IAOS 2026

EXTENDED ABSTRACT

Completeness-Aware Mortality Forecasting in Indonesia Using Remote Sensing and Geospatial Modelling

I Nyoman Setiawan[*†] and Suryo Adi Rakhmawan[*‡]

†Directorate of Statistical Analysis and Development, BPS-Statistics Indonesia
‡Directorate of Population and Employment Statistics, BPS-Statistics Indonesia
[*]Corresponding author. Email: inyomans1998@gmail.com; suryo.rakhmawan@bps.go.id

Abstract

Mortality forecasting is central to evidence-based planning for public health, social protection, and population ageing. However, in many national statistical systems including Indonesia, its operational value is often constrained by two structural limitations: (i) death registration that remains incomplete and temporally varying across areas, and (ii) the scarcity of timely, spatially granular covariates that can support subnational modelling and early warning. In the context of the data revolution, this study develops a scalable, quality-aware framework that integrates satellite-derived remote sensing signals with demographic count modelling to produce completeness-corrected mortality estimates, enhance forecasting performance, and generate actionable geospatial risk surfaces for policy targeting.

We compile a multi-source, multi-resolution dataset for Indonesia comprising annually aggregated registered deaths at the district/municipality (kabupaten/kota) level disaggregated by age group, corresponding exposure (mid-year population) by age group, and district-year completeness rates of death registration. To enrich the covariate space, we extract remote sensing features at the subdistrict (kecamatan) level and aggregate them to districts using age-relevant population weights, thereby aligning environmental and built-environment signals with the spatial distribution of populations at risk. Remote sensing indicators include vegetation proxies (e.g., NDVI), thermal stress measures, night-time lights, urban form/imperviousness, and land-cover dynamics, consistent with established guidance on linking Earth observation to health outcomes under spatial-temporal misalignment and privacy-protecting geolocation practices (Brown et al. 2014; Tucker 1979).

Methodologically, we treat registered deaths as underreported observations rather than ground truth. Observed death counts are modelled using a negative binomial specification with offsets for exposure and registration completeness, yielding principled estimates of underlying age-specific mortality rates under imperfect reporting. The proposed negative binomial hazard model is equivalent to a grouped-time piecewise exponential survival model extended with completeness correction and geospatial covariates. To accommodate nonlinearities, shocks, and complex spatial heterogeneity, we extend the demographic backbone (age and period components) with a hybrid residual-learning architecture in which gradient boosting models learn systematic residual variation from remote sensing and contextual covariates, an approach motivated by recent evidence that satellite imagery and deep learning can recover meaningful neighborhood features predictive of area-level mortality and enable interpretable embeddings (Levy et al. 2021; Jean et al. 2016; Suel et al. 2019). Spatial dependence is assessed and incorporated through spatial random effects, strengthening statistical coherence across neighbouring districts and improving stability in small-area estimation. Finally, corrected forecasts are operationalised into an early-warning monitoring layer by tracking deviations between expected and observed completeness-adjusted trajectories to flag emerging local anomalies.

Keywords: mortality forecasting, death registration completeness, remote sensing, Indonesia

1. Introduction

Mortality statistics are a foundational input for health planning, social protection, and understanding population ageing. Yet, in many countries, subnational mortality estimation and forecasting are constrained by incomplete and spatially heterogeneous civil registration and vital statistics (CRVS) systems. When completeness varies systematically across areas and time, naïve use of registered deaths can bias levels, distort spatial gradients, and undermine comparability.

In parallel, conventional subnational covariates are often temporally lagged, spatially coarse, or inconsistent across administrative units. Satellite remote sensing offers a complementary pathway: it provides timely, spatially consistent proxies for environmental and built-environment conditions plausibly linked to mortality risk, such as greenness (NDVI), thermal stress, urban intensity and land-cover change (Brown et al. 2014; Tucker 1979; Brown and Beurs 2008). A growing body of work demonstrates that satellite imagery combined with machine learning can extract interpretable neighborhood features and predict health-related outcomes at area level. For example, Levy et al. show that deep learning on satellite images can identify associations between residential neighborhood features and county-level mortality, supporting both prediction and interpretability via representation learning and post-hoc explanation tools (Levy et al. 2021). Related advances in satellite-based ML have been used to infer socioeconomic gradients (e.g., poverty) that are often tightly coupled with health risks and mortality (Jean et al. 2016; Suel et al. 2019).

This paper proposes a quality-aware framework for Indonesia that integrates: (i) completeness-corrected mortality modelling from aggregated death registration; (ii) satellite-derived covariates extracted at subdistrict level and aggregated to districts using population-at-risk weights; (iii) hybrid residual learning to capture nonlinearities and shocks; and (iv) explicit spatial modelling to stabilise small-area estimates. The outputs support both forecasting and geospatial policy targeting in official statistics.

2. Rationale

Indonesia provides a compelling case for modernising mortality forecasting because data quality is spatially uneven and temporally varying across districts. Improvements in observed death counts may reflect better reporting rather than genuine mortality change, motivating an approach that explicitly models the observation process through completeness adjustment.

From an official statistics perspective, an operational solution should satisfy four requirements: statistical coherence under imperfect registration; scalability for routine production; interpretability and transparency; and actionable subnational outputs (risk surfaces and anomaly alerts). We design the proposed framework to meet these requirements while explicitly addressing known challenges in integrating remote sensing with survey/administrative outcomes, including spatial-temporal resolution mismatch, aggregation/buffering choices, and privacy-protecting displacement of geolocated units (Brown et al. 2014).

