

ESTIMATING CONTAMINATED SOIL VOLUMES USING A GENERATIVE NEURAL NETWORK: A HYDROCARBON CASE IN FRANCE

Herbert RAKOTONIRINA ^(1,2), Paul HONEINE ⁽¹⁾, Olivier ATTEIA ⁽³⁾, Antonin VAN EXEM ⁽²⁾

⁽¹⁾Université de Rouen Normandie, LITIS Lab, Rouen, France

⁽²⁾TELLUX, France

⁽³⁾Université de Bordeaux, EPOC, Pessac, France

1. Improving Contaminated Soil Volume Estimation

- **Need for Estimation:** Essential for planning remediation, assessing costs and resources, and ensuring safety.
- **Current Methods:** Geostatistical methods like Sequential Gaussian Simulation (SGS) are standard due to their ability to model spatial data variations and estimate uncertainties as shown by [1].
- **Limitations:** These methods require assumptions of normal data distribution and a well-defined variogram for spatial continuity. They are also resource-heavy, particularly in 3D applications.
- **Proposed method:** We investigate the Deep Image Prior Spatial Interpolation (DIP-SI) approach [2] for an efficient map generation and extend it to contaminated soil volumes estimation. This method bypasses traditional geostatistical constraints, utilizes GPU acceleration for faster processing, and improves integration of 3D spatial information, enhancing both accuracy and result relevance.

2. A Generative Neural Network (DIP-GEN) for Soil Volumes Estimation

1. Perform spatial interpolation using the DIP-SI method as shown in Figure 1. The method utilizes a generative neural network function f that takes a random map Z as input. The loss function L minimizes the difference between the interpolated map $f(Z) = \hat{X}$ and the measured samples X , considering only known values via Hadamard multiplication with mask m :

$$L = \|\hat{X} - X \odot m\|$$

2. For each Z , \hat{X} varies. We conduct n full learning processes of f , varying Z each time, which results in n distinct maps $\{\hat{X}_1, \hat{X}_2, \dots, \hat{X}_n\}$.

3. Volume estimation is computed using thresholds T_{\min} and T_{\max} across the axes x , y , and z :

$$V = \sum_{x,y,z} P(\hat{X}_k(x,y,z); T_{\min}; T_{\max}) \times V_{\text{grid}} \quad \forall k \in [1, n],$$

where $P = 1$ if the value falls within the thresholds, and $P = 0$ otherwise. V_{grid} represents the grid cell volume.

4. The cumulative distribution of these volume estimations for each maps defines a probable volume range as referenced by [1].

