# Manifold embedding based on geodesic distance for non-stationary spatial estimation in higher-dimensions

Eungyu Park

Department of Geology, Kyungpook National University, 80 Daehak-ro, Daegu 41566, Republic of Korea (ORCID: 0000-0002-2293-4686)

Correspondence to: Eungyu Park (egpark@knu.ac.kr)

## Abstract

This technical note presents a significant advancement in geostatistical approaches for subsurface characterization by extending the manifold embedding method from two-dimensional to three-dimensional space. This development enables the capturing of complex, non-stationary spatial correlations within a higher-dimensional framework. Integrating a fourth dimension, representing varying spatial properties, the method transcends the limitations of conventional two-point geostatistics, particularly in modeling the intricate heterogeneity of subsurface media. Despite its computational intensity and the need for comprehensive secondary data, the manifold embedding approach is invaluable in high-stakes scenarios like geological repository design and contaminant remediation. It stands out as a pivotal tool in hydrogeology and related fields, offering unprecedented precision and detail in subsurface characterizations and modeling.

## 1. Introduction

Accurate spatial estimation is recognized as a foundational element for sustainable management of subsurface resources in hydrogeology. Within the confines of three-dimensional (3D) spaces, the complexity escalates due to the inherent heterogeneity and the non-stationary nature of geological media. It has been observed that traditional geostatistical methods often fail to adequately capture these complex characteristics, which can result in models that do not represent subsurface variability with the required accuracy, thus compromising the prediction of essential properties. In previous studies (Piao and Park, 2023; Park et al., 2024), challenges were addressed from a geological perspective by introducing manifold embedding within a two-dimensional (2D) framework. However, it has been identified that there is a critical need for accurately modeling spatial relationships in higher dimensions. Consequently, a novel geostatistical approach has been developed that surpasses the limitations of existing methodologies. This approach, grounded in the principles of manifold embedding within a four-dimensional (4D) space, has been shown to enhance the precision of spatial estimations and to offer a new perspective in the understanding and management of complex subsurface environments.

The progression to a 3D framework in the current study marks a substantial evolution from prior methodologies. It is widely accepted that traditional geostatistical methods, which are typically not process-based (Koltermann & Gorelick, 1996), strive to effectively represent subsurface property distributions. The adoption of any method that can enrich this representation is deemed beneficial (e.g., Matheron, 1963; Carle & Fogg, 1996; Strebelle, 2002; Park, 2010; Laloy et al., 2018). The extension of manifold embedding to 3D, as presented herein, facilitates a more comprehensive depiction of the subsurface, addressing the limitations observed with 2D constraints. This advancement is pivotal for the accurate modeling of hydrogeological phenomena

and enhances the existing geostatistical framework by accommodating the realistic complexities of subsurface environments (Høyer et al., 2017; Zhao & Illman, 2017).

Despite the advancements in geostatistical methods, traditional Euclidean-based approaches still often struggle to accurately capture the complex spatial behavior of geological properties, especially in extensive spatial domains characterized by pronounced directional changes and heterogeneities (Curriero, 2007; Piao & Park, 2023). While anisotropy has been recognized for its critical role in representing geological processes, such as flow direction in sedimentary environments, existing techniques have typically been constrained to stationary cases (Park et al., 2024). In these scenarios, the variations of anisotropy—pertaining to correlation lengths and orientations—are presupposed to be uniform throughout the domain, neglecting the dynamic and variable nature of geological formations (Cressie, 1993; Chilès, J.P., and Delfiner, 2012). Nonetheless, it has been posited in a few seminal works that the assumption of stationarity need not be a fundamental limitation of two-point statistics. The concept of manifold embedding, derived from differential geometry, emerges as a robust alternative, envisioning spatial properties as functions on a manifold, which inherently accounts for their non-stationarity and anisotropy. This conceptual shift enables a more sophisticated and authentic portrayal of geological phenomena, which is of paramount importance in extensive, 3D geological assessments.

Incorporating non-stationary estimation and simulation within this manifold embedding framework enriches our subsequent analyses, offering a diversity of interpretation previously unattainable with traditional methods. This advancement not only aids in capturing the dynamic and heterogeneous nature of geological processes but also enhances the geostatistical toolkit with a more adaptable and nuanced approach to subsurface modeling.

In this study, the theoretical extension of manifold embedding to 3D spaces is explored, illustrating its capability to more effectively model the intricate spatial variability characteristic of geological media. The limitations inherent in conventional geostatistical models—particularly their inadequacy in capturing directional variability and complex spatial patterns prevalent in geological structures—are addressed through this advancement. The progression of the manifold embedding framework from 2D to 3D equips practitioners with a more robust analytical tool, significantly enhancing the interpretation of the complex interplay between geological features and spatial attributes. Such an approach provides a pragmatic pathway for conducting sophisticated, 3D subsurface analyses. Representing a pioneering development in the field, this extension to 3D effectively bridges a vital gap in spatial estimation methodologies and is posited to yield more accurate, comprehensive, and nuanced geological models, thus contributing substantially to the advancement of hydrogeology.

## 2. Importance of manifold embedding in geostatistical approaches: a 2D review

Spatial estimation within hydrogeology frequently encounters the challenge of capturing the intricacies of the subsurface environment, a task that becomes exponentially more complex when advancing from 2D to 3D representations. The manifold corresponding to a 3D space extends into 4D, at which point, creating a visual representation that offers intuitive understanding to the observer becomes less feasible. This increase in complexity warrants a foundational review in a more conceptually accessible 2D space. Engaging with 2D concepts provides essential clarity and builds the intuition needed to grasp more abstract higher-dimensional spaces. Therefore, this section revisits the principles of manifold embedding within the well-understood confines of 2D, thereby laying the groundwork for the subsequent exploration into the complexities of 3D geostatistical modeling.

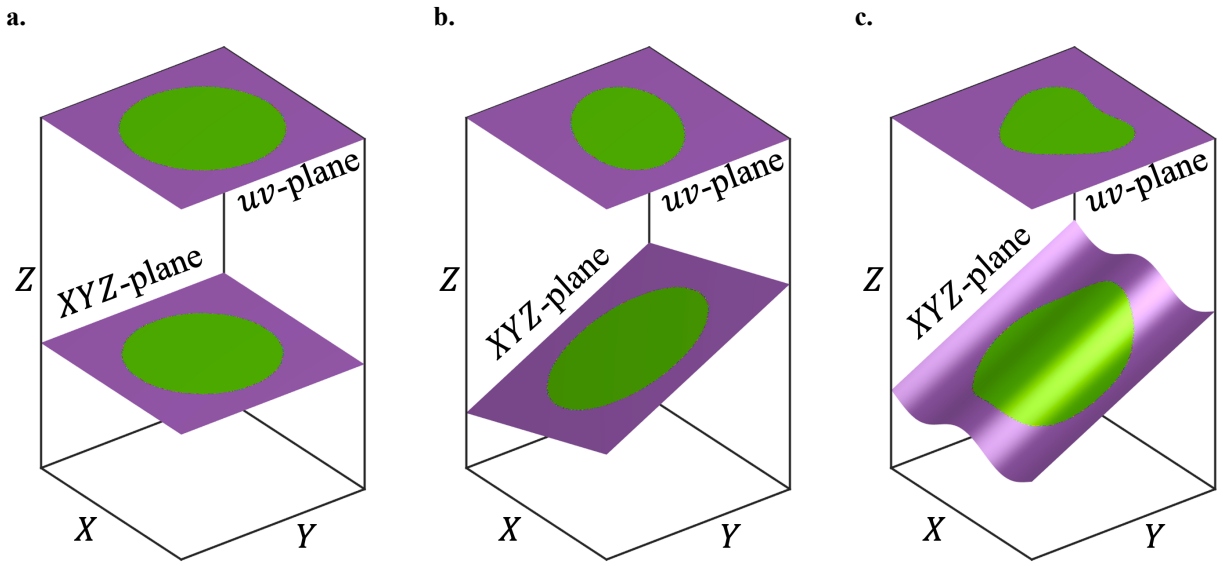



**Figure 1**. Evolution of manifold embedding from (a) isotropic and stationary 2D spatial correlations, through (b) introduction of static anisotropy, to (c) complex, non-stationary manifold with variable spatial correlations.

In Fig. 1, the stepwise increase in manifold complexity is illustrated, marking the transition from the simplicity of Euclidean space to the complexity of a non-Euclidean manifold. The '$uv$-plane' is identified as the actual 2D spatial domain under consideration, while the '$XYZ$-plane' embodies the embedded 3D manifold, defined by the coordinates $X = u$, $Y = v$, and $Z$ as a function of $u$ and $v$, denoted by $Z = f(u, v)$. In Fig. 1a, the $Z$ component is constant, reflecting the assumption of isotropic and stationary spatial correlations prevalent in conventional geostatistical models. Visually, this is represented by a green circle on the $uv$-plane, symbolizing the spatial correlation ellipse. It retains its shape and size when projected onto the embedded $XYZ$ manifold. This represents a scenario of uniform spatial correlations throughout the domain, aligning with the traditional Euclidean geostatistical framework that assumes static spatial relationships.

