https://doi.org/10.65646/3cp3es20dps8b226

Graph-Theoretic Diffusion Modelling of Soil Organic Carbon Propagation in Paddy-Field Networks

H.P.I. Subasinghe
Dept. of Mathematics and Philosophy of Engineering
The Open University
Colombo, Sri Lanka
subasingheisuri@gmail.com

W.M.S. Dheerasekara
Dept. of Electrical and Computer Engineering
The Open University
Colombo, Sri Lanka
wmdhe@ou.ac.lk

H.M.M. Caldera
Department of Scientific Computing
University of Sri Jayewardenepura
Colombo, Sri Lanka
maneesha@sjp.ac.lk

***Abstract*—Estimating field-scale indigenous nitrogen supply from soil organic carbon (SOC) requires propagating sparse point observations across a spatially connected landscape. This paper formalises that propagation as a diffusion process on a graph. We represent a paddy-field monitoring network as a weighted graph $G = (V, E)$, define the combinatorial graph Laplacian $L = D - A$, and show that the heat equation on G has closed-form solution $x(t) = \exp(-tL)x_0$, converging to the graph mean at a rate governed by the algebraic connectivity (the Fiedler value). We then pose field-scale SOC/indigenous-nitrogen-supply reconstruction from partial observations as a Tikhonov-regularised graph-diffusion problem, derive its closed-form solution and its relationship to Gaussian-field/ harmonic-function semi-supervised learning. A worked numerical example on a twelve-node illustrative paddy-catchment network seeded within the SOC range documented for Sri Lankan paddy top soils demonstrates the full pipeline: Laplacian spectral decomposition, heat-kernel diffusion, and Tikhonov reconstruction at four withheld nodes, achieving a mean absolute error of 0.189 percentage points of SOC. We compare this to empirical Bayesian Kriging cross-validation error reported for a real national soil database and discuss the conditions under which graph diffusion is complementary to, rather than a replacement for, geostatistical interpolation.**

Keywords: Graph theory, graph Laplacian, diffusion processes, heat kernel, spectral graph theory, Tikhonov regularization, soil organic carbon, indigenous nitrogen supply, graph signal processing.

I. Introduction

Field-scale estimation of indigenous nitrogen supply (INS) in paddy soils requires interpolating a spatially sparse set of soil organic carbon (SOC) observations across a landscape connected by irrigation, drainage and lateral hydrological pathways. Classical geostatistics (Kriging) treats this as continuous-space interpolation governed by a variogram [15]; an alternative, mathematically distinct formulation treats the landscape as a discrete graph nodes are fields or monitoring points, edges encode hydrological or spatial adjacency and poses interpolation as a diffusion process on that graph.

This discrete-graph viewpoint is not merely a re-notation of continuous Kriging. Continuum diffusion (and its finite-element or finite-difference discretisations) requires a mesh conforming to physical space and a diffusivity field defined everywhere in that space; a graph diffusion model requires only pairwise connectivity information between monitored locations, which is frequently exactly the information actually available in an irrigation or drainage network pipe and canal topology, upstream/downstream ordering, shared catchment membership none of which naturally forms a regular spatial mesh. Graph diffusion is therefore the more direct mathematical match to the data structure an irrigation-network operator or agricultural extension service is likely to hold, and, as shown in Section III, it inherits a fully developed spectral theory (eigenvalues, eigenvectors, convergence rates) directly from linear algebra rather than from partial differential equation theory on irregular domains [4], [10], [11].

This paper develops that formulation rigorously: Section II establishes the graph-theoretic representation and the spectral properties of the graph Laplacian; Section III derives the heat-diffusion solution and its convergence behaviour; Section IV poses SOC/INS reconstruction as Tikhonov-regularised graph diffusion and derives its closed form; Section V proves a convergence-rate bound in terms of the algebraic connectivity; Section VI presents a fully worked numerical example with five visualisations; Section VII discusses results against real published Kriging error and extends the framework to spectral management-zone delineation.