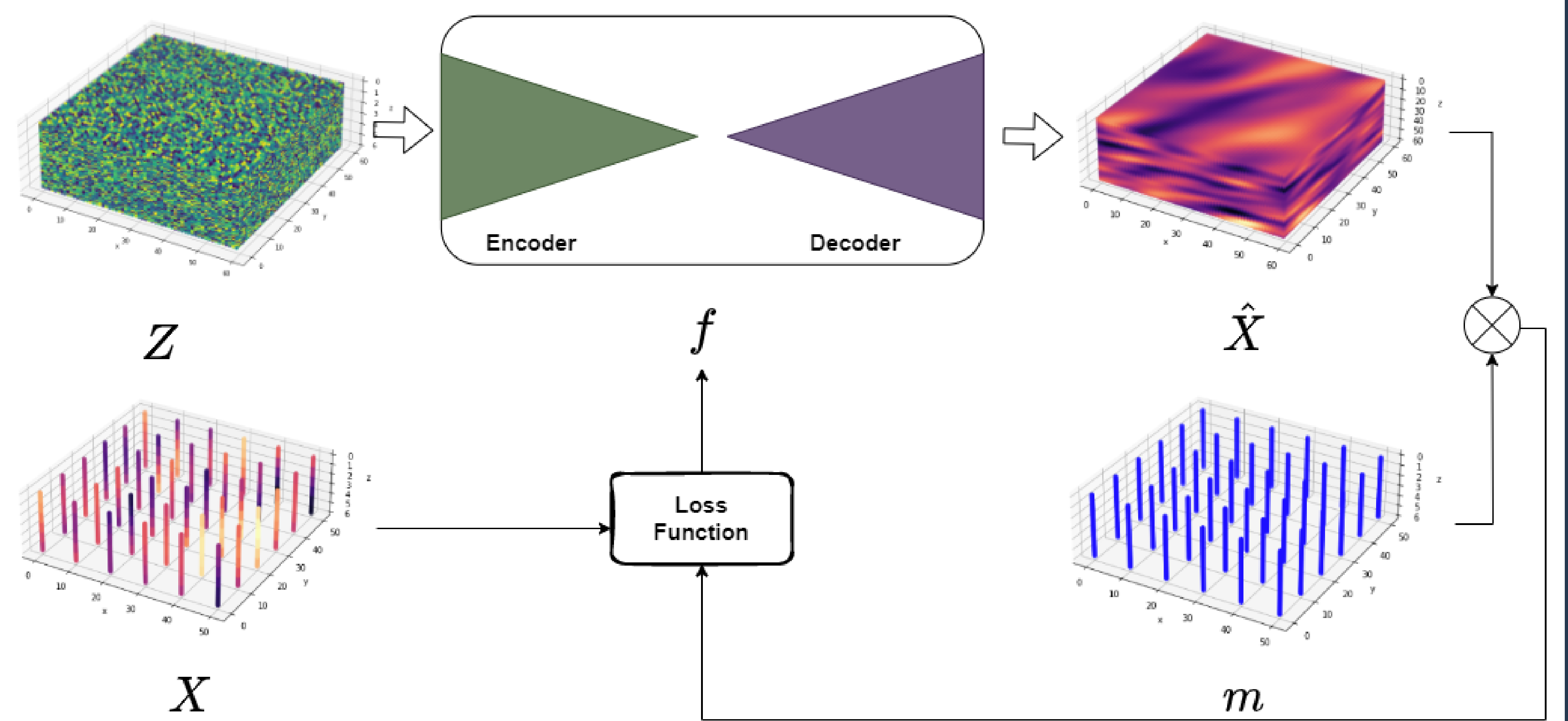


Figure 1: Representation of the learning process for spatial interpolation.

3. Case Study and Results

• The studied dataset comes from a site in France contaminated by total petroleum hydrocarbons (TPH), with six boreholes ($B1$ to $B6$) each 5 meters deep. The interpolation grid covers an area of 2612 m^2 , with a resolution of 2 m for the horizontal axes and 0.1 m for the vertical depth axis. The dataset originates from hyperspectral imaging analysis conducted by Tellux. Tellux develops Machine Learning algorithms that correlate spectral indices from hyperspectral imagery with TPH. Figure 2 represent the dataset.

• We compare our method against Ordinary Kriging (OK) (1) and the average maps of SGS, with test data from $B3$ and $B5$. And then, we assess volume estimates from our approach against those by SGS (2).