Moving to Fig. 1b, the function $Z = f(u, v)$ now represents a linear relationship with $u$ and $v$, signaling the emergence of static anisotropy. At this juncture, linear functions are employed to define $Z = f(u, v)$, transitioning the model towards an anisotropic paradigm. Spatial correlations

are now depicted as elliptical shapes, exhibiting both elongation and consistent directionality, which are indicative of the directional dependencies that are characteristic of geological processes. Despite the introduction of anisotropy, these spatial behaviors remain within the modeling capabilities of traditional geostatistical methods. Established variogram models can adequately capture such static anisotropic conditions, thus manifold embedding is not yet a necessity at this stage.

In Fig. 1c, the function $Z = f(u, v)$ represents a departure from simple analytical forms, signifying the presence of spatial non-stationarity. This is visually represented by notable distortions upon projection onto the $uv$-plane, an echo of the complex heterogeneity typical in geological media. Traditional geostatistical models, which rely heavily on affine transformations for rotation and scaling, fall short in capturing this level of complexity. Manifold embedding, on the other hand, provides a sophisticated framework that is adept at characterizing the nuanced spatial variability of subsurface properties. It adeptly models the dynamic nature of spatial correlations, which vary across different locations, thus capturing the true essence of the variable characteristics of geological processes and their spatial impact.

The progression depicted from Figs. 1a to 1c illuminates the limitations of conventional geostatistics when confronted with increasing spatial non-stationarity and manifold complexity. Affine transformations—characterized by rotation and elongation—are inadequate for capturing the intricate variations in the directionality of elongation, which become more pronounced at larger spatial scales as a result of complex geological processes influencing spatial correlations. Manifold embedding stands out as it adeptly models these spatial variabilities and associated geological complexities, which surpass the capabilities of elementary geometric transformations. This advanced method offers a robust framework designed to capture the dynamic behavior inherent in

geological non-stationarity, thereby significantly refining geostatistical analysis in non-stationary environments.

The progression illustrated from Figs. 1a to 1c encapsulates the essential paradigm shift required for advanced hydrogeological studies and geostatistical applications. Within these domains, the intricate phenomena of flow, transport, and migration are governed by the complex and heterogeneous nature of subsurface characteristics. This paradigm shift underscores the critical need for precise delineation and representation of these complex subsurface features. As methodologies evolve to encompass higher-dimensional spaces, such as 3D, the straightforward visual interpretation of embedded manifolds may no longer be practical. Nevertheless, the principle of manifold embedding remains an essential analytical tool. It adeptly captures the full spectrum of spatial correlations, which is crucial for delivering accurate characterizations and predictions within subsurface modeling.

## 3. Theoretical development

### 3.1 Intrinsic geometry and metric tensor formulation

While a manifold is an abstract mathematical construct without physical existence, its conceptual framework is particularly suited for geostatistical estimation. This section builds upon the insights from our exploration of 2D manifold embedding, extending the concept to address the complexities inherent in 3D Euclidean spaces. Conventional 3D Euclidean spaces, though useful, often fall short in capturing the non-stationary spatial behavior characteristic of geological properties. To bridge this gap, manifold embedding introduces a novel approach by constructing a 4D space $(X, Y, Z, \Omega)$ , enhancing the traditional 3D $(u, v, w)$ Euclidean space with an additional

dimension that encapsulates critical spatial variation information, thereby enabling more accurate modeling of subsurface complexities.

For readers primarily interested in the practical applications of these theoretical developments, it may be sufficient to understand the conceptual significance of adding the Ω dimension to better represent spatial variability. Those wishing to delve into the mathematical underpinnings are encouraged to engage with the detailed exposition in this section. However, for those focused on the practical aspects of the developed methodology, it is advisable to proceed to Section 4, which discusses the application of these principles in real-world geostatistical analysis.

In this refined theoretical construct, the coordinates $(u, v, w)$ re preserved to denote conventional 3D spatial locations. Through the incorporation of the Ω dimension, represented by the scalar field $f(u, v, w)$, the capacity of the model to encapsulate spatial property variations within the geological system is enhanced. This enhancement is not explicitly represented in the standard coordinate system. By introducing a fourth dimension, the embedding technique allows for more precise and intricate modeling of subsurface environments, accommodating the complexities and variations characteristic of geological media.

Additionally, in traditional 3D modeling, the natural coordinates $(u, v, w)$ lack inherent directional information about spatial correlation. This limitation is surmounted by the inclusion of $f(u, v, w)$, which integrates essential directional aspects necessary for modeling non-stationary processes. The function $f$ is rendered versatile, representing any spatially variable attribute, whether geological or hydrogeological, and enabling the depiction of non-stationary properties that evolve with location.

In geostatistical applications, manifold embedding has been found to be of particular value where a comprehensive understanding of variables such as hydraulic properties or lithology is critical for subsurface characterization. With the adoption of the 4D framework, it has been observed that spatial models exhibit a marked improvement in predictive capabilities. Such models are now equipped to accommodate directional dependencies and non-stationary behaviors, which are beyond the reach of conventional 3D frameworks.

The principal innovation of this methodology lies in the integration of a fourth dimension, Ω. It is this dimension that is instrumental in encapsulating the intricate spatial variation of properties. Such an advancement facilitates more refined interpretations and accurate modeling, which are indispensable in fields like hydrogeology, where a detailed understanding of the spatial variability of subsurface properties is paramount.

In the development of our 4D manifold embedding approach, we interpret the $XYZ\Omega$ space as a Riemannian manifold (Borovitskiy et al., 2020; Pereira et al., 2022), which is formalized as an inner product space denoted by $\mathcal{V}$. Through a specific mapping function, $\psi$, we relate our manifold $\mathcal{M}$ to $\mathcal{V}$, enabling the computation of geodesic distances essential for modeling non-stationary spatial correlations. The mapping $\psi(u, v, w)$ is defined as follows:

$$\psi(u, v, w) = \big(X(u, v, w), Y(u, v, w), Z(u, v, w), \Omega(u, v, w)\big), \tag{1}$$

where

$$\begin{aligned} X(u, v, w) &= u \\ Y(u, v, w) &= v \\ Z(u, v, w) &= w \\ \Omega(u, v, w) &= f(u, v, w) \end{aligned}, \tag{2}$$

and the function $f$, mapping from the product space $\mathcal{M} \times \mathcal{M}$ to the real numbers $\mathbb{R}$, is a manifold that encapsulates the spatial correlation for a specific location in the space $(u, v, w)$. Here, $f(u, v, w)$ is not merely a coordinate; it is a critical mathematical construct within our Riemannian framework that captures the essence of spatial variability and correlation at each point in our geological model.

Upon establishing the intrinsic geometry of the manifold through the mapping function $\psi$ as defined by Eqs. (1-2), we proceed to construct the metric tensor for space $\mathcal{M}$. This tensor, denoted as $g$, is a fundamental entity that characterizes the geometric structure of the manifold by quantifying the infinitesimal distances between points within the space. It is formally expressed as follows:

$$g = \begin{bmatrix} \frac{d\mathbf{r}}{du} \cdot \frac{d\mathbf{r}}{du} & \frac{d\mathbf{r}}{du} \cdot \frac{d\mathbf{r}}{dv} & \frac{d\mathbf{r}}{du} \cdot \frac{d\mathbf{r}}{dw} \\ \frac{d\mathbf{r}}{dv} \cdot \frac{d\mathbf{r}}{du} & \frac{d\mathbf{r}}{dv} \cdot \frac{d\mathbf{r}}{dv} & \frac{d\mathbf{r}}{dv} \cdot \frac{d\mathbf{r}}{dw} \\ \frac{d\mathbf{r}}{dw} \cdot \frac{d\mathbf{r}}{du} & \frac{d\mathbf{r}}{dw} \cdot \frac{d\mathbf{r}}{dv} & \frac{d\mathbf{r}}{dw} \cdot \frac{d\mathbf{r}}{dw} \end{bmatrix} \tag{3}$$

Eq. (3) encapsulates the inner products of differential position vectors, providing the necessary components to compute distances and angles in manifold $\mathcal{M}$. The metric tensor $g$ is pivotal in capturing the nuances of spatial relationships and is an indispensable tool in the quantification of non-Euclidean distances, thereby playing a critical role in the following geostatistical analyses.

