Proceedings of the Eleventh International Conference on Urban Pests 2025
AF-Borgen, Academic Society, Lund, Sweden
William H Robinson, editor

SPATIAL ANALYSIS OF BUILDING PEST DISTRIBUTION USING LAND USE DATA AND GIS

FUMI KIMURA, HIROKI MIYATA
Takenaka corporation

Abstract In Japan, architectural design preventing insect invasion into factories is highly desired due to high customer awareness of food safety. Insect contamination in food products is considered a serious incident that can jeopardize business management. Predicting the risk of insect invasion before construction allows us to improve design plans, such as site selection and loading dock locations, enhancing insect prevention. In this study, we attempted to evaluate the likelihood of insect invasion into the site by conducting light trap surveys and analyzing the data using a geographic information system (GIS). We conducted light trap surveys using LEDs at 10 sites in the Kanto region, Japan. We extracted land use information around each site by using GIS and linked the number of insects to land use ratio by regression analysis. Based on the obtained model, we estimated the number of insects arriving and visualized it on a map. In the light trap survey, 3,015 insects were captured, and the Chironomidae family was the most abundant, which accounted for 84% of the total. For this family, a model considering a 500-meter radius around each site and two latent variables explained the insect abundance most effectively. This variable was characterized by the proportion of rice fields. It was suggested that it was possible to estimate the amount of arriving insects quantitatively based on the surrounding land use ratios. Furthermore, visualizing the estimated amounts on a map indicated high-risk areas for insect invasion, suggesting that it can be applied to architectural planning.

Key words pest control, risk assessment, architectural planning, light trap, spatial analysis

Spatial analysis of insect pest distribution using land use data and GIS

○Fumi Kimura, Hiroki Miyata（TAKENAKA CO.）

Introduction

- ✓ Foreign matter contamination is the major factor in product quality deterioration for food and drug manufacturers. Especially in Japan, it causes serious damage to these manufacturers due to high customer awareness of food safety.
- ✓ Flying insects are highly mobile, so they frequently invade manufacturing facilities, causing foreign matter contamination.
- ✓ We think evaluating the extent of insect invasion before construction at premises allows us to improve design plans and avoid the contamination risk efficiently. Based on this concept, we finally conceived the GIS-based insect pest risk prediction system.

➢ In this study, models were developed to predict insect arrival numbers at premises based on surrounding land use information.

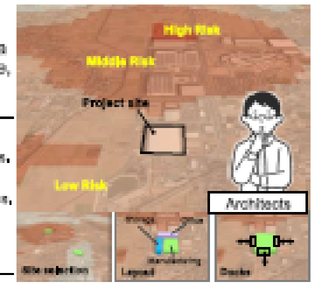
Conclusions

- ✓ We compared and evaluated methods for quantitatively estimating insect arrival numbers at premises by analyzing public data using GIS.
- ✓ It remains a challenge to collect data that accounts for variability and noise, and to improve model accuracy.

Buffer size	The models selected different distances. Multiple species may coexist, possibly with overlapping peaks for different insect groups.
Land Uses	Land uses generally known as sources for each insect group were selected as variables.
Temperature	Temperature influences capture numbers and may be creating noise in estimations based on land use parameters.

Prospects

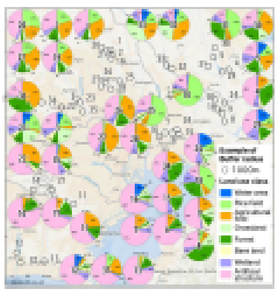
- Development of risk visualization tools for architectural schematic design.
- Development of more micro-scale models incorporating 3D obstacles.



Results

1. Surrounding landscape data

Land use data were obtained from public sources[*1], and area proportions within radius buffers[*3] around survey points were calculated. Models were created for each radius. Land use composition varied by location, with land use types increasing with larger radii.



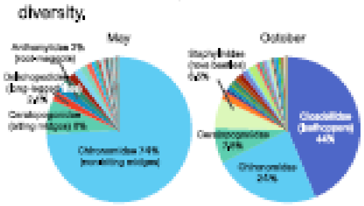
Land use area proportion within 1000m radius of survey points

2. Trap survey

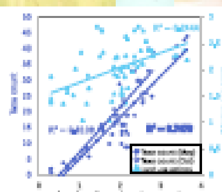
During two surveys, 21,135 insects were captured, with Chironomidae dominating May (74%) and Cicadellidae October (44%). Chironomidae commonly emerge from water and soil, while Cicadellidae are known as agricultural pests in rice and crop fields. Insect numbers correlated with taxonomic richness, not with land use diversity.