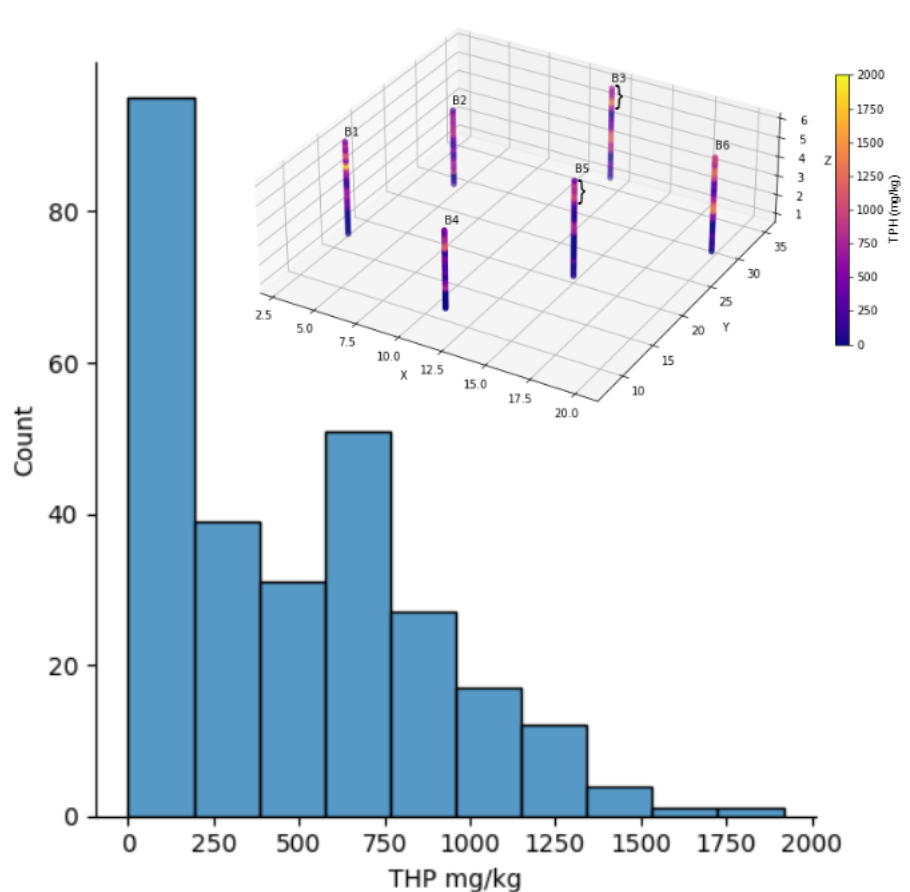


Figure 2: Spatial distribution of TPH concentrations for the 6 boreholes.

Cubic omnidirectional variogram for distance h :

$$\gamma(h) = 2.24 + \text{cubic}(10.74) \quad (1)$$

Spherical variogram with range set at directions NE , NW and Vertical:

$$\gamma(h) = 0.73 + \text{Sph}(18_{NE}, 15_{NW}, 6_V) \quad (2)$$

Table 1: RMSE, MAE and R2 score from test data

Methods	RMSE	MAE	R2
OK	321.44	261.92	0.35
SGS	256.34	184.31	0.49
DIP-GEN	233.53	163.40	0.58

Table 2: Estimation of the number of contaminated cells from test data

Methods	[0-500]	[500-2000]
V_{real}	10	23
SGS	[4-7]	[26-29]
DIP-GEN	[8-13]	[20-25]