Equation (3) introduces the metric tensor for manifold $\mathcal{M}$, which can be parametrically represented by the vector-valued function $\mathbf{r}$, as shown in Eq. (4):

$$\mathbf{r}(u, v, w) = \big(X(u, v, w), Y(u, v, w), Z(u, v, w), \Omega(u, v, w)\big). \tag{4}$$

Extending from the manifold representation in Eq. (4), the differentials $d\mathbf{r}/du$, $d\mathbf{r}/dv$, and $d\mathbf{r}/dw$ can be articulated as follows, in accordance with Eq. (3):

$$\begin{aligned}\frac{d\mathbf{r}}{du} &= \frac{dX}{du}\mathbf{e}_X + \frac{dY}{du}\mathbf{e}_Y + \frac{dZ}{du}\mathbf{e}_Z + \frac{d\Omega}{du}\mathbf{e}_\Omega = \mathbf{e}_X + \frac{df(u,v,w)}{du}\mathbf{e}_\Omega \\ \frac{d\mathbf{r}}{dv} &= \frac{dX}{dv}\mathbf{e}_X + \frac{dY}{dv}\mathbf{e}_Y + \frac{dZ}{dv}\mathbf{e}_Z + \frac{d\Omega}{dv}\mathbf{e}_\Omega = \mathbf{e}_Y + \frac{df(u,v,w)}{dv}\mathbf{e}_\Omega \\ \frac{d\mathbf{r}}{dw} &= \frac{dX}{dw}\mathbf{e}_X + \frac{dY}{dw}\mathbf{e}_Y + \frac{dZ}{dw}\mathbf{e}_Z + \frac{d\Omega}{dw}\mathbf{e}_\Omega = \mathbf{e}_Z + \frac{df(u,v,w)}{dw}\mathbf{e}_\Omega\end{aligned}, \tag{5}$$

where $\mathbf{e}_X = \partial\mathbf{r}/\partial X$, $\mathbf{e}_Y = \partial\mathbf{r}/\partial Y$, $\mathbf{e}_Z = \partial\mathbf{r}/\partial Z$, and $\mathbf{e}_\Omega = \partial\mathbf{r}/\partial\Omega$ represent the basis vectors of the tangent space at a given point on the manifold. These expressions, derived from the parameterizations established in Equation (2), enable the construction of the metric tensor that encapsulates the intrinsic geometric properties of the manifold.

Building upon the vectorial relationships delineated in Eq. (5), we now direct our attention to the metric coefficients, which serve as the fundamental descriptors of the manifold's intrinsic geometry. By incorporating the expressions from Eq. (5) into the structure of Eq. (3), we arrive at the components of the metric tensor $g$ as:

$$g = \begin{bmatrix} 1 + \left(\frac{df(u,v,w)}{du}\right)^2 & \frac{df(u,v,w)}{du}\frac{df(u,v,w)}{dv} & \frac{df(u,v,w)}{du}\frac{df(u,v,w)}{dw} \\ \frac{df(u,v,w)}{dv}\frac{df(u,v,w)}{du} & 1 + \left(\frac{df(u,v,w)}{dv}\right)^2 & \frac{df(u,v,w)}{dv}\frac{df(u,v,w)}{dw} \\ \frac{df(u,v,w)}{dw}\frac{df(u,v,w)}{du} & \frac{df(u,v,w)}{dw}\frac{df(u,v,w)}{dv} & 1 + \left(\frac{df(u,v,w)}{dw}\right)^2 \end{bmatrix}, \tag{6}$$

For brevity and clarity in subsequent discussions and derivations, we shall refer to the metric tensor in its matrix form as $g$:

$$g = \begin{bmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{bmatrix}, \tag{7}$$

where each element $g_{ij}$ of the metric tensor is the inner product of $d\mathbf{r}/di$ and $d\mathbf{r}/dj$ (for $i, j$ representing the natural coordinates $u, v, w$) and $g$ is a symmetric matrix (i.e., $g_{ij} = g_{ji}$). This tensor reflects the intrinsic curvature of the manifold and includes the derivatives of the function $f$, which encodes the spatial correlation.

### 3.2 Computing geodesic distances in 4D manifold embedding

In the context of manifold theory, geodesic distance is defined as the shortest path between two points on a curved surface, much like 'straight lines' are in Euclidean space. In Euclidean geometry, the shortest path is simply a straight line, representing the most direct distance between two points. In contrast, on a manifold possessing intrinsic curvature, geodesic paths embody the manifold's complex geometry. By applying the metric tensor as established in Eqs. (6) or (7), we can demonstrate that distances within the $uvw$-plane correspond directly to those within the $XYZ\Omega$ manifold. This isometry between the $uvw$-plane and the $XYZ\Omega$ manifold signifies that, despite the curvature present, they retain equivalent geometric properties. Recognizing this isometric relationship is crucial as it allows geostatistical methods to adapt to the dynamic geological processes that shape subsurface media.

The geodesic distance, denoted $d_g$ , represents the shortest path between any two points on a manifold, capturing the essence of the curvature of the space. The computation of this distance for points $(u_0, v_0, w_0)$ and $(u_1, v_1, w_1)$ in curved space is typically achieved by evaluating the integral:

$$d_g = \int \left\| \frac{d\mathbf{r}}{d\lambda} \right\| d\lambda \tag{8}$$

The integrand in Eq. (8) involves the norm of the differential position vector $d\mathbf{r}/d\lambda$, which, when squared, can be expressed in matrix form using the metric tensor $g$ as presented in Eq. (7). This is mathematically represented by:

$$\left\|\frac{d\mathbf{r}}{d\lambda}\right\|^2 = \begin{bmatrix}\frac{du}{d\lambda} & \frac{dv}{d\lambda} & \frac{dw}{d\lambda}\end{bmatrix}\begin{bmatrix}g_{11} & g_{12} & g_{13}\\ g_{21} & g_{22} & g_{23}\\ g_{31} & g_{32} & g_{33}\end{bmatrix}\begin{bmatrix}\frac{du}{d\lambda}\\ \frac{dv}{d\lambda}\\ \frac{dw}{d\lambda}\end{bmatrix}. \tag{9}$$

The trajectory connecting two subsurface points, $(u_0, v_0, w_0)$ and $(u_1, v_1, w_1)$, can be elegantly described using parametric equations. The parameter $\lambda$, ranging from 0 to 1, linearly interpolates the coordinates between these two points as:

$$\begin{aligned} u(\lambda) &= u_0 + (u_1 - u_0)\lambda \\ v(\lambda) &= v_0 + (v_1 - v_0)\lambda \\ w(\lambda) &= w_0 + (w_1 - w_0)\lambda \end{aligned}. \tag{10}$$

From this formulation of Eq. (10), the derivatives with respect to $\lambda$ are simply the differences in the respective coordinates: $du/d\lambda = u_1 - u_0$, $dv/d\lambda = v_1 - v_0$, and $dw/d\lambda = w_1 - w_0$.

Substituting these derivatives into the norm of $\|d\mathbf{r}/d\lambda\|^2$, we obtain:

$$\begin{aligned}\left\|\frac{d\mathbf{r}}{d\lambda}\right\|^2 &= (u_1 - u_0)^2 g_{11} + (v_1 - v_0)^2 g_{22} + (w_1 - w_0)^2 g_{33} \\ &+2[(u_1 - u_0)(v_1 - v_0)g_{12} + (u_1 - u_0)(w_1 - w_0)g_{13} + (v_1 - v_0)(w_1 - w_0)g_{23}]\,.\end{aligned} \tag{11}$$

The arc length, representing the geodesic distance that correlates two data locations, is then determined by integrating the norm as:

$$d_g(u_1, v_1, w_1; u_0, v_0, w_0) = \int_0^1 \left\|\frac{d\mathbf{r}}{d\lambda}\right\| d\lambda\,. \tag{12}$$

The numerical integration of the geodesic distances, as formulated in Eq. (12), is accomplished using the Gauss-Legendre quadrature method. This approach is carefully chosen to complement the intricate geometry of the manifold in question. Notably, the complexity of the manifold—such as high frequency fluctuations—necessitates a higher number of sampling points (abscissae) to ensure precision. Consequently, there exists an inherent balance to be struck between the precision of the results and the computational demands of the process. Optimizing the efficiency of these calculations is essential, as it has profound implications for the scalability and utility of the geostatistical methods being developed.

While the concept of a 4D manifold may be abstract and imperceptible, the practical application of geostatistical analysis in this study is grounded within our 3D reality. The methodologies elucidated in Sections 3.1 and 3.2 simplify the theoretical 4D model into a tangible 3D computational problem. This simplification, which is central to the manifold embedding technique, constitutes the essence of this study. It facilitates the integration of advanced theoretical concepts into actionable geostatistical analyses, effectively navigating the complexities that higher-dimensional models present.

