International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 8, 2025
https://theaspd.com/index.php

Spatio-Temporal Analysis Of The Association Between Ambient Air Pollution And Breast Cancer Incidence In Southern India: A Bayesian Modeling Approach

Remya Nath P[1], Aleyamma Mathew[2], Sruthi S[3], Beela Sarah Mathew[4] , Preethi Sara George[5*]
[1,3] Research Scholar, Manipal Academy of Higher Education, Manipal, Karnataka, India.
[1,2,3,5] Division of Cancer Epidemiology & Biostatistics, Regional Cancer Centre, Thiruvananthapuram, India.
[4]Department of Radiation Oncology, Regional Cancer Centre, Thiruvananthapuram, India
*Corresponding Author: Preethi Sara George (preeethircc@gmail.com)

Abstract
Breast cancer (BC) is the most common cancer among women in India, and in Thiruvananthapuram, Kerala, its incidence has shown a steady increase over recent decades. While reproductive and lifestyle factors remain well-established determinants, environmental exposures such as air pollution are emerging as potential contributors, yet limited research in India has investigated this relationship using advanced spatial epidemiological approaches. In this study, data on BC cases from the Population-Based Cancer Registry between 2005 and 2019 were analyzed and aggregated into five 3-year periods. Air pollutant concentrations were estimated using inverse distance weighting, and spatial clustering was examined through Global and Local Moran's I, Getis-Ord Gi, and Kulldorff's spatio-temporal scan statistic. To evaluate associations, Bayesian spatio-temporal models incorporating Conditional Autoregressive priors were applied, and competing models, including AR(1), AR(2), and a separated spatial structure, were compared using deviance information criterion (DIC) and Watanabe-Akaike information criterion (WAIC). A total of 4,390 BC cases were reported over the study period, with crude incidence rates rising from 36.8 to 60.8 per 100,000. Spatial clustering was particularly evident in urban areas. Significant associations were found for NO_2 and $PM_{2.5}$, with relative risks of 1.22 and 1.20, respectively. The findings highlight the importance of environmental monitoring and targeted interventions.*

Keywords: *Air pollutants, Bayesian conditional autoregressive models, Breast cancer, Monte Carlo Markov chain, Spatial autocorrelation, Spatio-temporal.*

1. INTRODUCTION

Breast cancer is the most frequently diagnosed cancer and the principal cause of cancer-related mortality among women worldwide. In 2022, an estimated 2.3 million new cases were reported globally, representing about 11.6% of all cancer diagnoses (Ferlay et al., 2024). The global burden is rising, particularly in low- and middle-income countries (LMICs), driven by demographic transitions as well as shifts in reproductive and lifestyle patterns.

In India, breast cancer has emerged as a significant public health concern. Between 1990 and 2016, the age-standardized incidence rate (ASR) increased by 39.1%, reflecting both improved detection capabilities and a real rise in risk exposures (Dhillon et al., 2018). Among Indian states, Kerala stands out for its elevated life expectancy, rapid urbanization, and superior cancer registration, and exhibits disproportionately high rates of breast cancer. For instance, in Thiruvananthapuram, Kerala's capital, the ASR for breast cancer has been reported at 35.6 per 100,000 women, with a crude incidence rate (CR) of 47 per 100,000 and an annual growth rate of approximately 5.1%. Moreover, breast cancer accounts for nearly 31% of all female cancers in this region, with approximately 35% of cases occurring in women under 50 years old suggesting early onset and potential influence from non-traditional or environmental risk factors (Mathew et al., 2016; NCRP, 2020).

A variety of established risk factors influence breast cancer incidence, including non-modifiable factors such as increasing age, family history, and pathogenic gene variants (e.g. BRCA1/2), as well as reproductive traits such as early menarche, late menopause, delayed childbirth, and curtailed breastfeeding. Equally important are modifiable lifestyle factors physical inactivity, overweight/obesity, alcohol intake, and poor dietary regimes that have been repeatedly linked to rising breast cancer risk (Faryabi et al., 2023). Beyond individual-level determinants, population-level factors such as socioeconomic status, urbanization, and environmental exposures are increasingly recognized as significant contributors to the spatial heterogeneity in breast cancer incidence (Song et al., 2023).

Ambient environmental pollutants, particularly those affecting air quality, have surfaced as plausible emerging risk factors for hormone-sensitive cancers including breast cancer. Pollutants such as NO_2, $PM_{2.5}$, PM_{10}, SO_2, NO_x, NH_3, ozone, and CO are under growing scrutiny due to their ubiquity and detrimental impacts on human health. While the respiratory and cardiovascular harms of air pollution are well established, an expanding body of evidence now hints at potential associations between chronic exposure to ambient air pollution and breast cancer risk, mediated via mechanisms like oxidative stress, systemic inflammation, endocrine modulation, and genotoxicity (Duboeuf et al., 2024; Le Provost et al., 2024; Wei et al., 2021; White et al., 2024). Despite this growing interest, uncertainty remains—particularly in LMIC contexts, where data are often sparse and long-term exposure histories less well characterised.

Due to the spatial and temporal complexity of both environmental exposures and disease incidence, advanced statistical methods are needed to untangle these relationships. Conventional regression models may overlook spatial dependency or temporal trends, leading to bias or reduced precision. In contrast, Bayesian spatio-temporal (BST) models, especially those employing the conditional autoregressive (CAR) prior, offer a robust alternative. CAR models adeptly capture spatial autocorrelation among geographically adjacent units by incorporating neighbourhood structures (Wah et al., 2020), while autoregressive temporal components (e.g. AR(1), AR(2)) smooth and model longitudinal trends (Rushworth et al., 2014; Lee et al., 2018). Additionally, frameworks that separate spatial risk surfaces by time period yet adjust for overarching temporal patterns so called separated spatial models allow nuanced estimation of evolving disease dynamics (Napier et al., 2016).

Such methodologies have been successfully applied in environmental epidemiology for mapping cancer risk, detecting high-risk clusters, and illuminating spatio-temporal trends often yielding deeper insights compared to static spatial or time-series analysis alone (Cramb et al., 2015; Raei et al., 2018). Applying these tools to the Indian context, particularly using district-level data, promises to inform targeted public health interventions and environmental policy.

Given this background, our study aimed to evaluate the association between long-term exposure to ambient air pollutants and breast cancer incidence in Thiruvananthapuram district, using Bayesian spatio-temporal conditional autoregressive modeling. We leveraged population-based breast cancer data from the Thiruvananthapuram Cancer Registry over a 15-year period (2005–2019).

2. MATERIALS AND METHODS

2.1 Study Area and Data Source

Thiruvananthapuram district, located in southern Kerala (approximately at 8°21′–8°40′ N, 76°47′–77°01′ E), covers around 307.55 km² and comprises 29 administrative units, with 83% urban population as per the 2011 Census of India. The district functions under the Population-Based Cancer Registry (PBCR) of the National Cancer Registry Programme, Indian Council of Medical Research (ICMR).

Breast cancer incidence data spanning 2005–2019 were extracted from the registry and aggregated into five three-year temporal intervals: 2005–07, 2008–10, 2011–13, 2014–16, and 2017–19. Population denominators for each administrative unit and time period were estimated using the difference distribution method, which utilized the 2011 census population data and interpolated forward and backward based on the observed growth rate from 2001 to 2011.

