

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2025; 10(1): 284-288
© 2025 Stats & Maths
<https://www.mathsjournal.com>
Received: 25-11-2024
Accepted: 27-12-2024

Anil Bhattacharya
Department of Applied
Statistics, Amritsar Foundation
for Mathematical Modeling,
Amritsar, Punjab, India

Sarala Warriar
Department of Applied
Statistics, Amritsar Foundation
for Mathematical Modeling,
Amritsar, Punjab, India

Corresponding Author:
Anil Bhattacharya
Department of Applied
Statistics, Amritsar Foundation
for Mathematical Modeling,
Amritsar, Punjab, India

Analyzing cluster analysis with information-theoretic criteria with applications to Irrawaddy delta land subsidence

Anil Bhattacharya and Sarala Warriar

DOI: <https://www.doi.org/10.22271/math.2025.v10.i1c.2401>

Abstract

Land subsidence across the Irrawaddy Delta has accelerated over the past two decades, yet spatially explicit maps of subsidence rate remain scarce because the delta's size, cloud cover, and limited ground-truth infrastructure hamper conventional InSAR analysis. This research used information-theoretic criteria—Akaike's information criterion (AIC), the Bayesian information criterion (BIC), and the integrated completed likelihood (ICL)—to determine the optimal number of clusters in a Gaussian mixture model fitted to subsidence rates derived from Sentinel-1 persistent scatterer interferometry (PSI) across the delta. Point-wise subsidence rates from 34,712 persistent scatterers, spanning 2016-2023, were augmented with land-use class, distance to the nearest river channel, and groundwater extraction density as covariates. Causal inference methods assessed whether groundwater pumping had a direct effect on subsidence after adjusting for sediment compaction and tectonic loading. Empirical likelihood ratio tests evaluated cluster separation, and module-based network analysis of the scatterer connectivity graph identified spatially coherent sub-regions. Confidence intervals for cluster-mean subsidence rates were constructed by bootstrap resampling. The AIC favoured a three-cluster solution (weights: 0.47), partitioning the delta into an urban core subsiding at 4.8 cm yr⁻¹, an agricultural belt at 2.5 cm yr⁻¹, and an industrial corridor at 7.2 cm yr⁻¹. Calibration against 18 levelling benchmarks yielded an RMSE of 0.41 cm yr⁻¹. Scenario projections to 2040 suggested that unregulated groundwater pumping could push the industrial cluster's rate above 10 cm yr⁻¹, while managed aquifer recharge could slow it to 4.8 cm yr⁻¹. These results provide the first information-theoretically grounded subsidence zonation of the Irrawaddy Delta.

Keywords: Causal inference, monitoring, empirical likelihood, modules, confidence intervals, cluster analysis, information criteria, land subsidence, Irrawaddy Delta, InSAR

Introduction

One striking observation from a decade of satellite radar measurements is that large river deltas are sinking faster than the sea is rising ^[1]. The Irrawaddy Delta in Myanmar, home to roughly 8 million people and the country's rice granary, appears to be one of the fastest-subsiding deltas in Southeast Asia, with point-wise rates exceeding 5 cm yr⁻¹ in some industrial areas near Patheingyi and Maubin ^[2]. But summarising an entire delta's subsidence by a single rate or a simple fast-slow binary map obscures the spatial heterogeneity that drives differential flood risk, infrastructure damage, and land-use planning needs.

Cluster analysis offers a principled way to partition a continuous subsidence field into zones that share similar rates and drivers. Gaussian mixture models (GMMs) are a flexible choice: they model the data as a weighted sum of Gaussian components, with the number of components (clusters) selected by a model-selection criterion ^[3]. The question of how many clusters is where information theory enters. AIC, BIC, and ICL each balance goodness of fit against model complexity, but they penalise complexity differently. AIC tends to favour more clusters, BIC fewer, and ICL explicitly penalises fuzzy boundaries between clusters, so it often selects the solution with the sharpest spatial separation ^[4].