### 3.3 Implementation and efficacy of geodesic kernels in manifold embedded spatial analysis

The concept of geodesic distance inherently captures spatial affinity, which can be adeptly integrated into a multitude of spatial interpolation methods, such as nearest neighbor algorithms, inverse distance weighting, and clustering algorithms, to refine our understanding of spatial relationships. Although geodesic distances are versatile in application, this study has chosen Gaussian process regression (GPR) to align with the commonly preferred methods in the field (i.e.,

simple kriging), utilizing it to demonstrate the implementation and benefits of our manifold embedding approach.

To facilitate our Gaussian process analysis for property distribution, two geodesic kernels of the Gaussian and exponential kernels are employed. In general, the Gaussian kernel indicates a gradual decay of spatial similarity with incremental distance, suggesting a predominant influence of proximal points. In contrast, the Exponential kernel denotes a more precipitous decrease in similarity as the distance increases, accentuating a distinct demarcation between proximate and remote points. Specifically, the Gaussian kernel we adopt is defined as follows (Jayasumana et al., 2015; Piao and Park, 2023):

$$k(\mathbf{x}_1, \mathbf{x}_2) = \exp\left(-\frac{d_g(\mathbf{x}_1; \mathbf{x}_0)^2}{2\rho^2}\right), \tag{13}$$

and the exponential kernel is introduced as:

$$k(\mathbf{x}_1, \mathbf{x}_2) = \exp\left(-\frac{d_g(\mathbf{x}_1; \mathbf{x}_0)}{\rho}\right), \tag{14}$$

where $\rho$ is the parameter that dictates the scale of correlation on the manifold, influencing the spatial reach over which data points are considered to be significantly correlated.

The geodesic distance $d_g(\mathbf{x}_1; \mathbf{x}_0)$, derived from Equations (12), adheres to the essential metric properties: non-negativity ($d_g(\mathbf{x}_1; \mathbf{x}_0) \geq 0$), symmetry ($d_g(\mathbf{x}_1; \mathbf{x}_0) = d_g(\mathbf{x}_0; \mathbf{x}_1)$), and the triangle inequality ($d_g(\mathbf{x}_0; \mathbf{x}_1) + d_g(\mathbf{x}_1; \mathbf{x}_2) \geq d_g(\mathbf{x}_0; \mathbf{x}_2)$) for all points within manifold $\mathcal{M}$.

This theoretical advancement in geostatistical analysis is a practical application of isometric embedding. By extending traditional 3D Euclidean spaces to include a fourth dimension, Ω, we enrich the spatial model to more accurately encapsulate critical variations. This aligns with the

principles of isometric embedding, ensuring spatial relationships are preserved, as per Curriero (2007). Though discussions by Feragen et al. (2015) highlight the careful consideration required in non-Euclidean spaces, our empirical validation within a scaled-down domain supports the positive definiteness of the covariance matrix derived from the exponential kernel.

The validation process provides a preliminary basis for the applicability of a kernel to larger domains. However, it is acknowledged that this extrapolation must be approached with caution. Adequate verification as the domain size increases is essential to ensure that the scalability of the model is empirically supported. This level of scrutiny is critical to maintaining the reliability and computational practicability of geostatistical models, especially when considering the dynamic geological processes in subsurface media.

### 3.4 Practical implications of manifold embedding for geological characterization

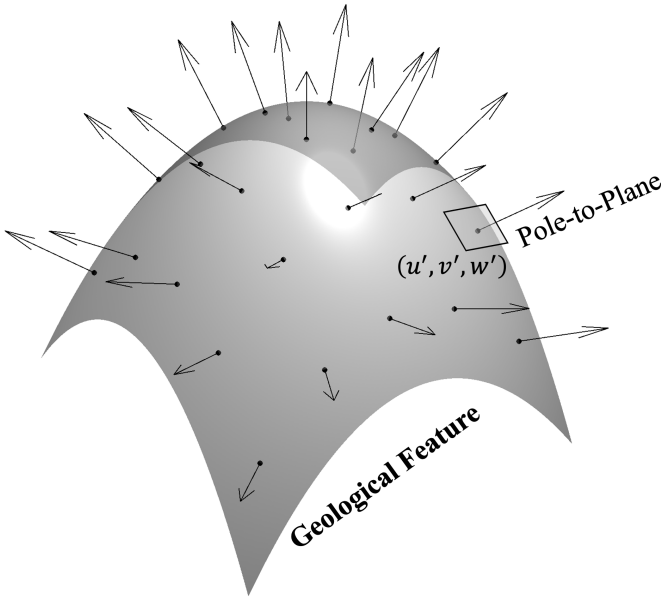
In the established manifold embedding construct, as defined by Eq. (6) and Eq. (12), the gradients $df/du$, $df/dv$, and $df/dw$ maintain a vital connection with the metric tensor $g$. These gradients provide insights into the spatial orientation of geological structures at any specific point $(u', v', w')$. In a spatially variable system guided by two-point statistics, properties demonstrate pronounced correlations in directions orthogonal to the gradient vector $(df/du, df/dv, df/dw)$. This orthogonal characteristic suggests that geological features naturally extend in orientations where the rate of change of a property is minimized, aligning with the established principles of spatial correlation within the traditional geostatistical framework, and introducing the additional complexity of non-stationarity.

The work of Park et al. (2024) provides an in-depth exploration of this relationship, offering a detailed exposition on the inherent connection between the directional orientation of spatial

correlations in geological features and the level-set curves that define the curvature of geological layers at any given horizon. This key concept forms the bedrock of the manifold embedding method, which significantly augments geostatistical analysis by incorporating the non-stationary behavior present within geological data. The present study takes these foundational ideas, originally confined to a 2D setting as presented by Park et al. (2024), and extends them into a 3D framework. This extension allows for the establishment of a direct correlation with the three-dimensional geometric characteristics of geological features.

To bring the theoretical constructs of manifold embedding into alignment with the practical aspects of geological surveying, one may consider the role of pole to plane measurements derived from strike and dip observations routinely recorded during fieldwork. These empirical measurements serve as real-world representations of the gradients discussed earlier. They are oriented perpendicularly to the strike and dip directions, analogous to the gradient vector $(df/du, df/dv, df/dw)$, which signifies the most substantial change in the value of a property. Consequently, the spatial correlation extending in the direction counter to this gradient vector becomes observable in the field as the orientation of the pole to the geological plane. This provides a pragmatic approach to deciphering subsurface structures.

Incorporating the manifold embedding approach as detailed in Eqs. (7) and (12), it is crucial to acknowledge that while the gradients $df/du$, $df/dv$, and $df/dw$, derived from the metric tensor $g$, impart directional insights at a given point $(u', v', w')$ on the manifold. However, these gradients do not inform the magnitude of spatial correlation, a limitation intrinsic to pole to plane measurements derived from strike and dip readings. Such measurements provide orientation data but fall short of quantifying the correlation scale.




340

341 **Figure 2**. Representation of a geological feature with pole to plane vectors indicating the
342 orientation of spatial correlations at point $(u', v', w')$, demonstrating the application of manifold
343 embedding in modeling geological structures.

344 Refer to Fig. 2, which illustrates a geological dome comprising multiple planar features amenable
345 to pole to plane measurement. Data from borehole logs or 3D seismic profiles serve as exemplary
346 3D representations of such features. The pole to plane vector, $\mathbf{p}$, at a measurement location
347 $(u', v', w')$ on a planar geologic feature $L$, is expressed as:

348 $$\mathbf{p}_{u',v',w'} = \left[ -\frac{dL}{du}\Big|_{u',v',w'} \quad -\frac{dL}{dv}\Big|_{u',v',w'} \quad -\frac{dL}{dw}\Big|_{u',v',w'} \right]^{\mathrm{T}}, \tag{15}$$

349 with its magnitude normalized to unity:

350 $$\sqrt{\left(\frac{dL}{du}\Big|_{u',v',w'}\right)^2 + \left(\frac{dL}{dv}\Big|_{u',v',w'}\right)^2 + \left(\frac{dL}{dw}\Big|_{u',v',w'}\right)^2} = 1\,. \tag{16}$$

351 This vector $\mathbf{p}$ encapsulates solely the orientation of the spatial correlation at the location of
352 investigation, necessitating additional methods to ascertain the magnitude of the correlation scale.

353 Such an understanding is imperative for the precise modeling of geological non-stationarity and
354 for the enhanced interpretation of subsurface structures as part of geostatistical analyses.