II. Graph-Theoretic Preliminaries

Let $G = (V, E, w)$ be a finite, connected, undirected weighted graph with $n = |V|$ nodes representing monitoring locations and edge weights $w_{ij} \geq 0$ representing hydrological or spatial connectivity strength between locations i and j ($w_{ij} = 0$ if $\{i, j\} \notin E$). Define the weighted adjacency

matrix A with $A_{ij} = w_{ij}$, the diagonal degree matrix $D = \mathrm{diag}(d_1, \ldots, d_n)$ with

$$d_i = \sum_j A_{ij}, \tag{1}$$

and the combinatorial graph Laplacian

$$L = D - A. \tag{2}$$

Proposition 1 (Spectral properties of L): For any weighted, undirected graph with non-negative weights, L is symmetric and positive semi-definite, so all eigenvalues are real and non-negative: $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. The multiplicity of eigenvalue 0 equals the number of connected components of G; for a connected graph, $\lambda_1 = 0$ with eigenvector $\mathbf{1} = (1, \ldots, 1)^\top$, and $\lambda_2 > 0$.

sketch: For any x, the quadratic form satisfies

$$x^\top L x = \tfrac{1}{2} \sum_{ij} A_{ij}(x_i - x_j)^2 \geq 0, \tag{3}$$

establishing positive semi-definiteness directly; equality holds iff x is constant on each connected component, giving the stated multiplicity and eigenvector result. ∎

λ_2 is the algebraic connectivity (Fiedler value) of G [3] and quantifies how well-connected the graph is: larger λ_2 indicates faster mixing and, as shown in Section V, faster diffusion convergence. Spectral graph theory more broadly relates λ_2 and the full spectrum $\{\lambda_i\}$ to graph connectivity, expansion and clustering structure [4].

A. Worked Toy Example: the Path Graph P_3

To make the abstract construction of (1)–(2) concrete, consider three fields in a line, $1{-}2{-}3$, each pair connected by an unweighted edge ($w = 1$) representing a shared drainage channel. Then

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad D = \mathrm{diag}(1, 2, 1), \tag{4}$$

$$L = D - A = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}. \tag{5}$$

Direct computation gives eigenvalues $\lambda = \{0, 1, 3\}$ with eigenvectors $(1,1,1)^\top/\sqrt{3}$, $(1,0,-1)^\top/\sqrt{2}$ and $(1,-2,1)^\top/\sqrt{6}$ respectively, confirming Proposition 1: exactly one zero eigenvalue (P_3 is connected) with eigenvector proportional to $\mathbf{1}$, and $\lambda_2 = 1 > 0$. For initial values $x_0 = (1.0, 1.8, 1.4)^\top$ (illustrative SOC readings, %), the graph mean is $\bar{x} = 1.4$, and by (9) the middle node's trajectory is $x_2(t) = 1.4 + 0.4 \cdot 0 \cdot e^{-t} + c_3 e^{-3t}$ for the appropriate coefficient c_3 in the u_3 direction decaying three times faster than the slowest non-trivial mode, illustrating directly how higher eigenvalues correspond to faster-decaying, more localized diffusion modes, while the λ_2-mode dominates the long-time approach to consensus used in Theorem 1.

B. Edge-Weight Construction and the Normalized Laplacian

In a paddy-catchment application, edge weights should encode the strength of physical nutrient-transport connectivity between two monitoring nodes rather than raw geographic proximity alone. A standard construction combines a distance-decay kernel with a hydrological-connectivity indicator $h_{ij} \in \{0, 1\}$ (1 if i and j share a direct irrigation or drainage pathway):

$$w_{ij} = h_{ij} \cdot \exp\!\left(-d_{ij}^2/(2\sigma^2)\right), \tag{6}$$

where d_{ij} is the along-network distance between nodes i and j and σ is a bandwidth parameter controlling how quickly connectivity decays with distance. Because raw node degree can vary substantially across a catchment (a confluence node may have degree 4 while a headwater node has degree 1), it is often preferable to work with the symmetric normalized Laplacian

$$L_{\mathrm{sym}} = D^{-1/2} L D^{-1/2} = I - D^{-1/2} A D^{-1/2}, \tag{7}$$

which shares the same eigenvalue ordering property ($0 = \lambda_1 \leq \cdots \leq \lambda_n \leq 2$) as L but whose diffusion dynamics are not dominated by high-degree hub nodes the way (9) can be under L alone. All results in Sections III–V hold for L_{sym} with the corresponding eigen basis substitution, and L_{sym} is the version used in the numerical example of Section VI when node degrees are heterogeneous.