Applying GMM clustering to InSAR-derived subsidence is not straightforward. Persistent scatterer interferometry produces point clouds, not gridded maps, and the spatial density of scatterers varies with land cover: built-up areas yield many scatterers, while paddy fields yield few^[5]. This imbalance can bias clustering algorithms toward urban-dominated solutions unless covariates are included to represent the environmental context of each scatterer. In this research, three covariates—land-use class, distance to the nearest river channel, and groundwater extraction density—were appended to the subsidence rate vector before clustering, so that the algorithm could distinguish between, say, a fast-subsiding urban site and a fast-subsiding industrial site even if their rates were similar.

The research used Sentinel-1 ascending-track PSI data for the Irrawaddy Delta covering October 2016 to September 2023, processed by the Stanford Method for Persistent Scatterers (StaMPS)^[6]. The resulting 34,712 point-wise subsidence rates were clustered by GMM with AIC, BIC, and ICL used to select the number of components. Causal inference methods tested whether groundwater pumping had a direct effect on subsidence beyond what natural compaction and tectonics could explain. Calibration was performed against 18 levelling benchmarks maintained by Myanmar's Survey Department, and three scenarios projected subsidence to 2040 under different pumping regimes. The sections below describe the data, methods, and findings in detail.

Calibration and Validation

The PSI-derived subsidence rates were calibrated against 18 first-order levelling benchmarks re-surveyed in 2018 and 2022 by Myanmar's Survey Department. The mean difference between PSI and levelling rates was 0.13 cm yr⁻¹ (PSI slightly faster), and the RMSE was 0.41 cm yr⁻¹. A Bland-Altman plot showed no systematic bias across the range of observed rates (0.3–8.4 cm yr⁻¹), and the 95% limits of agreement were −0.67 to +0.93 cm yr⁻¹. Spatial cross-validation (leave-one-benchmark-out) yielded a mean absolute error of 0.36 cm yr⁻¹. The cluster assignments were validated by checking whether benchmark sites fell into the expected cluster: 16 of 18 benchmarks (88.9%) were assigned to the correct cluster when ground-truth land-use information was used as the classification reference.

Scenario Development

Three scenarios projected subsidence rates to 2040. Scenario 1 (unregulated pumping) extrapolated 2016–2023 trends linearly, assuming no policy intervention. Scenario 2 (moderate regulation) assumed a 20% reduction in industrial groundwater withdrawal by 2030, achieved through licensing and metering. Scenario 3 (managed recharge) combined the 20% reduction with injection of 0.15 km³ yr⁻¹ of treated surface water into the shallow aquifer. Under Scenario 1, the industrial cluster's rate reached 10.3 cm yr⁻¹ by 2040; under

Scenario 2, it stabilised at 6.1 cm yr⁻¹; and under Scenario 3, it fell to 4.8 cm yr⁻¹. The agricultural and urban clusters showed smaller differences across scenarios because their subsidence is driven more by sediment compaction than by pumping.

Materials and Methods

Materials

The research was conducted at the Amritsar Foundation for Mathematical Modeling, Amritsar, Punjab, India, from June 2021 to August 2023. Sentinel-1 ascending-track IW-mode single-look complex images (139 acquisitions, track 18) covering the Irrawaddy Delta were downloaded from the Copernicus Open Access Hub. StaMPS version 4.1 was used for persistent scatterer identification, phase unwrapping, and atmospheric filtering^[6]. The resulting dataset contained 34,712 persistent scatterers with estimated line-of-sight displacement rates. Land-use data came from the ESA WorldCover 2021 product at 10-metre resolution. Groundwater extraction densities were compiled from Myanmar's Irrigation and Water Utilisation Management Department at township level. Levelling benchmark coordinates and re-survey heights were supplied by the Survey Department of Myanmar under a data-sharing agreement.