Ambient air pollution data including $PM_{2.5}$, PM_{10}, NO_2, SO_2, NO_x, NH_3, ozone (μg/m³), and CO (mg/m³) were obtained from Kerala State Pollution Control Board (KSPCB) and Central Pollution Control Board (CPCB) monitoring stations. Population density, operationalized as population divided by administrative area, was included as an additional covariate.

2.2 Spatial Statistical Methods

2.2.1 Spatial Structure and Interpolation

Administrative boundary shapefiles for the district were sourced from the DIVA-GIS database and projected in WGS 1984 UTM Zone 40N. Point-level pollutant concentration data were geocoded using Google Earth, and Inverse Distance Weighting (IDW) spatial interpolation was applied to estimate annual pollutant surface values for unmonitored locations (Wong et al., 2004). Neighbourhood structure for spatial modeling was defined using Queen contiguity, where administrative units sharing borders or corners were considered neighbours. A row-standardized adjacency matrix was constructed to capture spatial autocorrelation.

2.2.2 Exploratory Spatial Data Analysis (ESDA)

ESDA was performed separately for each three-year time period to assess changes in spatial clustering of breast cancer incidence over time. At the global level, Moran's I (Moran, 1950) was computed to detect overall spatial clustering, and Getis-Ord General G index (Getis and Ord, 1992) was used to identify whether clustering tended to be of high values (hot spots) or low values (cold spots). At the local level, Anselin Local Moran's I (Anselin, 1995) was calculated to identify significant local clusters and outliers, while Getis-Ord Gi* statistic (Getis and Ord, 1992) was used to map localized hotspots or cold spots of breast cancer incidence.

Additionally, bivariate Moran scatterplots were generated to examine spatial correlation between cancer incidence and neighbouring air pollution levels; both variables were row-standardized and standardized to mean zero and unit variance. Spatial overlay of interpolated pollutant surfaces and cancer clustering maps allowed visual assessment of potential ecological associations.

2.2.3 Space-Time Cluster Detection

To detect statistically significant space–time clusters, the Kulldorff's space–time scan statistic (Kulldorff and Nagarwalla, 1995) was applied using a discrete Poisson model, appropriate for count data. Given very small populations in certain administrative units, the scanning radius was defined in terms of population coverage rather than Euclidean distance. The most likely cluster was identified by maximum log-likelihood ratio (LLR), and secondary clusters were considered if they reached statistical significance. P-values for LLRs were estimated via 9,999 Monte Carlo simulations (Han et al., 2016).

2.2.4 Spatio-Temporal Modeling

The Poisson Generalized Linear Model (Poisson GLM) was initially used to assess the effects of air pollutants on BC incidence, without considering spatial or temporal random effects. To account for spatio-temporal autocorrelation, BST models were employed. These models incorporated random effects (ψ_{kt}) to capture spatial and temporal dependencies, with CAR priors used for smoother and more realistic estimates. The general formulation of spatio-temporal model is as follows:

$$Y_{kt} \sim Poisson(E_{kt}\theta_{kt}) \quad \text{for } k = 1, \dots, K,\ t = 1, \dots, T \qquad [1]$$
$$ln\ (\theta_{kt}) = X_{kt}^{c}\beta + \psi_{kt}$$

Where Y_{kt} denote disease count in a given area k during period t, E_{kt} is the expected count, calculated using internal standardization, X_{kt} represent the covariates and θ_{kt} is the underlying risk of disease in the given area k during the year t. For modelling cancer incidence, BST CAR models with various spatio-temporal structure of the random effects were used, including first order autoregressive (Rushworth et al., 2014) and second order autoregressive (Lee et al., 2018) and separated spatial models (Napier et al., 2016) (Table 1).

2.2.5 Model Comparison and Sensitivity Analysis

Model fitting was performed using Markov Chain Monte Carlo (MCMC) simulation with 2,200,000 iterations, discarding the first 200,000 as burn-in. The remaining samples were thinned by a factor of 1,000, yielding 6,000 posterior samples (2,000 per chain). Convergence was assessed using trace plots and the potential scale reduction factor. The posterior median and 95% credible interval (CI) were reported. Model fit was evaluated using the Deviance Information Criterion (DIC) and Watanabe-Akaike Information Criterion (WAIC) (Vranckx et al., 2021). Sensitivity analysis assessed three prior distributions: Inverse-Gamma (1, 0.01), (0.5, 0.005), and (0.001,0.001) (Jang et al., 2007).

3. RESULTS

3.1 Descriptive Epidemiology and Environmental Exposure

During the 15-year study period (2005–2019), a total of 4,390 breast cancer cases were recorded across the 29 administrative units of Thiruvananthapuram district. The crude incidence rates (CR) per 100,000 population demonstrated a consistent upward trend over the five consecutive three-year periods. Rates increased from 36.8 in 2005–2007 to 43.0 in 2008–2010, 47.8 in 2011–2013, 55.9 in 2014–2016, and peaked at 60.8 in 2017–2019. On average, over half of the administrative units reported CR values exceeding 50 during each time interval, indicating a widespread rise in incidence across the district.

Environmental exposure data, averaged over latency periods relevant to breast cancer development, revealed notable concentrations of ambient air pollutants. The mean levels of pollutants were: NO_2 21 µg/m³, SO_2 7.1 µg/m³, PM_{10} 54 µg/m³, $PM_{2.5}$ 72 µg/m³, NO_x 10.4 µg/m³, NH_3 2.53 µg/m³, ozone 24 µg/m³, and CO 0.4 mg/m³. The district's average population density was 303 persons per square kilometre. These exposures exhibited significant heterogeneity across regions, suggesting possible spatial variation in environmental risk.

3.2 Exploratory Spatial Data Analysis (ESDA)

Spatial distribution maps (Figure 1) illustrated substantial heterogeneity in breast cancer incidence across the district over time. In the early period (2005–2007), incidence patterns appeared dispersed, with few distinct clusters. From 2008–2010 onwards, emerging high-incidence zones were noted, particularly in central and coastal units. This trend intensified in subsequent periods, with the 2017-2019 map showing stable, well-defined clusters of elevated incidence, indicating persistent high-risk areas.

The Global Moran's I statistic confirmed increasing spatial autocorrelation across time. Moran's I rose from 0.13 ($Z = 1.40$, $p = 0.160$) in 2005–2007 to 0.45 ($Z = 4.14$, $p < 0.001$) in 2017–2019 (Table 2). Spatial dependence became statistically significant starting in 2008–2010 ($p < 0.05$), indicating that high-incidence areas were increasingly surrounded by similar high-risk units. Parallel findings were obtained with the Getis–Ord General G statistic, where Z-scores rose from 2.85 ($p = 0.004$) to 3.30 ($p = 0.001$), confirming persistent hot spots of incidence (Table 2).

Hotspot analysis using Getis–Ord Gi* (Figures 2a–2e) revealed that early hotspots (2005–2007) were limited and scattered (RR = 1.98). The number and intensity of hotspots grew during 2008–2010 (RR = 2.12) and stabilized into clusters in central and coastal zones during 2011–2013 (RR = 2.16) and 2014–2016 (RR = 2.49). By 2017–2019, several adjacent units formed statistically significant high-risk clusters (RR = 2.42; $p < 0.05$).