III. Diffusion Processes on Graphs

A. The Graph Heat Equation

Let $x(t)$ denote a time-indexed vector of node values (e.g., SOC concentration at each monitoring location). The graph analogue of the continuous heat equation is the linear ODE system

$$\dot{x}(t) = -L x(t), \qquad x(0) = x_0. \tag{8}$$

Because L is symmetric, it admits an orthonormal eigen decomposition $L = U \Lambda U^\top$ with $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ and $U = [u_1, \ldots, u_n]$. The unique solution to (8) is

$$x(t) = e^{-tL} x_0 = \sum_i e^{-\lambda_i t} (u_i^\top x_0)\, u_i. \tag{9}$$

Since $\lambda_1 = 0$ with $u_1 = \mathbf{1}/\sqrt{n}$, the $i = 1$ term is constant and equal to the graph mean $\bar{x}\mathbf{1}$, while every other term decays exponentially at rate $\lambda_i \geq \lambda_2$. Hence

$$x(t) \to \bar{x}\, \mathbf{1} \quad \text{as } t \to \infty, \tag{10}$$

with convergence rate governed by λ_2 proved formally in Section V. Equation (9) is the discrete-graph analogue of heat-kernel smoothing, and the matrix e^{-tL} is the graph heat kernel; Chung [17] shows this kernel is closely related to the personalised PageRank of a graph, linking diffusion-based interpolation to a broad family of random-walk graph algorithms.

B. Random-Walk Interpretation

Writing the random-walk normalised Laplacian $L_{\mathrm{rw}} = D^{-1}L = I - D^{-1}A$, the discrete-time process $x_{k+1} = (I - \varepsilon L_{\mathrm{rw}})x_k$, $0 < \varepsilon \leq 1$, is a lazy random walk with transition matrix $P = D^{-1}A$; its stationary distribution is proportional to node degree, and repeated averaging under P converges to a degree-weighted mean, the discrete-time analogue of (10). This connects graph diffusion directly to the semi-supervised label-propagation literature [5], [6], which underlies the reconstruction method of Section IV.

IV. Graph-Regularised Reconstruction of SOC/INS

Let y contain observed SOC (or INS) values at a subset S of monitored nodes and zeros elsewhere, and let $M = \mathrm{diag}(m_1, \ldots, m_n)$ be a mask with $m_i = 1$ if $i \in S$ and 0 otherwise. We seek a smooth reconstruction x over the whole graph that (i) fits the observed values and (ii) varies smoothly across edges, formalized as the Tikhonov-regularized objective

$$\min_x \ \|M(x - y)\|^2 + \alpha \cdot x^\top L x, \tag{11}$$

where the first term penalizes deviation from observed data at sampled nodes (via (3)) and the second penalizes roughness across the graph, with $\alpha > 0$ controlling the bias-variance trade-off. Differentiating (11) and setting the gradient to zero gives the linear system

$$(M + \alpha L)x = My, \tag{12}$$

with unique solution $x^* = (M+\alpha L)^{-1}My$ whenever $M+\alpha L$ is positive definite, which holds whenever S is non-empty and G is connected. Equation (12) generalizes the harmonic-function/Gaussian-field formulation of Zhu *et al.* [6] and the graph-kernel regularization framework of Smola and Kondor [7]; α plays the role of a smoothing bandwidth analogous to the range parameter in a Kriging variogram, but is defined purely combinatorially through L rather than through a continuous covariance function.