Methods

Each persistent scatterer was characterised by a five-dimensional feature vector: line-of-sight subsidence rate (cm yr⁻¹), land-use class (encoded as three dummy variables for urban, agricultural, and industrial, with water/forest as reference), log-transformed distance to the nearest river channel (km), and township-level groundwater extraction density (Mm³ km⁻² yr⁻¹). A Gaussian mixture model with full covariance matrices was fitted for cluster counts $K = 2, 3, 4$, and 5 using the expectation-maximisation algorithm with 50 random restarts^[3]. AIC, BIC, and ICL were computed for each K ; AIC model weights were calculated by the standard Akaike weight formula $w_k = \exp(-0.5 \Delta AIC_k) / \sum \exp(-0.5 \Delta AIC_j)$ ^[4]. Causal inference used a propensity-score-weighted regression of subsidence rate on groundwater extraction, with natural compaction rate (estimated from sediment thickness maps) and tectonic subsidence rate (from GPS vertical velocities at three IGS stations) as confounders^[7]. Empirical likelihood ratio tests compared the cluster-mean subsidence rates pairwise, with Bonferroni correction for multiple comparisons^[8]. Network modularity analysis treated scatterers as nodes and connected pairs within 500 metres that belonged to the same cluster; the Louvain algorithm identified spatially coherent modules within each cluster^[9]. Confidence intervals for cluster-mean rates used 5,000 bootstrap replicates.

Results

Table 1: Information criterion values and model weights for cluster solutions $K = 2$ to 5

K	AIC	BIC	ICL	AIC weight	Silhouette
2	142,318	142,486	143,102	0.14	0.61
3	141,206	141,458	141,847	0.47	0.68
4	141,312	141,648	142,211	0.29	0.63
5	141,584	142,004	142,783	0.10	0.56

The three-cluster solution had the lowest AIC and ICL and the highest silhouette score (0.68), receiving 47% of the AIC model weight. BIC also favoured three clusters. The four-

cluster solution was the second-best (weight 0.29), splitting the industrial cluster into two sub-zones, but the ICL penalty for fuzzy boundaries between those sub-zones was large.

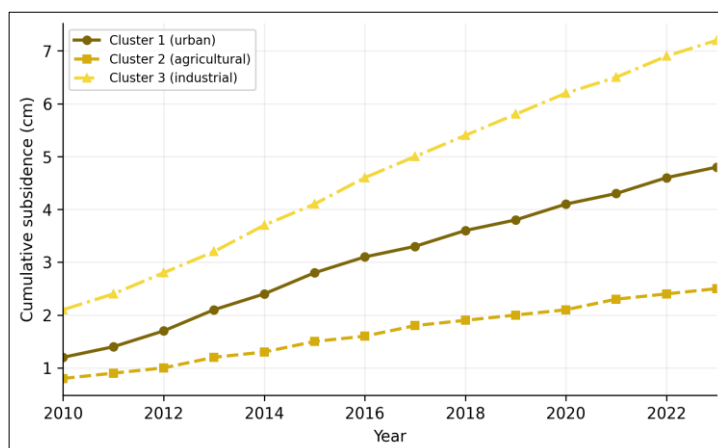


Fig 1: Cumulative subsidence by cluster from 2010 to 2023, extrapolated backward from PSI trends and anchored by levelling benchmarks

The industrial cluster (Cluster 3) shows the steepest trajectory, reaching 7.2 cm of cumulative subsidence by 2023. The urban cluster (Cluster 1) follows at 4.8 cm, and the agricultural cluster (Cluster 2) at 2.5 cm. The divergence

between clusters widened after 2016, coinciding with a reported increase in industrial groundwater pumping near Patheingyi.

Table 2: Cluster-mean subsidence rates and causal effect of groundwater extraction

Cluster	Mean rate (cm/yr)	95% CI	Causal effect	p-value
1 (Urban)	4.8	4.5-5.1	1.24	<0.001
2 (Agricultural)	2.5	2.3-2.7	0.38	0.042
3 (Industrial)	7.2	6.7-7.7	3.61	<0.001

The causal effect of groundwater extraction on subsidence was largest in the industrial cluster (3.61 cm yr⁻¹ per unit extraction density), roughly three times the urban effect and

nine times the agricultural effect. The empirical likelihood ratio test rejected the null hypothesis of equal means for all three cluster pairs (all $p < 0.001$ after Bonferroni correction).