Cluster-outlier analysis using Anselin Local Moran's I (Figures 2f–2j) supported these findings, showing an increasing number of high-high clusters surrounded by similar high-risk units. Log-likelihood ratios (LLR) for these clusters were consistently high, further confirming statistical significance.

3.3 Air Pollution Distribution and Spatial Association

Spatial maps of average pollutant concentrations (Figure S1) revealed substantial heterogeneity in the distribution of NO_2, $PM_{2.5}$, SO_2, and PM_{10} across administrative units. Pollution hotspots largely coincided with urbanized and densely populated regions, overlapping with high-incidence cancer clusters.

Bivariate Moran's I scatterplots (Figure 3) assessed spatial correlation between NO_2 and lagged breast cancer incidence. In all five periods, a positive spatial correlation was observed, where regions with elevated NO_2 levels were surrounded by neighbouring areas with higher cancer incidence. The correlation strengthened over time: from a weak, emerging pattern in 2005–2007 (Figure 3a) to clear clustering in 2008–2013 (Figures 3b–3c), and more pronounced groupings in 2014–2019 (Figures 3d–3e). These findings suggest potential spillover effects, where pollutant exposures in one area influence risk in adjacent locations.

3.4 Space-Time Cluster Analysis

Kulldorff's space-time scan statistic identified a significant high-incidence period spanning 2014–2019 (LLR = 51.02, $p = 0.007$), during which breast cancer risk was 36% higher compared to other intervals (RR = 1.36). The most likely spatio-temporal cluster was localized in urban regions (Figure S2), exhibiting a relative risk (RR) of 2.10, indicating that women residing in these areas during this period had more than double the risk compared to those outside the cluster. This finding was supported by a highly significant LLR = 238.5 ($p = 0.0005$), confirming the presence of a persistent and intense spatio-temporal hotspot.

3.5 Bayesian Spatio-Temporal Model Estimates

Table 3 summarizes parameter estimates and 95% credible intervals (CI) for covariates from three Bayesian spatio-temporal models: AR(1), AR(2), and the separated spatial model. The separated spatial model also provided RR for covariates.

- NO_2 consistently showed a strong, positive association with breast cancer incidence across all models. Under the separated spatial model, the estimate was 0.47 (95% CI: 0.41–0.54), translating to an RR of 1.22 (95% CI: 1.11–1.34), indicating a 22% increased risk for each unit rise in NO_2.
- $PM_{2.5}$ was also significantly associated, with an estimate of 0.22 (95% CI: 0.16–0.28) (RR = 1.20, 95% CI: 1.08–1.32).
- PM_{10} showed a weaker and inverse association (estimate: 0.04; RR = 0.87, 95% CI: 0.81–0.93), suggesting a modest protective effect or potential confounding interaction.
- Population density was positively associated (estimate: 0.10; RR = 1.11, 95% CI: 0.98–1.25), although the CI overlapped unity, implying marginal statistical significance.
- Other pollutants (SO_2, NO_x, NH_3, CO, ozone) had weak or non-significant associations. For example, CO had an RR = 1.28 (95% CI: 0.73–2.24), indicating considerable uncertainty.

These results highlight NO_2 and $PM_{2.5}$ as dominant environmental risk factors for breast cancer in this region. Moreover, the spatial ($\rho = 0.42$) and temporal ($\alpha = 0.50$) dependence parameters were relatively

high, indicating the presence of both spatial and temporal autocorrelation, even after adjusting for covariate effects. Notably, the temporal autocorrelation (α) exceeded the spatial component (ρ), suggesting stronger temporal continuity in risk patterns across the years.

3.6 Model Fit Comparison

Model performance was evaluated using DIC and WAIC, where lower values indicate better fit. The separated spatial model outperformed AR(1) and AR(2) models, with the lowest DIC = 692.81 and WAIC = 676.70. AR(1) showed moderate fit (DIC = 696.20; WAIC = 684.77), while AR(2) performed worst (DIC = 725.20; WAIC = 726.16). The close DIC/WAIC values between AR(1) and the separated spatial model suggest that short-term temporal autocorrelation contributes, but spatial heterogeneity drives model accuracy.

Figure 4 maps posterior RR estimates across periods, revealing dynamic changes in spatial risk. Initially (2005–2007), most units had RR near or below 1. From 2008–2010, central and southern regions emerged as high-risk zones (RR > 1.1). By 2011–2016, these areas expanded and intensified, with some regions exceeding RR > 1.2. In 2017–2019, high-risk clusters became more concentrated and defined, reflecting persistent and localized risk elevation.

3.7 Sensitivity Analysis of Hyperpriors

Table 4 reports sensitivity analyses under three inverse-gamma priors: IG(0.5,0.0005), IG(1,0.01), and IG(0.001,0.001). Across all specifications, the separated spatial model consistently achieved the lowest DIC and WAIC, confirming its robustness. For example, under IG(0.5,0.0005), DIC was 693.76 and WAIC 677.90, outperforming AR(1) (DIC = 699.26) and AR(2) (DIC = 724.75). Even with more informative priors, the separated model maintained superiority. These findings indicate that results were not sensitive to prior choice, strengthening confidence in model conclusions.

4. DISCUSSION

This study provides compelling evidence of a growing spatial and temporal burden of BC in Thiruvananthapuram Kerala, with a significant association between disease incidence and ambient air pollution. Over the 15-year study period, crude incidence rates increased markedly, from 36.8 to 60.8 per 100,000 population, confirming an alarming upward trend. Spatial analyses demonstrated persistent and intensifying clusters in urban areas, confirmed by progressively increasing Moran's I and significant Getis–Ord Gi* statistics. These patterns suggest the influence of geographically structured risk factors, including environmental exposures and urbanization-related dynamics.

Consistently elevated breast cancer risks were observed in urban zones, where crude incidence rates exceeded 100 per 100,000 population during certain intervals. These findings align with earlier reports from the PBCR, which recorded a crude rate of 57.5 per 100,000 in 2016, and with previous research from the same region that identified disproportionately high incidence in urban areas (George and Mathew, 2016). The present study strengthens these observations by applying advanced spatio-temporal statistical methods to identify persistent high-risk clusters that closely correspond to areas of high population density and elevated pollutant exposure. Parallel trends have been reported for other cancers in this region. For instance, thyroid cancer showed high-risk clusters in coastal belts, with 56.9% of cases concentrated in these areas (George et al., 2023). This consistency across cancer types supports the hypothesis that urbanization and environmental exposures may jointly influence cancer patterns.

The findings reveal strong and statistically significant associations between breast cancer incidence and exposure to $PM_{2.5}$ and NO_2. The Bayesian spatio-temporal separated spatial model estimated relative risks of 1.20 for $PM_{2.5}$ and 1.22 for NO_2, indicating a meaningful elevation in risk with increased exposure. These results are consistent with previous studies in other contexts. For example, Hystad et al. (2015) in Canada reported a positive association between long-term NO_2 exposure and breast cancer, particularly among premenopausal women exposed to traffic-related pollution. Similarly, Goldberg et al. (2019) found that even low levels of NO_2 increased breast cancer risk, and Poulsen et al. (2023) in Denmark demonstrated a significant association between $PM_{2.5}$ exposure and increased incidence. These findings collectively suggest that fine particulate matter and nitrogen dioxide may play a role in breast cancer etiology.