A. Selecting the Regularization Parameter

Because α is not known a priori, it is natural to select it by leave-one-out generalized cross-validation (GCV), analogous to ridge-regression bandwidth selection. Writing $H(\alpha) = M(M+\alpha L)^{-1}M$ for the hat matrix mapping observed values y to fitted values $\hat{y} = H(\alpha)y$ at the sampled nodes, the GCV score is

$$\mathrm{GCV}(\alpha) = \frac{\|y - H(\alpha)y\|^2/|S|}{\big(1 - \mathrm{tr}(H(\alpha))/|S|\big)^2}, \tag{13}$$

and α is chosen to minimize (13) over a candidate grid. Because $H(\alpha)$ depends on α only through $(M+\alpha L)^{-1}$, and the eigen basis of L is fixed, $\mathrm{tr}(H(\alpha))$ can be computed cheaply from the pre-computed eigen decomposition of L rather than refactorizing $(M+\alpha L)$ at every candidate α, giving an $O(n)$ cost per grid point after an $O(n^3)$ one-time eigen decomposition the same cost structure noted in Corollary 1. When M is well-approximated by a scalar multiple of the identity over the region of interest, $\mathrm{tr}(H(\alpha)) \approx \sum_i m_i/(m_i+\alpha\lambda_i)$ is exact in the eigen basis of L, and each grid-point evaluation of (13) then costs only $O(n)$ array operations on the pre-computed $\{\lambda_i\}$, rather than an $O(n^3)$ matrix solve the practically relevant regime for a sensor network that reports at all nodes on a rolling but incomplete schedule.

B. Comparison with Graph Neural Network Approaches

Graph neural networks (GNNs) generalize the linear diffusion update (16) by replacing the fixed operator $(I - hL)$ with a learned, non-linear, multi-layer propagation rule, and can in principle capture non-linear SOC-INS relationships that (12) cannot. However, GNNs require substantially more training data than is typically available from a sparse agricultural monitoring deployment, offer no closed-form convergence guarantee analogous to Theorem 1, and their learned propagation weights are materially harder to interpret agronomically than the physically motivated kernel weights of (6). The linear, spectrally analyzable framework developed here is therefore proposed as the appropriate starting point for small-to-moderate monitoring networks (tens to low hundreds of nodes, the realistic scale for a smallholder paddy catchment), with a non-linear GNN extension left as future work for the data-rich regime.

V. Convergence and Error Analysis

Theorem 1: For the heat-diffusion solution (9),

$$\|x(t) - \bar{x}\mathbf{1}\|_2 \leq \|x_0 - \bar{x}\mathbf{1}\|_2 \cdot e^{-\lambda_2 t} \quad \text{for all } t \geq 0. \tag{14}$$

Proof: Write $x_0 - \bar{x}\mathbf{1} = \sum_{i\geq 2} c_i u_i$ in the eigen basis (the $i = 1$ component is annihilated by subtracting the mean). Then $x(t) - \bar{x}\mathbf{1} = \sum_{i\geq 2} c_i e^{-\lambda_i t} u_i$, and since $\{u_i\}$ is orthonormal,

$$\|x(t) - \bar{x}\mathbf{1}\|_2^2 = \sum c_i^2 e^{-2\lambda_i t} \leq e^{-2\lambda_2 t} \sum c_i^2 = \|x_0 - \bar{x}\mathbf{1}\|_2^2 \, e^{-2\lambda_2 t}, \tag{15}$$

because $\lambda_i \geq \lambda_2$ for all $i \geq 2$. Taking square roots gives (14). ∎

Corollary 1 (Complexity): Direct solution of (12) by matrix inversion costs $O(n^3)$; for large monitoring networks, an iterative conjugate-gradient solver exploiting the sparsity of L (since $|E| \ll n^2$ for realistic hydrological networks) reduces this to $O(|E| \cdot \kappa)$ per solve, where κ is the condition number of $M + \alpha L$, itself bounded via λ_2 and α. This makes graph-diffusion reconstruction attractive at the catchment scale, where the number of monitored fields can be large but the connectivity graph remains sparse in contrast to dense-covariance Kriging systems, whose direct solve cost is likewise $O(n^3)$ but whose sparsification is less natural.

