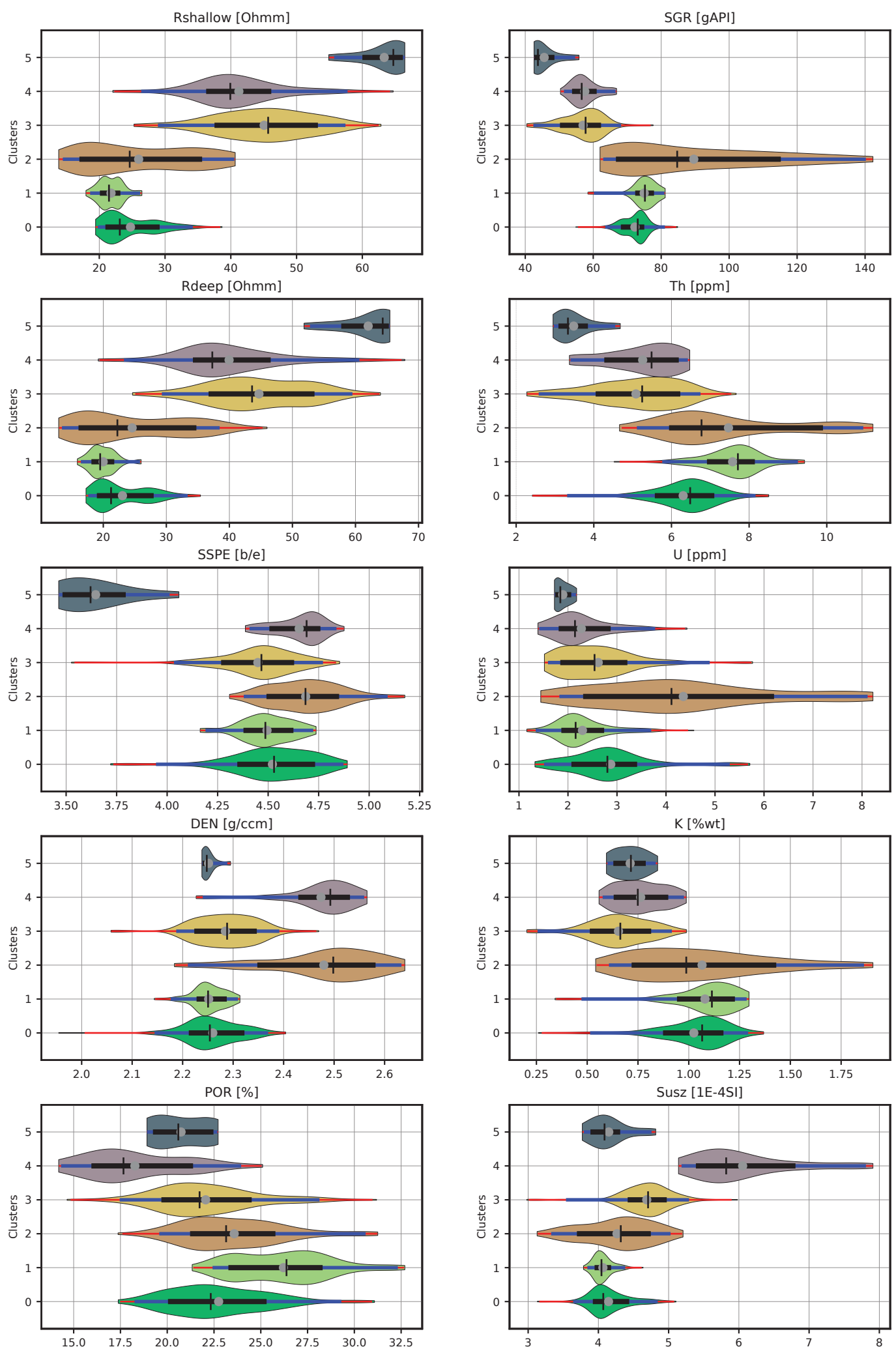


Figure S9: Violin plot of the individual log data distribution of the clusters of 5068_1_C, log with the corresponding unit is displayed as the title of each subplot, clusters are labeled on the y-axis form 0-5, and color-coded. Each violin plot includes the position of the mean (grey dot), the median (centered vertical black line), and the 1-, 2-, and 3- σ -ranges are displayed on the horizontal (1 σ : black; 2 σ : black + red; 3 σ : black + red + blue).