355 The present state of methodology does not provide the means to directly quantify the magnitude
356 of the correlation scale from pole to plane measurements. The absence of this metric poses
357 significant challenges in the comprehensive characterization of spatial correlations within
358 geological features. Preliminary insights suggest that the thickness of geological formations
359 observed at boreholes might serve as a proxy for the correlation scale, offering a potential avenue
360 for future research. This hypothesis posits that the extent of geological features could be indicative
361 of the magnitude of spatial correlation. The pursuit of methodologies to substantiate the correlation
362 scale from the observed thickness of geological features forms a promising direction for future
363 studies, aimed at bridging the existing gap in geostatistical analysis.

364 To provide a tangible understanding of how pole to plane information translates into estimations
365 of geological properties, consider the following simplifying assumption: the magnitude of the
366 correlation scale is held constant, and a scalar factor $\beta$ is introduced to link the manifold gradient
367 vector $(df/du, df/dv, df/dw)$ with the pole to plane information $(-dL/du, -dL/dv, -dL/$
368 $dw)$:

369 $$\begin{bmatrix} \frac{df}{du} & \frac{df}{dv} & \frac{df}{dw} \end{bmatrix}^{\mathrm{T}} = \beta \begin{bmatrix} \frac{\partial L}{\partial u} & \frac{\partial L}{\partial v} & \frac{\partial L}{\partial w} \end{bmatrix}^{\mathrm{T}} . \quad (17)$$

370

## 4. Results

### 4.1 Incorporating pole-to-plane information in 3D estimation

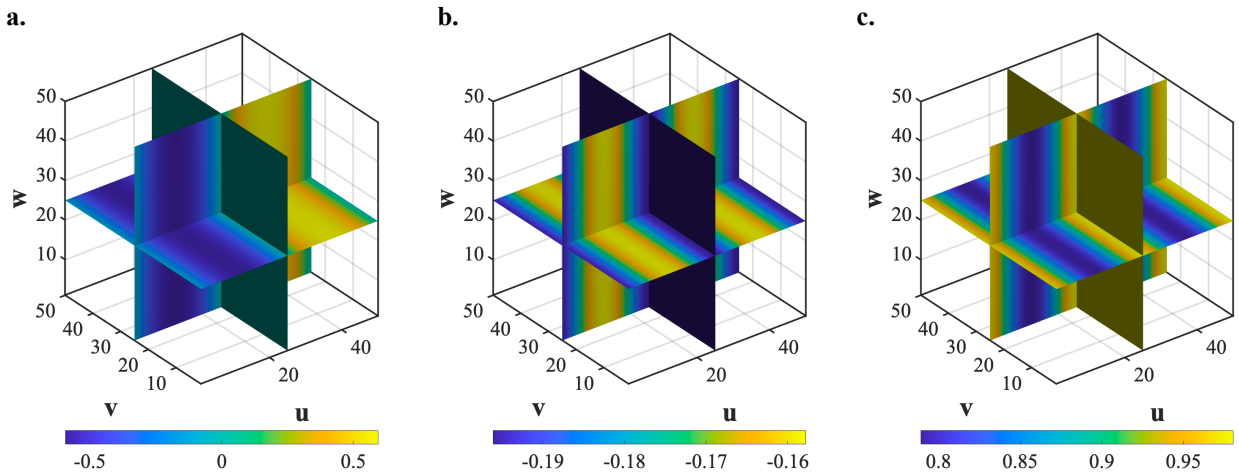



**Figure 3**. Cross-sectional representations of the pole to plane field in a hypothetical 3D domain at sections ($u = 25$, $v = 25$, and $w = 25$), visualizing the spatial derivatives $dL/du$, $dL/dv$, and $dL/dw$, according to the defined functions and constraint of Eq. (16).

In this hypothetical case, the property distributed across the 3D domain is defined as an abstract, unitless measure and estimated. Additionally, the spatial scales presented in this case are grid-based coordinates devoid of physical units. This construct is devised for representing the developed manifold embedding techniques to focus the unique capability of the developed method and bears no direct correlation to any specific, quantifiable physical attributes.

Figures 3a, 3b, and 3c demonstrate the incorporation of the pole-to-plane field to a hypothetical property distribution within a 3D domain. The 3D domain of consideration is consistently scaled between 1 and 50 along the $w$-, $v$-, and $u$-directions. These figures depict slices of $dL/du$, $dL/dv$, and $dL/dw$ at three central cross-sections ($u = 25$, $v = 25$, and $w = 25$), respectively. For the purpose of this hypothetical scenario, we define the following arbitrary functions at any given location $(u, v, w)$ in the domain as the the local pole-to-plane information:

388 $$\frac{dL}{du} = -\frac{6\pi}{C}\sin\left(\frac{2\pi u}{50}\right), \frac{dL}{dv} = -\frac{5}{C}, \text{and } \frac{dL}{dw} = \frac{25}{C},$$

389 where the normalizing constant $C = \sqrt{dL/du^2 + dL/dv^2 + dL/dw^2}$ following Eq. (16).

390 To compose the geodesic kernel adopted in GPR that reflects the pole-to-plane information,
391 Gaussian kernel of Eq. (13) was used. For the interested readers of the GPR estimation theory,
392 refer Piao and Park (2023).

393 For the estimation, a single conditioning data of 4.725 given at the location $(28, 43, 21)$ was used.
394 In addition, the scalar factor $\beta$ of Eq. (17) set to 10, and the correlation scale $\rho$ of Eq. (13) is
395 postulated to be 120. For the numerical integration of Eq. (12), five abscissae were used
396 considering the gently changing pole-to-plane fields (Fig. 3). In the estimation, small
397 regularization factor of $1\times10^{-2}$ was applied for the estimation stability.

398 In the depicted scenario, Fig. 4 offers a visual representation of the regression outcomes, which
399 resembles plunging geological fold often observed in a textbook. However, it is note worthy that
400 such a shape cannot be estimated by conventional methods grounded in two-point statistics, and
401 the unique capability of the developed method is evident. Figure 4a demonstrates a 3D rendering
402 of the property distribution across the domain, with the remaining 1/8th of the volume
403 (corresponding to from 1 to 25 along the $u$- and $v$-directions, and from 26 to 50 along $w$-direction)
404 intentionally cropped off to show the internal estimation result. Figure 4b reveals the lower half of
405 the domain along the $w$-axis, with arrows indicating the pole-to-plane direction at $w = 25$. As
406 shown in the figure, the pole-to-plane arrows are symmetrical with $u = 25$ as the axis. Figures 4c
407 to 4e, conversely, present 2D cross-sectional views at the constant planes of $w = 25$, $v = 25$, and
408 $u = 25$, respectively, which resembles toy examples of plunging fold shown in the textbook.

These visualizations are pivotal for the explication of spatial variability and offer a quantitative portrayal of the theoretical constructs applied within a geological framework. The graphical representations serve as a bridge between the abstract manifold theory developed in the present study and the tangible geostatistical application when spatial non-stationarity exist in the spatial correlation structure, enhancing the interpretability of subsurface characterizations.