The biological plausibility of these associations is supported by mechanistic studies. Airborne pollutants can induce oxidative stress, DNA damage, hormonal disruption, and chronic inflammation all of which are implicated in carcinogenesis (Darbre, 2018; Rodgers et al., 2018). However, not all studies have reported consistent associations. For example, Andersen et al. (2017) and Hart et al. (2016) did not observe clear links, possibly due to differences in exposure assessment, study design, or confounding

control. Such variability underscores the need for methodologically rigorous approaches, as employed in this study.

The present findings corroborate evidence from a recent investigation in the same district that reported a significant association between $PM_{2.5}$ and lung cancer, particularly among older males (>60 years) (Sruthi et al., 2024). This suggests that air pollution may have a broader impact on cancer incidence beyond site-specific risks. Moreover, the current results expand upon earlier pilot data (George and Mathew, 2016), which documented a concentration of breast cancer cases in urban localities, providing statistically robust evidence of spatial clustering and environmental risk factors using advanced modeling techniques.

A major novelty of this research lies in the application of BST CAR models to investigate breast cancer risk in India. To our knowledge, this is the first study in the country to apply such models for analysing the spatial association between breast cancer and air pollution. These models explicitly account for spatial autocorrelation (dependence between neighboring areas) and temporal correlation (dependence across time), reducing estimation bias and improving precision.

Among the three tested models AR(1), AR(2), and separated spatial, the separated spatial model performed best (DIC = 692.81; WAIC = 676.70), highlighting the dominant role of spatial heterogeneity in breast cancer distribution. Elevated spatial ($\rho = 0.42$) and temporal ($\alpha = 0.50$) dependence parameters confirmed persistent high-risk patterns and possible spatial spillover effects, where pollution exposure in one area may influence adjacent regions. This modeling approach aligns with international advancements. Louzada et al. (2021) and Tesema et al. (2023) have shown that Bayesian frameworks enhance risk estimation by borrowing strength across neighboring regions. Similarly, Napier et al. (2016) demonstrated that CAR models can incorporate dynamic spatial variability by using distinct neighbourhood matrices for different time periods. Yin et al. (2022) further improved this approach by allowing the model to estimate the neighborhood matrix itself, facilitating the detection of clusters and discontinuities. In the present study, these methodological strengths allowed for precise delineation of high-risk areas and robust estimation of pollutant effects.

The detection of persistent high-risk clusters in urban areas has significant public health implications. Urban environments in Kerala, like many developing regions, experience higher levels of vehicular emissions, industrial pollution, and population density, factors that may collectively elevate breast cancer risk. In addition to pollution, urbanization is associated with lifestyle factors such as lower physical activity, dietary changes, and reproductive behaviour, which may contribute to increased risk (Moss et al., 2017). Although these factors could not be directly analysed in this study due to data limitations, they warrant consideration in future research and prevention strategies.

While this study employs advanced statistical methods and provides high-resolution risk estimates, several limitations should be acknowledged. (1) data on Lack of individual-level risk factors such as reproductive and lifestyle variables (e.g., age at menarche, parity, obesity, alcohol consumption) were unavailable at the population level, preventing adjustment for potential confounders.(2) Ambient air pollution levels were derived from a limited number of monitoring stations and interpolated across space. Although interpolation techniques capture broad trends, they may not fully represent fine-scale variability, especially in under-monitored rural areas. (3) The analysis was confined to a relatively small region primarily encompassing urban zones. While this allowed for detailed spatial resolution, it limits generalizability to other settings. (4) Breast cancer development involves long latency periods. Although averaged pollutant data were considered, the exact exposure windows influencing disease onset remain uncertain. Despite these limitations, the consistency of findings across multiple analytical approaches supports the robustness of conclusions.

The study's strengths include (1) integration of spatial and temporal analysis within a Bayesian framework, minimizing bias and improving precision.(2)Use of multiple model comparisons (DIC, WAIC) to ensure selection of the most appropriate model.(3)Application of complementary spatial methods (Moran's I, Getis–Ord Gi*, Kulldorff's scan statistics), providing convergent evidence of clustering.(4)Incorporation of high-resolution environmental exposure data, enabling detailed assessment of pollutant–disease associations.

5. CONCLUSION

This study provides compelling evidence of a spatial and temporal relationship between ambient air pollution and breast cancer incidence in Thiruvananthapuram district, Kerala. By applying advanced Bayesian spatio-temporal models, we identified persistent high-risk clusters in urban areas and

demonstrated strong associations between long-term exposure to NO_2 and $PM_{2.5}$ with increased breast cancer risk.

The findings underscore the importance of continued cancer surveillance in pollution-prone urban areas and support the need for region-specific interventions to reduce exposure. Future research should aim to expand analyses to larger and more diverse geographic areas, incorporate individual-level risk factors to control for confounding, investigate biological mechanisms underlying pollutant-related carcinogenesis and evaluate the impact of environmental policies on cancer incidence. In conclusion, the integration of environmental data with advanced geostatistical modeling offers powerful insights into the environmental determinants of cancer, guiding targeted public health actions and contributing to the global understanding of air pollution's role in breast cancer etiology.

Acknowledgments: The authors express sincere gratitude to the Indian Council of Medical Research, Government of India (R. l l013/ 13/2021-GIA/HR) for the financial support for this study. We acknowledge NCDIR, ICMR, Government of India, for the financial support given to PBCR Thiruvananthapuram, Kerala State Pollution Control Board for the air pollution data, and Centre for Environment and Development (CED) for the technical support.

Funding details: Funding support was obtained from the Indian Council of Medical Research (R. l l013/ 13/2021-GIA/HR) and the National Centre for Disease Informatics and Research of ICMR for PBCR. The funding sources had no role in the design, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Ethical Approval: The study was approved by the Institutional Review Board of Regional Cancer Centre, Thiruvananthapuram (IRB No.12/2019/18,12/2021/17,12/2023/31).

Data and Codes Availability Statement

De-identified GIS data are stored in the Division of Cancer Epidemiology & Biostatistics, RCC, Thiruvananthapuram, and the authors of this team can access them. De-identified data of the present study may be shared on the basis upon requests to the corresponding author.