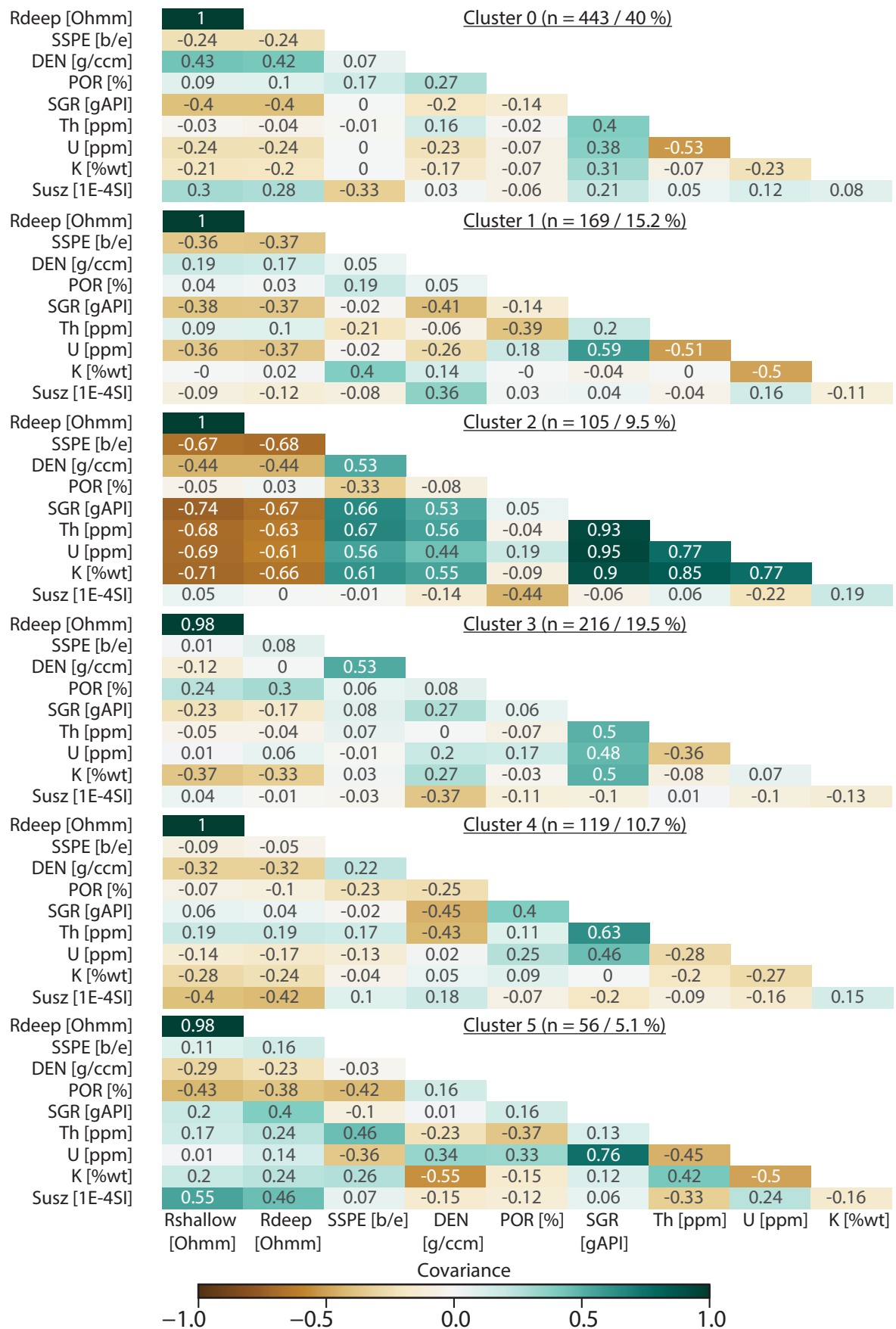


Figure S10: Triangular covariance matrix without self-correlation heatmap of the clusters (CC0-CC5) of 5068_1_A from the combined clustering of 5068_1_A and C, indicating the variance between each log (1-05 Strong positive correlation; 0.5-0.25 medium to weak positive correlation; 0.25- -0.25 -> weak to no positive or negative correlation; -0.25 - -0.5 medium to weak negative correlation; -0.5 - -1 -> medium to strong negative correlation).

