Enhancing Wastewater Surveillance with Regression Kriging

Odey R. Mofokeng
With: Inger Fabris-Rotelli, Raeesa Docrat, Pravesh Debba, Jenny Holloway, Nontembeko Dudeni-Tlhone, Jabulani Jele, Lisa Schaefer (Née Burke), Wouter le Roux, David Rose, Rene Stander and Nada Abdelatif
School of Actuarial Science and Statistics, University of the Witwatersrand, Johannesburg, South Africa

Introduction and Problem Statement

Traditional spatial methods like standard Kriging often fail in complex urban areas because they ignore environmental and demographic factors.

Key Challenges:

- Data misalignment between health districts and WWTP drainage areas
- Standard Kriging ignores key environmental covariates
- The need for accurate predictions to support public health decisions

Our Solution: A **Regression Kriging with Spatial Error Model (RK-SEM)** that integrates wastewater data, COVID-19 cases, and environmental predictors for more accurate spatial predictions.

Methodology

Our RK-SEM framework addresses spatial misalignment through a three-stage approach:

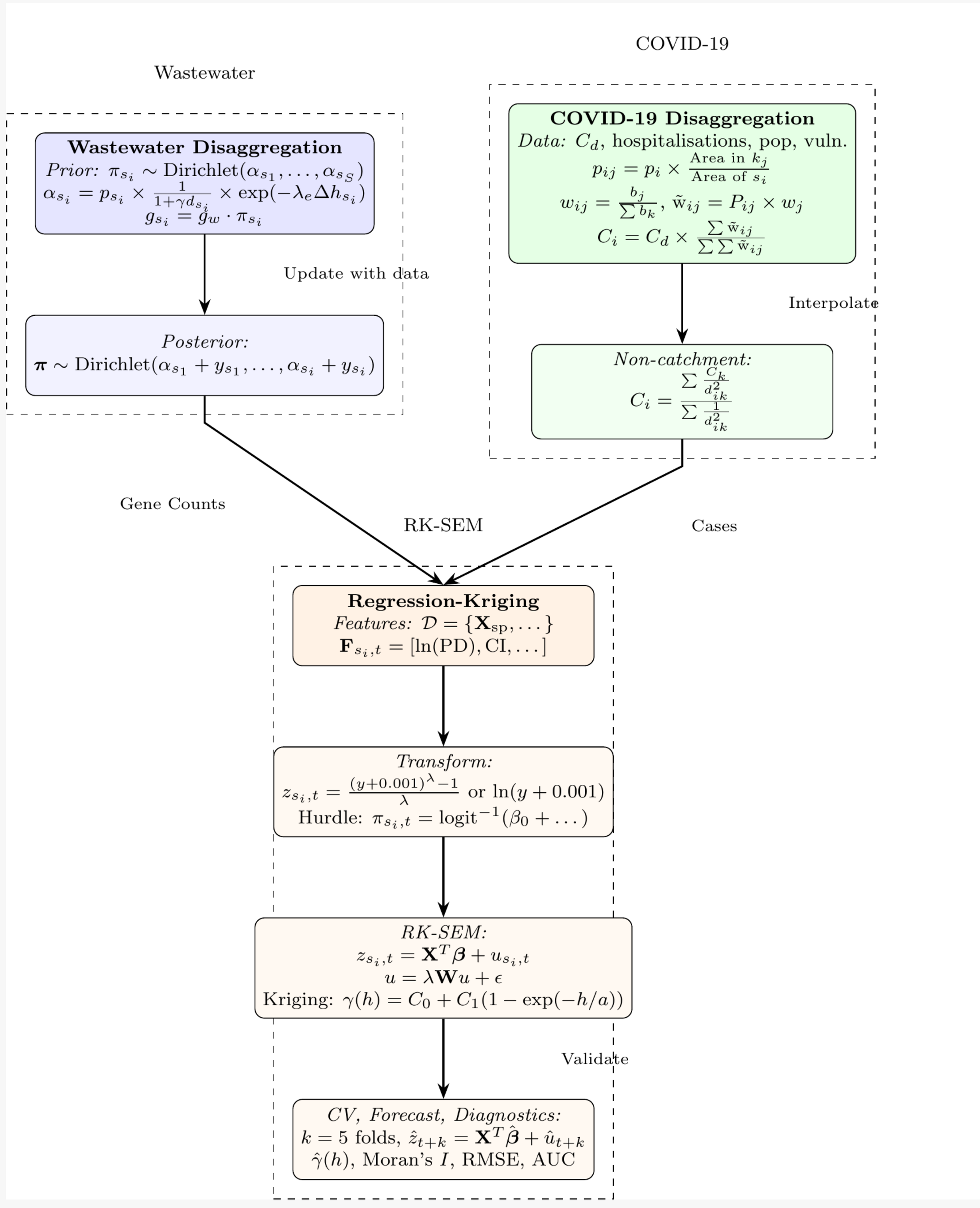

Figure Methodology flowchart showing the integrated approach to wastewater surveillance

The model captures both covariate effects and spatial dependencies in the residuals.