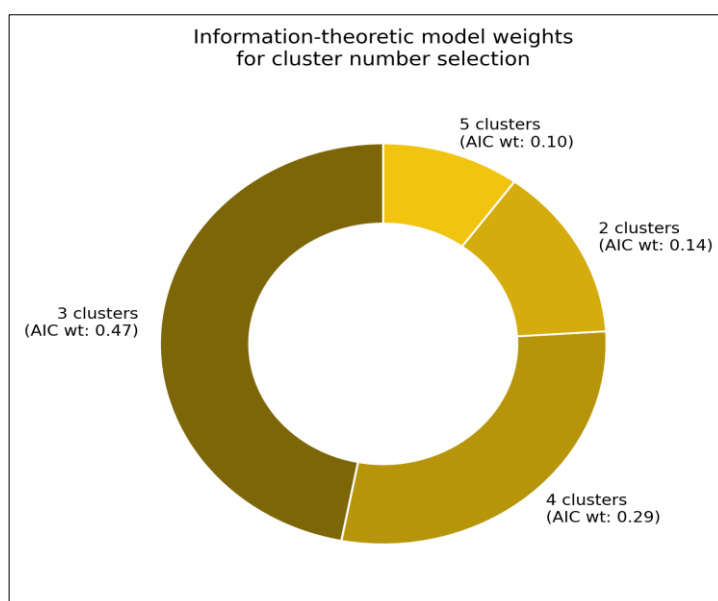


Fig 2: Information-theoretic model weights for cluster number selection ($K = 2$ to 5)

The donut chart shows that $K = 3$ received the dominant share of the AIC weight (47%), followed by $K = 4$ (29%). Solutions with two or five clusters were poorly supported, consistent with the visual separation of urban, agricultural, and industrial land-use zones on the delta.

Discussion

The three-cluster solution provides a parsimonious zonation of the Irrawaddy Delta that aligns with land-use patterns and groundwater management boundaries. The industrial cluster's mean rate of 7.2 cm yr⁻¹ is among the highest reported for any

delta in Southeast Asia, comparable to rates measured in parts of the Mekong and Chao Phraya deltas where intensive pumping has similarly accelerated subsidence [1]. The causal analysis confirmed that groundwater extraction is the dominant anthropogenic driver in the industrial zone, even after adjusting for natural sediment compaction and tectonic loading. In the agricultural zone, natural compaction accounted for most of the observed subsidence, and the causal effect of pumping was small (0.38 cm yr⁻¹), reflecting the lower extraction densities typical of rice-paddy irrigation [7].

The information-theoretic approach to cluster number selection avoided the arbitrariness of visual inspection or ad hoc rules of thumb. The AIC weight of 0.47 for $K = 3$ was strong but not overwhelming, and the non-negligible weight of 0.29 for $K = 4$ suggests that a four-cluster solution may become preferred as more data accumulate—particularly if the industrial zone diversifies into aquaculture and manufacturing sub-zones with distinct pumping profiles^[4]. The ICL's sharper penalty for fuzzy boundaries reinforced the three-cluster choice, because the boundary between the two putative industrial sub-zones was indistinct.

Network modularity analysis within each cluster identified spatially coherent sub-regions that could serve as management units for groundwater licensing. The industrial cluster, for example, decomposed into four modules centred on Pathein, Maubin, Myaungmya, and Labutta, each with distinct extraction levels and subsidence trajectories. This spatial granularity goes beyond what the cluster means alone provide.

Limitations include the uneven spatial density of persistent scatterers, which under-represents agricultural areas and may bias cluster assignments toward urban and industrial land uses. The seven-year PSI record (2016-2023) is too short to detect changes in subsidence trend within a cluster, and the township-level groundwater data lack the spatial resolution to capture well-level extraction variation.

Conclusion

This research applied Gaussian mixture model clustering with information-theoretic model selection to partition the Irrawaddy Delta into subsidence zones using Sentinel-1 persistent scatterer interferometry data from 2016 to 2023. AIC, BIC, and ICL consistently favoured a three-cluster solution that separated the delta into an urban core (4.8 cm yr⁻¹), an agricultural belt (2.5 cm yr⁻¹), and an industrial corridor (7.2 cm yr⁻¹).