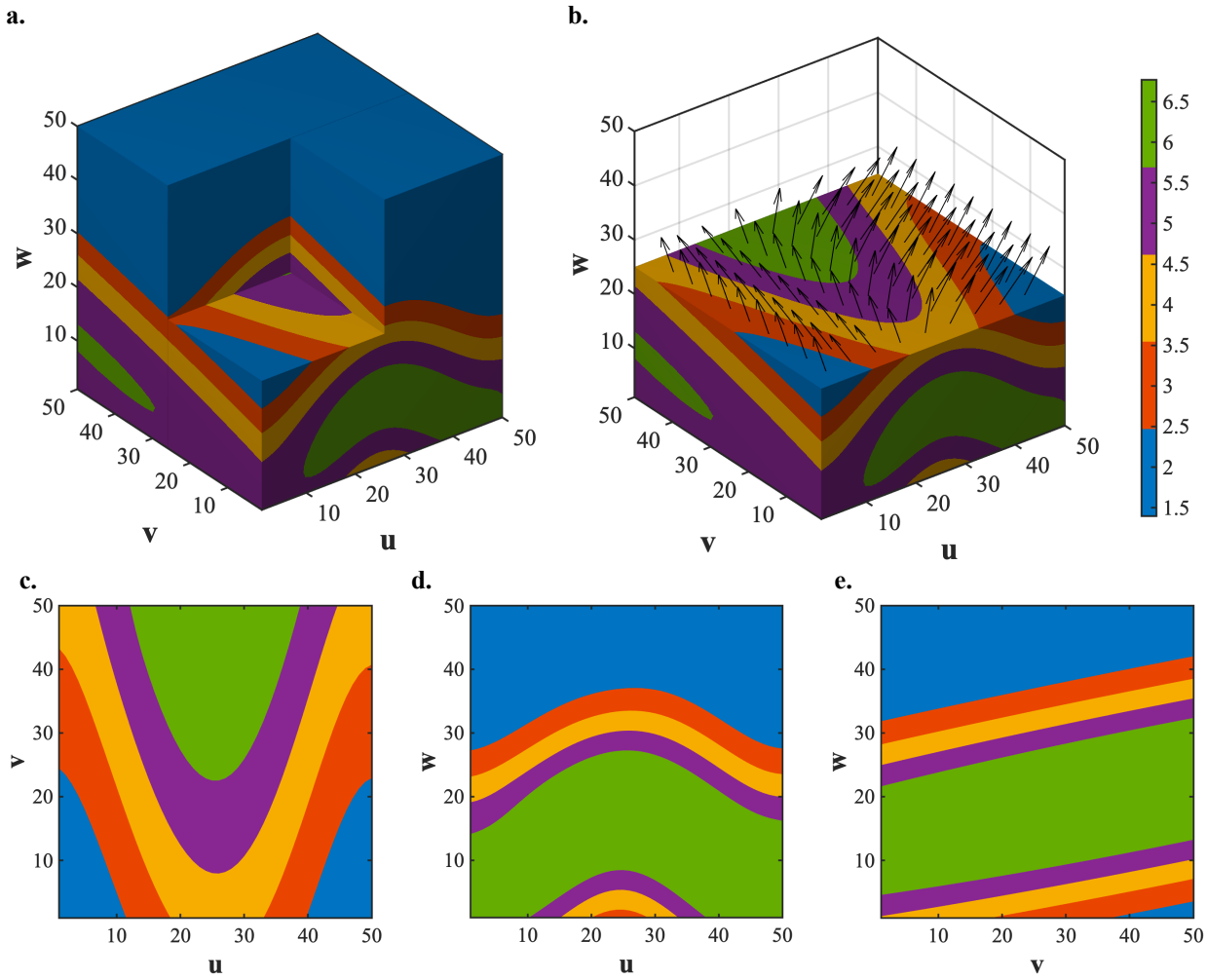



**Figure 4**. Gaussian process regression visualizations of property distribution within a 3D domain, illustrating a plunging geological fold structure: (a) shows the complete property distribution, (b) represents the lower half of the domain along the $w$-axis with pole to plane arrows, and (c-e) display cross-sectional views at constant planes of $w = 25$, $v = 25$, and $u = 25$, respectively.

In practice, as opposed to the idealized scenario previously discussed, pole-to-plane data are often gathered in a discrete manner, originating from limited field surveys or sparse sampling in 3D geophysical data (seismic profiles). This discrete nature necessitates the use of interpolation techniques to construct a continuous spatial representation throughout the 3D domain. However, it is crucial to acknowledge that the integrity of the original orientation data may be compromised if the data acquired from these measurements or samplings are not sufficiently dense. Furthermore, the interpolation process in 3D can be a computationally intensive task, introducing potential constraints on computational efficiency. Such considerations underscore the importance of striking an optimal balance between the resolution of data and the computational resources required in geostatistical modeling endeavors.

### 4.2 Geostatistical simulations and the hydraulic estimates

The developed manifold embedding method, diverging from conventional two-point statistics, adeptly visualizes geological dome shapes as illustrated in Fig. 5. This technique overcomes the computational complexities associated with geodesic distance calculations and transcends the limitations of traditional search algorithms in non-Euclidean spaces. With the impracticality of exhaustive sequential estimations across extensive domains, the algorithm employs GPR within a sequential simulation, adopting a random path to approach unexplored locations and weaving in randomness to capture spatial uncertainty. Upon estimating an initial subset of data, the GPR-based approach continues into a deterministic estimation phase for the remaining locations, mirroring the process presented in Fig. 4. This strategy ensures a robust representation of uncertainty while managing the computational load inherent in large spatial domains.

For representing geological structures (i.e., geological dome), functions were defined for any point within the domain by coordinates $(u, v, w)$. The gradients of these functions are:

$$\frac{dL}{du} = -3.75(u - 50)\exp\left[\frac{(u-50)^2 + (v-50)^2}{2 \cdot 40^2}\right]$$

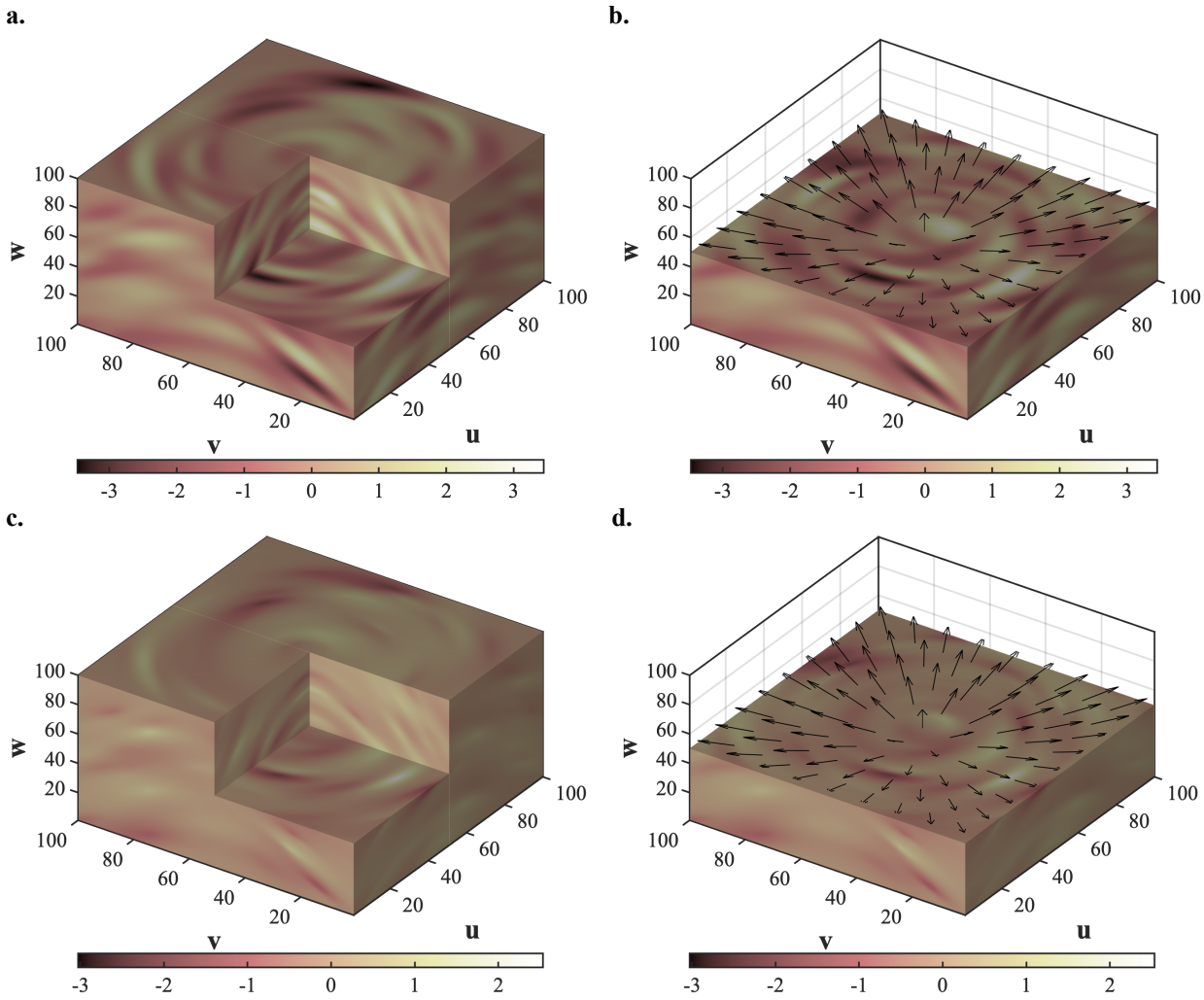
$$\frac{dL}{dv} = -3.75(v - 50)\exp\left[\frac{(u-50)^2 + (v-50)^2}{2 \cdot 40^2}\right], \text{and}$$

$$\frac{dL}{dw} = -\frac{40}{C}$$

Here, $C$ is the normalization factor from the magnitude of these gradients, ensuring consistency with Equation (16). In Figs. 5a and 5b, the manifold gradient vectors, derived from Eq. (17), are combined with GPR using the geodesic Gaussian kernel from Eq. (13) for unconditional simulation. In contrast, Figs. 5c and 5d utilizes the exponential kernel from Eq. (14) in the simulation process. Both cases employ a scalar factor $\beta$ of 4 and a correlation scale $\rho$ of 12 for calculating geodesic distances as per Eq. (12) where regularization factor of $1\times10^{-2}$ was used for the estimation (Piao and Park, 2023). For this development, 50 synthetic data points were generated unconditionally at random locations within the domain, assuming a normal distribution with mean 0 and variance 1. Subsequently, 450 locations were estimated stochastically in a sequential manner, while the remaining points were estimated deterministically.