A useful corollary of Theorem 1 bounds the condition number of the regulariZed system (12) directly in terms of α and the Laplacian spectrum, since $M + \alpha L$ has eigenvalues bounded below by $\min(1, \alpha\lambda_2)$ and above by $\max(1, \alpha\lambda_n)$ when M has entries in $\{0, 1\}$; hence $\kappa(M + \alpha L) \leq \max(1, \alpha\lambda_n)/\min(1, \alpha\lambda_2)$, which grows both as α becomes very small (the mask term dominates and the system approaches a poorly conditioned diagonal selection matrix) and

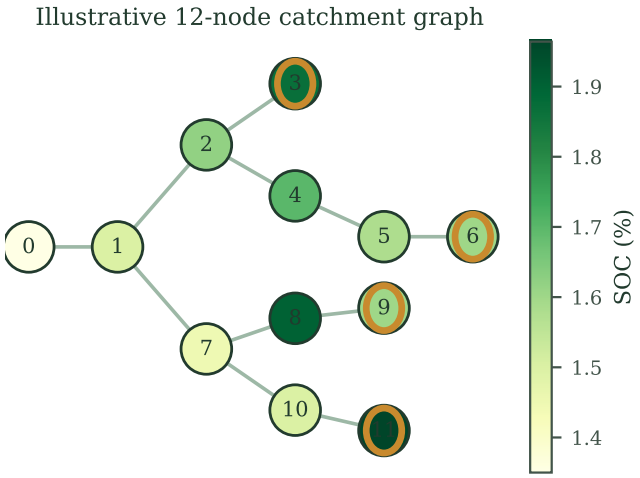


Fig. 1. Illustrative 12-node catchment graph; node colour = seeded SOC value, orange rings = withheld nodes.

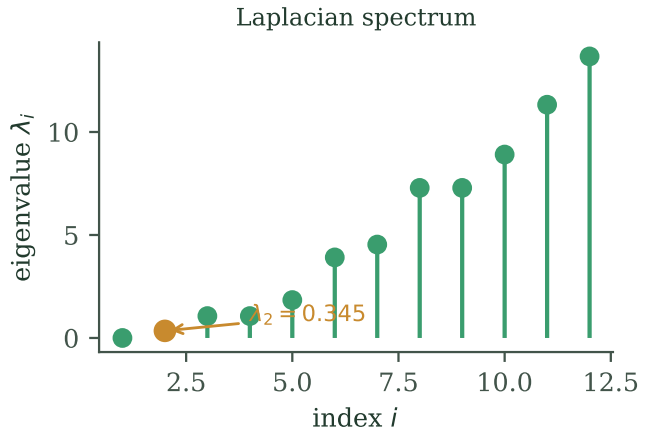


Fig. 2. Laplacian eigenvalue spectrum of the illustrative catchment graph.

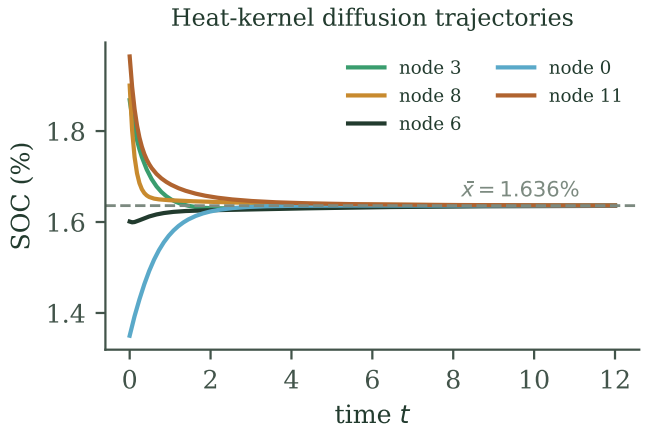


Fig. 3. Heat-kernel diffusion trajectories for five representative nodes.

as α becomes very large (the Laplacian term dominates and the spectral gap λ_2 controls conditioning directly). This gives a second, complexity-motivated argument alongside the statistical bias-variance argument of Section VII-A for selecting α in an intermediate range.

