

Advances and Perspectives in Modern Geostatistics for Energy, Natural Resources, and Environmental Studies

Dario Grana^{a,*}, Leonardo Azevedo^b, Jo Eidsvik^c,
J. Jaime Gmez-Hernandez^d, Alessandra Menafoglio^e, Joao Narciso^b,
Javier Rodrigo-Ilarri^d, Amilcar Soares^b, Valeria Todaro^f,
and Andrea Zanini^f

^a University of Wyoming, Laramie, WY, USA. Email: dgrana@uwyo.edu

^b Universidade de Lisboa, Portugal.

^c Norwegian University of Science and Technology (NTNU), Norway.

^d Universitat Politcnica de Valncia, Spain.

^e Politecnico di Milano, Italy.

^f Universit degli Studi di Parma, Italy.

*Corresponding author: dgrana@uwyo.edu

Abstract

Geostatistics plays a central role in exploration and management of energy, natural resources, and environment, where sparse and/or indirect measurements must be translated into predictive models of the subsurface. Despite decades of advances, the field faces a set of pressing challenges that demand conceptual and methodological developments. This work summarizes recent advances over the past decade and discusses perspectives and research directions in the field of modern geostatistics, which refers to the integration of classical spatial statistics with statistical learning and physics-informed methods. The growing availability of data and computational tools promotes the integration with artificial intelligence and machine learning and it offers new opportunities, but it also raises questions of interpretability and robustness. Capturing geological structures, heterogeneity,

and complexity remains a fundamental issue, particularly when using data-driven methods that must balance realism with computational tractability. Emerging hybrid workflows, combining classical geostatistics, deep generative models, and physics-informed learning, hold promise but require rigorous approaches to uncertainty quantification to be trusted in decision making. An additional challenge lies in bridging the gap between scales of a broad range of data, reconciling high-resolution models with the heterogeneous spatial representativity of measurements. These range from sparse point samples to densely sampled borehole data, broad-scale geophysical surveys, and full-coverage remote sensing outputs. This diversity of data types and applications shapes which geostatistical methods are applicable and how uncertainty must be represented. Beyond methodological advances, there is a need to promote geostatistical thinking in new application domains, from carbon storage and hydrogen energy systems to sustainable mining, to environmental and climate studies. Addressing these challenges will define the trajectory of geostatistics in the coming decade, positioning it as a key tool of sustainable resource and environmental management.

Keywords: Geostatistics; Spatial modeling; Uncertainty quantification; Deep learning; Earth modeling; Energy resources; Natural resources; Environmental modeling

1 Introduction

Geostatistics for energy, natural resources, and environmental studies refers to the use of spatial statistics techniques to describe, analyze, and predict surface and subsurface properties that are essential for exploration and management of Earth’s resources. Classical geostatistics explicitly accounts for the spatial correlation among measurements and aims to model the spatial pattern and structures of Earth’s properties by interpolating and simulating model variables at unsampled locations. Therefore, geostatistics provides a quantitative framework for handling the inherent incompleteness of surface and subsurface data and consistently accounts for uncertainties.

Geostatistical methods have become crucial across a wide range of applications. For example, in groundwater studies, geostatistical methods are adopted to model aquifer properties, predict flow patterns, and track the spread of contaminants. In oil and gas exploration, geostatistics is used to

map porosity, permeability, and fluid saturation from well logs and geophysical data, monitor reservoir conditions and production, and quantify uncertainty in reserves and production forecasts. In mining, it is applied for ore grade estimation and tonnage calculations from limited drill core samples and it supports mine planning and economic evaluation.

The importance of geostatistics stems from the fact that measured data in energy, natural resources, and environmental settings are sparse and often indirect. Data collection, such as acquiring core samples, well log data, and geophysical surveys in the field and conducting rock and fluid analysis in the laboratory, involves high costs, making it impractical to achieve exhaustive coverage. These datasets span a wide spectrum of spatial representativity: from small-volume, sparsely distributed point samples such as drill cores, through densely sampled borehole arrays, to spatially extensive but coarser geophysical measurements such as seismic or electromagnetic surveys, and finally to high-coverage remote sensing and satellite products. This variety of data types directly conditions the choice of geostatistical method, the scale at which uncertainty can be quantified, and how indirect measurements are integrated in spatial modeling best practices. Geostatistics provides the framework to extract information from these incomplete datasets, producing predictive models that guide exploration strategies and management decisions. The field now faces a critical transition where traditional methods intersect with artificial intelligence (AI) and deep generative models.

The methodological foundations of geostatistics include Kriging that provides the best linear unbiased predictor for spatial interpolation, stochastic sampling that generates multiple statistical realizations of the properties of interest, and geostatistical inversion that integrates measured data with geological knowledge to predict geological properties. The common element across these approaches is the ability to quantify uncertainty and provide decision-makers with probabilistic rather than deterministic insights.

Throughout this paper, modern geostatistics refers to the integration of classical spatial statistics with statistical learning and physics-informed methods that has shaped the methodological developments of geostatistics over the past decade. Recent developments tackle complex subsurface and environmental modeling challenges through object-oriented and functional approaches, deep learning frameworks, generative models, and hybrid AI-geostatistical workflows. Object-oriented and functional approaches now enable the analysis of non-Euclidean and high-dimensional data structures, such as curves, surfaces, and distributional data, while advanced covariance mod-

100 eling captures spatial dependencies in heterogeneous systems (Delicado et al.,
 101 2010; Pigoli et al., 2016; Menafoglio and Secchi, 2017; Martínez-Hernández
 102 and Genton, 2020; Menafoglio, 2025). Deep learning frameworks, including
 103 physics-informed and connectivity-aware models, enhance classical geostatistical
 104 methods by incorporating physical constraints, such as geological structures
 105 or flow paths, improving realism of model predictions (Zhang, 2022;
 106 Zhang et al., 2023; Miele and Azevedo, 2024; Yan et al., 2025). Generative
 107 models, including latent diffusion models, variational autoencoders, and generative
 108 adversarial networks, allow the reconstruction of realistic geological
 109 facies and reservoir systems from sparse measurements, often outperforming
 110 traditional geostatistical simulations by preserving variability, uncertainty,
 111 and spatial continuity (Mosser et al., 2017; Chan and Elsheikh, 2019; Azevedo
 112 et al., 2020; Miele et al., 2024; Lee et al., 2025; Ovanger et al., 2025). Additionally,
 113 hybrid AI-geostatistical workflows support high-resolution downscaling,
 114 joint inversion of geophysical datasets, and physics-constrained data
 115 simulation, to improve accurate, uncertainty-aware subsurface and environmental
 116 modeling (Tahmasebi et al., 2020; Bai and Tahmasebi, 2021; Gomes
 117 et al., 2025).

118 This work examines geostatistical advances over the past decade across
 119 several application domains, namely groundwater, energy systems, climate
 120 modeling, environmental monitoring, and mining operations, and it addresses
 121 fundamental challenges in scale integration, uncertainty quantification, and
 122 incorporation of geological realism into data-driven workflows. The remainder
 123 of the paper is organized into five main sections. Section 2 reviews
 124 the methodological foundations, beginning with classical geostatistical
 125 approaches and progressing to modern physics-informed and data-driven
 126 methods. Section 3 presents methodological advances, including stochastic inverse
 127 techniques, deep generative models, and physics-informed neural networks.
 128 Section 4 illustrates the application of these methods across groundwater,
 129 energy, climate, environmental, and mining studies. Section 5 discusses key
 130 foundational challenges that remain in spatial modeling. Section 6 outlines
 131 future perspectives and emerging research directions.

132 2 Methodological Foundations

133 2.1 Classical Geostatistical Methods

134 Classical geostatistical approaches, such as interpolation and sampling, as
135 well as stochastic inversion, optimization, and data assimilation, play a key
136 role in energy, natural resource, and environmental applications. These
137 methods provide complementary tools for extracting information from sparse
138 and noisy data, and together they form a methodological toolbox for geo-
139 science and engineering studies (Isaaks and Srivastava, 1989; Kitanidis, 1997;
140 Goovaerts, 1997; Wackernagel, 2003; Caers, 2011; Chiles and Delfiner, 2012;
141 Pyrcz and Deutsch, 2014).

142 Interpolation methods, such as Kriging, aim to predict values at unsam-
143 pled locations using nearby observations. While simple deterministic ap-
144 proaches such as inverse distance weighting or splines have long been used,
145 geostatistical interpolation via Kriging remains the most rigorous framework
146 as it explicitly accounts for spatial correlation structures (Kitanidis, 1997;
147 Chiles and Delfiner, 2012). Interpolation is central to applications for Earth’s
148 resources exploration and environmental studies. Examples of applications
149 include ore grade mapping in mining, where concentrations between drill
150 cores must be predicted, or reservoir characterization, where porosity and
151 permeability fields are inferred between wellbore measurements (Caers, 2011;
152 Pyrcz and Deutsch, 2014). Sampling methods, by contrast, are designed to
153 represent variability and uncertainty in subsurface properties. Geostatistical
154 simulations, such as two-point and multiple-point geostatistics, are widely
155 employed to generate multiple realizations of geological models, each dif-
156 ferent from one another but consistent with prior geological knowledge and
157 available data (Goovaerts, 1997; Mariethoz and Caers, 2014). This ensemble
158 of realizations provides a probabilistic model of reserves, flow behavior, or
159 contaminant transport, and it allows practitioners and stakeholders to as-
160 sess risk and quantify uncertainty (Caers, 2011; Scheidt et al., 2018). For
161 example, Figure 1 illustrates the application of Kriging and sequential Gaus-
162 sian simulation for estimating elevation maps in Yellowstone National Park,
163 conditioned on direct measurements and temperature data.

164 A natural extension of classical Kriging to richer data structures is Func-
165 tional Kriging, which treats observations at each location as entire functions
166 rather than scalar or multivariate values (Ramsay and Silverman, 2005).
167 Conceptually, Functional Kriging aims to extend spatial prediction to sit-

168 uations where the observed object at each location is more complex than
 169 a scalar value. Functional Kriging extends classical Kriging principles to
 170 infinite-dimensional spaces, enabling spatial interpolation of curves, time se-
 171 ries, or other functional objects. However, when dealing with constrained
 172 or compositional functional data, such as probability density functions or
 173 positive-valued curves, Functional Kriging approaches may fail to respect in-
 174 herent geometric constraints, and an appropriate geometric framework for
 175 distributional data is required (Van den Boogaart et al., 2014; Pawlowsky-
 176 Glahn et al., 2015; Menafoglio, 2021).

177 Integrating interpolation and sampling with stochastic inversion, opti-
 178 mization, and data assimilation extends the reach of geostatistics by trans-
 179 forming sparse and indirect observations into predictive surface and subsur-
 180 face models and it provides the methodological foundation for modeling and
 181 decision-making processes in energy, natural resources, and environmental
 182 studies. Inverse problems are commonly studied in geoscience applications
 183 where the properties of interest cannot be measured directly but must be
 184 predicted from indirect observations. Seismic inversion is a classical exam-
 185 ple, where rock properties are estimated from recorded seismic waveforms.
 186 Because such problems are typically ill-posed, they require regularization and
 187 probabilistic formulations to ensure stability and realism (Tarantola, 2005;
 188 Grana et al., 2021). Optimization methods aim to identify the best model
 189 or operational strategy under given objectives, constraints, and uncertainty
 190 models. For instance, in petroleum engineering, optimization is central to
 191 well placement, production scheduling, and history matching, where model
 192 parameters are tuned to minimize the mismatch between observed and sim-
 193 ulated data. Similarly, in mining, optimization underpins the sequencing of
 194 ore extraction to maximize net present value while balancing grade and costs
 195 (Oliver et al., 2008; Emerick and Reynolds, 2013). Data assimilation pro-
 196 vides a dynamic framework for updating models with new observations as
 197 they become available. In reservoir engineering, data assimilation methods
 198 are used to update flow models in near real time with production and pressure
 199 data, improving forecasts and decision-making. Similar techniques are ap-
 200 plied to monitor carbon dioxide sequestration or hydrogen storage projects,
 201 where time-lapse geophysical and borehole measurements are assimilated to
 202 track plume migration and ensure safe containment. Data assimilation is
 203 also widely used in weather and climate modeling, where the integration of
 204 real-time measurements improves short-term forecasts of environmental con-
 205 ditions (Evensen, 2009; Oliver and Chen, 2011; Carrassi et al., 2018; Evensen

et al., 2022).

2.2 Modern Geostatistics: Physics-Informed and Data-Driven Methods

Modern geostatistical methods focus on integrating classical spatial statistics with advanced object-oriented, data-driven, and AI techniques to address complex subsurface and environmental modeling challenges. Recent advances in statistical deep learning for spatial and spatio-temporal data (Wikle and Zammit-Mangion, 2023; Zhan and Datta, 2025; Menafoglio, 2025) provide comprehensive frameworks for this integration, while maintaining the spatial correlation structures fundamental to geostatistics.

Object-oriented spatial statistics is a framework for analyzing complex spatial data structures, demonstrating how geometric principles enable the modeling of spatially distributed functional data while addressing non-stationarity and non-Euclidean domains (Menafoglio, 2025). Figure 2 shows an example of Kriging for object data in an aquifer where spatially distributed particle-size density measurements from boreholes are interpolated to predict soil types, and an example of Kriging for manifold data in a large estuarine ecosystem, illustrating the spatial prediction of the association between aquatic variables. Recent approaches also allow for multiple-point geostatistics of functional processes (Ojo and Genton, 2025). On the other hand, machine learning models, such as convolutional neural networks, provide a tool to handle large datasets and nonlinear predictions, capturing complex spatial patterns and distributions (Liu et al., 2023a; Wang et al., 2025), but often face challenges in interpretability and uncertainty representation.

Beyond these methods, physics-informed models and deep learning approaches, combined with classical geostatistical methods, improve the realism of spatial models by incorporating physical constraints and spatial dependence directly into learning architectures. Deep Kriging (Chen et al., 2020; Peli et al., 2022) explicitly embeds spatial correlation into neural network architectures, while Bayesian deep learning methods (Kirkwood et al., 2022) maintain uncertainty quantification as well as accurate predictive capabilities. Rather than modifying the network architecture, spatial covariance structures from classical geostatistics can be imposed as priors or regularizers during training, combining statistical rigor with learning-based flexibility (Levy et al., 2023).