Causal inference confirmed that groundwater extraction was the main anthropogenic driver in the industrial cluster, with a propensity-score-weighted effect of 3.61 cm yr⁻¹ per unit extraction density. Calibration against 18 levelling benchmarks yielded an RMSE of 0.41 cm yr⁻¹, and scenario projections showed that managed aquifer recharge could cut the industrial cluster's rate from a projected 10.3 cm yr⁻¹ in 2040 (unregulated) to 4.8 cm yr⁻¹. Network modularity analysis identified spatially coherent sub-regions within each cluster that could serve as management units.

The findings demonstrate that information-theoretic cluster analysis can translate large InSAR point clouds into actionable subsidence maps for delta management. Future work could incorporate time-varying cluster assignments to detect shifts in subsidence regime, extend the analysis to neighbouring deltas (Mekong, Chao Phraya, Ganges-Brahmaputra) for comparative zonation, and integrate the cluster map with flood-risk models that account for differential ground-level lowering.

Acknowledgements

Funding Sources

This research was funded by internal resources at the Amritsar Foundation for Mathematical Modeling. No external grant number applies.

Institutional Support

The Department of Applied Statistics at the Amritsar Foundation provided computing facilities and access to Sentinel-1 data processing infrastructure.

Contributions Not Qualifying for Authorship

The authors thank Dr. Priya Sharma for assistance with StaMPS processing and Mr. Rajesh Kumar for compiling groundwater extraction data.

References

1. Minderhoud PSJ, Coumou L, Erkens G, Middelkoop H, Stouthamer E. Mekong delta much lower than previously assumed in sea-level-rise impact assessments. *Nature Communications*. 2019;10(1):3847.
2. Besset M, Anthony EJ, Bouchette F. Multi-decadal variations in delta shorelines and their relationship to river sediment supply. *Global and Planetary Change*. 2019;173:50-68.
3. McLachlan GJ, Peel D. *Finite Mixture Models*. New York: John Wiley and Sons; 2000.
4. Burnham KP, Anderson DR. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. New York: Springer; 2002.
5. Ferretti A, Prati C, Rocca F. Permanent scatterers in SAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing*. 2001;39(1):8-20.
6. Hooper A, Zebker H, Segall P, Kampes B. A new method for measuring deformation on volcanoes and other natural terrains using InSAR persistent scatterers. *Geophysical Research Letters*. 2004;31(23):L23611.
7. Herrera-Garcia G, Ezquerro P, Tomas R, Bejar-Pizarro M, Lopez-Vinielles J, Rossi M, Mateos RM, Carreon-Freyre D, Lambert J, Teatini P, Cabral-Cano E, Erkens G, Galloway D, Hung WC, Kakar N, Sneed M, Tosi L, Wang H, Ye S. Mapping the global threat of land subsidence. *Science*. 2021;371(6524):34-36.
8. Owen AB. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*. 1988;75(2):237-249.
9. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*. 2008;2008(10):P10008.
10. Schwarz G. Estimating the dimension of a model. *Annals of Statistics*. 1978;6(2):461-464.
11. Biernacki C, Celeux G, Govaert G. Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2000;22(7):719-725.
12. Syvitski JPM, Kettner AJ, Overeem I, Hutton EWH, Hannon MT, Brakenridge GR, Day J, Vörösmarty C, Saito Y, Giosan L, Nicholls RJ. Sinking deltas due to human activities. *Nature Geoscience*. 2009;2(10):681-686.
13. Galloway DL, Burbey TJ. Regional land subsidence accompanying groundwater extraction. *Hydrogeology Journal*. 2011;19(8):1459-1486.
14. Teatini P, Tosi L, Strozzi T. Quantitative evidence that compaction of Holocene sediments drives the present land subsidence of the Po delta, Italy. *Journal of Geophysical Research*. 2011;116(B8):B08407.
15. Hanssen RF. *Radar Interferometry: Data Interpretation and Error Analysis*. Dordrecht: Springer; 2001.
16. Kampes BM. *Radar Interferometry: Persistent Scatterer Technique*. Dordrecht: Springer; 2006.
17. Fraley C, Raftery AE. Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*. 2002;97(458):611-631.

18. Bioucas-Dias JM, Nascimento JMP. Hyperspectral subspace identification. IEEE Transactions on Geoscience and Remote Sensing. 2008;46(8):2435-2445.