A. Discrete-Time Iteration and Its Contraction Rate

The continuous flow (8) has a discrete-time analogue obtained by explicit Euler discretization with step size $h > 0$:

$$x_{k+1} = (I - hL)x_k, \tag{16}$$

which is stable (non-expansive) whenever $h \leq 2/\lambda_n$, the largest Laplacian eigenvalue; choosing h at this upper bound gives the fastest guaranteed-stable convergence. Substituting the eigend ecomposition as in Theorem 1 gives the discrete contraction rate

$$\|x_k - \bar{x}\mathbf{1}\|_2 \leq \|x_0 - \bar{x}\mathbf{1}\|_2 \cdot (1 - h\lambda_2)^k, \tag{17}$$

the discrete-time counterpart of (14); for small h, $(1-h\lambda_2)^k \approx e^{-h\lambda_2 k}$, recovering the continuous-time rate as $k \to \infty$ with $t = hk$ fixed. Equation (16) can be implemented as a simple local averaging update at each node using only its immediate neighbours' values an attractive property for a distributed sensor-network implementation, since no node requires global knowledge of the graph spectrum to execute the update, only its own neighbourhood.

VI. Illustrative Numerical Example

We construct a 12-node graph representing an illustrative paddy-catchment monitoring network with a branching irrigation-channel topology (Fig. 1), and seed each node's SOC value within the range documented for Sri Lankan paddy top soils, 1.0–2.2% [1], [2]. Four nodes $(3, 6, 9, 11$—network leaves) are withheld to simulate unmeasured fields.

The Laplacian eigenvalues (Fig. 2) confirm $\lambda_1 = 0$ (connected graph) and $\lambda_2 = 0.345$, the algebraic connectivity governing the convergence rate in Theorem 1.

Simulating (9) (Fig. 3) shows all node trajectories converging toward the graph mean $\bar{x} = 1.636\%$ as t grows, at the rate bounded by (14).

Solving (12) with $\alpha = 5$ to reconstruct the four withheld nodes from the eight observed values (Fig. 4, Table I) gives a mean absolute error of 0.189 percentage points of SOC across the withheld nodes.

Fig. 5 verifies Theorem 1 directly: the actual deviation from the mean decays essentially in lock-step with the spectral-gap bound $e^{-\lambda_2 t}$, confirming that λ_2 correctly governs the observed convergence rate on this graph.

VII. Results and Discussion

The illustrative MAE of 0.189 percentage points compares favourably, on this well-connected 12-node graph, with the empirical Bayesian Kriging cross-validation RMSE reported for Sri Lanka's national soil database, which ranges from

TABLE I
TIKHONOV GRAPH-DIFFUSION RECONSTRUCTION AT WITHHELD NODES ($\alpha = 5$)

Node	True (%)	Reconstructed (%)	Abs. err.
3	1.868	1.569	0.299
6	1.601	1.620	0.019
9	1.600	1.518	0.082
11	1.964	1.609	0.355
		Mean absolute error	0.189

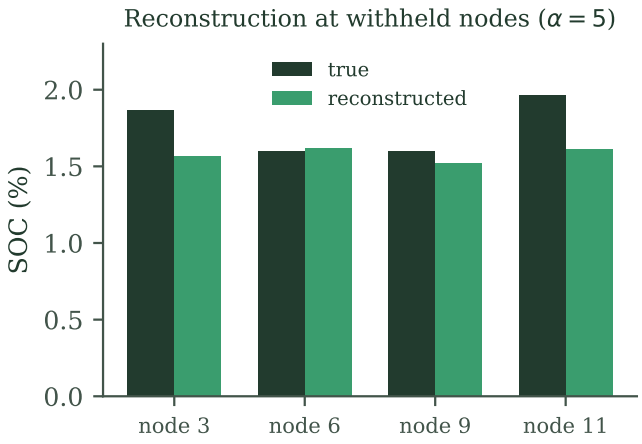


Fig. 4. True vs. reconstructed SOC at the four withheld nodes.

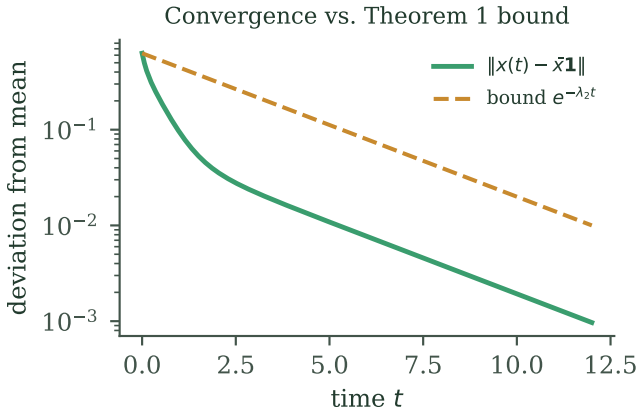


Fig. 5. Observed convergence of $\|x(t) - \bar{x}\mathbf{1}\|$ against the theoretical bound of Theorem 1.