241 Deep generative models, such as generative adversarial networks, vari-
 242 ational autoencoders, and latent diffusion models, offer an alternative by
 243 learning the distribution of geological patterns from training images or prior
 244 model ensembles, and then producing new realizations conditioned on sparse
 245 field measurements at inference time. Training-image based geostatistical
 246 inversion using spatial generative adversarial networks (Mosser et al., 2017;
 247 Laloy et al., 2018) demonstrates that adversarial networks can learn complex
 248 geological patterns and generate conditional realizations. U-Net generative
 249 adversarial networks have shown effectiveness for subsurface modeling and
 250 seismic inversion (Zhang et al., 2021; Bürkle et al., 2023). Physics-informed
 251 extensions embed physical constraints into generative architectures to pro-
 252 duce models that are both geologically realistic and physically consistent
 253 (Zhang et al., 2023; Miele and Azevedo, 2024; Lee et al., 2025), and have
 254 been applied to high-dimensional inversion problems for subsurface charac-
 255 terization, combining data-driven efficiency with physics-based constraints
 256 (Wang et al., 2021; Guo et al., 2023).

257 The main challenge in integrating classical methods with modern data-
 258 driven and physics-informed approaches lies in maintaining interpretability,
 259 reliable uncertainty quantification, and geological plausibility. Hybrid pa-
 260 rameterizations offer promising pathways. For example, deep generative
 261 models learn low-dimensional representations of complex geological patterns,
 262 and classical data assimilation methods are then applied in the learned latent
 263 space (Canchumuni et al., 2019; Liu et al., 2024). This strategy preserves
 264 the uncertainty quantification rigor of traditional methods while exploiting
 265 the pattern recognition capabilities of deep learning.

266 These integration strategies reflect a paradigm shift. Classical geosta-
 267 tistical methods operate under explicit assumptions such as stationarity, a
 268 parametric covariance model, and a Gaussian (or otherwise fully specified)
 269 error structure. As a consequence, uncertainty is interpretable and traceable
 270 to well-understood theoretical foundations, and interpolated models, con-
 271 ditional realizations, and uncertainty quantification follow directly within
 272 this framework. Modern extensions are motivated by the need to reproduce
 273 spatial patterns of greater geometric and statistical complexity than para-
 274 metric covariance models can readily capture. A deep generative model or a
 275 physics-informed neural network is not constrained to Gaussian or stationary
 276 behavior. Indeed, it can learn long-range connectivity, curvilinear channels,
 277 and non-linear multi-point dependencies directly from training data or phys-
 278 ical simulators. However, the uncertainty model is no longer fully specified

279 by assumption and must instead be inferred, often from the same data and
280 architecture that produced the prediction.

281 Making these implicit assumptions explicit is the first step toward princi-
282 pled method selection. The assumptions of classical Kriging, namely second-
283 order stationarity, a parametric covariance model, and/or a multivariate
284 Gaussian distribution, can be tested with variogram analysis and normal-
285 ity checks, and their violation produces predictable and documented effects
286 on interpolation quality. The assumptions introduced by modern extensions
287 are often harder to test and less transparent in their implications. Gen-
288 erative adversarial networks and variational autoencoders assume that the
289 training images or training datasets are statistically representative of the tar-
290 get domain, so that bias in the training data propagates into the generated
291 realizations. Convolutional neural networks assume translational invariance
292 of spatial patterns, which is violated wherever subsurface geometry varies
293 with depth or across structural domains. Physics-informed neural networks
294 assume that the embedded governing equations adequately describe the rel-
295 evant processes, and might produce inaccurate results when boundary con-
296 ditions or constitutive relationships are poorly constrained. Deep Kriging
297 and Bayesian deep learning methods assume that the chosen posterior ap-
298 proximation, whether variational, Monte Carlo dropout, or ensemble-based,
299 is a reliable proxy for the true predictive uncertainty, that is rarely verified
300 in practice. Table 1 summarizes these methods, their implicit assumptions,
301 and the conditions under which each is most appropriate.

302 **2.3 Uncertainty in Classical and Modern Geostatistics**

303 The integration of classical and modern methods changes not only how spa-
304 tial models are built but also how their uncertainty is interpreted. Two
305 conceptually distinct sources of uncertainty arise in geostatistical modeling:
306 variability inherent in the phenomenon, and lack of knowledge about the
307 model or data (Caers, 2011). These two notions, often termed aleatoric and
308 epistemic uncertainty in the machine learning literature, have clear physi-
309 cal interpretations in the geoscience context and behave differently under
310 classical and ML/AI-based workflows.

311 Aleatoric uncertainty refers to the variability inherent in the phenomenon
312 itself, including the natural spatial heterogeneity of subsurface properties,
313 the random component of a geological process, or measurement noise that
314 cannot be eliminated regardless of how many observations are collected. In

315 classical geostatistics, aleatoric uncertainty is represented by the nugget effect
316 in the variogram and by the spread of conditional simulation realizations. It
317 is the uncertainty that remains even with a perfect model and unlimited
318 computational resources.

319 Epistemic uncertainty, by contrast, arises from lack of knowledge, from
320 having too few data, from using an imperfect model structure, or from making
321 assumptions that may not hold. In classical Kriging, epistemic uncertainty is
322 partially captured by the Kriging variance, which decreases as more data are
323 added. In Bayesian inversion, it is encoded in the posterior distribution over
324 model parameters. Epistemic uncertainty is reducible in principle: collecting
325 more data, improving the model, or relaxing restrictive assumptions can all
326 reduce it.

327 The distinction is critical in practical applications because the two types
328 of uncertainty often require different responses Scheidt et al. (2018). Aleatoric
329 uncertainty must be communicated to decision-makers as an irreducible fea-
330 ture of the model, while epistemic uncertainty motivates further data acqui-
331 sition and model refinement and is the primary target of value-of-information
332 analysis. In classical geostatistical workflows the two sources are relatively
333 well separated by construction. In ML/AI-based workflows the separation be-
334 comes less clear: for example, GANs or diffusion models generate an ensemble
335 of realizations whose spread reflects both the natural variability encoded in
336 the training data and the limitations of the learned model. Improving the
337 fundamental understanding of these contributions is an active research prob-
338 lem, and methods such as Bayesian neural networks, Monte Carlo dropout,
339 and deep ensembles each provide only approximate answers.

340 **3 Methodological Advances**

341 This section reviews three directions of methodological advancements in geo-
342 statistical modeling. Stochastic inverse methods extend classical Bayesian
343 and ensemble frameworks to new data types and joint inversion regimes.
344 Deep generative models introduce learned representations of geological pat-
345 terns as alternatives to parametric covariance priors. Physics-informed neural
346 networks embed governing equations into network training for data-efficient
347 forward modeling and inversion. Together they illustrate how the integration
348 of classical spatial statistics with statistical learning and physics-informed
349 methods has shaped the methodological advancements of the past decade.

3.1 Stochastic Inverse Methods

Stochastic inversion has been a primary tool for probabilistic subsurface characterization, and the past decade has seen substantial extensions to new data types, joint inversion frameworks, and geophysical applications. This subsection reviews these developments. Stochastic inversion methods estimate subsurface properties by generating and updating multiple realizations rather than a single deterministic model, allowing explicit quantification of nonuniqueness and uncertainty (Tarantola, 2005). For example, in seismic inversion, data measured at the surface are linked to subsurface rock and fluid properties through nonlinear physical models, and stochastic inversion iteratively samples and perturbs model realizations until synthetic responses match the observed data. Approaches include stochastic sampling algorithms which approximate the posterior distribution by exploring the model space through iterative sampling, as well as stochastic optimization algorithms which update geostatistical realizations (Grana et al., 2021).

Stochastic methods have been increasingly used in seismic inversion to handle complex, non-Gaussian priors and to integrate geostatistical information into the inference process. MCMC methods can jointly infer multiple rock and fluid properties by embedding spatially informed priors derived from geostatistical modeling (Zunino et al., 2015). Several extensions of MCMC have been developed to improve sampling efficiency in high-dimensional settings, including Hamiltonian Monte Carlo, adaptive Metropolis schemes, and parallel tempering. Together, these methods preserve geological spatial correlation while providing both predictions and uncertainty for reservoir characterization and decision making.

Beyond MCMC, alternative stochastic methods are also used in high-dimensional inversion settings, including simulated annealing, genetic algorithms, gradual deformation methods, and ensemble-based approaches (Grana et al., 2021). These methods can be combined with dimensionality reduction to make sampling tractable in large-scale problems; classical reduction methods such as principal component analysis can be replaced or complemented by machine-learning parameterizations such as deep convolutional autoencoders, in which the autoencoder learns a compact representation of geostatistical realizations that is then updated through ensemble data assimilation (Liu and Grana, 2020). Figure 3 shows the seismic history matching workflow for carbon sequestration monitoring using ensemble smoother with multiple data assimilation (ES-MDA) with a deep convolutional autoencoder.

387 Stochastic inversion has also been applied to non-seismic data. Prob-
 388 abilistic joint inversion frameworks combine frequency-domain electromag-
 389 netic and DC resistivity data using Bayesian inference, improving resolution
 390 and reducing ambiguity through multi-method integration (Narciso et al.,
 391 2025). Geostatistical priors have been embedded into Bayesian inversion
 392 of near-surface petrophysical properties, using Monte Carlo sampling and
 393 ensemble smoother updates conditioned on electrical resistivity and seismic
 394 measurements, preserving geological spatial correlation while quantifying un-
 395 certainty (Grana et al., 2022).

396 Stochastic approaches have been extended to hydrogeophysical inversion,
 397 a framework that integrates hydrogeology and geophysics to estimate sub-
 398 surface properties that control groundwater flow and transport (Zhou et al.,
 399 2014). The ES-MDA has been used to estimate contaminant source location
 400 and release history in groundwater systems (Todaro et al., 2021).

401 Data-driven methods such as artificial neural networks offer an alter-
 402 native to stochastic inversion by approximating forward operators directly
 403 (Gramacy, 2020), motivating the deep generative and physics-informed ap-
 404 proaches discussed in the following subsections.

405 3.2 Deep Generative Models

406 Deep generative models address two limitations of classical geostatistical sim-
 407 ulation: reproducing complex geological patterns that two-point covariance
 408 models cannot capture, and conditioning on diverse data types within a uni-
 409 fied framework.

410 Generative adversarial networks (GANs) learn to generate realistic sam-
 411 ples from a training distribution by training two networks in opposition: a
 412 generator that produces synthetic samples and a discriminator that tries to
 413 distinguish them from real data. Applied to geological modeling, GANs can
 414 capture complex spatial patterns from training images and reproduce them
 415 at scales ranging from pore geometries (Mosser et al., 2017) to nonstation-
 416 ary sedimentary channels and elastic properties (Azevedo et al., 2020). A
 417 key challenge is conditioning the generated realizations on sparse direct mea-
 418 surements such as well data, which has been addressed by integrating condi-
 419 tioning constraints into the GAN training or sampling procedure (Chan and
 420 Elsheikh, 2019).

421 Variational autoencoders (VAEs) are generative models that learn a low-
 422 dimensional latent space where the latent variables follow a known prior dis-

423 tribution, typically Gaussian, from which new samples can be drawn and de-
424 coded into realistic outputs. Unlike GANs, this probabilistic structure allows
425 samples to be drawn and conditioned within a Bayesian framework, making
426 VAEs well suited to data assimilation workflows. The latent space also pro-
427 vides a compact parameterization of complex geological models, which can be
428 combined with ensemble-based data assimilation methods such as the ensem-
429 ble smoother to condition facies realizations on observed data (Canchumuni
430 et al., 2019).

431 Diffusion models generate samples by learning to reverse a gradual nois-
432 ing process: training data are progressively corrupted with noise over many
433 steps, and a neural network is trained to invert this process step by step,
434 transforming random noise into realistic samples. Latent diffusion models
435 perform this process in a compressed latent space rather than in the original
436 data space, which substantially reduces computational cost. In geological
437 modeling, latent diffusion has been used to conditionally simulate reservoir
438 facies from sparse well and seismic measurements (Lee et al., 2025). Figure 4
439 shows a schematic representation of conditional simulation of reservoir facies
440 models using latent diffusion.

441 Deep generative models are commonly evaluated against classical two-
442 point geostatistical methods, which can reproduce simple covariance struc-
443 tures but struggle with curvilinear or multi-point patterns. Comprehensive
444 statistical comparisons show that diffusion models can reproduce training
445 image statistics, maintain spatial continuity, and honor conditioning data
446 more accurately than two-point methods (Ovanger et al., 2025). Beyond
447 model generation alone, recent work has unified generation, forward surrogate
448 modeling, and inversion within a single architecture, demonstrating strong
449 generalization across hydrogeological modeling tasks (Zhan et al., 2025).

450 3.3 Physics-Informed Neural Networks

451 Physics-informed neural networks (PINNs) embed governing equations di-
452 rectly into neural network training by adding the residuals of those equa-
453 tions as terms in the loss function. This makes PINNs particularly valuable
454 in geoscience applications where data are sparse, expensive, or noisy, be-
455 cause the physics provides constraint where data do not (Raissi et al., 2019;
456 Karniadakis et al., 2021).

457 In subsurface characterization, PINNs have been applied to a range of
458 forward modeling problems, including seismic wave propagation, tomography

459 and full waveform inversion (RashtBehesht et al., 2022; Gou et al., 2023; Ren
 460 et al., 2024), groundwater flow simulation and transient unconfined aquifer
 461 systems (Tartakovsky et al., 2020; Secci et al., 2024), and single-phase and
 462 multi-phase fluid flow in heterogeneous reservoirs (Zhang, 2022; Zhang et al.,
 463 2023). A key advantage of PINNs over traditional numerical solvers lies in
 464 their meshless formulation, which handles complex boundary conditions such
 465 as absorbing boundaries in semi-infinite domains and enables incorporation of
 466 a priori geological knowledge (Karniadakis et al., 2021; RashtBehesht et al.,
 467 2022).

468 In geophysical inverse problems, PINNs serve as efficient surrogate models
 469 that rapidly evaluate forward operators during inversion, reducing computa-
 470 tional cost while maintaining physical consistency. Differentiable end-to-end
 471 PINN architectures, in which subsurface properties are reparameterized by
 472 neural networks coupled to differentiable forward models, enable gradient-
 473 based multi-objective optimization for joint inversion of multiple geophysical
 474 datasets, with applications to monitoring of carbon dioxide sequestration
 475 (Liu et al., 2023b). Bayesian PINNs (B-PINNs) extend the PINN framework
 476 to probabilistic inference, combining physics-informed priors with methods
 477 such as variational inference or Hamiltonian Monte Carlo to produce poste-
 478 rior distributions over model parameters and predictions rather than point
 479 estimates (Yang et al., 2021; Li et al., 2024). This enables rigorous quantifica-
 480 tion of both aleatoric and epistemic uncertainties in geophysical inverse prob-
 481 lems. Ensemble-based Bayesian inference strategies have also been developed
 482 to avoid the computational cost of traditional MCMC sampling (Pensoneault
 483 and Zhu, 2024).