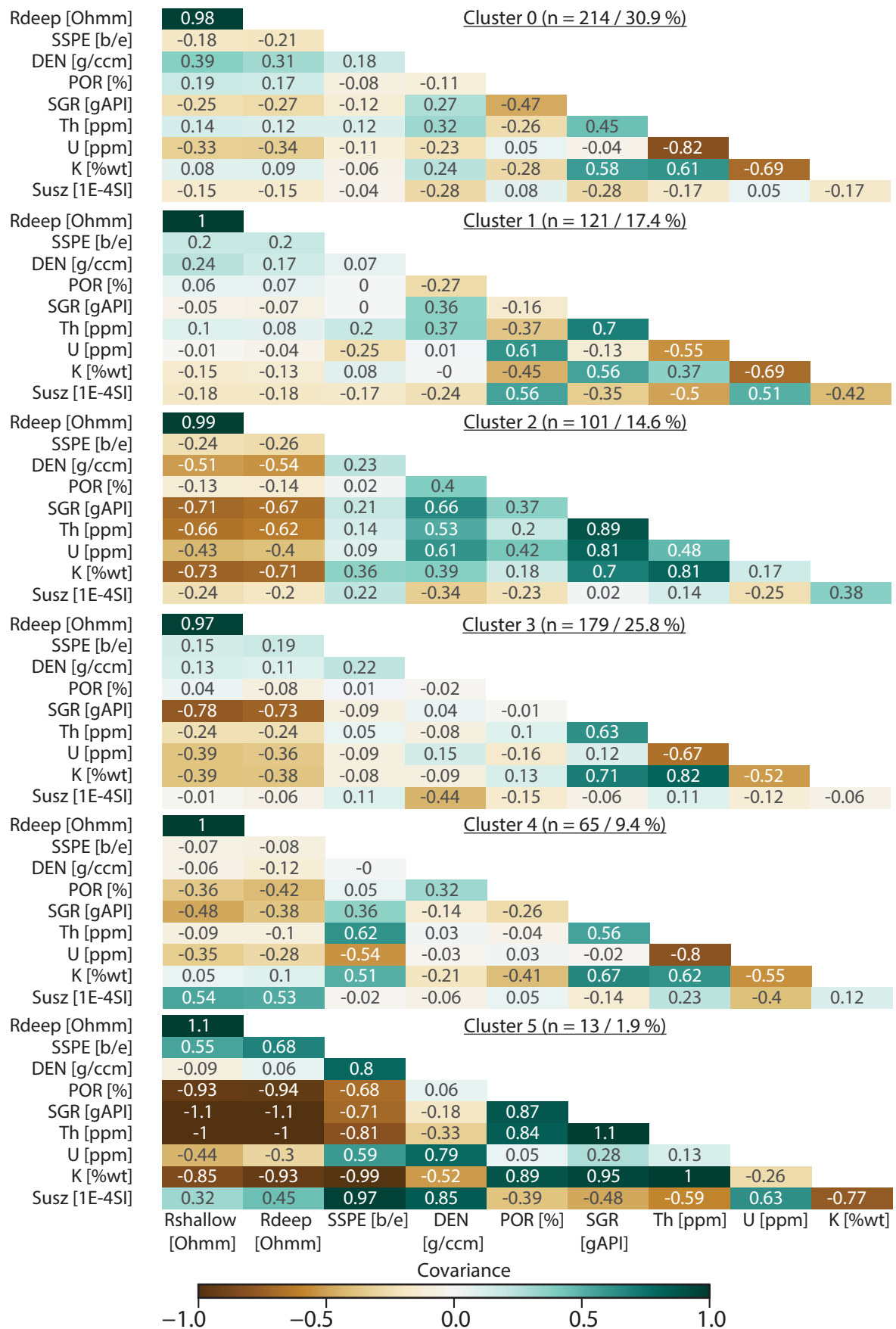


Figure S11: Triangular covariance matrix without self-correlation heatmap of the clusters (CC0-CC5) of 5068_1_C from the combined clustering of 5068_1_A and C, indicating the variance between each log (1-05 Strong positive correlation; 0.5-0.25 medium to weak positive correlation; 0.25- -0.25 -> weak to no positive or negative correlation; -0.25 - -0.5 medium to weak negative correlation; -0.5 - -1 -> medium to strong negative correlation).