3. Data

3.1 Outcome and exposure

We construct annual district-level registered deaths disaggregated by age group, denoted $D^{obs}_{j,t,a}$ for district j, year t, and age group a, and exposure $E_{j,t,a}$ from mid-year population by age group (person-years approximation).

3.2 Completeness

District-year completeness rates $c_{j,t}$ quantify the proportion of true deaths captured by registration. We treat completeness as an input to the observation model to correct undercount and improve comparability across space and time.

3.3 Remote sensing covariates and aggregation

Remote sensing covariates are extracted at subdistrict (kecamatan) level and aggregated to districts using population-at-risk weights:

$$X^{RS}_{j,t,a} = \sum_{i \in j} w_{i,j,t,a} \, X^{RS}_{i,t}, \qquad w_{i,j,t,a} = \frac{E_{i,t,a}}{\sum_{i' \in j} E_{i',t,a}}.$$

This design follows established remote sensing–health linkage practice by aligning environmental signals with where populations at risk reside, while acknowledging that geolocation practices may incorporate privacy-preserving displacement in some survey contexts (Brown et al. 2014).

4. Methodology

4.1 Completeness-aware hazard model

We treat registered deaths as underreported observations. Observed deaths follow a negative binomial model with offsets for exposure and completeness:

$$D^{obs}_{j,t,a} \sim \mathrm{NegBin}(\mu_{j,t,a}, \phi), \tag{1}$$

$$\log \mu_{j,t,a} = \log E_{j,t,a} + \log c_{j,t} + \eta_{j,t,a}. \tag{2}$$

Here, $\exp(\eta_{j,t,a})$ is the underlying grouped-time hazard (age-specific mortality rate). This formulation is equivalent to a piecewise exponential survival model where hazard is approximately constant within each age interval and year.

4.2 Demographic backbone and geospatial covariates

We specify a transparent demographic structure (age and period components), augment it with remote sensing covariates, and include spatial effects to improve stability for small areas.

4.3 Hybrid residual learning

To capture nonlinearities and shocks, we add a residual-learning layer using gradient boosting (GBM/XGBoost). This follows the broader logic of combining interpretable structural models with ML to learn systematic residual variation, while preserving an official-statistics-friendly backbone (Levy et al. 2021; Maharana and Nsoesie 2018).

4.4 Spatial dependence and coherence

We assess spatial dependence (e.g., Moran's I on residuals) and incorporate spatial random effects to borrow strength across neighbouring districts, reducing spurious volatility and improving coherence of risk maps.

4.5 Forecasting and early warning

We evaluate using out-of-time validation and systematic ablation. We operationalise forecasts into an early-warning layer by monitoring deviations between expected and observed completeness-adjusted trajectories, analogous in spirit to process monitoring approaches used to detect subtle shifts in demographic indicators.

5. Results and Discussion

This section reports (i) spatial heterogeneity of completeness and its implications for bias; (ii) incremental gains from completeness correction, remote sensing covariates, residual learning, and spatial effects; (iii) age-specific vulnerability patterns, especially for elderly mortality; and (iv) early-warning signals for emerging local anomalies.

6. Conclusion

By modelling registered deaths as underreported observations and integrating subdistrict-derived remote sensing signals, hybrid ML augmentation, and spatial coherence, the proposed framework supports completeness-corrected mortality estimation and forecasting for official statistics. The outputs include age-specific corrected rates, improved district-level forecasts, subdistrict-informed risk surfaces, and an anomaly monitoring layer to support proactive demographic and public health decision-making.

References

Brown, Molly E., and Kirsten de Beurs. 2008. Evaluation of multi-sensor semi-arid crop season parameters based on ndvi and rainfall. *Remote Sensing of Environment* 112:2261–2271.

Brown, Molly E., Kathryn Grace, Gerald Shively, Kiersten B. Johnson, and Mark Carroll. 2014. Using satellite remote sensing and household survey data to assess human health and nutrition response to environmental change. *Population and Environment* 36:48–72. https://doi.org/10.1007/s11111-013-0201-0.

Jean, Neal, Marshall Burke, Michael Xie, W. Matthew Davis, David B. Lobell, and Stefano Ermon. 2016. Combining satellite imagery and machine learning to predict poverty. *Science* 353:790–794. https://doi.org/10.1126/science.aaf7894.

Levy, Joshua J., Rebecca M. Lebeaux, Anne G. Hoen, Brock C. Christensen, Louis J. Vaickus, and Todd A. MacKenzie. 2021. Using satellite images and deep learning to identify associations between county-level mortality and residential neighborhood features proximal to schools: a cross-sectional study. *Frontiers in Public Health* 9:766707. https://doi.org/10.3389/fpubh.2021.766707.

Maharana, Ashutosh, and Elaine O. Nsoesie. 2018. Use of deep learning to examine the association of the built environment with prevalence of neighborhood adult obesity. *JAMA Network Open* 1:e181535. https://doi.org/10.1001/jamanetworkopen.2018.1535.

Suel, Esra, John W. Polak, James E. Bennett, and Majid Ezzati. 2019. Measuring social, environmental and health inequalities using deep learning and street imagery. *Scientific Reports* 9:6229. https://doi.org/10.1038/s41598-019-42036-w.

Tucker, Compton J. 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment* 8:127–150.