484 Despite promising results, challenges remain in PDE convergence and in
 485 scaling to high-dimensional problems. Advances in Fourier feature embed-
 486 dings, domain decomposition, and adaptive training strategies are expand-
 487 ing the applicability of PINNs to increasingly complex geoscience problems
 488 (Song and Wang, 2022; Ding et al., 2025). Neural operator methods, which
 489 learn mappings between function spaces rather than between fixed inputs
 490 and outputs, have also been combined with finite-volume discretizations and
 491 physics-informed training to enable parametric learning of subsurface flow
 492 (Yan et al., 2025).

493 Physics-informed principles also extend beyond strict PINN formulations
 494 to other neural architectures, including generative models such as physics-
 495 informed GANs that embed physical constraints into the generator and pro-
 496 duce realizations consistent with both observed data and governing physics

497 (Miele and Azevedo, 2024). Figure 5 shows the seismic inversion workflow us-
498 ing a physics-informed W-Net GAN combined with multidimensional scaling
499 for model selection.

500 4 Applications

501 In this section, the application of modern geostatistical methods is illustrated
502 across key domains where spatial modeling and uncertainty quantification are
503 critical. Specifically, five areas of applications are presented: 1) groundwater
504 management, where understanding aquifer heterogeneity and flow dynam-
505 ics is essential for sustainable management; 2) energy exploration, including
506 hydrocarbon, geothermal, and subsurface storage systems, where character-
507 izing reservoirs and facies distributions guides exploration and production;
508 3) climate and drought modeling, where spatio-temporal modeling supports
509 the assessment and prediction of extreme events such as droughts; 4) en-
510 vironmental studies, where geostatistical techniques enable the monitoring
511 and mapping of contaminants, land-use impacts, and other environmental
512 variables; and 5) mining operations, where sparse, expensive drilling data
513 are transformed into probabilistic models of ore grade and tonnage distribu-
514 tions.

515 4.1 Groundwater management

516 Groundwater systems are highly heterogeneous, with hydraulic conductivity
517 and porosity varying across multiple spatial scales, which complicates the
518 prediction of flow and transport. Field methods such as pumping tests pro-
519 vide reliable but sparse information, and geostatistical models extend these
520 measurements by treating hydraulic properties as spatial random fields with
521 prescribed correlation structures.

522 Hydrogeophysical data, particularly electrical resistivity tomography (ERT),
523 provide an additional and independent constraint on subsurface characteriza-
524 tion, since the resistivity of the rock-fluid system depends on porosity, satura-
525 tion, and pore fluid conductivity, properties that also control hydraulic con-
526 ductivity. ERT surveys yield apparent resistivity values that require inverse
527 modeling to estimate the true spatial distribution of subsurface resistivity,
528 and several inversion strategies have been developed for this purpose. For
529 example, Fagandini et al. (2024) couple ERT with geostatistical groundwater

inversion to better resolve aquifer heterogeneity, with the latter specifically targeting contaminant plume delineation by combining hydrogeophysics with geostatistical data assimilation. Figure 6 shows the estimation of a contaminant plume from hydrogeophysical data measured in a controlled sandbox experiment mimicking a contaminated aquifer, using ES-MDA.

By enforcing constraints such as the Laplacian of hydraulic head for steady-state flow, PINNs generate solutions consistent with governing equations and remain effective under complex boundary conditions, heterogeneous or anisotropic aquifers, and transient unconfined flow. Recent work has applied PINNs to transient single-phase and two-phase Darcy flows in heterogeneous reservoirs (Zhang, 2022; Zhang et al., 2023), transient unconfined groundwater flow (Secci et al., 2024), and finite-volume-based parametric learning of subsurface flow (Yan et al., 2025), producing fast, reliable surrogates for forward and inverse groundwater modeling.

4.2 Energy exploration

Energy exploration relies on characterizing subsurface reservoirs from sparse and indirect data: well logs and core samples provide high-resolution information at a few locations, while seismic, electromagnetic, and other geophysical surveys offer broad spatial coverage at lower resolution. Reservoir properties such as porosity, permeability, and fluid saturation must be inferred between wells, integrated with the geophysical data, and propagated through forward models of fluid flow to support discovery, production planning, and monitoring of subsurface storage. Stochastic sampling and optimization methods provide flexible tools for the resulting high-dimensional inverse problems, though often at high computational cost for large datasets (Grana et al., 2021).

Stochastic sampling and optimization methods, such as ES-MDA, have been applied to large-scale seismic inversion and history matching in reservoir characterization, where they provide multiple geostatistical realizations conditioned on observed data and quantify uncertainty for production forecasting and decision making. Thurin et al. (2019) and Gineste et al. (2020) extend ensemble methods to full waveform inversion in frequency and time domain respectively. To address the high dimensionality of these applications, ensemble methods have also been combined with deep parameterizations that compress the model space before assimilation, using variational autoencoders (Canchumuni et al., 2019) or deep convolutional autoencoders

566 (Liu and Grana, 2020).

567 Deep learning has also been applied directly to seismic inversion for
568 impedance estimation, velocity estimation, and full waveform inversion, pro-
569 viding fast alternatives to traditional iterative inversion. Das et al. (2019) use
570 convolutional neural networks to estimate acoustic impedance from seismic
571 data, and Fabien-Ouellet and Sarkar (2020) apply deep recurrent networks
572 to seismic velocity estimation. Wu et al. (2021) introduce a semi-supervised
573 GAN framework for impedance inversion, and Sun et al. (2023) reformulate
574 full waveform inversion as an optimization over neural network weights, pro-
575 viding an implicit representation of the velocity model. Liu et al. (2024)
576 combine autoencoder-based model parameterization with Stein variational
577 gradient descent to integrate geophysical data with geological priors and
578 quantify uncertainty more effectively. Figure 7 shows an intergrated work-
579 flow for seismic inversion at well locations and along seismic lines, combining
580 machine learning parameterization with advanced sampling.

581 4.3 Climate and Drought

582 Climate extremes, particularly droughts, severely affect agriculture, water
583 resources, ecosystems, and society. Because droughts develop slowly and
584 exhibit multi-scale variability, their characterization relies on indices that
585 condense complex hydro-climatic processes into numerical measures (Zargar
586 et al., 2011). The standardized precipitation index, based on precipita-
587 tion anomalies, is the most widely used; more advanced metrics such as
588 the drought severity index and the standardized precipitation evapotran-
589 spiration index incorporate temperature and water balance to account for
590 climate-warming effects.

591 Geostatistical and spatio-temporal models play a central role in map-
592 ping drought severity from sparse station data and reconstructing drought
593 evolution over time. Varouchakis et al. (2018) apply spatial statistical anal-
594 yses to identify hydrological years, return periods, and spatial variability in
595 precipitation records. Machine learning extends these methods to drought
596 forecasting by learning nonlinear relationships between climatic predictors
597 and drought indices: Gomes et al. (2025) show that random forest meth-
598 ods can predict short-term drought severity using the consecutive dry days
599 index, effectively forecasting moderate to severe events during the driest
600 months. Figure 8 shows the classification of meteorological station data into
601 three drought-severity categories and their forecasting using a random forest

602 model.

603 A growing research direction combines geostatistics, data assimilation,
604 and machine learning to merge climate model outputs, in-situ records, and
605 satellite-based indices. Such hybrid approaches improve reconstruction of
606 past droughts and enhance prediction under climate-change scenarios (Park
607 et al., 2016; Diaz et al., 2020; He et al., 2023; Seo and Lee, 2023). Spatio-
608 temporal models can further support early-warning systems tailored to agri-
609 culture, water supply, and energy sectors.

610 4.4 Environmental studies

611 Environmental geostatistics addresses spatial assessment of contamination,
612 exposure, and ecosystem quality, where measurements are typically sparse,
613 costly, and influenced by anthropogenic and natural sources. Established
614 application areas include soil contamination mapping for heavy metals such
615 as cadmium and arsenic, air quality interpolation from sparse monitoring
616 networks, marine pollution assessment, and radioactive contamination char-
617 acterization. These applications share a common challenge: sparse, irregular
618 observations must be combined with auxiliary information and physical un-
619 derstanding to produce spatially explicit maps of contamination and risk
620 (Goovaerts, 2008).

621 Rapid urbanization and agricultural intensification are reshaping land-
622 scape dynamics and hydrological systems, generating significant pressures
623 on groundwater resources and ecosystem services. Spatial analyses of land
624 use and land cover change reveal how urban sprawl disrupts aquifer recharge,
625 fragments ecosystems, and degrades environmental quality, while intensified
626 agriculture increases agrochemical inputs and creates overlapping contamina-
627 tion risks. Rodrigo-Illarri et al. (2025) use spatial analysis to document how
628 urban expansion in the Valencia plain aquifer reduces aquifer recharge and
629 environmental value over time. Figure 9 shows the expansion of urban areas,
630 the decline of natural-rainfall-dependent agriculture, and the corresponding
631 drop in environmental value.

632 Pesticide and agrochemical contamination is a particularly challenging
633 case for environmental monitoring, because intensive land management prac-
634 tices such as flood-based irrigation can accelerate chemical infiltration far
635 beyond what sparse direct measurements can capture. Integrated geostatistical
636 approaches combine these limited observations with mechanistic trans-
637 port models and spatially explicit land use data to assess contamination risk.

For agrochemical contamination specifically, deterministic models such as the pesticide root zone model can generate concentration fields under different land use scenarios, which geostatistical methods then interpolate to quantify uncertainty and identify contamination hot spots (Tariq, 2023). The resulting probabilistic risk maps support sustainable land and water management in data-limited environments.

4.5 Mining operations

Mining remains a core application area of geostatistics, supporting mineral resource estimation, ore reserve classification, and mine planning from sparse and expensive drilling data (Isaaks and Srivastava, 1989; Emery and Sguret, 2021). Modern mining faces new challenges, including lower ore grades, deeper and more complex deposits, and operational pressures for near real-time decision making, all of which demand robust uncertainty quantification. The data context is asymmetric: drill cores provide high-resolution but spatially sparse measurements, while remote sensing and geophysical surveys offer broad coverage at lower resolution, and production-monitoring sensors generate streaming data that must be assimilated as they arrive.

Data science methods now complement classical geostatistical approaches across the mining workflow. Neural networks, random forests, and support vector machines have been used to capture nonlinear geological relationships, and deep learning architectures have been developed for specific mining tasks: Jordo et al. (2022) apply GANs to ore-type modeling in complex geological environments, and Avalos and Ortiz (2023) combine multivariate geostatistical simulation with deep Q-learning to optimize mining decisions under uncertainty. The role of geostatistical simulation across the life cycle of a mineral deposit, from exploration to closure, is reviewed by Deutsch (2023). At larger scales, remote sensing combined with machine learning has expanded prospectivity mapping and alteration zone detection (Shirmard et al., 2022; Davies et al., 2025), while at operational scales near real-time model updating from production sensors remains a key challenge (Neves et al., 2018; Benndorf, 2020). Figure 10 shows the joint distribution of X-ray fluorescence and laboratory grades used for near real-time estimation of ore grade distributions.

The energy transition is accelerating interest in critical minerals such as lithium, cobalt, nickel, copper, and rare earth elements, whose deposits often involve complex geology and multielement geochemical correlations.

Modern workflows increasingly rely on multivariate techniques, hierarchical models, and joint simulation to honor these geochemical relationships. Sustainability considerations further require integrating geological, geochemical, geophysical, and geospatial data to optimize extraction while minimizing water use, land disturbance, and community impacts. Current developments center on hybrid frameworks linking machine learning-based prospectivity, physics-informed or ensemble geostatistical resource estimation, stochastic optimization for mine planning (Dimitrakopoulos, 2018), and real-time data assimilation during production.

5 Outstanding Challenges

Despite the recent advances in the field of classical geostatistics and modern extensions based on deep learning and data-driven approaches, spatial modeling of surface and subsurface properties still faces several challenges. In the following, the main fundamental and methodological challenges are discussed.

5.1 Scale Integration and Data Assimilation

The scale mismatch between dense geophysical surveys and sparse borehole measurements, introduced in earlier sections, remains a persistent open challenge for geostatistical modeling. Bridging this mismatch is not solved by either data type alone: geophysical data require inversion to recover physical parameters, while borehole data require interpolation to provide spatial coverage (Grana et al., 2021).

Addressing this scale mismatch requires multi-resolution modeling strategies, upscaling and downscaling techniques, and careful treatment of measurement footprints. Advances in hierarchical modeling and multi-scale stochastic simulation provide a promising path to reconcile the scales, ensuring that geostatistical predictions remain both realistic and consistent with the physics of geophysical data (Linde et al., 2015). The integration of multiple data sources through joint inversion frameworks is an active research direction (Guo et al., 2020; Colombo et al., 2021; Hu et al., 2023; Ma et al., 2025), but important gaps remain. While joint inversion and multi-resolution strategies have reduced some of these discrepancies, merging information across scales remains far from straightforward.

Modern extensions integrating machine learning and AI offer new tools to learn scale-dependent relationships directly from data, but without geostatistical constraints they may reproduce patterns that are statistically plausible yet geologically unrealistic. Hybrid approaches that embed spatial correlation structures, measurement footprints, and physical consistency into data-driven models provide a more robust pathway for scale integration than simply increasing model complexity.

5.2 Geological Realism and Heterogeneity

Capturing realistic geological heterogeneity while maintaining computational tractability remains a fundamental challenge for geostatistical modeling. Traditional two-point methods are computationally efficient but cannot reproduce complex geological patterns or long-range connectivity (Hu and Chugunova, 2008; Tahmasebi et al., 2012; Mariethoz and Caers, 2014). Multiple-point statistics and training-image approaches improve realism but introduce their own difficulties, particularly in the selection and conditioning of training images and in updating realizations within an inversion framework, motivating recent extensions that combine MPS with deep learning (Avalos and Ortiz, 2020).

One direction of recent work uses machine learning to enhance classical geostatistical frameworks rather than replace them. High-order sequential simulation can be performed in reproducing kernel Hilbert spaces to capture statistical structure beyond two-point covariance (Yao et al., 2020), and generative networks combined with gradual deformation can produce facies realizations that honor both training-image patterns and classical update mechanisms (Feng et al., 2025). These hybrid approaches retain interpretable geostatistical structure while exploiting the pattern-recognition capacity of deep models. A second direction relies on deep generative models to reproduce complex geological patterns directly: GANs can perform conditional facies simulation (Song et al., 2021) and produce three-dimensional geological models from sparse borehole data combined with prior knowledge (Lyu et al., 2024), capturing structures such as channel systems, fractured networks, and carbonate facies distributions. Ensuring geological plausibility and interpretability of these learned models remains an open question, particularly when training data are limited or unrepresentative of the target geological setting.

The emergence of physics-guided generative models offers a promising

743 pathway for capturing geological complexity while preserving physical con-
744 sistency, but raises questions about how to incorporate geological knowledge
745 into the learning objective and ensure that generated realizations remain
746 physically meaningful. Hybrid workflows that combine classical geostatistical
747 priors with deep generative architectures represent one direction for
748 addressing this gap.