Capture results and examples of insects

	May	October
Number of taxa	Order 8 Family 44	Order 12 Family 59
Total insect	2,705	18,428
Numbers per site	2–839	2–4,482



Proportion of captured insects by taxonomic group



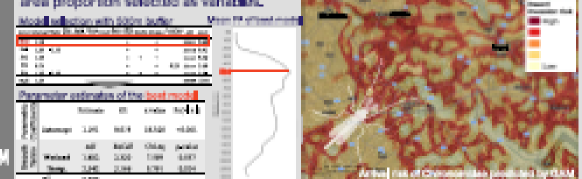
Correlation between capture numbers and taxonomic richness / land use diversity

3. Statistical analysis

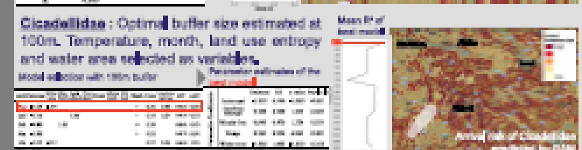
Parameter estimation and capture number prediction were performed for Chironomidae and Cicadellidae using GAM and RF models. Results were visualized as risk maps showing predicted insect arrivals at any location.

GAM

Chironomidae : Optimal buffer size estimated at 500m. Temperature and wetland area proportion selected as variables.



Cicadellidae : Optimal buffer size estimated at 100m. Temperature, month, land use entropy and water area selected as variables.



RF

Chironomidae : Optimal buffer estimated at 300m. Temperature, forest, and rice fields were identified as key variables. R² was slightly higher than GAM, with lower predicted capture numbers.

R^2	0.393
Number of variables	7

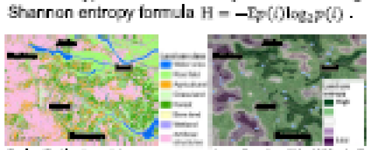
Cicadellidae : Optimal buffer estimated at 400m. Temperature, month and rice fields were identified as key variables. R² was slightly higher than GAM, with higher predicted capture numbers.

R^2	0.508
Number of variables	8

Materials & Methods

1. Surrounding landscape data

Land use types from the public data[*1] were reclassified from the original 14 categories into 8 categories[*2]. Buffers with radii of 50-3000m[*3] were created around trap locations. For each buffer, we calculated the proportion of each land use type and the diversity evaluated using Shannon entropy formula $H = -\Sigma p(i)\log_2 p(i)$.



Reclassified land use data　　Land use diversity within 1000m buffer

*1　Land use Map (Geospatial Information Authority of Japan)
*2　'Water area', 'Rice field', 'Agricultural field', 'Grassland', 'Forest', 'Bare land', 'Wetland' and 'Artificial structure'.
*3　50m, 100m, 200m, 300m, 400m, 500m, 600m, 700m, 800m, 900m, 1000m, 1100m, 1200m, 1300m, 1400m, 1500m, 2000m, 2500m, 3000m

2. Trap survey

Trap surveys were conducted at 23 sites in May 2024 and 37 sites in October (+ new 14) in Kanto region, using simple light traps.
Each trap consisted of white boards with LED lights (40W-equivalent) and sticky papers on both sides. Lights operated for 3 hours within 17:00-21:00, including sunset time.
Captured specimens were identified to family level when possible, or to order level if identification was not feasible[*4].

*4 Ephemeroptera, Plecoptera, Diptera, Trichoptera, Lepidoptera : total of 163 specimens

Survey sites　　Simple Light Trap

3. Statistical analysis

Parameter estimation and capture number prediction were conducted for the most abundant taxonomic groups using Generalized Additive Models (GAM) and Random Forest (RF).

Common variable settings		Explanation
Response variable	Log-transformed capture count	Value obtained by log-transforming the number of captured specimens.
Explanatory variables	Land use proportion	Proportion of each land use type within the buffer.
	Temperature	Mean temperature at the time of survey (27°C used for prediction).
	Month	Categorized into two categories: May and Oct (Oct used for prediction).
	Land use entropy	Land use diversity index (richness and evenness) calculated using H=−Σp(i)log₂p(i).

Generalized Additive Model (GAM)

Probability distribution	Negative binomial distribution	Link function	log
Full model	log(capture count) ~ Σ s(Land use proportion*) + s(Temperature) + s(Land use entropy) + Month *Water, Rice fields, Agricultural field, Grassland, Forest, Bare land, and Wetland		

- ➢ For each buffer size, variables were selected via stepwise method to minimize AIC.
- ➢ K-fold cross-validation (CV) of best models provided R² and mean predicted captures.
- ➢ The model with highest mean R² across buffer sizes was adopted.

Random Forest (RF)

Decision trees in Random Forest	1,000
Number of variables	Number of variables minimizing OOB error rate.

➢ For each buffer size, models were developed and evaluated using CV, selecting the one with highest mean R² across all buffer size.