Study Area and Data

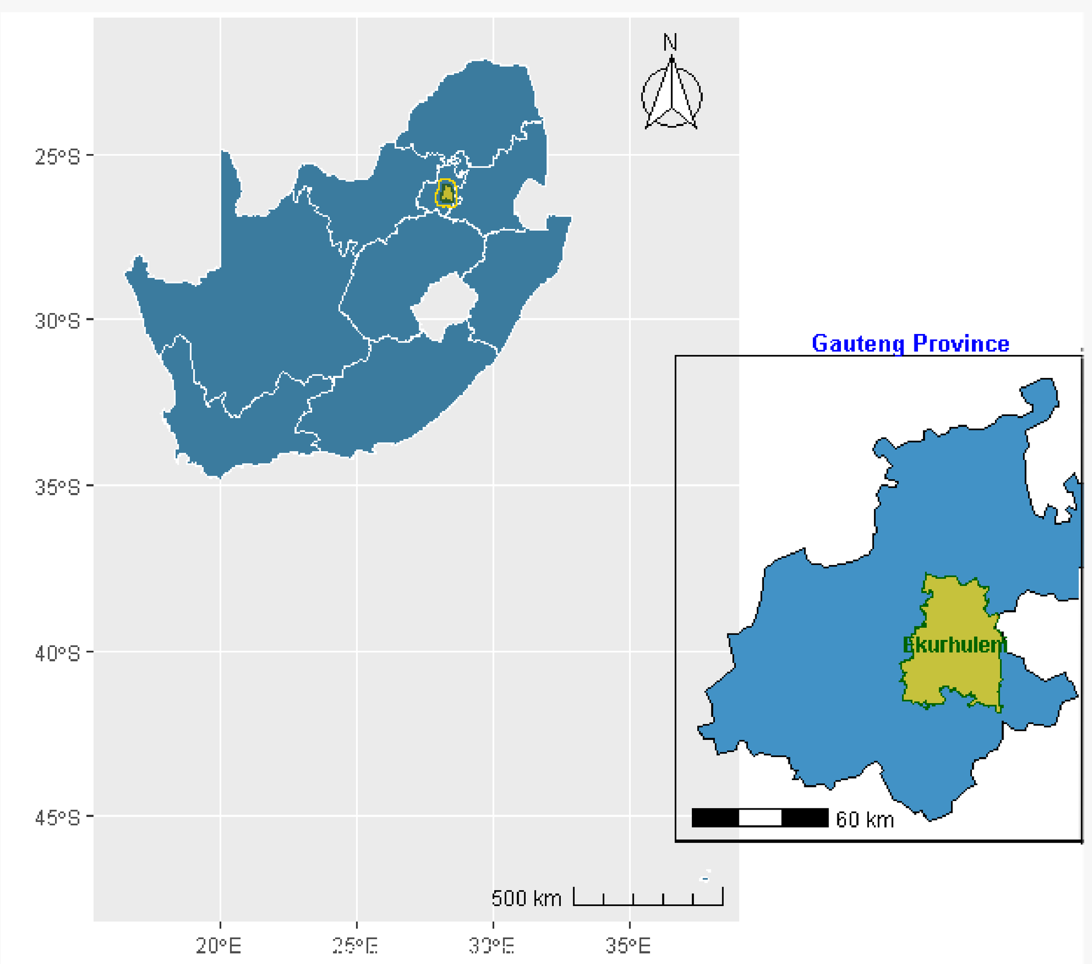

Figure Ekurhuleni Metropolitan Municipality: A complex urban environment with diverse settlement patterns and 19 wastewater treatment plants

Why Ekurhuleni?

- Mix of formal urban centres, informal settlements, and rural farmland
- Major economic hub with mining, manufacturing, and logistics sectors
- 19 wastewater treatment plants (8 sampled for this study)
- Represents typical urban challenges in South African municipalities

Data Sources:

- Wastewater SARS-CoV-2 gene counts from ERWAT treatment plants
- COVID-19 case data from municipality health records
- Hospital accessibility and catchment areas
- Population density and vulnerability indices

Model Performance

Our RK-SEM framework significantly outperforms both baseline methods and sophisticated alternatives across multiple epidemiological phases.

Table Performance comparison on 2021-01-01 (transition period)

Metric	RK-SEM	Rolling Avg.	LOCF	Spatial Kriging	Overall Avg.
MAE	**0.0133**	0.1546	0.2066	0.2548	0.1847
RMSE	**0.0185**	0.2547	0.3056	0.3556	0.2865
sMAPE	**21.16**	35.01	50.07	25.09	48.01
R^2	**0.9045**	0.2254	0.2037	0.5056	0.3885

Key Findings:

- RK-SEM error metrics are 11-19 times lower than baseline methods
- Explains 40-50% more variance (R^2) than best baseline
- Consistent superior performance across peak, transition, and post-COVID phases