749 **5.3 Uncertainty Quantification in Complex Models**

750 Uncertainty quantification becomes substantially more challenging when deep
751 learning and physics-informed models are part of the workflow, because the
752 explicitly-specified uncertainty structure of classical geostatistics is no longer
753 guaranteed.

754 Hybrid workflows that combine classical geostatistical models with physics-
755 informed machine learning and deep generative approaches offer one route
756 to recover probabilistic rigor while exploiting the flexibility of AI. Bayesian
757 formulations, ensemble methods, and variational inference can be combined
758 with complex geostatistical priors, integrating deep architectures including
759 probabilistic and Bayesian neural networks within rigorous probabilistic frame-
760 works (Zhang and Curtis, 2021; Kirkwood et al., 2022; Romero et al., 2025).
761 Simulation-based and neural inference approaches (Cranmer et al., 2020) take
762 a different angle by training deep networks to learn complex likelihood mod-
763 els or posteriors directly, with recent applications to carbon dioxide storage
764 monitoring from well and seismic data (Gahlot et al., 2025).

765 Conformal prediction is a second promising direction, providing distribution-
766 free prediction sets with finite-sample coverage (Fontana et al., 2023). Unlike
767 classical parametric approaches, which typically require restrictive assump-
768 tions on the target process, conformal methods are nonparametric and can
769 be applied on top of existing statistical, machine learning, or AI-based tech-
770 niques, including high-dimensional and complex models (Messoudi et al.,
771 2021; Diquigiovanni et al., 2022; Bortolotti et al., 2024). In the framing
772 introduced in Section 2.3, conformal prediction offers a way to restore statis-
773 tical control over flexible models. Recent attempts to generalize conformal
774 prediction to spatial contexts have been made (Mao et al., 2024; Lou et al.,
775 2025) and this line of research is expected to remain active in the near fu-
776 ture. Across all of these directions, the goal is to recover the probabilistic
777 rigor of classical geostatistics within the flexibility of modern data-driven and
778 physics-informed methods.

779 5.4 Integration with Artificial Intelligence and Hybrid 780 Methods

781 The integration of artificial intelligence with geostatistical frameworks has
782 opened a broad set of opportunities. AI and big data are reshaping the geo-
783 statistical and Bayesian formalisms that have defined the field for decades
784 (De Iaco et al., 2022; Sinding-Larsen et al., 2025). For example, deep neural
785 networks have been embedded directly into spatial prediction tasks through
786 architectures that fold spatial dependence into the neural network rather than
787 treating it as an external correction. Beyond these conceptual integrations,
788 machine learning is also being used for targeted tasks within otherwise clas-
789 sical workflows, including acceleration of variogram inference for very large
790 datasets, generation or augmentation of training images for multiple-point
791 geostatistics when outcrop analogues are scarce, dimensionality reduction for
792 stochastic inversion, and detection of spatial anomalies in big data.

793 A particularly promising hybrid strategy is one in which classical and
794 AI methods are applied sequentially rather than as replacements for one an-
795 other. A classical geostatistical model is first fitted to the data under its full
796 set of explicit assumptions, yielding a prediction and a well-characterized
797 uncertainty structure. The ML or AI component is then applied not to
798 the original data but to the residuals to model the discrepancy between the
799 classical model predictions and the observations. If hybrid models identify
800 systematic structure in these residuals, such as non-Gaussian tails, direc-
801 tional anisotropy that varies spatially, or long-range connectivity not cap-
802 tured by the variogram, this provides concrete, interpretable evidence that
803 the classical model assumptions are violated. Sequential strategies preserve
804 statistical control over the primary component of the prediction, so that the
805 classical geostatistical predictions and their associated variance remain fully
806 interpretable. The added value of the ML component is directly measurable
807 through residual variance explained and predictive performance on held-out
808 data. Extending this principle to deep neural networks operating on geosta-
809 tistical residuals is a promising research direction that combines geological
810 realism from flexible models with traceable uncertainty from classical spatial
811 statistics.

812 Across these directions, the goal is to combine the strengths of both tradi-
813 tions: using geostatistical principles to anchor models in physical and prob-
814 abilistic settings while exploiting AI’s capacity to learn from large, complex
815 datasets. Such hybrid approaches aim for geological realism, computational

816 efficiency, and decision-relevant uncertainty estimates.

817 **5.5 Big Data and Spatio-Temporal Processes**

818 The constant growth of data from 3D seismic surveys, distributed acoustic
819 sensing using fiber optics, remote sensing platforms, and monitoring networks
820 has transformed geostatistical modeling. These big datasets hold unprece-
821 dented potential for subsurface characterization but also pose challenges in
822 storage, scalability, and model calibration. Machine learning and deep learn-
823 ing methods have been increasingly applied to tasks on these large datasets,
824 including event detection on distributed acoustic sensing recordings (Walter
825 et al., 2020; Yu et al., 2024b) and three-dimensional microseismic event loca-
826 tion and high-resolution seismic imaging (Yang et al., 2024; Ma et al., 2024).
827 At the same time, the dimension of the datasets enables geostatistical meth-
828 ods to move beyond interpolation and sampling, toward real-time monitoring,
829 adaptive sampling, and dynamic updating of models under uncertainty.

830 Monitoring and calibration data are often costly, making it essential to
831 determine not only how to integrate large datasets but which data are worth
832 acquiring. Value-of-information frameworks provide a principled way to eval-
833 uate the expected benefit of new observations relative to their expense, help-
834 ing prioritize what data to collect and where to deploy monitoring efforts
835 (Eidsvik et al., 2015). This shift emphasizes strategic data acquisition rather
836 than simply increasing data volume, ensuring that geostatistical models are
837 updated efficiently and in a decision-relevant manner.

838 Many problems in Earth, energy, and environmental studies involve pro-
839 cesses that evolve over time, such as reservoir production, carbon dioxide
840 plume migration, aquifer contamination, and geothermal circulation. Tradi-
841 tional geostatistics is primarily spatial, but dynamic systems require coupled
842 spatio-temporal models that integrate monitoring data as they become avail-
843 able. The challenge is to extend geostatistical methods to time-dependent
844 processes in a way that is both computationally feasible and physically con-
845 sistent, while accounting for uncertainties in evolving states. Advances in
846 sequential data assimilation, ensemble methods, and physics-informed spatio-
847 temporal geostatistics represent promising directions (Cappello et al., 2022;
848 Wikle and Zammit-Mangion, 2023; Yu et al., 2024a), but bridging theory
849 with field-scale applications remains a critical frontier.

850 Across all of these directions, the practical realization of geostatistical
851 methods at scale will depend on continued advances in computational infras-

852 structure. High-performance computing, parallel algorithms, and dimensionality-
853 reduction techniques are essential for handling massive datasets, running
854 ensemble-based spatio-temporal models, and supporting adaptive workflows
855 where models are updated in near real time. Co-developing geostatistical
856 methodology with the computational tools needed to deploy it at field scale
857 is itself an open research frontier.

858 **6 Future Perspectives**

859 The future of geostatistics will be shaped by both the evolving demands of
860 society and continued methodological innovation. The following sections out-
861 line key perspectives on advancing computational and theoretical frontiers,
862 emerging application domains in energy transitions, and fostering collabora-
863 tion and effective communication across disciplines and with decision-makers.

864 **6.1 Methodological Frontiers**

865 In the coming decade, several methodological developments are poised to
866 transform geostatistics. Advances in deep generative modeling, such as dif-
867 fusion models, normalizing flows, and hybrid architectures, will increasingly
868 be tailored to spatial data and geological constraints, enhancing the realism
869 and flexibility of statistical realizations and potentially providing a class of
870 pre-trained foundational models for multiple purposes and applications. At
871 the same time, scalability will become a central focus, with new algorithms
872 designed to process the ever-growing volume of data generated by distributed
873 sensors, satellite images, and continuous monitoring systems. Multi-scale in-
874 tegration will also gain prominence, with techniques that connect fine-scale
875 local measurements with broad-coverage regional and global data across the
876 application domains discussed in this review.

877 These methodological developments pose new fundamental questions about
878 the scope and target of the models. In the classical view of numerical Earth
879 science modeling (Oreskes et al., 1994), a model is a structured set of assump-
880 tions, and its predictions are valid insofar as those assumptions hold, which
881 implies that discrepancies between model and data must be explained by
882 identifying which assumptions fail. Classical geostatistics fits this view, be-
883 cause assumptions are explicit, the uncertainty model is fully specified, and
884 out-of-sample performance can be traced back to identifiable components.

885 The growing use of ML/AI in geostatistics introduces a less clear relationship
886 between models and reality. Deep generative models or a physics-informed
887 neural networks may produce prediction that matches very accurately the
888 training data but they assumptions are not always clearly identified nor geo-
889 logically justified. Yet, the classical models may be so inadequate that even a
890 less interpretable improvement is operationally valuable. Hybrid workflows,
891 conformal prediction methods, and Bayesian deep learning frameworks are
892 all partial responses to this challenge.

893 **6.2 Sustainable Energy Transitions**

894 Geostatistics has traditionally been applied to resource estimation in min-
895 ing, petroleum, and groundwater studies, but its potential reaches far be-
896 yond these areas. As energy systems transition toward renewable sources
897 and carbon neutrality, geostatistics faces new application domains requiring
898 methodological adaptation. Carbon capture and storage demands charac-
899 terization of storage capacity, injectivity, and containment assurance, which
900 can be predicted and quantified from the integration of sparse well data with
901 geophysical monitoring. Hydrogen storage in subsurface formations presents
902 similar challenges, with additional complexities related to gas-water-rock in-
903 teractions and the sealing capability of the cap rock. Geothermal energy
904 development depends on geostatistical characterization of temperature, per-
905 meability, and fracture networks, and the integration of thermal monitoring
906 with hydrogeological and geophysical observations will be essential for sus-
907 tainable reservoir management.

908 In parallel, mining and traditional energy resource exploration continue
909 to evolve toward environmental sustainability, waste reduction, and real-time
910 adaptive extraction. Modern sensor technologies and continuous monitoring
911 enable dynamic updating of ore-body and reservoir models through sequen-
912 tial data assimilation, while integration with environmental and social im-
913 pact modeling requires geostatistical characterization of contaminant trans-
914 port and remediation scenarios. The breadth of these emerging applications,
915 from carbon neutrality to critical mineral extraction, positions geostatistics
916 as a core methodology for the data challenges of the energy transition.

917 6.3 Community and Collaboration

918 The advancement of geostatistics depends critically on collaboration and
919 knowledge sharing across disciplines, including geophysics, hydrology, energy
920 resources engineering, mining, climatology, oceanography, environmental sci-
921 ence, and machine learning. Proactively positioning geostatistics within
922 emerging application domains, from carbon capture and storage to climate-
923 change adaptation, will increase its visibility, attract interdisciplinary collab-
924 orations, and highlight its ability to handle sparse, uncertain, and spatially
925 correlated data that characterize many sustainability-related challenges.

926 Several concrete community practices support this advancement. Stan-
927 dardized benchmarking datasets and comparative studies enable objective
928 evaluation of competing methods and guide practitioners toward appropri-
929 ate tool selection. Educational initiatives, such as hackathons and dedicated
930 geospatial modeling courses, promote geostatistical thinking and a working
931 understanding of spatial correlation, uncertainty quantification, and data in-
932 tegration. Open-source software accelerates innovation and fosters trust in
933 new methods. Regular exchange between the geostatistics, machine learning,
934 and applied geoscience communities will help interdisciplinary projects and
935 ideas develop.

936 An equally critical dimension is the communication of results to poli-
937 cymakers and stakeholders. Decision-makers often need clear, actionable
938 information rather than technical details about modeling uncertainty. Geo-
939 statistical risk maps can be translated into more intuitive categories of risk
940 (low, moderate, high), accompanied by confidence intervals or traffic-light
941 color schemes to facilitate understanding. Coupling geostatistical results
942 with narratives about anthropogenic practices and long-term sustainability
943 can strengthen the case for adaptive Earth management policies.

944 7 Conclusion

945 Modern geostatistics stands at a juncture where decades of methodological
946 development in spatial statistics and inverse theory meet recent advances in
947 deep learning and physics-informed computation. The field has expanded
948 from its traditional domains in mining and petroleum engineering to ad-
949 dress pressing challenges in groundwater management, climate adaptation,
950 and environmental monitoring. Key advances reviewed here include stochas-

951 tic inversion methods for integrating geophysical and direct measurements,
 952 deep generative models including GANs, variational autoencoders, and dif-
 953 fusion models for geological model generation, and physics-informed neural
 954 networks for forward modeling and inversion. These methodological devel-
 955 opments have been illustrated through applications to hydrogeophysical in-
 956 version, energy resources characterization and monitoring, spatio-temporal
 957 modeling of climate extremes, environmental assessment, and mining opera-
 958 tions.

959 Critical challenges remain: bridging scales from high-resolution models
 960 to coarse geophysical data, capturing geological heterogeneity realistically
 961 while maintaining computational tractability, ensuring rigorous uncertainty
 962 quantification in hybrid classical-AI workflows, integrating massive datasets
 963 and spatio-temporal processes, and developing interpretable methods that
 964 incorporate domain knowledge. Addressing these challenges requires the in-
 965 tegration of classical geostatistical principles with modern computational ap-
 966 proaches, with a focus on geological realism, physical consistency, and quan-
 967 tified uncertainty. Hybrid workflows that combine classical sampling and
 968 optimization with deep learning show strong potential for advancing surface
 969 and subsurface characterization on large datasets.

970 The future of geostatistics over the coming decade will be defined by the
 971 community’s ability to harness emerging technologies while preserving statis-
 972 tical rigor and domain knowledge integration. This trajectory will position
 973 geostatistics as an essential enabler of sustainable energy systems, resilient
 974 water resources, responsible mining, and effective climate adaptation strate-
 975 gies, providing a scientifically rigorous and transparent framework that en-
 976 ables informed decisions across energy, natural resources, and environmental
 977 applications.

978 Glossary

979 **Aleatoric uncertainty:** The irreducible variability inherent in a phenomenon
 980 or its observations, including natural spatial heterogeneity and mea-
 981 surement noise that cannot be reduced by collecting more data.

982 **Bayesian neural network:** A neural network in which weights are treated
 983 as probability distributions rather than point values, enabling uncer-
 984 tainty quantification of predictions.

985 **Conditional simulation:** A stochastic simulation in which generated re-
986 alizations honor observed data at their measurement locations while
987 reproducing the modeled spatial variability.

988 **Conformal prediction:** A nonparametric framework that produces predic-
989 tion sets with finite-sample coverage guarantees, applicable on top of
990 any underlying statistical, machine learning, or AI-based model.