Disclosure Statement: The authors declare that they have no conflict of interest.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used Chatgpt in order to improve language and readability. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

6. REFERENCES

- Andersen ZJ, Ravnskjær L, Andersen KK, Loft S, Brandt J, Becker T, Ketzel M, Hertel O, Lynge E, Bräuner EV, 2017. Long-term Exposure to Fine Particulate Matter and Breast Cancer Incidence in the Danish Nurse Cohort Study. Cancer Epidemiol Biomarkers Prev 26(3),428-430; doi:10.1158/1055-9965.EPI-16-0578
- Anselin L, 1995. Local Indicators of Spatial Association—LISA. Geographical Analysis 27(2),93-115; doi:10.1111/j.1538-4632.1995.tb00338.x
- Cramb SM, Baade PD, White NM, Ryan LM, Mengersen KL, 2015. Inferring lung cancer risk factor patterns through joint Bayesian spatio-temporal analysis. Cancer Epidemiol 39(3),430-439; doi:10.1016/j.canep.2015.03.001
- Darbre PD, 2018. Overview of air pollution and endocrine disorders. Int J Gen Med 11,191-207; doi:10.2147/IJGM.S102230
- Dhillon PK, Mathur P, Nandakumar A, et al, 2018. The burden of cancers and their variations across the states of India: the Global Burden of Disease Study 1990–2016. The Lancet Oncology 19(10),1289-1306; doi:10.1016/S1470-2045(18)30447-9
- Duboeuf M, Amadou A, Coudon T, Grassot L, Ramel-Delobel M, Faure E, Salizzoni P, Gulliver J, Severi G, Mancini FR, Fervers B, Praud D, 2024. Long-term exposure to air pollution at residential and workplace addresses and breast cancer risk: A case-control study nested in the French E3N-Générations cohort from 1990 to 2011. European Journal of Cancer 210,114293; doi:10.1016/j.ejca.2024.114293
- Faryabi R, Daneshi S, Davarani ER, Yusefi AR, Arabpour M, Ezoji K, Movahed E, Daneshi-Maskooni M, Hussaini SM, 2023. The assessment of risk factors and risk perception status of breast cancer in Northern Iran. BMC Womens Health 23(1),268; doi:10.1186/s12905-023-02422-z
- Ferlay J, Ervik M, Lam F, Laversanne M, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I, Bray F, 2024. Global Cancer Observatory: Cancer Today. Lyon, France: International Agency for Research on Cancer. https://gco.iarc.who.int/today
- George PS, Mathew A, 2016. Spatial Distribution of Cancer in Trivandrum, India Using Geographic Information System: Pilot Study 9(3); doi:10.5829/idosi.ajcr.2016.9.3.10582
- George PS, Mathew A, Sruthi S, Kurup SA, Kattoor J, Vinod TR, 2023. Spatiotemporal pattern of thyroid cancer incidence in South Kerala: A population-based study. Indian J Public Health 67(2),240-246; doi:10.4103/ijph.ijph_1419_22

- Getis, A, Ord JK, 1992. The analysis of spatial association by use of distance statistics. Geographical Analysis 24(3),189–206; doi:10.1111/j.1538-4632.1992.tb00261.x
- Goldberg MS, Villeneuve PJ, Crouse D, To T, Weichenthal SA, Wall C, Miller AB, 2019. Associations between incident breast cancer and ambient concentrations of nitrogen dioxide from a national land use regression model in the Canadian National Breast Screening Study. Environ Int 133(Pt B),105182; doi:10.1016/j.envint.2019.105182
- Han J, Zhu L, Kulldorff M, Hostovich S, Stinchcomb DG, Tatalovich Z, Lewis DR, Feuer EJ, 2016. Using Gini coefficient to determining optimal cluster reporting sizes for spatial scan statistics. Int J Health Geogr 15(1),27; doi:10.1186/s12942-016-0056-6
- Hart JE, Bertrand KA, DuPre N, James P, Vieira VM, Tamimi RM, Laden F, 2016. Long-term Particulate Matter Exposures during Adulthood and Risk of Breast Cancer Incidence in the Nurses' Health Study II Prospective Cohort. Cancer Epidemiol Biomarkers Prev 25(8),1274-1276; doi:10.1158/1055-9965.EPI-16-0246
- Hystad P, Villeneuve PJ, Goldberg MS, Crouse DL, Johnson K, 2015. Canadian Cancer Registries Epidemiology Research Group. Exposure to traffic-related air pollution and the risk of developing breast cancer among women in eight Canadian provinces: a case-control study. Environ Int 74,240-248; doi:10.1016/j.envint.2014.09.004
- Jang MJ, Lee Y, Lawson AB, Browne WJ, 2007. A comparison of the hierarchical likelihood and Bayesian approaches to spatial epidemiological modelling. Environmetrics 18,809-821; doi:10.1002/env.877
- Kulldorff M, Nagarwalla N, 1995. Spatial disease clusters: detection and inference. Stat Med 14(8),799-810; doi:10.1002/sim.4780140809
- Lee D, Rushworth A, Napier G, 2018. Spatio-Temporal Areal Unit Modeling in R with Conditional Autoregressive Priors Using the CARBayesST Package. Journal of Statistical Software 84,1-39; doi:10.18637/jss.v084.i09
- Le Provost B, Parent MÉ, Villeneuve PJ, Waddingham CM, Brook JR, Lavigne E, Dugandzic R, Harris SA, 2024. Residential exposure to ambient fine particulate matter (PM2.5) and nitrogen dioxide (NO2) and incident breast cancer among young women in Ontario, Canada. Cancer Epidemiol 92,102606; doi:10.1016/j.canep.2024.102606
- Louzada F, Nascimento DC do, Egbon OA, 2021. Spatial Statistical Models: An Overview under the Bayesian Approach. Axioms 10(4),307; doi:10.3390/axioms10040307
- Mathew A, George PS, Arjunan A, Augustine P, Kalavathy M, Padmakumari G, Mathew BS, 2016. Temporal Trends and Future Prediction of Breast Cancer Incidence Across Age Groups in Trivandrum, South India. Asian Pac J Cancer Prev 17(6),2895-2899
- Moran P, 1950. Notes on Continuous Stochastic Phenomena. Biometrika 37:17-23
- Moss JL, Liu B, Feuer EJ, 2017. Urban/Rural Differences in Breast and Cervical Cancer Incidence: The Mediating Roles of Socioeconomic Status and Provider Density. Womens Health Issues 27(6),683-691; doi:10.1016/j.whi.2017.09.008
- Napier G, Lee D, Robertson C, Lawson A, Pollock KG, 2016. A model to estimate the impact of changes in MMR vaccine uptake on inequalities in measles susceptibility in Scotland. Stat Methods Med Res 25(4),1185-1200; doi:10.1177/0962280216660420
- National Cancer Registry Programme, 2020. Report of National Cancer Registry Programme. ICMR-NCDIR. https://www.ncdirindia.org/All_Reports/Report_2020/default.aspx
- Poulsen AH, Hvidtfeldt UA, Sørensen M, Pedersen JE, Ketzel M, Brandt J, Geels C, Christensen JH, Raaschou-Nielsen O, 2023. Air pollution with NO2, PM2.5, and elemental carbon in relation to risk of breast cancer- a nationwide case-control study from Denmark. Environ Res 216(Pt 3),114740; doi:10.1016/j.envres.2022.114740
- Raei M, Schmid VJ, Mahaki B, 2018. Bivariate spatiotemporal disease mapping of cancer of the breast and cervix uteri among Iranian women. Geospatial Health 13(1); doi:10.4081/gh.2018.645
- Rodgers KM, Udesky JO, Rudel RA, Brody JG, 2018. Environmental chemicals and breast cancer: An updated review of epidemiological literature informed by biological mechanisms. Environ Res 160,152-182; doi:10.1016/j.envres.2017.08.045
- Rushworth A, Lee D, Mitchell R, 2014. A spatio-temporal model for estimating the long-term effects of air pollution on respiratory hospital admissions in Greater London. Spat Spatiotemporal Epidemiol 10,29-38; doi:10.1016/j.sste.2014.05.001
- Song M, Huang X, Wei X, Tang X, Rao Z, Hu Z, Yang H, 2023. Spatial patterns and the associated factors for breast cancer hospitalization in the rural population of Fujian Province, China. BMC Women's Health 23(1),247; doi:10.1186/s12905-023-02336-w
- S S, Mathew A, K M JK, P RN, Sankar A, T R V, George PS, 2024. Assessment of spatial variation in lung cancer incidence and air pollutants: spatial regression modeling approach. Int J Environ Health Res 1-15; doi:10.1080/09603123.2024.2362844
- Tesema GA, Tessema ZT, Heritier S, Stirling RG, Earnest A, 2023. A Systematic Review of Joint Spatial and Spatiotemporal Models in Health Research. Int J Environ Res Public Health 20(7),5295; doi:10.3390/ijerph20075295
- Vranckx M, Neyens T, Faes C, 2021. The (in)stability of Bayesian model selection criteria in disease mapping. Spatial Statistics 43:100502; doi:10.1016/j.spasta.2021.100502
- Wah W, Ahern S, Earnest A, 2020. A systematic review of Bayesian spatial-temporal models on cancer incidence and mortality. Int J Public Health 65(5),673-682; doi:10.1007/s00038-020-01384-5
- Wei W, Wu BJ, Wu Y, Tong ZT, Zhong F, Hu CY, 2021. Association between long-term ambient air pollution exposure and the risk of breast cancer: a systematic review and meta-analysis. Environ Sci Pollut Res Int 28(44), 63278-63296; doi:10.1007/s11356-021-14903-5
- White AJ, Fisher JA, Sweeney MR, Freedman ND, Kaufman JD, Silverman DT, Jones RR, 2024. Ambient fine particulate matter and breast cancer incidence in a large prospective US cohort. JNCI: Journal of the National Cancer Institute 116(1), 53-60; doi:10.1093/jnci/djad170
- Wong DW, Yuan L, Perlin SA, 2004. Comparison of spatial interpolation methods for the estimation of air quality data. J Expo Anal Environ Epidemiol 14(5), 404-415; doi:10.1038/sj.jea.7500338
- Yin X, Napier G, Anderson C, Lee D, 2022. Spatio-temporal disease risk estimation using clustering-based adjacency modelling. Statistical Methods in Medical Research 31(6), 1184–1203; doi:10.1177/09622802221084131