Figure 5 showcases the 3D distribution of a hypothetical property delineated using Gaussian and exponential kernels, consistently scaled between 1 and 100 along the $w$-, $v$-, and $u$-directions. The figures capture the spatial non-stationarity manifested through variable orientations and elongation patterns across different areas. Figs. 5a and 5c offer a comprehensive view of the domain, while Figs. 5b and 5d zoom into the lower half along the w-axis, highlighting the pole to plane vectors at $w = 50$ with radial symmetry around the center axes at $u = 50$ and $v = 50$. Such directional

460 vectors, often derived from field surveys or 3D geophysical explorations, are interpolated to
461 reconstruct geological features, with their availability directly influencing the precision of the
462 geological characterization. The Gaussian kernel simulation outlines a mean of –0.042 and a
463 standard deviation of 0.779, with the property values spanning from –3.663 to 3.443 and
464 approximating a normal distribution.




465 **Figure 5**. Comparative 3D distribution of a hypothetical property across a domain using Gaussian
466 (a, b) and exponential (c, d) kernels, highlighting the impact of kernel choice on spatial
467 heterogeneity and property distribution. In the figure, (a) and (c) show full domain distributions,
468 while (b) and (d) display lower halves with pole to plane vectors at $w = 50$.

469

Conversely, the results from the exponential kernel, depicted in Figs. 5c and 5d, present the entire domain and a sectional view, respectively. The exponential kernel yields a more condensed distribution, evidenced by a reduced standard deviation of 0.388, indicating a rapid decay of similarity with increased distance. This leads to more subdued structural features across the domain, with less extensive lengths and widths relative to those from the Gaussian simulation. Despite this, the mean property value stands at –0.041, with a range from –3.029 to 3.268, showing a distribution breadth akin to that of the Gaussian kernel.

Assuming the hypothetical property illustrated in Fig. 5 represents log-transformed hydraulic conductivity values of a confined aquifer, we can deduce the groundwater head distribution and particle transport pathlines as demonstrated in Fig. 6. For this simulation, MODFLOW-2005 was employed, using 3D voxels measuring 100×100×10 meters ($\Delta u \times \Delta v \times \Delta w$) along the $u$-, $v$-, and $w$-directions. The transmissivity ($T$) within the domain was calculated by the exponentiation of the property values, expressed as $\mathbf{T} = \exp(2\mathbf{z})$, with $\mathbf{z}$ representing the log-transformed transmissivity (Fig. 5). Consequently, the transmissivity range for the Gaussian kernel scenario spans from 0.0007 to 978.5 $m^2$ $day^{-1}$, while the exponential kernel yields a range from 0.0023 to 689.5 $m^2$ $day^{-1}$. In this simplified flow model, specified head boundaries were set at 1100 m for $u = 1$ and $u = 100$, with no-flow boundaries at $v = 1$ and $v = 100$. Additionally, a fully penetrating extraction well with a constant head of 1050 m was posited at the center of the domain at $u = 50$ and $v = 50$.

Figures 6a and 6c depict the drawdown distributions resulting from a centrally located constant head well, modeled on the simulations using the Gaussian and exponential kernels, respectively. The drawdown contours in these figures are influenced by the permeability variations within the

aquifer, where they either conform to low permeability barriers or follow pathways of higher permeability. Consequently, the drawdown patterns display complex geometries that deviate from the symmetric ellipsoidal shapes typically observed in homogeneous aquifers. The complexity is particularly accentuated in Fig. 6a, which reflects the extensive and varied property structures associated with the Gaussian kernel as observed in Figs. 5a and 5b. Conversely, the drawdown in Fig. 6c, influenced by the exponential kernel, exhibits a more diffuse spread.

Particle path streamlines provide a nuanced representation of subsurface heterogeneity, effectively illustrating how the flow interacts with varying permeability structures. In Figs. 6b and 6d, backward particle path streamlines emanate from the central wellbore, utilizing the specific discharge $\mathbf{q} = \mathbf{T}/\Delta w \times \nabla \mathbf{H}$, where $\mathbf{T}$ is transmissivity distribution, $\Delta w$ is the voxel thickness along the $w$-direction, and $\nabla \mathbf{H}$ represents the gradient of the hydraulic head distribution in Figs. 6a and 6c. In the streamline estimations, 400 particles are randomly distributed within narrow interval of $u = 50$ and $v = 50$ along the whole thichness of the confined aquifer (from 0 to 100 along $w$-axis) so that the particle starts from the central abstraction well into the domain. These streamlines converge upon the wellbore, predominantly influenced by dome-shaped permeability structures shown in Fig. 5. Consequently, they highlight a zone of influence extending from the wellbore to the lower portion of the constant head boundaries, suggesting that the water quality and quantity of the centrally located well are significantly impacted by inputs from the lower regions of the boundaries. The streamline dispersion contrast between Figs. 6a and 6c is notable; streamlines from the Gaussian kernel simulation exhibit a more eccentric pattern with a less dispersed distribution (Fig. 6b). In comparison, those from the exponential kernel display a relatively uniform dispersion (Fig. 6d). These disparities underscore how the inherent spatial characteristics modeled by each kernel can define the flow dynamics within the domain.

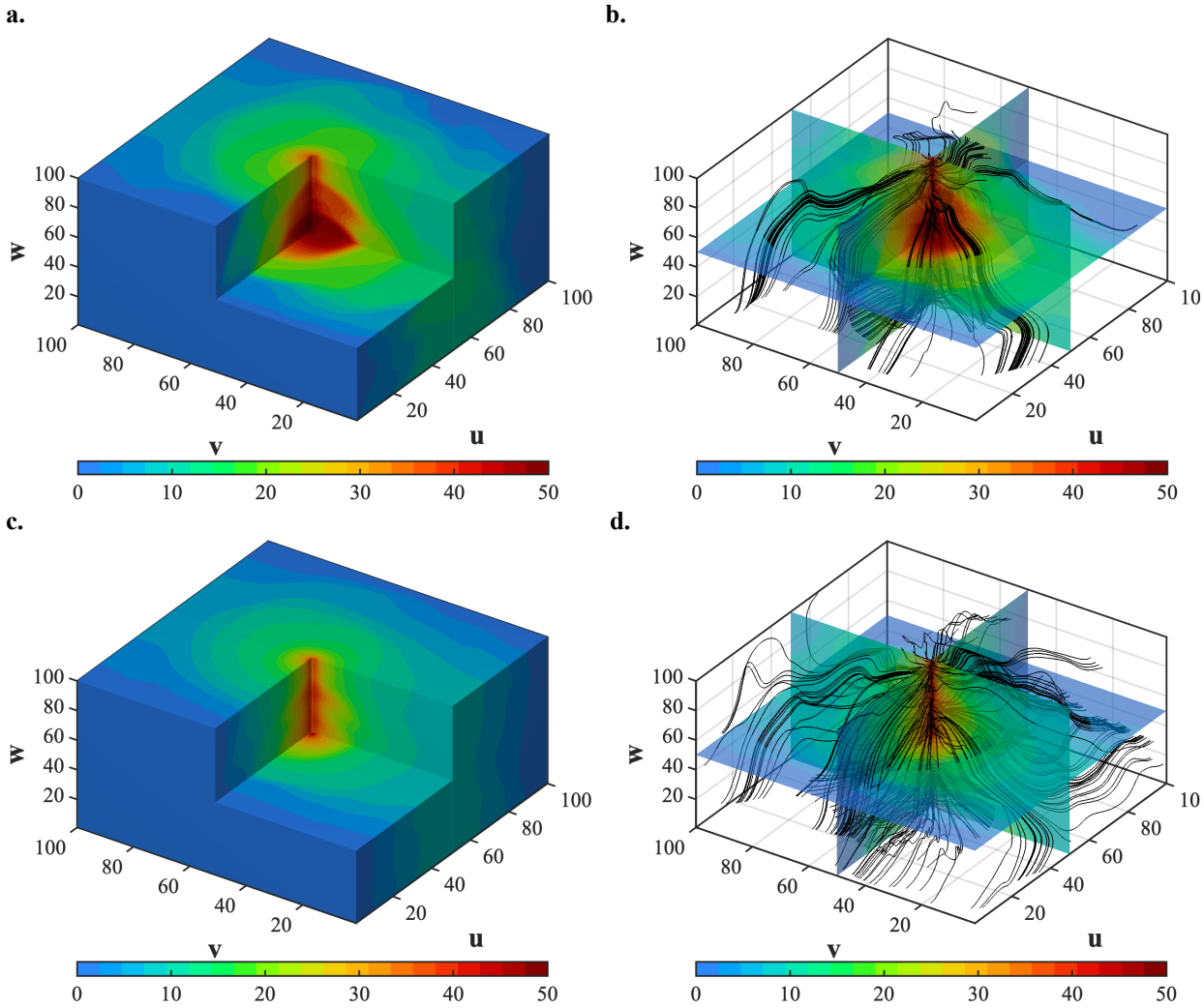



**Figure 6**. Hydraulic head distribution and particle pathlines in a confined aquifer modeled with MODFLOW-2005, illustrating the effects of hydraulic conductivity values derived from Gaussian (a, b) and exponential (c, d) kernels. In the figure, (a, c) show drawdown distributions around a central well, while (b, d) depict pathlines indicating the influence of permeability variations on flow patterns, with 400 particles starting near the wellbore and highlighting zones of influence and dispersion differences between kernels.