991 **Deep learning:** A class of machine-learning methods based on multilayer
992 neural networks capable of learning complex, high-dimensional rela-
993 tionships from large datasets.

994 **Diffusion model:** A generative model that learns to synthesize data by re-
995 versing a gradual noise-adding diffusion process, enabling high-fidelity
996 sampling of complex spatial patterns.

997 **Ensemble smoother:** A data-assimilation method that updates an ensem-
998 ble of model realizations using observations, typically in a single or a
999 small number of smoothing steps.

1000 **Epistemic uncertainty:** The reducible uncertainty arising from limited
1001 data, imperfect model structure, or modeling assumptions, which can
1002 in principle be reduced by acquiring more data or improving the model.

1003 **Functional Kriging:** A Kriging method that treats observations as func-
1004 tional data and models spatial dependence of these functions rather
1005 than scalar values.

1006 **Generative Adversarial Network (GAN):** A deep generative model com-
1007 posed of a generator and discriminator trained adversarially to produce
1008 realistic synthetic samples.

1009 **Geostatistical simulation:** A stochastic modeling approach that gener-
1010 ates multiple equally probable realizations of a spatial variable, pre-
1011 serving spatial continuity and quantifying uncertainty.

1012 **Geostatistics:** A branch of spatial statistics that models, analyzes, and pre-
1013 dicts spatially correlated variables using probabilistic and interpolation
1014 methods.

1015 **Hydrogeophysics:** The integration of geophysical measurements with hy-
1016 drological modeling to characterize subsurface processes such as ground-
1017 water flow, storage, and contamination.

1018 **Inverse problem:** A problem in which unknown model parameters are esti-
1019 mated from indirect or noisy observations, often ill-posed and requiring
1020 regularization or prior information.

1021 **Kriging:** A best linear unbiased interpolation method that predicts values
1022 at unsampled locations using spatial correlation structures encoded in
1023 a variogram or covariance model.

1024 **Markov Chain Monte Carlo (MCMC):** A family of algorithms that draws
1025 samples from complex probability distributions by constructing Markov
1026 chains whose stationary distribution is the target posterior.

1027 **Multiple-point statistics (MPS):** A class of geostatistical methods that
1028 extracts higher-order spatial statistics from a training image to repro-
1029 duce complex geological patterns beyond two-point covariance models.

1030 **Physics-Informed Neural Network (PINN):** A neural network that em-
1031 beds physical laws, expressed as differential equations or constraints,
1032 directly into its loss function to guide training and ensure physical con-
1033 sistency.

1034 **Stochastic inversion:** An inversion approach that estimates a probability
1035 distribution of subsurface properties rather than a single deterministic
1036 model, using random sampling or ensemble methods.

1037 **Training image:** A reference geological model, often derived from outcrop
1038 analogues or forward simulations, used in multiple-point statistics to
1039 capture the spatial patterns reproduced in generated realizations.

1040 **Variational Autoencoder (VAE):** A generative neural network that learns
1041 a low-dimensional latent representation of data by optimizing a varia-
1042 tional lower bound on the likelihood.

1043 **Variogram:** A function describing the spatial correlation structure of a ran-
1044 dom field as a function of separation distance and direction.

1045 Author Contributions

1046 **Conceptualization:** All authors contributed to the conceptualization in
1047 dedicated discussion session at the GeoNERDs conference.

1048 **Writing original draft:** Grana.

1049 **Writing review and editing:** All authors.

1050

1051 References

1052 Avalos, S. and Ortiz, J. M. (2020). Recursive convolutional neural net-
1053 works in a multiple-point statistics framework. *Computers & geosciences*,
1054 141:104522.

1055 Avalos, S. and Ortiz, J. M. (2023). Multivariate geostatistical simulation and
1056 deep q-learning to optimize mining decisions. *Mathematical Geosciences*,
1057 55(5):673692.

1058 Azevedo, L., Paneiro, G., Santos, A., and Soares, A. (2020). Generative ad-
1059 versarial network as a stochastic subsurface model reconstruction. *Com-
1060 putational Geosciences*, 24(4):1673–1692.

1061 Bai, T. and Tahmasebi, P. (2021). Accelerating geostatistical modeling us-
1062 ing geostatistics-informed machine learning. *Computers & Geosciences*,
1063 146:104663.

1064 Benndorf, J. (2020). *Closed loop management in mineral resource extraction:
1065 turning online geo-data into mining intelligence*. Springer Nature.

1066 Bortolotti, T., Casu, F., Virelli, M., Tapete, D., Siciliani De Cumis, M.,
1067 Menafoglio, A., and Vantini, S. (2024). Remote monitoring of ground
1068 motion exploiting functional data analysis and nonparametric prediction
1069 on multi-temporal dinsar data. In *IAF Earth Observation Symposium*,
1070 pages 967–975.

1071 Bürkle, P. Y., Azevedo, L., and Vellasco, M. (2023). Deep physics-aware
1072 stochastic seismic inversion. *Geophysics*, 88(1):R11–R24.

1073 Caers, J. (2011). *Modeling uncertainty in the earth sciences*. John Wiley &
1074 Sons.

- 1075 Canchumuni, S, W. A., Emerick, A. A., and Pacheco, M. A. C. (2019).
1076 Towards a robust parameterization for conditioning facies models using
1077 deep variational autoencoders and ensemble smoother. *Computers & Geo-*
1078 *sciences*, 128:87–102.
- 1079 Cappello, C., De Iaco, S., and Palma, M. (2022). Computational advances
1080 for spatio-temporal multivariate environmental models. *Computational*
1081 *Statistics*, 37(2):651–670.
- 1082 Carrassi, A., Bocquet, M., Bertino, L., and Evensen, G. (2018). Data assim-
1083 ilation in the geosciences: An overview of methods, issues, and perspectives.
1084 *Wiley Interdisciplinary Reviews: Climate Change*, 9(5):e535.
- 1085 Chan, S. and Elsheikh, A. H. (2019). Parametric generation of conditional
1086 geological realizations using generative neural networks. *Computational*
1087 *Geosciences*, 23:925–952.
- 1088 Chen, W., Li, Y., Reich, B. J., and Sun, Y. (2020). Deepkriging: Spatially
1089 dependent deep neural networks for spatial prediction. *arXiv preprint*
1090 *arXiv:2007.11972*.
- 1091 Chiles, J.-P. and Delfiner, P. (2012). *Geostatistics: modeling spatial uncer-*
1092 *tainty*. John Wiley & Sons.
- 1093 Colombo, D., Turkoglu, E., Li, W., and Rovetta, D. (2021). Coupled physics-
1094 deep learning inversion. *Computers & Geosciences*, 157:104917.
- 1095 Cranmer, K., Brehmer, J., and Louppe, G. (2020). The frontier of
1096 simulation-based inference. *Proceedings of the National Academy of Sci-*
1097 *ences*, 117(48):30055–30062.
- 1098 Das, V., Pollack, A., Wollner, U., and Mukerji, T. (2019). Convolutional
1099 neural network for seismic impedance inversion. *Geophysics*, 84(6):R869–
1100 R880.
- 1101 Davies, R. S., Trott, M., Georgi, J., and Farrar, A. (2025). Artificial intelli-
1102 gence and machine learning to enhance critical mineral deposit discovery.
1103 *Geosystems and Geoenvironment*, 4(2):100361.
- 1104 De Iaco, S., Hristopulos, D. T., and Lin, G. (2022). Special issue: Geostatis-
1105 tics and machine learning. *Mathematical Geosciences*, 54(3):459465.

- 1106 Delicado, P., Giraldo, R., Comas, C., and Mateu, J. (2010). Statistics for
1107 spatial functional data: some recent contributions. *Environmetrics: The*
1108 *official journal of the International Environmetrics Society*, 21(3-4):224–
1109 239.
- 1110 Deutsch, C. V. (2023). The place of geostatistical simulation through the life
1111 cycle of a mineral deposit. *Minerals*, 13(11):1400.
- 1112 Diaz, V., Perez, G. A. C., Van Lanen, H. A., Solomatine, D., and
1113 Varouchakis, E. A. (2020). An approach to characterise spatio-temporal
1114 drought dynamics. *Advances in Water Resources*, 137:103512.
- 1115 Dimitrakopoulos, R., editor (2018). *Advances in Applied Strategic Mine Plan-*
1116 *ning*. Springer International Publishing, Cham, Switzerland.
- 1117 Ding, Y., Chen, S., Miyake, H., and Li, X. (2025). Physics-informed neural
1118 networks with fourier features for seismic wavefield simulation in time-
1119 domain nonsmooth complex media. *IEEE Transactions on Geoscience*
1120 *and Remote Sensing*, 63:113.
- 1121 Diquigiovanni, J., Fontana, M., and Vantini, S. (2022). Conformal prediction
1122 bands for multivariate functional data. *Journal of Multivariate Analysis*,
1123 189:104879.
- 1124 Eidsvik, J., Mukerji, T., and Bhattacharjya, D. (2015). *Value of information*
1125 *in the earth sciences: Integrating spatial modeling and decision analysis*.
1126 Cambridge University Press.
- 1127 Emerick, A. A. and Reynolds, A. C. (2013). Ensemble smoother with multiple
1128 data assimilation. *Computers & Geosciences*, 55:3–15.
- 1129 Emery, X. and Sguret, S. A. (2021). *Geostatistics for the Mining Industry:*
1130 *Applications to Porphyry Copper Deposits*. CRC Press / Taylor & Francis
1131 Group, Boca Raton / London.
- 1132 Evensen, G. (2009). *Data assimilation: the ensemble Kalman filter*. Springer.
- 1133 Evensen, G., Vossepoel, F. C., and Van Leeuwen, P. J. (2022). *Data as-*
1134 *simulation fundamentals: A unified formulation of the state and parameter*
1135 *estimation problem*. Springer Nature.

- 1136 Fabien-Ouellet, G. and Sarkar, R. (2020). Seismic velocity estimation: A
1137 deep recurrent neural-network approach. *Geophysics*, 85(1):U21–U29.
- 1138 Fagandini, C., Todaro, V., Escada, C., Azevedo, L., Gómez-Hernández, J.,
1139 and Zanini, A. (2024). Coupled hydrogeophysical inversion through en-
1140 semble smoother with multiple data assimilation and convolutional neural
1141 network for contaminant plume reconstruction. *Stochastic Environmental*
1142 *Research and Risk Assessment*, 38(11):4227–4242.
- 1143 Feng, R., Grana, D., and Mosegaard, K. (2025). Geostatistical facies sim-
1144 ulation based on training image using generative networks and gradual
1145 deformation. *Mathematical Geosciences*, pages 1–24.
- 1146 Fontana, M., Zeni, G., and Vantini, S. (2023). Conformal prediction: a
1147 unified review of theory and new challenges. *Bernoulli*, 29(1):1–23.
- 1148 Gahlot, A. P., Orozco, R., Yin, Z., Bruer, G., and Herrmann, F. J. (2025). An
1149 uncertainty-aware digital shadow for underground multimodal co2 storage
1150 monitoring. *Geophysical Journal International*, 242(1):ggaf176.
- 1151 Gineste, M., Eidsvik, J., and Zheng, Y. (2020). Ensemble-based seismic
1152 inversion for a stratified medium. *Geophysics*, 85(1):R29–R39.
- 1153 Gomes, M., Castro, A., Durão, R., and Soares, A. (2025). Geostatistical
1154 predictive model of drought severity: A case study of southern portugal.
1155 *Mathematical Geosciences*, pages 1–20.
- 1156 Goovaerts, P. (1997). *Geostatistics for natural resources evaluation*. Oxford
1157 university press.
- 1158 Goovaerts, P. (2008). Geostatistical analysis of health and exposure data:
1159 State of the art and perspectives. *Epidemiology*, 19(6):S20–S21.
- 1160 Gou, R., Zhang, Y., Zhu, X., and Gao, J. (2023). Bayesian physics-informed
1161 neural networks for the subsurface tomography based on the eikonal equa-
1162 tion. *IEEE Transactions on Geoscience and Remote Sensing*, 61:112.
- 1163 Gramacy, R. B. (2020). *Surrogates: Gaussian process modeling, design, and*
1164 *optimization for the applied sciences*. Chapman and Hall/CRC.
- 1165 Grana, D., Mukerji, T., and Doyen, P. (2021). *Seismic reservoir modeling:*
1166 *Theory, examples, and algorithms*. John Wiley & Sons.

- 1167 Grana, D., Parsekian, A. D., Flinchum, B. A., Callahan, R. P., Smeltz, N. Y.,
1168 Li, A., Hayes, J. L., Carr, B. J., Singha, K., Riebe, C. S., and Holbrook,
1169 W. S. (2022). Geostatistical rock physics inversion for predicting the spatial
1170 distribution of porosity and saturation in the critical zone. *Mathematical*
1171 *Geosciences*, 54(8):1315–1345.
- 1172 Guo, Q., Liu, M., and Luo, J. (2023). Predictive deep learning for high-
1173 dimensional inverse modeling of hydraulic tomography in gaussian and
1174 non-gaussian fields. *Water Resources Research*, 59(10):e2023WR035408.
- 1175 Guo, R., Yao, H. M., Li, M., Ng, M. K. P., Jiang, L., and Abubakar, A.
1176 (2020). Joint inversion of audio-magnetotelluric and seismic travel time
1177 data with deep learning constraint. *IEEE Transactions on Geoscience and*
1178 *Remote Sensing*, 59(9):7982–7995.
- 1179 He, X., Li, Y., Liu, S., Xu, T., Chen, F., Li, Z., Zhang, Z., Liu, R., Song,
1180 L., Xu, Z., et al. (2023). Improving regional climate simulations based on
1181 a hybrid data assimilation and machine learning method. *Hydrology and*
1182 *Earth System Sciences*, 27(7):1583–1606.
- 1183 Hu, L. and Chugunova, T. (2008). Multiple-point geostatistics for model-
1184 ing subsurface heterogeneity: A comprehensive review. *Water resources*
1185 *research*, 44(11).
- 1186 Hu, Y., Wei, X., Wu, X., Sun, J., Chen, J., Huang, Y., and Chen, J. (2023).
1187 A deep learning-enhanced framework for multiphysics joint inversion. *Geo-*
1188 *physics*, 88(1):K13–K26.
- 1189 Isaaks, E. H. and Srivastava, R. M. (1989). *An Introduction to Applied*
1190 *Geostatistics*. Oxford University Press, New York.
- 1191 Jordo, H., Azevedo, L., Sousa, A. J., and Soares, A. (2022). Generative
1192 adversarial network applied to ore type modeling in complex geological
1193 environments. *Mathematical Geosciences*, 54(7):1165–1182.
- 1194 Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., and
1195 Yang, L. (2021). Physics-informed machine learning. *Nature Reviews*
1196 *Physics*, 3(6):422–440.