Table 1. Summary of the spatio-temporal structure for the six models. W, spatial neighbourhood matrix; D, temporal neighbourhood matrix; τ^2, $\tau_1^2 \ldots, \tau_T^2, \tau_N^2$, variance parameters; ρ, spatial dependence term; α , α_1,α_2 , temporal dependence parameters.

	Structure	Prior
AR(1)	$\psi_{kt} = \phi_{kt}$ = Spatial effect for each time	$\phi_t \mid \phi_{t-1} \sim N(\alpha_1 \phi_{t-1}, \tau^2 Q(\rho, W)^{-1})$ $t\epsilon$ 2,3, … , T $\phi_1 = N(0, \tau^2 Q(\rho, W)^{-1})$ $\tau^2 \sim Inverse - gamma(1, 0.01)$ $\rho, \alpha_1 \sim U(0,1)$
AR(2)	$\psi_{kt} = \phi_{kt}$ = Spatial effect for each time	$\phi_t \mid (\phi_{t-1}, \phi_{t-2}) \sim N(\alpha_1 \phi_{t-1} + \alpha_2 \phi_{t-2}, \tau^2 Q(\rho, W)^{-1})$ $t\epsilon$ 3, … , T $\phi_1 and\ \phi_2 \sim N(0, \tau^2 Q(\rho, W)^{-1})$ $\tau^2 \sim Inverse - gamma(1, 0.01)$ $\rho, \alpha_1, \alpha_2 \sim U(0,1)$
Separated spatial	$\psi_{kt} = \phi_{kt} + \delta_t$ = Separate spatial effect(for each time)+temporal effect (for all areas)	$\phi_{kt} \mid (\phi_{-kt,} W) \sim N\left(\frac{\rho \sum_{j=1}^{K} w_{kj} \phi_{jt}}{\rho \sum_{j=1}^{K} w_{kj} + 1 - \rho}, \frac{\tau_t^2}{\rho \sum_{j=1}^{K} w_{kj} + 1}\right)$ $\delta_t \mid \delta_{-t} \sim N\left(\frac{\alpha \sum_{j=1}^{T} d_{tj} \delta_j}{\alpha \sum_{j=1}^{T} d_{tj} + 1 - \alpha}, \frac{\tau_N^2}{\alpha \sum_{j=1}^{T} d_{tj} + 1 - \alpha}\right)$ $\tau_1^2, \ldots, \tau_T^2, \tau_N^2 \sim \text{Inverse} - \text{gamma}(1, 0.01)$ $\rho, \alpha \sim \text{U}(0,1)$

Table 2. Global Moran's I and Getis-Ord General G Spatial Autocorrelation of Breast Cancer incidence during 2005-2019.

Year	Global Moran's I			Getis-Ord General G		High/Low Cluster
	Morans's I	Z Score	p-value	Z score	p-value	
2005-2007	0.13	1.40	0.160	2.85	0.004	High cluster
2008-2010	0.21	2.10	0.036	2.94	0.003	High cluster
2011-2013	0.27	2.54	0.011	4.98	0.0005	High cluster
2014-2016	0.35	3.28	0.001	2.78	0.005	High cluster
2017-2019	0.45	4.14	0.000	3.30	0.001	High cluster

Table 3. Parameter estimates with 95% CI from Bayesian Spatio-Temporal Models and RR from Separated Spatial Model.

Covariate	**AR(1) Estimate (95% CI)**	**AR(2) Estimate (95% CI)**	**Separated Spatial Estimate (95% CI)**	**RR Separated Spatial Model (95% CI)**
Population Density	0.10 (−0.02, 0.23)	0.08 (0.02, 0.16)	0.10 (0.03, 0.16)	1.11 (0.98, 1.25)
SO_2	0.04 (−0.08, 0.23)	0.02 (−0.06, 0.33)	0.01 (−0.07, 0.07)	0.75 (0.63, 0.80)

NO_2	0.48 (0.38, 0.58)	0.47 (0.36, 0.53)	0.47 (0.41, 0.54)	1.22 (1.11, 1.34)
PM_{10}	0.03 (−0.06, 0.21)	0.02 (−0.02, 0.20)	0.04 (−0.07, 0.21)	0.87 (0.81, 0.93)
$PM_{2.5}$	0.18 (0.08, 0.27)	0.20 (0.10, 0.27)	0.22 (0.16, 0.28)	1.20 (1.08, 1.32)
NO_x	0.02 (−0.11, 0.17)	0.04 (−0.03, 0.14)	0.01 (−0.08, 0.09)	1.02 (0.89, 1.18)
NH_3	−0.06 (−0.22, 0.09)	−0.11 (−0.24, 0.02)	−0.04 (−0.15, 0.08)	0.93 (0.80, 1.08)
CO	0.23 (−0.42, 0.81)	0.30 (−0.05, 0.65)	0.16 (−0.22, 0.58)	1.28 (0.73, 2.24)
Ozone	0.01 (−0.07, 0.09)	0.04 (−0.02, 0.10)	0.01 (−0.09, 0.10)	1.01 (0.93, 1.09)
Model Fit Statistics	**AR(1)**	**AR(2)**	**Separated Spatial**	
ρ	0.49	0.61	0.42	
α	-	-	0.50	
α_1	0.82	0.9	-	
α_2	-	0.12	-	
DIC	696.2	725.2	**692.81** (Lowest)	
WAIC	684.77	726.16	**676.70** (Lowest)	

Note: CI, credible interval; RR, Relative RiskAR(1), first-order temporal autoregressive model; AR(2), second-order temporal autoregressive model; RR, Relative Risk; DIC, Deviance Information Criterion; WAIC, Watanabe-Akaike Information Criterion; SO_2, Sulphur Dioxide; NO_2, Nitrogen Dioxide; PM_{10}, Particulate Matter <10μm; $PM_{2.5}$, Particulate Matter <2.5μm; NO_x, Nitrogen Oxides; NH_3, Ammonia; CO, Carbon Monoxide.