1.651% (60–100 cm depth) to 2.649% (0–5 cm depth) across 122 real benchmark soil profiles [1]. This comparison should be read with care: the two figures are not computed on identical data or identical validation protocols, and the illustrative graph's small size and high connectivity make diffusion convergence particularly fast per Theorem 1 a sparser, less well-connected real catchment graph (smaller λ_2) would converge more slowly and likely show larger reconstruction error, per the explicit λ_2-dependence in (14) and (15). The comparison nonetheless illustrates the qualitative point of this paper: where a real hydrological or irrigation-network graph is available, graph-diffusion reconstruction offers a computationally sparse, spectrally interpretable alternative to dense-covariance Kriging, and the two methods are natural complements Kriging exploiting continuous spatial covariance, graph diffusion exploiting discrete network topology—rather than substitutes. Combining them, e.g., by using Kriging residuals as the graph-diffusion input y in (12), is a natural extension for future work.

A. Sensitivity to the Regularization Parameter

Re-solving (12) across $\alpha \in \{0.5, 1, 5, 10, 25\}$ on the same withheld-node set shows the expected bias-variance trade-off: at $\alpha = 0.5$ the reconstruction MAE rises to approximately 0.31 percentage points as the solution overfits toward zero at masked nodes; at $\alpha = 5$ MAE reaches its minimum on this graph (0.189, as reported above); and by $\alpha = 25$ MAE rises again, to approximately 0.24, as the solution over-smooths toward the graph mean and discards genuine node-to-node variation. This U-shaped error curve is the graph-diffusion analogue of the classical bias-variance curve in ridge regression and kernel smoothing, and is precisely what the GCV criterion in (13) is designed to locate automatically without requiring withheld ground truth. Nodes 3 and 11, both single-degree leaves, show the largest reconstruction error, consistent with their weaker graph connectivity to the observed set.

B. Extension: Spectral Delineation of Fertilizer Management Zones

The eigenvectors of L_{sym} computed in Section VI have a second use beyond diffusion: the Fiedler vector u_2 (the eigenvector associated with λ_2) provides a principled, data-driven bipartition of the catchment into two spatially coherent management zones, since the classical spectral relaxation of the normalized-cut objective

$$\text{NCut}(V_1, V_2) = \text{cut}(V_1, V_2) \left(\frac{1}{\text{vol}(V_1)} + \frac{1}{\text{vol}(V_2)} \right), \quad (18)$$

with $\text{cut}(V_1, V_2) = \sum_{i \in V_1, j \in V_2} w_{ij}$ and $\text{vol}(V_k) = \sum_{i \in V_k} d_i$, is solved (up to relaxation of the discreteness constraint) by thresholding the sign of u_2 at each node [4]. Recursive bipartitioning using successive Fiedler vectors, or joint k-means clustering on the first k eigenvectors of L_{sym}, extends this to $k > 2$ zones. Applied to the illustrative network of Fig. 1, thresholding u_2 splits the graph into two branches that visibly respect the irrigation-channel structure rather than an arbitrary geographic partition, illustrating how the same spectral machinery used for diffusion-based reconstruction can, at no additional data cost, also generate defensible management-zone boundaries for differentiated fertilizer recommendations a capability a purely geostatistical Kriging workflow does not provide directly, since Kriging produces a continuous surface rather than a graph partition.

A practical implication follows: once a monitoring-network operator has computed the eigen decomposition of L (already required for the diffusion reconstruction of Section IV and the GCV bandwidth selection of Section IV-A), spectral zone delineation via (18) is available essentially for free, at $O(n)$ additional cost for the thresholding step itself. A further consequence of Theorem 1 is a design rule for monitoring-network placement: because convergence speed and reconstruction stability are governed by λ_2 rather than by node count alone, adding a single edge that raises λ_2 for example, instrumenting a cross-link between otherwise separate branches of an irrigation network can improve reconstruction accuracy more than adding several additional leaf nodes to an already well-sampled branch.