From an environmental management standpoint, the intricacies of 3D flow patterns must be carefully considered when delineating a capture zone of a well, ensuring that the true nature of the hydraulic property distribution is incorporated, particularly when determining well locations and pumping rates. Similarly, the design of high-risk geological repositories, such as those for spent nuclear fuel or $CO_2$ sequestration, requires a comprehensive understanding of these complex flow

dynamics to prevent adverse outcomes. In such a situation, the insentive of the developed method is evident and unique among the two-point statistics. Therefore, it is crucial to conduct an extensive assessment that not only contemplates all potential scenarios but also integrates advanced non-stationary analytical techniques like those presented in this study, complementing conventional stationary geostatistical methods.

## 5. Discussions and limitations

In the realm of subsurface characterization, the discussion around manifold embedding within geostatistical models underscores its transformative impact despite certain limitations. While the method demands extensive computational resources due to the intricate calculations of geodesic distances, and thus poses scalability challenges, its precision in modeling spatial variability in subsurface media is unparalleled. As demonstrated by a hydraulic simulation in this study, without incorporation of actual spatial pattern of the hydraulic property distribution often showing non-stationary characteristics, the simulated results and the analysis is mere a rough possibility having weak bonding to actual situation.

One limitation of the method is the reliance on a considerable amount of secondary information, such as in-depth geophysical data and thorough field surveys. This requirement for rich directional data is critical for accurately depicting spatial correlations and the heterogeneity of subsurface structures, representing a significant advancement over conventional geostatistical models.

The adaptability of traditional local search algorithms, which are integral for efficient conditioning in geostatistical models, also falls short in this context. These algorithms, traditionally developed around Euclidean distances, are not yet equipped to handle the complex geometry associated with

manifold embedding, suggesting an open field for future algorithmic development to better align with these advanced methodologies.

However, the potential of the manifold embedding method comes to the fore in high-stakes or high-risk applications. In scenarios such as subsurface contaminant source and plume remediation design, geological repositories for spent fuel, $CO_2$ storage reservoirs, and subsurface exploration for energy or mineral resources, the in-depth and accurate portrayal of local and global subsurface heterogeneity that can be obtained by the proposed method is crucial. For these applications, the computational intensity and requirement for extensive data are secondary considerations compared to the need for precision and detail in characterization.

The proposed manifold embedding method offers a refined level of detail in the representation of subsurface characteristics, which is not commonly achievable with traditional two-point geostatistical approaches. This enhanced representational capacity is especially critical in applications where the spatial complexity of subsurface features has a direct impact on the phenomena of interest. While comparisons to conventional methods are necessary for a holistic understanding, it must be acknowledged that the advanced capabilities of manifold embedding in capturing detailed spatial variability set a new benchmark in the field. This is not to diminish the utility of traditional methods, which continue to offer value in many contexts, but rather to highlight the progression in methodological sophistication afforded by manifold embedding. Such advancements are crucial for advancing the accuracy and reliability of subsurface characterizations in complex geological scenarios.

In summary, while manifold embedding presents challenges in terms of computational demand and data requirements other than the mathematical and computational complexity, its ability to

accurately capture non-stationary and complex spatial correlations makes it an invaluable tool in hydrogeology and geostatistics. Its application is particularly justified in contexts where the risks are high, and the need for precise and detailed subsurface characterization is paramount. The benefits of this method, in terms of the enhanced understanding and management of subsurface environments, significantly outweigh its limitations.

## 6. Summary and Conclusions

This study extends conventional geostatistics into the domain of manifold embedding, enabling 3D subsurface characterization with a level of detail beyond the capabilities of traditional methods. Manifold embedding incorporates spatial variability through an additional dimension, capturing non-stationary properties essential for accurate geological modeling.

Despite the computational intensity and substantial data requirements of the method, the precision of the spatial correlations it provides is vital for high-stakes applications like contaminant remediation and resource exploration. While current search algorithms are limited in addressing the complex geometry of the manifold, this presents an opportunity for developing advanced algorithms attuned to these sophisticated models.

In summary, manifold embedding is a transformative approach in hydrogeology and geostatistics, offering significant advancements in understanding subsurface environments. Its comprehensive representation of spatial heterogeneity justifies the increased computational effort, setting a new standard for detailed and precise subsurface analysis.

## Acknowledgement

The authors declare that they have no conflict of interest. This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MSIT) (NRF-2020R1A2C2013606).

## Open Research

All software programs were written in MATLAB. All the executable software used in this study are available through a public data repository once the manuscript is accepted for publication. All the data used in this study are available through a public data repository once the manuscript is accepted for publication.

## References

Borovitskiy, V., Terenin, A., & Mostowsky, P. (2020). Matérn Gaussian processes on Riemannian manifolds. Advances in Neural Information Processing Systems, 33, 12426-12437.

Carle, S. F., & Fogg, G. E. (1996). Transition probability-based indicator geostatistics. Mathematical geology, 28, 453-476.

Chiles, J. P., & Delfiner, P. (2012). Geostatistics: modeling spatial uncertainty (Vol. 713). John Wiley & Sons.

Cressie, N. (2015). Statistics for spatial data. John Wiley & Sons.

Curriero, F. C. (2006). On the use of non-Euclidean distance measures in geostatistics. Mathematical Geology, 38, 907-926.

Feragen, A., Lauze, F., & Hauberg, S. (2015). Geodesic exponential kernels: When curvature and linearity conflict. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3032-3042).

Høyer, A.-S., Vignoli, G., Hansen, T. M., Vu, L. T., Keefer, D. A., and Jørgensen, F.: Multiple-point statistical simulation for hydrogeological models: 3-D training image development and conditioning strategies, Hydrol. Earth Syst. Sci., 21, 6069–6089, https://doi.org/10.5194/hess-21-6069-2017, 2017.

Jayasumana, S., Hartley, R., Salzmann, M., Li, H., & Harandi, M. (2015). Kernel methods on Riemannian manifolds with Gaussian RBF kernels. IEEE transactions on pattern analysis and machine intelligence, 37(12), 2464-2477.

Koltermann, C. E., & Gorelick, S. M. (1996). Heterogeneity in sedimentary deposits: A review of structure-imitating, process-imitating, and descriptive approaches. Water Resources Research, 32(9), 2617-2658.

Laloy, E., Hérault, R., Jacques, D., & Linde, N. (2018). Training-image based geostatistical inversion using a spatial generative adversarial neural network. Water Resources Research, 54(1), 381-406.

Matheron, G. (1963). Principles of geostatistics. Economic geology, 58(8), 1246-1266.

Pereira, M., Desassis, N., & Allard, D. (2022). Geostatistics for large datasets on Riemannian manifolds: a matrix-free approach. arXiv preprint arXiv:2208.12501.

626 Park, E. (2010). A multidimensional, generalized coupled Markov chain model for surface and
627 subsurface characterization. Water Resources Research, 46(11).

628 Piao, J., & Park, E. (2023). Enhancing estimation accuracy of nonstationary hydrogeological fields
629 via geodesic kernel-based Gaussian process regression. Journal of Hydrology, 626, 130150.

630 Park, E., Piao, J., Jun, H., Kim, Y.-S., Suk, H., & Han, W. S. (2024). Manifold Embedding Based
631 on Geodesic Distance for Non-stationary Subsurface Characterization Using Secondary
632 Information. ESS Open Archive, DOI: 10.22541/essoar.170688766.69001702/v1.

633

634 Zhao, Z., & Illman, W. A. (2017). On the importance of geological data for three-dimensional
635 steady-state hydraulic tomography analysis at a highly heterogeneous aquifer-aquitard
636 system. Journal of Hydrology, 544, 640-657.