Table 4. Choice of hyperprior in sensitivity analysis.

	Choice of Hyperprior								
	Inverse-Gamma (0.5, 0.0005)			Inverse-Gamma (1, 0.01)			Inverse-Gamma (0.001, 0.001)		
	AR (1)	AR (2)	Separated spatial	AR (1)	AR (2)	Separated spatial	AR (1)	AR (2)	Separated spatial
DIC	699.26	724.75	693.76	696.20	725.20	692.81	696.13	725.52	693.56
WAIC	690.34	726.12	677.90	684.77	726.16	676.70	690.84	726.65	678.20

Note: AR(1), first order temporal autoregressive; AR(2), second order autoregressive; DIC, deviance information criterion; WAIC, Watanabe-Akaike Information Criterion.

Figure1. Spatial distribution of breast cancer incidence across 29 units in Thiruvananthapuram for the years (a) 2005–07, (b) 2008–10, (c) 2011–13, (d) 2014–16, and (e) 2017–19.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 8, 2025
https://theaspd.com/index.php

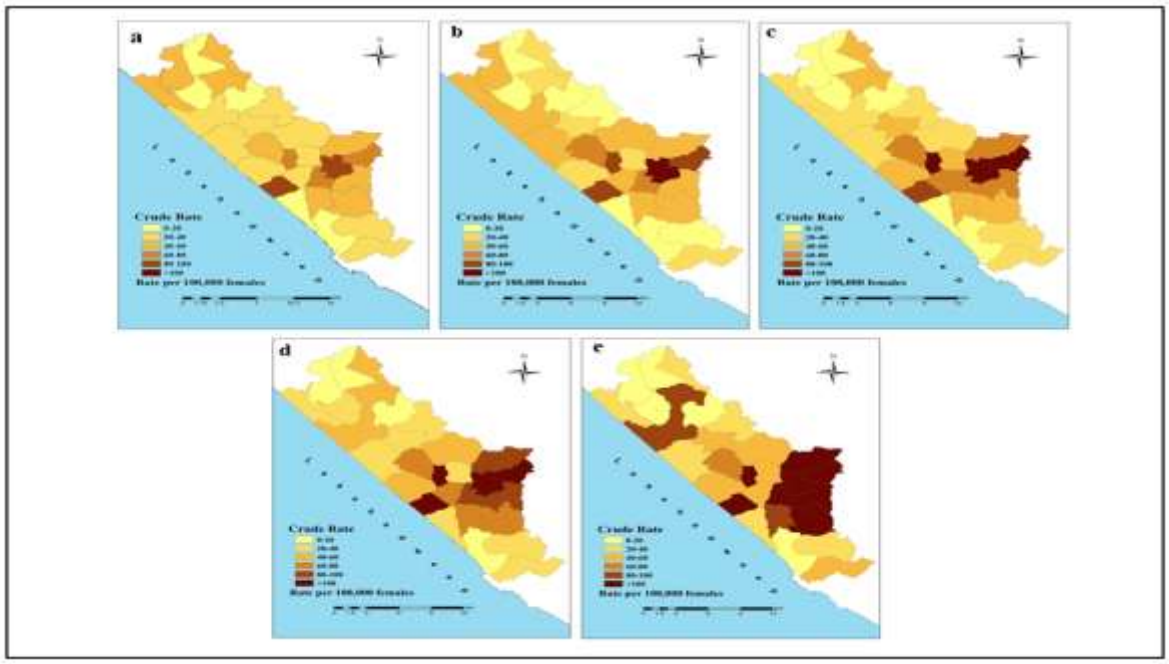

Figure 2. Hotspot analysis of breast cancer incidence in Thiruvananthapuram for the years (a) 2005–07, (b) 2008–10, (c) 2011–13, (d) 2014–16, and (e) 2017–18; and cluster & outlier analysis for the years (f) 2005–07, (g) 2008–10, (h) 2011–13, (i) 2014–16, and (j) 2017–18. LLR, Log-likelihood ratio; RR, Relative Risk.

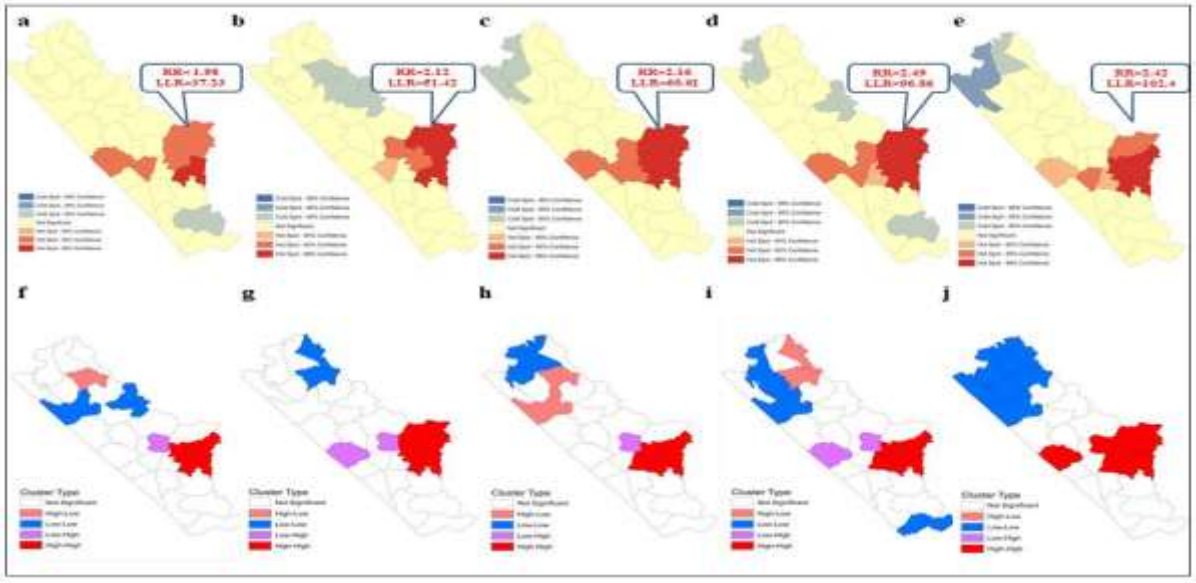

Figure 3. Spatial distribution of average concentrations of air pollutants across 29 units in Thiruvananthapuram for 2017–2019: (a) NO_2, (b) $PM_{2.5}$, (c) SO_2, and (d) PM_{10}.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 8, 2025
https://theaspd.com/index.php