- 1197 Kirkwood, C., Economou, T., Pugeault, N., and Odbert, H. (2022). Bayesian
1198 deep learning for spatial interpolation in the presence of auxiliary infor-
1199 mation. *Mathematical Geosciences*, 54(3):507–531.
- 1200 Kitanidis, P. K. (1997). *Introduction to Geostatistics: Applications in Hy-*
1201 *drogeology*. Cambridge University Press.
- 1202 Laloy, E., Hérault, R., Jacques, D., and Linde, N. (2018). Training-image
1203 based geostatistical inversion using a spatial generative adversarial neural
1204 network. *Water Resources Research*, 54(1):381–406.
- 1205 Lee, D., Ovanger, O., Eidsvik, J., Aune, E., Skauvold, J., and Hauge, R.
1206 (2025). Latent diffusion model for conditional reservoir facies generation.
1207 *Computers & Geosciences*, 194:105750.
- 1208 Levy, S., Laloy, E., and Linde, N. (2023). Variational Bayesian inference with
1209 complex geostatistical priors using inverse autoregressive flows. *Computa-*
1210 *tional Geosciences*, 171:105263.
- 1211 Li, P., Grana, D., and Liu, M. (2024). Bayesian neural network and bayesian
1212 physics-informed neural network via variational inference for seismic petro-
1213 physical inversion. *Geophysics*, 89(6):M185–M196.
- 1214 Linde, N., Renard, P., Mukerji, T., and Caers, J. (2015). Geological realism
1215 in hydrogeological and geophysical inverse modeling: A review. *Advances*
1216 *in Water Resources*, 86:86–101.
- 1217 Liu, L., Prodanović, M., and Pyrcz, M. J. (2023a). Impact of geostatistical
1218 nonstationarity on convolutional neural network predictions. *Computa-*
1219 *tional Geosciences*, 27(1):35–44.
- 1220 Liu, M. and Grana, D. (2020). Petrophysical characterization of deep saline
1221 aquifers for co2 storage using ensemble smoother and deep convolutional
1222 autoencoder. *Advances in Water Resources*, 142:103634.
- 1223 Liu, M., Grana, D., and Mukerji, T. (2024). Geostatistical inversion for
1224 subsurface characterization using stein variational gradient descent with
1225 autoencoder neural network: An application to geologic carbon sequestra-
1226 tion. *Journal of Geophysical Research: Solid Earth*, 129(7).

- 1227 Liu, M., Vashisth, D., Grana, D., and Mukerji, T. (2023b). Joint inversion
1228 of geophysical data for geologic carbon sequestration monitoring: A differ-
1229 entiable physics-informed neural network model. *Journal of Geophysical*
1230 *Research: Solid Earth*, 128(3):e2022JB025372.
- 1231 Lou, X., Luo, P., and Meng, L. (2025). Geoconformal prediction: a model-
1232 agnostic framework for measuring the uncertainty of spatial prediction.
1233 *Annals of the American Association of Geographers*, pages 1–28.
- 1234 Lyu, B., Wang, Y., and Shi, C. (2024). Multi-scale generative adversarial
1235 networks (gan) for generation of three-dimensional subsurface geological
1236 models from limited boreholes and prior geological knowledge. *Computers*
1237 *and Geotechnics*, 170:106336.
- 1238 Ma, J., Deng, Y., Li, X., Guo, R., Zhou, H., and Li, M. (2025). Recent
1239 advances in machine learning-enhanced joint inversion of seismic and elec-
1240 tromagnetic data. *Surveys in Geophysics*, 46(1):197–225.
- 1241 Ma, L., Han, L., and Feng, Q. (2024). Deep learning for high-resolution
1242 seismic imaging. *Scientific Reports*, 14(1):10319.
- 1243 Mao, H., Martin, R., and Reich, B. J. (2024). Valid model-free spatial predic-
1244 tion. *Journal of the American Statistical Association*, 119(546):904–914.
- 1245 Mariethoz, G. and Caers, J. (2014). *Multiple-point geostatistics: stochastic*
1246 *modeling with training images*. John Wiley & Sons.
- 1247 Martínez-Hernández, I. and Genton, M. G. (2020). Recent developments
1248 in complex and spatially correlated functional data. *Brazilian Journal of*
1249 *Probability and Statistics*, 34(2):204 – 229.
- 1250 Menafoglio, A. (2021). *Spatial Statistics for Distributional Data in Bayes*
1251 *Spaces: From Object-Oriented Kriging to the Analysis of Warping Func-*
1252 *tions*, pages 207–224. Springer International Publishing, Cham.
- 1253 Menafoglio, A. (2025). Moving object-oriented spatial statistics beyond sta-
1254 tionary and euclidean paradigms. In *International Workshop on Functional*
1255 *and Operatorial Statistics*, pages 393–400. Springer.
- 1256 Menafoglio, A. and Secchi, P. (2017). Statistical analysis of complex and
1257 spatially dependent data: a review of object oriented spatial statistics.
1258 *European journal of operational research*, 258(2):401–410.

- 1259 Messoudi, S., Destercke, S., and Rousseau, S. (2021). Copula-based
1260 conformal prediction for multi-target regression. *Pattern Recognition*,
1261 120:108101.
- 1262 Miele, R. and Azevedo, L. (2024). Physics-informed W-Net GAN for the
1263 direct stochastic inversion of fullstack seismic data into facies models. *Sci-*
1264 *entific Reports*, 14(1):5122.
- 1265 Miele, R., Levy, S., Linde, N., Soares, A., and Azevedo, L. (2024). Deep
1266 generative networks for multivariate fullstack seismic data inversion using
1267 inverse autoregressive flows. *Computers & Geosciences*, 188:105622.
- 1268 Mosser, L., Dubrule, O., and Blunt, M. J. (2017). Reconstruction of three-
1269 dimensional porous media using generative adversarial neural networks.
1270 *Physical Review E*, 96(4):043309.
- 1271 Narciso, J., Verhegge, J., and Van De Vijver, E. (2025). Geostatistical joint
1272 inversion of frequency-domain electromagnetic data and direct current re-
1273 sistivity data for near-surface modelling. *Scientific Reports*, 15(1).
- 1274 Neves, J., Pereira, M. J., Pacheco, N., and Soares, A. (2018). Updat-
1275 ing mining resources with uncertain data. *Mathematical Geosciences*,
1276 51(7):905924.
- 1277 Ojo, O. T. and Genton, M. G. (2025). Functional multiple-point simulation.
1278 *Computers & Geosciences*, 195:105767.
- 1279 Oliver, D. S. and Chen, Y. (2011). Recent progress on reservoir history
1280 matching: a review. *Computational Geosciences*, 15(1):185–221.
- 1281 Oliver, D. S., Reynolds, A. C., and Liu, N. (2008). *Inverse theory for*
1282 *petroleum reservoir characterization and history matching*. Cambridge
1283 University Press.
- 1284 Oreskes, N., Shrader-Frechette, K., and Belitz, K. (1994). Verification, val-
1285 idation, and confirmation of numerical models in the earth sciences. *Sci-*
1286 *ence*, 263(5147):641–646.
- 1287 Ovanger, O., Lee, D., Eidsvik, J., Hauge, R., Skauvold, J., and Aune, E.
1288 (2025). A statistical study of latent diffusion models for geological facies
1289 modeling. *Mathematical Geosciences*, pages 1–25.

- 1290 Park, S., Im, J., Jang, E., and Rhee, J. (2016). Drought assessment and mon-
1291 itoring through blending of multi-sensor indices using machine learning ap-
1292 proaches for different climate regions. *Agricultural and forest meteorology*,
1293 216:157–169.
- 1294 Pawlowsky-Glahn, V., Egozcue, J. J., and Tolosana-Delgado, R. (2015). *Mod-
1295 eling and analysis of compositional data*. John Wiley & Sons.
- 1296 Peli, R., Menafoglio, A., Cervino, M., Dovera, L., and Secchi, P. (2022).
1297 Physics-based residual kriging for dynamically evolving functional ran-
1298 dom fields. *Stochastic Environmental Research and Risk Assessment*,
1299 36(10):3063–3080.
- 1300 Pensoneault, A. and Zhu, X. (2024). Efficient bayesian physics informed neu-
1301 ral networks for inverse problems via ensemble kalman inversion. *Journal
1302 of Computational Physics*, 508:113006.
- 1303 Pigoli, D., Menafoglio, A., and Secchi, P. (2016). Kriging prediction for
1304 manifold-valued random fields. *Journal of Multivariate Analysis*, 145:117–
1305 131.
- 1306 Pyrcz, M. J. and Deutsch, C. V. (2014). *Geostatistical reservoir modeling*.
1307 Oxford university press.
- 1308 Raissi, M., Perdikaris, P., and Karniadakis, G. (2019). Physics-informed
1309 neural networks: A deep learning framework for solving forward and in-
1310 verse problems involving nonlinear partial differential equations. *Journal
1311 of Computational Physics*, 378:686707.
- 1312 Ramsay, J. O. and Silverman, B. W. (2005). *Functional data analysis*.
1313 Springer.
- 1314 RashtBehesht, M., Huber, C., Shukla, K., and Karniadakis, G. E. (2022).
1315 Physicsinformed neural networks (pinns) for wave propagation and full
1316 waveform inversions. *Journal of Geophysical Research: Solid Earth*, 127(5).
- 1317 Ren, P., Rao, C., Chen, S., Wang, J.-X., Sun, H., and Liu, Y. (2024). Seis-
1318 micnet: Physics-informed neural networks for seismic wave modeling in
1319 semi-infinite domain. *Computer Physics Communications*, 295:109010.

- 1320 Rodrigo-Illarri, J., Romero-Hernández, C. P., Salazar-Galán, S., and Rodrigo-
1321 Clavero, M. (2025). Temporal and spatial analysis of the environmental
1322 state of the valencia plain aquifer area using the weighted environmental
1323 index (wei). *Sustainability*, 17(13):5921.
- 1324 Romero, J., Heidrich, W., and Ravasi, M. (2025). Bayesian seismic inversion
1325 with implicit neural representations. *Geophysical Journal International*,
1326 page ggaf249.
- 1327 Scheidt, C., Li, L., and Caers, J., editors (2018). *Quantifying Uncertainty in*
1328 *Subsurface Systems*. John Wiley & Sons, Hoboken, NJ.
- 1329 Secci, D., Godoy, V. A., and Gómez-Hernández, J. (2024). Physics-informed
1330 neural networks for solving transient unconfined groundwater flow. *Com-*
1331 *puters & Geosciences*, 182:105494.
- 1332 Seo, J. Y. and Lee, S.-I. (2023). Probabilistic evaluation of drought propa-
1333 gation using satellite data and deep learning model: from precipitation to
1334 soil moisture and groundwater. *IEEE Journal of Selected Topics in Applied*
1335 *Earth Observations and Remote Sensing*, 16:6048–6061.
- 1336 Shirmard, H., Farahbakhsh, E., Müller, R. D., and Chandra, R. (2022). A
1337 review of machine learning in processing remote sensing data for mineral
1338 exploration. *Remote Sensing of Environment*, 268:112750.
- 1339 Sinding-Larsen, R., Eidsvik, J., Ellefmo, S., and Pyrcz, M. J. (2025). In-
1340 troduction to the special issue: How do big data, ai, and ml chal-
1341 lenge geostatistical and bayesian formalisms? *Mathematical Geosciences*,
1342 57(6):10151020.
- 1343 Song, C. and Wang, Y. (2022). Simulating seismic multifrequency wave-
1344 fields with the fourier feature physics-informed neural network. *Geophys-*
1345 *ical Journal International*, 232(3):15031514.
- 1346 Song, S., Mukerji, T., and Hou, J. (2021). Gansim: Conditional facies sim-
1347 ulation using an improved progressive growing of generative adversarial
1348 networks (gans). *Mathematical Geosciences*, 53(7):1413–1444.
- 1349 Sun, J., Innanen, K., Zhang, T., and Trad, D. (2023). Implicit seismic full
1350 waveform inversion with deep neural representation. *Journal of Geophys-*
1351 *ical Research: Solid Earth*, 128(3):e2022JB025964.

- 1352 Tahmasebi, P., Hezarkhani, A., and Sahimi, M. (2012). Multiple-point geo-
1353 statistical modeling based on the cross-correlation functions. *Computa-*
1354 *tional Geosciences*, 16(3):779–797.
- 1355 Tahmasebi, P., Kamrava, S., Bai, T., and Sahimi, M. (2020). Machine learn-
1356 ing in geo-and environmental sciences: From small to large scale. *Advances*
1357 *in Water Resources*, 142:103619.
- 1358 Tarantola, A. (2005). *Inverse problem theory and methods for model param-*
1359 *eter estimation*. SIAM.
- 1360 Tariq, A. (2023). Quantitative comparison of geostatistical analysis of inter-
1361 polation techniques and semivariogram spatial dependency parameters for
1362 soil atrazine contamination attribute. In *Geoinformatics for Geosciences*,
1363 pages 261–279. Elsevier.
- 1364 Tartakovsky, A. M., Marrero, C. O., Perdikaris, P., Tartakovsky, G. D., and
1365 BarajasSolano, D. (2020). Physicsinformed deep neural networks for learn-
1366 ing parameters and constitutive relationships in subsurface flow problems.
1367 *Water Resources Research*, 56(5).
- 1368 Thurin, J., Brossier, R., and Métivier, L. (2019). Ensemble-based uncertainty
1369 estimation in full waveform inversion. *Geophysical Journal International*,
1370 219(3):1613–1635.
- 1371 Todaro, V., D’Oria, M., Tanda, M. G., and Gómez-Hernández, J. J. (2021).
1372 Ensemble smoother with multiple data assimilation to simultaneously es-
1373 timate the source location and the release history of a contaminant spill
1374 in an aquifer. *Journal of Hydrology*, 598:126215.
- 1375 Van den Boogaart, K. G., Egozcue, J. J., and Pawlowsky-Glahn, V. (2014).
1376 Bayes hilbert spaces. *Australian & New Zealand Journal of Statistics*,
1377 56(2):171–194.
- 1378 Varouchakis, E. A., Corzo, G. A., Karatzas, G. P., and Kotsopoulou, A.
1379 (2018). Spatio-temporal analysis of annual rainfall in crete, greece. *Acta*
1380 *Geophysica*, 66(3):319–328.
- 1381 Wackernagel, H. (2003). *Multivariate geostatistics: an introduction with ap-*
1382 *plications*. Springer Science & Business Media.