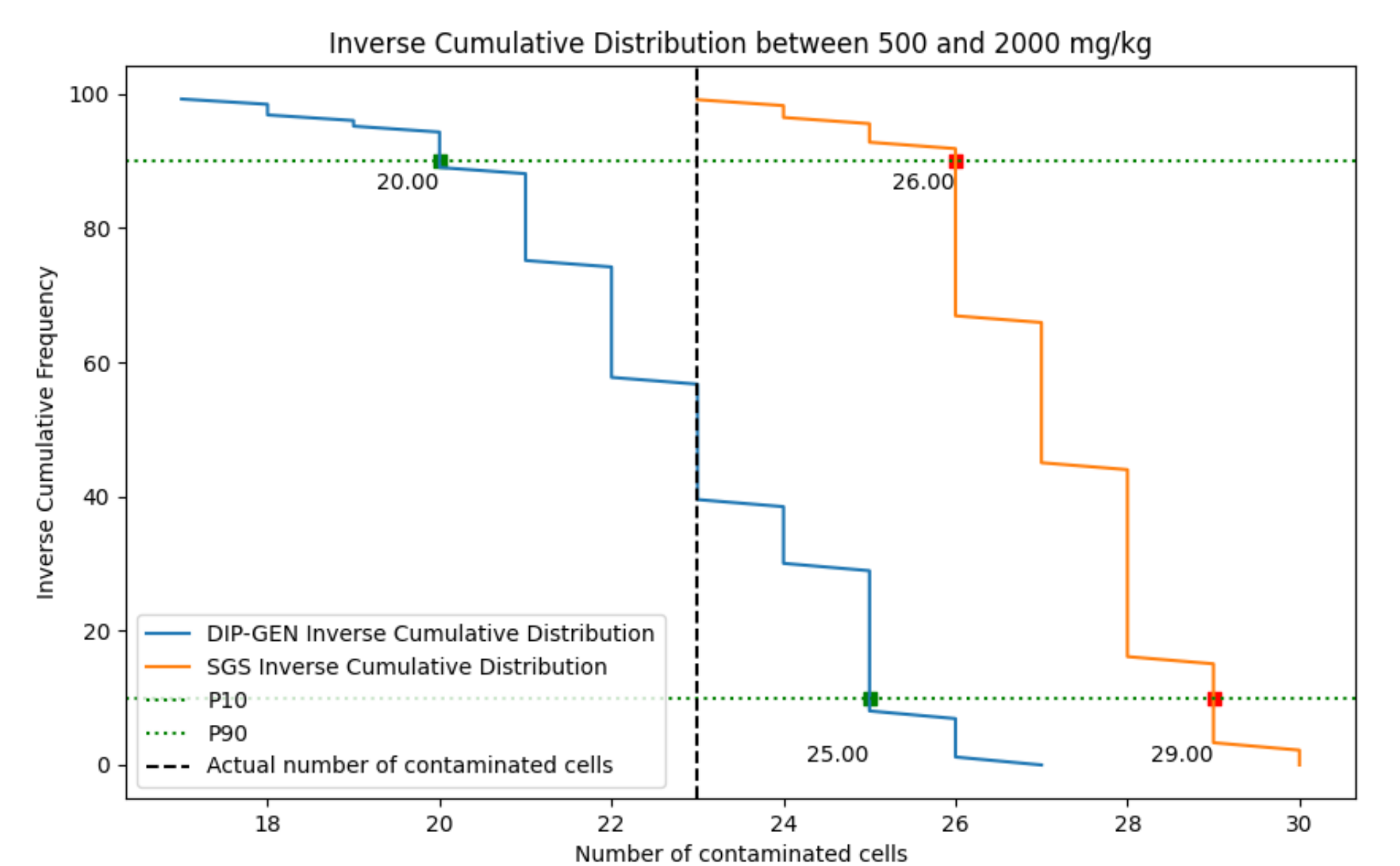


Figure 3: Cumulative distribution of test data between 500 and 2000 mg/kg

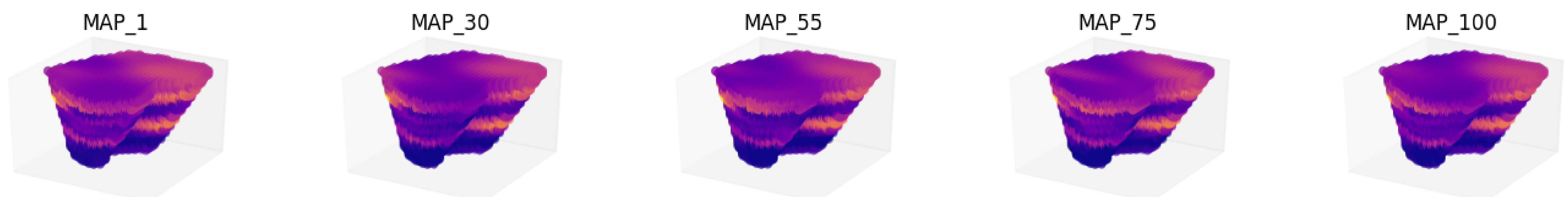


Figure 4: Five maps generated by the DIP-GEN among the 100 maps

4. Conclusion

- The proposed method employs a generative neural network for spatial interpolation. This method is inspired by a prior training-free image reconstruction technique.
- By using a Deep Learning approach, we overcome limitations associated with traditional geostatistical methods. Additionally, our method leverages GPU computing, enabling faster estimations.
- We demonstrate that the proposed method achieves better metric scores than OK and the average maps of SGS for spatial interpolation. It also provides more accurate estimates of contaminated soil volumes.

[1] Helene Demougeot-Renard and Chantal De Fouquet. "Geostatistical approach for assessing soil volumes requiring remediation: Validation using lead-polluted soils underlying a former smelting works". In: *Environmental science & technology* 38.19 (2004), pp. 5120–5126.

[2] Herbert Rakotonirina et al. "Spatial Interpolation and Conditional Map Generation Using Deep Image Prior for Environmental Applications". In: *Mathematical Geosciences* (2024), pp. 1–26.