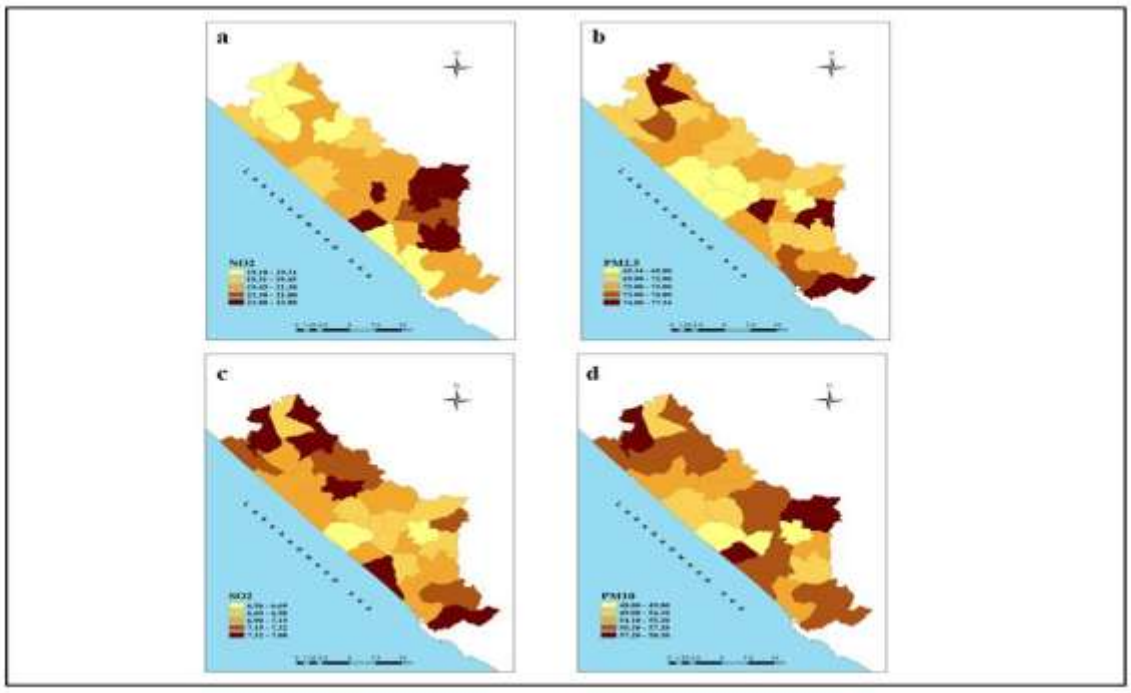

Figure 4. Bivariate Moran's I scatter plot of lagged breast cancer incidence rate vs NO_2 concentration for the years (a) 2005–07, (b) 2008–10, (c) 2011–13, (d) 2014–16, and (e) 2017–19 across 29 geographic regions of Thiruvananthapuram taluk.

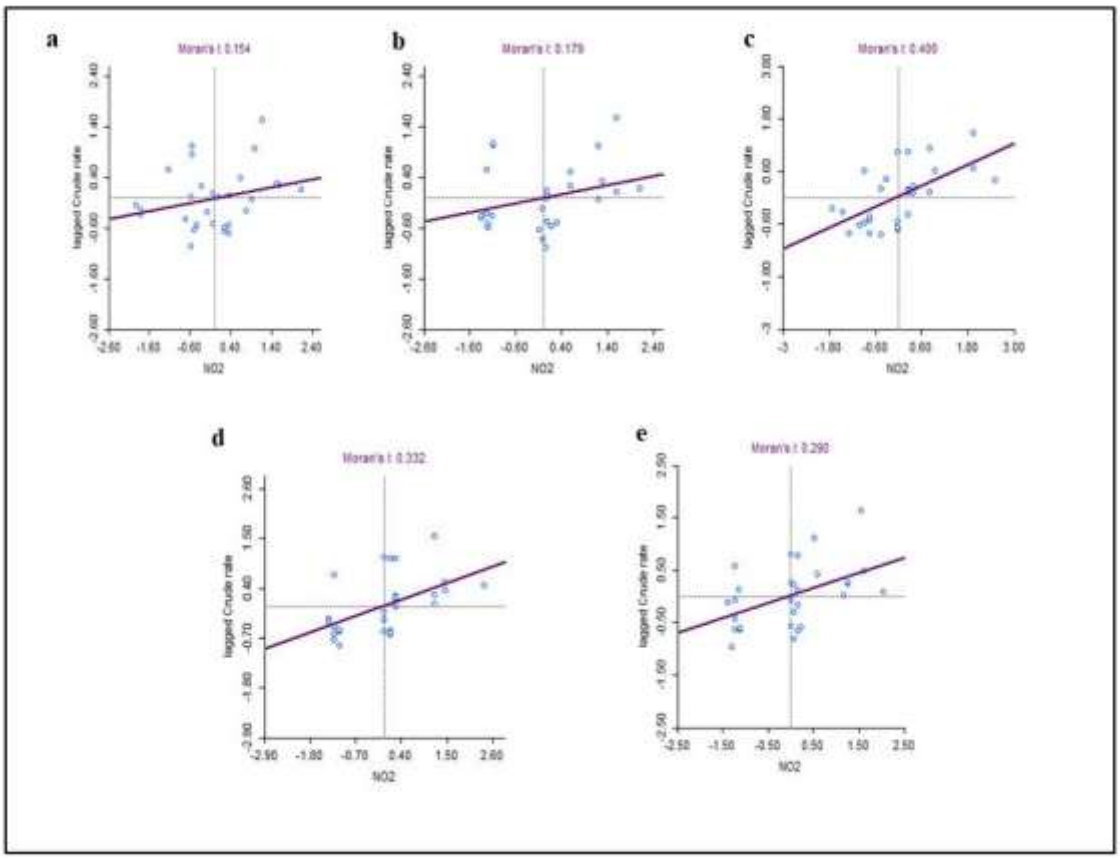

Figure 5. Spatio-temporal cluster analysis of breast cancer incidence across 29 units in Thiruvananthapuram using SaTScan. LLR: Log-likelihood ratio; RR: Relative Risk.

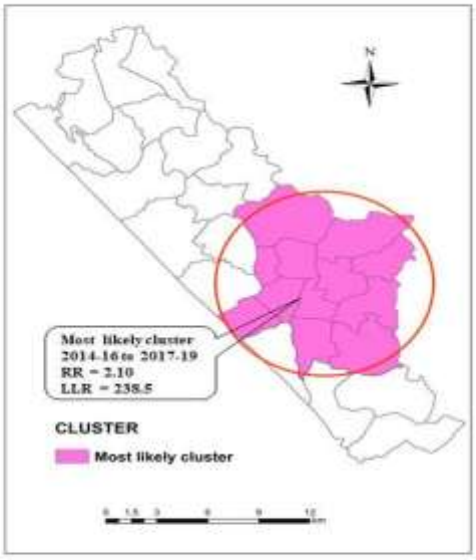

Figure 6. Relative risk estimates of breast cancer obtained using the Bayesian spatio-temporal Conditional Autoregressive (CAR) Separated Spatial model for the years (a) 2005–07, (b) 2008–10, (c) 2011–13, (d) 2014–16, and (e) 2017–19.