- 1383 Walter, F., Gräff, D., Lindner, F., Paitz, P., Köpfl, M., Chmiel, M.,
1384 and Fichtner, A. (2020). Distributed acoustic sensing of microseismic
1385 sources and wave propagation in glaciated terrain. *Nature communica-*
1386 *tions*, 11(1):2436.
- 1387 Wang, N., Chang, H., and Zhang, D. (2021). Deep-learning-based inverse
1388 modeling approaches: A subsurface flow example. *Journal of Geophysical*
1389 *Research: Solid Earth*, 126(2):e2020JB020549.
- 1390 Wang, Q., Parker, P. A., and Lund, R. (2025). Spatial deep convolutional
1391 neural networks. *Spatial Statistics*, 66:100883.
- 1392 Wikle, C. K. and Zammit-Mangion, A. (2023). Statistical deep learning
1393 for spatial and spatiotemporal data. *Annual Review of Statistics and Its*
1394 *Application*, 10(1):247–270.
- 1395 Wu, B., Meng, D., and Zhao, H. (2021). Semi-supervised learning for seis-
1396 mic impedance inversion using generative adversarial networks. *Remote*
1397 *Sensing*, 13(5):909.
- 1398 Yan, X., Lin, J., Ju, Y., Zhang, Q., Zhang, Z., Zhang, L., Yao, J., and
1399 Zhang, K. (2025). A finite-volume based physics-informed Fourier neural
1400 operator network for parametric learning of subsurface flow. *Advances in*
1401 *Water Resources*, 205:105087.
- 1402 Yang, L., Meng, X., and Karniadakis, G. E. (2021). B-pinns: Bayesian
1403 physics-informed neural networks for forward and inverse pde problems
1404 with noisy data. *Journal of Computational Physics*, 425:109913.
- 1405 Yang, Y., Saad, O. M., and Alkhalifah, T. (2024). Deep learning-based 3d
1406 microseismic event direct location using simultaneous surface and borehole
1407 data. *Journal of Geophysical Research: Machine Learning and Computa-*
1408 *tion*, 1(4):e2024JH000365.
- 1409 Yao, L., Dimitrakopoulos, R., and Gamache, M. (2020). High-order sequen-
1410 tial simulation via statistical learning in reproducing kernel hilbert space.
1411 *Mathematical Geosciences*, 52(5):693–723.
- 1412 Yu, M., Huang, Q., and Li, Z. (2024a). Deep learning for spatiotemporal
1413 forecasting in earth system science: a review. *International Journal of*
1414 *Digital Earth*, 17(1):2391952.

- 1415 Yu, P., Zhu, T., Marone, C., Elsworth, D., and Yu, M. (2024b). Dasevent-
1416 net: Ai-based microseismic detection on distributed acoustic sensing data
1417 from the utah forge well 16a (78)-32 hydraulic stimulation. *Journal of*
1418 *Geophysical Research: Solid Earth*, 129(9):e2024JB029102.
- 1419 Zargar, A., Sadiq, R., Naser, B., and Khan, F. I. (2011). A review of drought
1420 indices. *Environmental Reviews*, 19(NA):333–349.
- 1421 Zhan, C., Dai, Z., Jiao, J. J., Soltanian, M. R., Yin, H., and Carroll, K. C.
1422 (2025). Toward artificial general intelligence in hydrogeological modeling
1423 with an integrated latent diffusion framework. *Geophysical Research Let-*
1424 *ters*, 52(3):e2024GL114298.
- 1425 Zhan, W. and Datta, A. (2025). Neural networks for geospatial data. *Journal*
1426 *of the American Statistical Association*, 120(549):535–547.
- 1427 Zhang, C., Song, X., and Azevedo, L. (2021). U-net generative adversar-
1428 ial network for subsurface facies modeling. *Computational Geosciences*,
1429 25(1):553–573.
- 1430 Zhang, X. and Curtis, A. (2021). Bayesian full-waveform inversion with
1431 realistic priors. *Geophysics*, 86(5):A45–A49.
- 1432 Zhang, Z. (2022). A physics-informed deep convolutional neural network for
1433 simulating and predicting transient Darcy flows in heterogeneous reser-
1434 voirs without labeled data. *Journal of Petroleum Science and Engineering*,
1435 211:110179.
- 1436 Zhang, Z., Yan, X., Liu, P., Zhang, K., Han, R., and Wang, S. (2023). A
1437 physics-informed convolutional neural network for the simulation and pre-
1438 diction of two-phase Darcy flows in heterogeneous porous media. *Journal*
1439 *of Computational Physics*, 477:111919.
- 1440 Zhou, H., Gómez-Hernández, J. J., and Li, L. (2014). Inverse methods in
1441 hydrogeology: Evolution and recent trends. *Advances in Water Resources*,
1442 63:22–37.
- 1443 Zunino, A., Mosegaard, K., Lange, K., Melnikova, Y., and Mejer Hansen, T.
1444 (2015). Monte carlo reservoir analysis combining seismic reflection data
1445 and informed priors. *Geophysics*, 80(1):R31–R41.

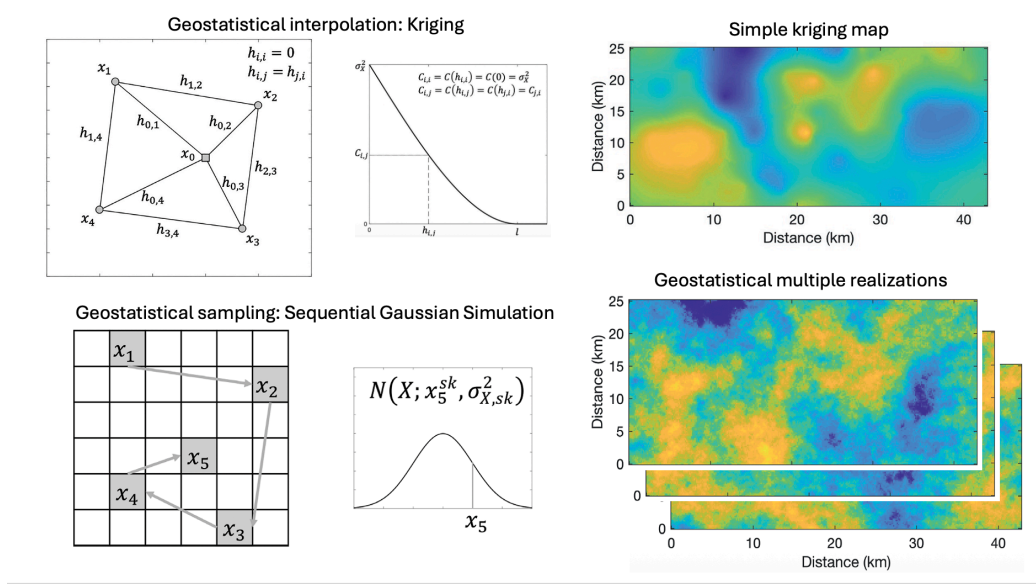
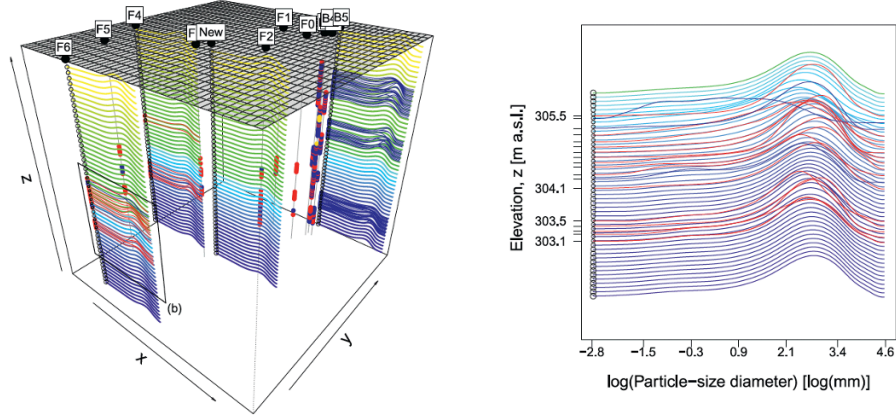


Figure 1: Classical geostatistical methods, such as Kriging and sequential Gaussian simulation, applied to the interpolation and stochastic simulation of elevation maps in Yellowstone National Park conditioned on direct measurements and temperature data. Modified after Grana et al. (2021).

Kriging for object data



Kriging of manifold data

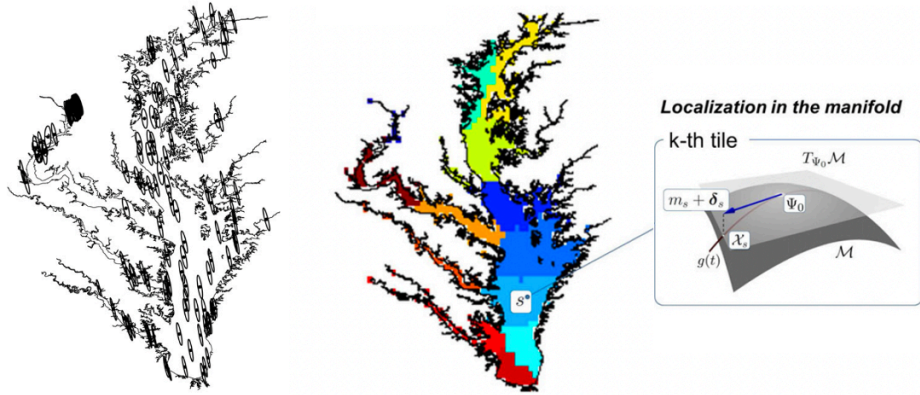


Figure 2: Examples of Kriging for object data and manifold data: (Top) Kriging of particle-size density functions for the characterization of types; and (Bottom) Kriging of manifold data in an estuarine ecosystem for the spatial prediction of hydrologic variables. Modified after Menafoglio (2021).

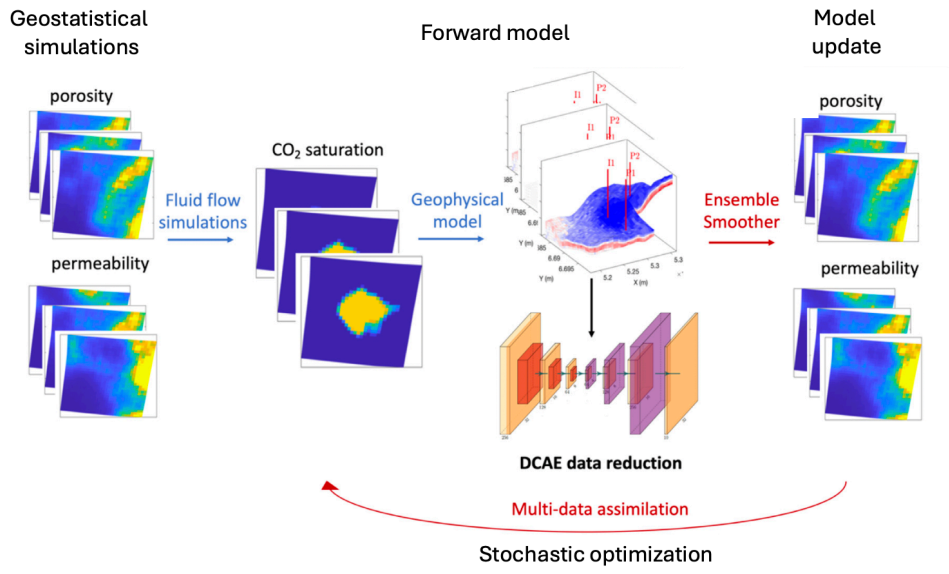
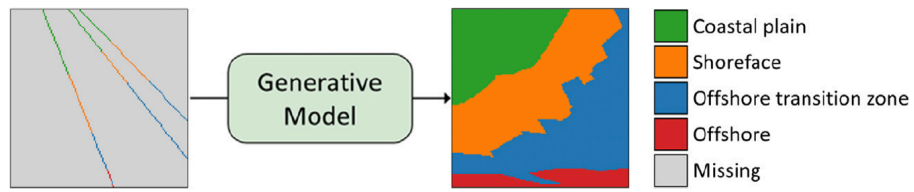
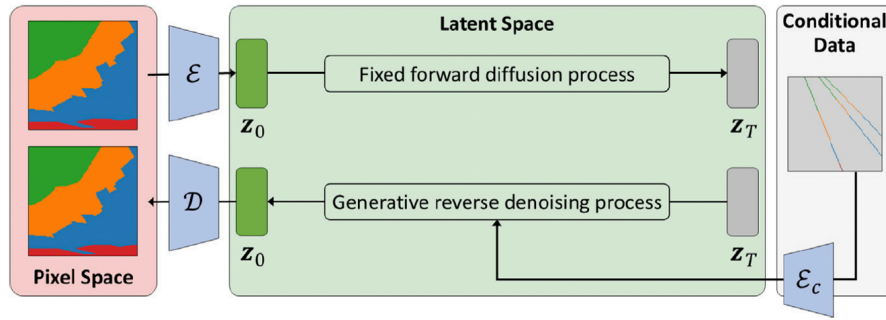


Figure 3: Seismic history matching workflow for carbon sequestration monitoring studies using the ensemble smoother multiple data assimilation (ES-MDA) with the deep convolutional autoencoder. The workflow combines geostatistical realizations, forward seismic modeling, dimensionality reduction, and iterative data assimilation. Modified after Liu and Grana (2020).



Conditional reservoir model generation



Latent Diffusion Model for reservoir modeling

Figure 4: Schematic representation of conditional simulation of reservoir facies models conditioned on sparse observations using latent diffusion model. Modified after Lee et al. (2025).

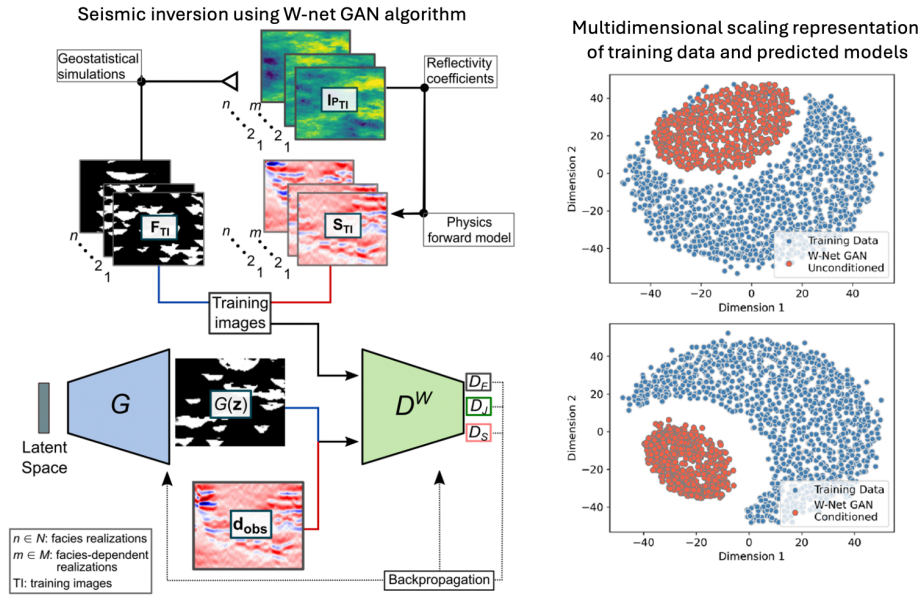


Figure 5: Seismic inversion workflow using W-net generative adversarial network (GAN) combined with multidimensional scaling for model selection. The workflow embeds physical constraints in the GAN to ensure consistency with wave-propagation physics and seismic measurements. Modified after Miele and Azevedo (2024).