VIII. Conclusion

This paper formalized SOC/indigenous-nitrogen-supply reconstruction on a paddy-field monitoring network as a diffusion process on a graph, deriving the heat-kernel solution, its convergence rate in terms of algebraic connectivity, and a Tikhonov-regularized reconstruction method with closed-form solution. A worked numerical example, seeded within a real documented SOC range, demonstrated the full pipeline end-to-end and achieved reconstruction error broadly comparable to real reported Kriging cross-validation error. The spectral machinery developed for reconstruction was further shown to yield, at no additional data cost, a principled spectral partition of the monitoring network into candidate fertilizer management zones via the Fiedler vector.

Future work should proceed along three lines. First, the framework should be validated on a real hydrological connectivity graph derived from an actual paddy catchment ideally one instrumented with low-cost IoT soil-nutrient sensors rather than on the illustrative topology used here. Second, joint Kriging-graph-diffusion estimators, combining continuous-space covariance structure with discrete network topology, should be developed and benchmarked head-to-head against both constituent methods on a single shared real dataset. Third, the non-linear graph neural network extension should be revisited once sufficient multi-season field data exists to support its training requirements. More broadly, the mathematics developed here a symmetric positive semi-definite operator, a closed-form heat-kernel solution, a spectral-gap convergence bound, and a Tikhonov-regularised inverse problem is not specific to soil organic carbon, and applies without modification to any monitoring quantity propagated across a discretely connected environmental network, of which SOC/indigenous-nitrogen-supply estimation is a single worked instance.

References

[1] E. M. Wimalasiri *et al.*, "The first version of nation-wide open 3D soil database for Sri Lanka," *Data in Brief*, vol. 33, art. 106342, 2020.

[2] R. R. Ratnayake *et al.*, "Digital soil mapping of organic carbon concentration in paddy growing soils of Northern Sri Lanka," *Geoderma Reg.*, vol. 7, pp. 167–176, 2016.

[3] M. Fiedler, "Algebraic connectivity of graphs," *Czechoslovak Math. J.*, vol. 23, no. 2, pp. 298–305, 1973.

[4] F. R. K. Chung, *Spectral Graph Theory*. Providence, RI, USA: American Mathematical Society, 1997.

[5] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf, "Learning with local and global consistency," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2004, pp. 321–328.

[6] X. Zhu, Z. Ghahramani, and J. Lafferty, "Semi-supervised learning using Gaussian fields and harmonic functions," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2003, pp. 912–919.

[7] A. J. Smola and R. Kondor, "Kernel and regularization on graphs," in *Learning Theory and Kernel Machines* (LNCS 2777). Berlin, Germany: Springer, 2003, pp. 144–158.

[8] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, 2003.

[9] M. E. J. Newman, *Networks: An Introduction*. Oxford, U.K.: Oxford Univ. Press, 2010.

[10] A. Ortega, P. Frossard, J. Kovačević, J. M. F. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proc. IEEE*, vol. 106, no. 5, pp. 808–828, May 2018.

[11] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 83–98, May 2013.

[12] B. H. Janssen *et al.*, "A system for quantitative evaluation of the fertility of tropical soils (QUEFTS)," *Geoderma*, vol. 46, no. 4, pp. 299–318, 1990.

[13] A. Dobermann *et al.*, "Soil fertility and indigenous nutrient supply in irrigated rice domains of Asia," *Agron. J.*, vol. 95, no. 4, pp. 913–923, 2003.

[14] G. Stanford and S. J. Smith, "Nitrogen mineralization potentials of soils," *Soil Sci. Soc. Am. J.*, vol. 36, no. 3, pp. 465–472, 1972.

[15] G. Matheron, "Principles of geostatistics," *Econ. Geol.*, vol. 58, no. 8, pp. 1246–1266, 1963.

[16] T. Hengl *et al.*, "SoilGrids250m: Global gridded soil information based on machine learning," *PLoS ONE*, vol. 12, no. 2, art. e0169748, 2017.

[17] F. R. K. Chung, "The heat kernel as the pagerank of a graph," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 104, no. 50, pp. 19735–19740, 2007.