Spatial Prediction Accuracy

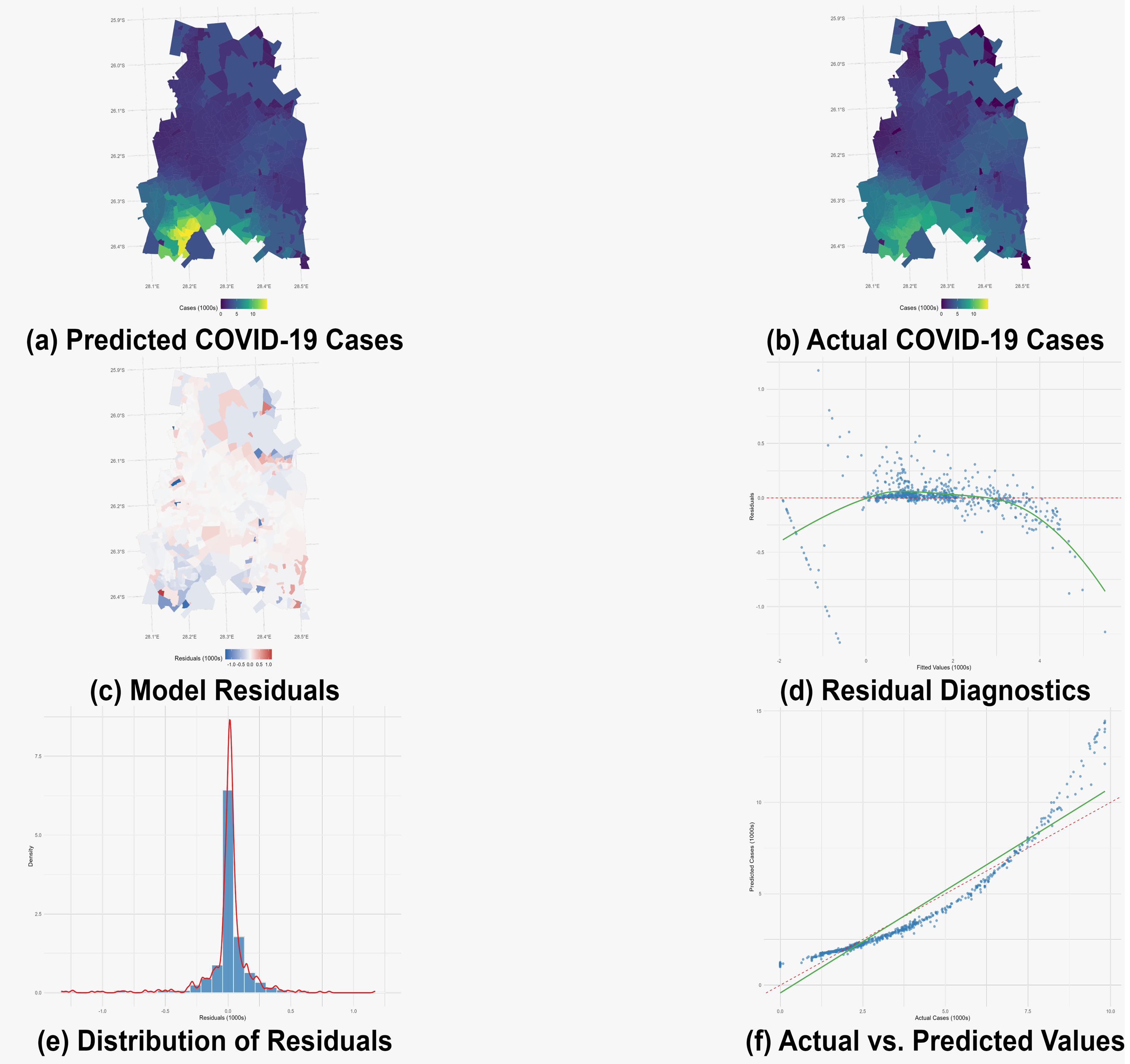

Figure RK-SEM predictions closely match actual COVID-19 case distribution, effectively capturing spatial gradients and clusters across Ekurhuleni subplaces

Comparison with Spatial Linear Mixed Model (SLMM):

Table RK-SEM vs SLMM performance across epidemiological phases

	2021-01-01		2021-12-24		2023-01-01	
Metric	SLMM	**RK-SEM**	SLMM	**RK-SEM**	SLMM	**RK-SEM**
MAE	0.0164	**0.0133**	7.87	**4.68**	0.686	**0.576**
R^2	0.860	**0.904**	0.449	**0.690**	0.740	**0.849**

The performance gap is largest during complex peak periods, where RK-SEM's MAE is 40% lower and R^2 is 24% higher than SLMM.

Public Health Impact: Outbreak Detection

Table Outbreak detection capability (Sensitivity = ability to detect real outbreaks)

	2021-01-01		2021-12-24		2023-01-01	
Metric	SLMM	**RK-SEM**	SLMM	**RK-SEM**	SLMM	**RK-SEM**
Sensitivity	0.825	**0.836**	0.158	**0.842**	0.751	**0.842**
Specificity	1.000	**1.000**	0.863	**0.855**	1.000	**1.000**
F1 Score	0.911	**0.911**	0.205	**0.749**	0.752	**0.914**

Critical Finding: RK-SEM's outbreak detection capability remains robust during challenging peak periods, while SLMM's sensitivity collapses to 0.16, missing 84% of real outbreaks.

Conclusions and Impact

- **RK-SEM framework** integrates wastewater data, covariates, and spatial structure
- **Outperforms baselines and SLMM**, especially during complex periods
- **Robust outbreak detection** with high sensitivity for early warnings
- **Strong spatial autocorrelation** ($\lambda \approx 0.95$) shows disease clustering
- **Key drivers:** Case acceleration and healthcare access influence spread
- **Enables proactive interventions** in South African urban settings

2555438@students.wits.ac.za