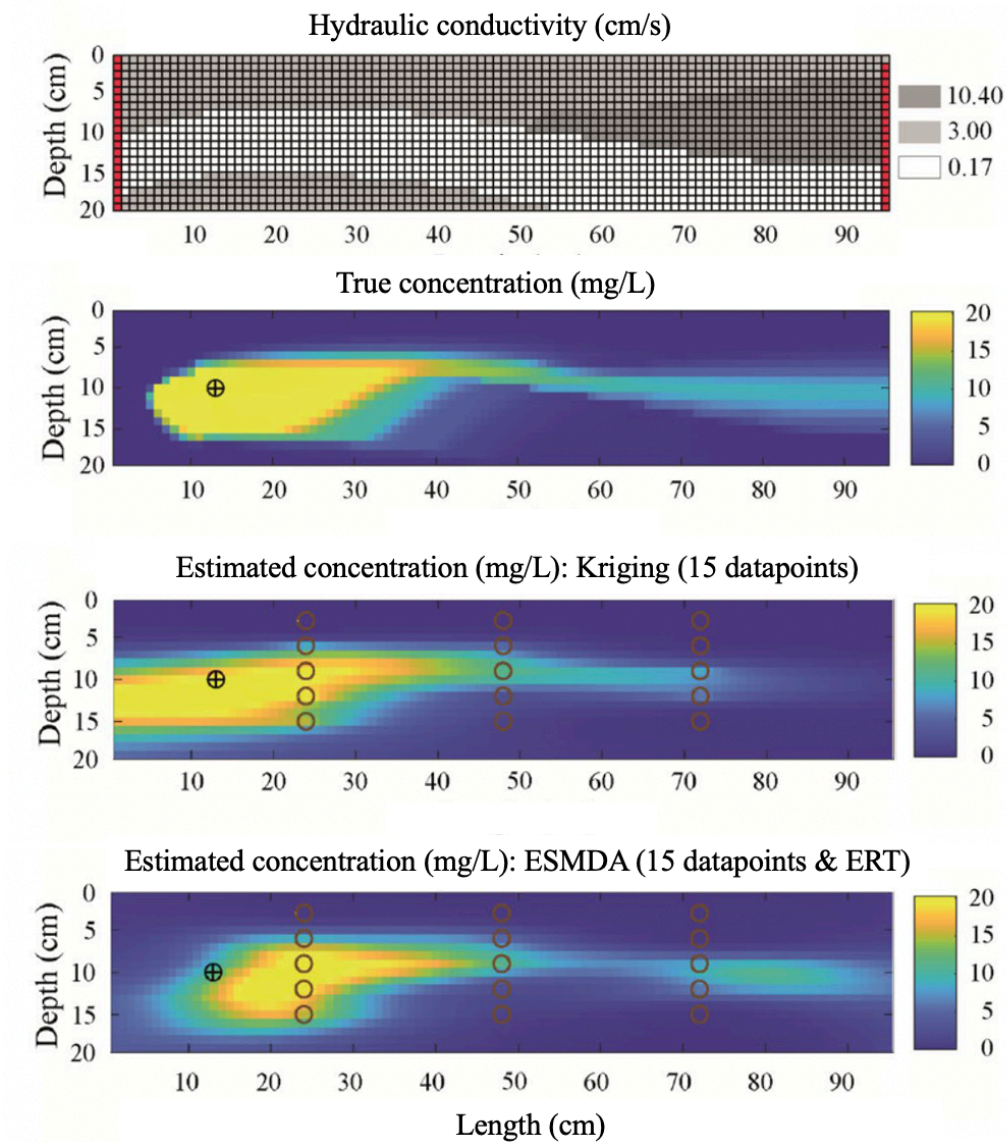


Figure 6: Estimation of contaminant plume in controlled sandbox experiment using Kriging interpolation of concentration data and joint inversion of concentration and electrical resistivity data using ensemble smoother with multiple data assimilation (ESMDA). Modified after Fagandini et al. (2024).

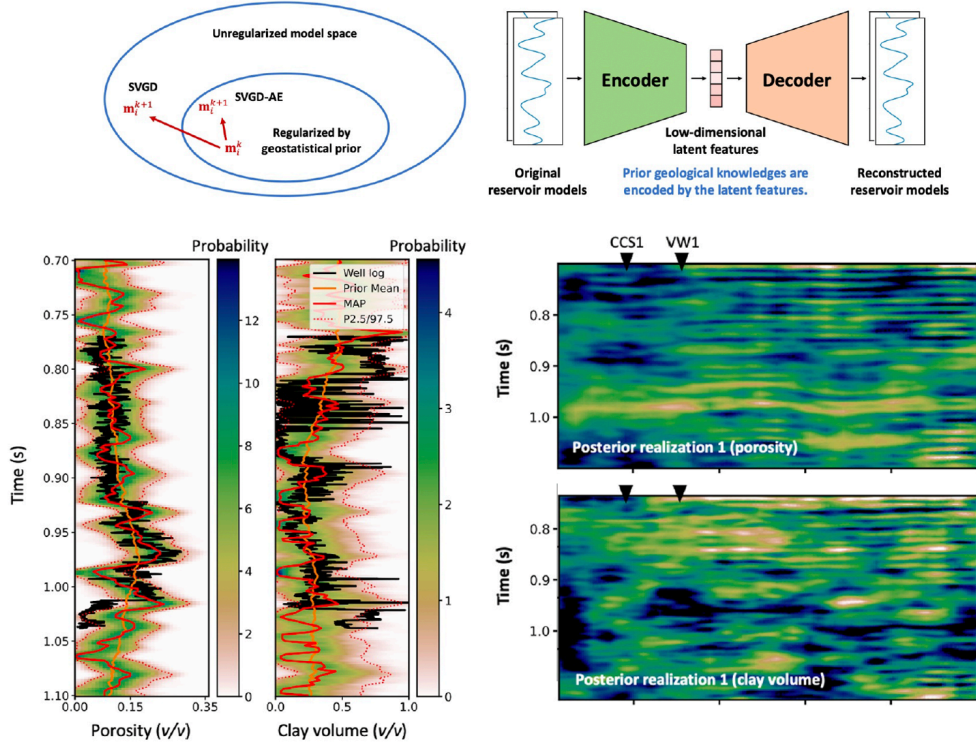


Figure 7: Stochastic inversion of seismic data for the prediction of petroelastic properties: (Top) Schematic overview of stochastic inversion based on the Stein variational gradient descent algorithm within the regularized space and of the auto-encoder for model reparameterization; and (bottom) application to seismic inversion at the well location and along a seismic line. Modified after Liu et al. (2024).

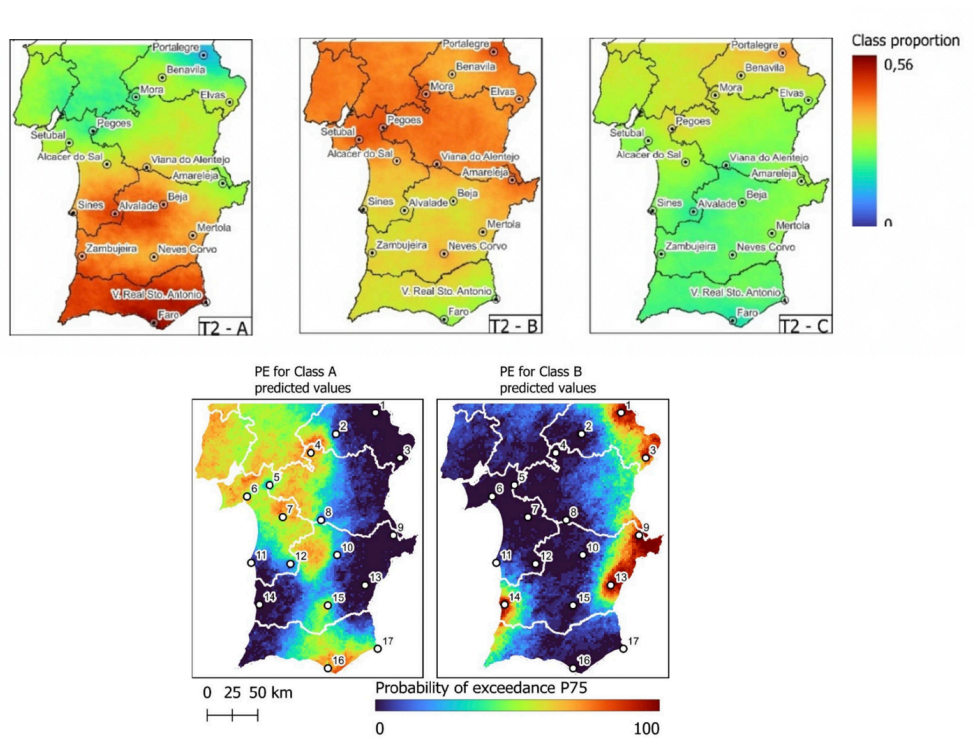


Figure 8: Classification of meteorological station data in Southern Portugal for the classification and forecasting drought severity using consecutive dry days index and random forest method. Modified after Gomes et al. (2025).

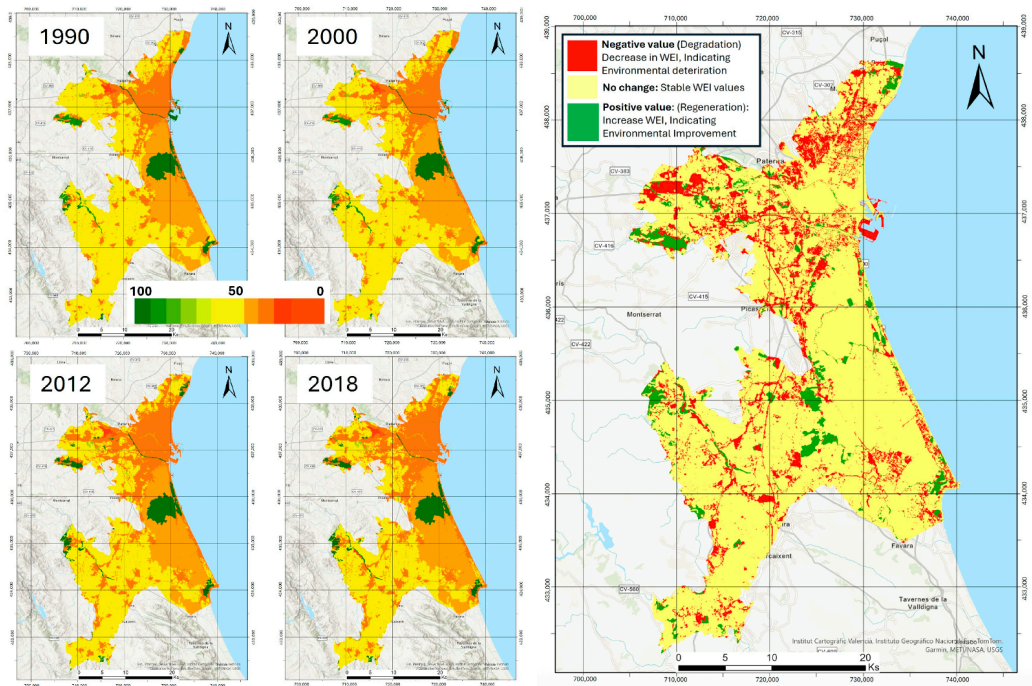


Figure 9: Temporal evolution of land use and environmental quality in the Valencia plain aquifer system, showing the water and environmental index across the years with green areas indicating environmental improvement, yellow stable values, and orange/red areas indicating degradation. Modified after Rodrigo-Illarri et al. (2025).

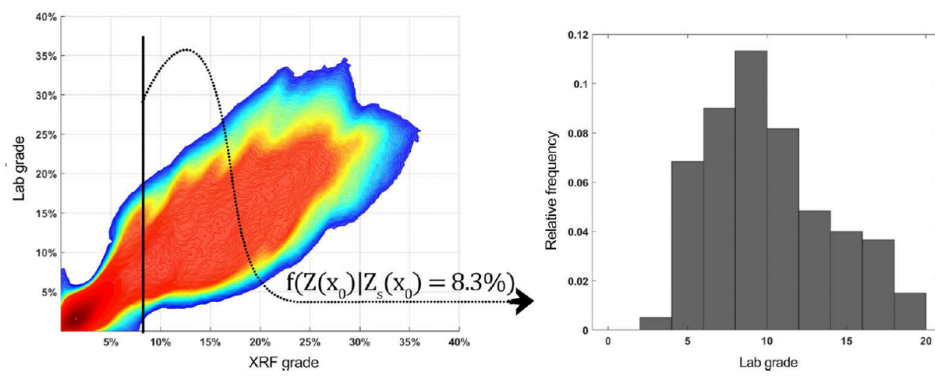


Figure 10: Joint distribution of X-ray fluorescence and lab grades and estimation of ore grade distribution. Modified after Neves et al. (2018).

Table 1: Summary of main geostatistical and ML/AI methods reviewed in this paper: key implicit assumptions, primary data requirements, and recommended conditions of application.

Method	Key implicit assumptions	Data requirements	When to use
Kriging / Variogram-based Sequential Simulation	Stationarity; Gaussian or transform-to-Gaussian; two-point covariance sufficient	Variograms; conditioning hard data	Well-understood geology; uncertainty must be fully transparent and analytically traceable
Multiple-point statistics	Training image representative of target; patterns stationary across domain	Training image; conditioning hard data	Complex non-Gaussian spatial patterns; training image available from analogues or simulations
Functional Kriging	Spatial stationarity of functional objects; appropriate geometry of data space	Dense functional observations (e.g., curves) per location	Spatial prediction of profiles, time series, or distributional objects at each location
Generative Adversarial Networks	Training data representative of target domain; pattern translational stationarity	Large training image or simulation ensemble	High geological complexity; training data abundant; visual/connectivity realism priority
Variational Autoencoders	Latent space smoothness; training data representative	Training ensemble	Low-dimensional parameterization needed for data assimilation or history matching
Diffusion models	Training distribution representative; conditioning network well trained	Large, diverse training ensemble	State-of-the-art geological realism; computational budget permits training
Physics-Informed Neural Networks (PINNs)	Governing physics equations adequate; boundary conditions known; domain homogeneous enough	Sparse labeled data; known physical equations	Data-scarce settings; physical consistency priority; forward model expensive