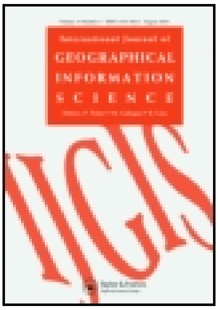


International Journal of Geographical Information Science
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tgis20

# Estimating the spatial pattern of human-caused forest fires using a generalized linear mixed model with spatial autocorrelation in South Korea

Hanbin Kwak [a] , Woo-Kyun Lee [a] , Joachim Saborowski [b] , Si-Young Lee [c] , Myoung-Soo Won [d] , Kyo-Sang Koo [d] , Myung-Bo Lee [d] & Su-Na Kim [a]

[a] Department of Environmental Science and Ecological Engineering, Korea University, Seoul, Republic of Korea

[b] Chair of Ecoinformatics, Biometrics and Forest Growth and Chair of Ecosystem Modelling, Georg-August-University Göttingen, Göttingen, Germany

[c] Department of Disaster Prevention and Safety Engineering, Kangwon National University, Samcheck-si, Gangwon-do, Republic of Korea

[d] Division of Forest Disaster Management, Korea Forest Research Institute, Seoul, Republic of Korea

Available online: 20 Feb 2012

*International Journal of Geographical Information Science*
i*First*, 2012, 1–14

Taylor & Francis
Taylor & Francis Group



# Estimating the spatial pattern of human-caused forest fires using a generalized linear mixed model with spatial autocorrelation in South Korea

Hanbin Kwak[a], Woo-Kyun Lee[a]*, Joachim Saborowski[b], Si-Young Lee[c], Myoung-Soo Won[d], Kyo-Sang Koo[d], Myung-Bo Lee[d] and Su-Na Kim[a]

[a]*Department of Environmental Science and Ecological Engineering, Korea University, Seoul, Republic of Korea;* [b]*Chair of Ecoinformatics, Biometrics and Forest Growth and Chair of Ecosystem Modelling, Georg-August-University Göttingen, Göttingen, Germany;* [c]*Department of Disaster Prevention and Safety Engineering, Kangwon National University, Samcheck-si, Gangwon-do, Republic of Korea;* [d]*Division of Forest Disaster Management, Korea Forest Research Institute, Seoul, Republic of Korea*

(*Received 24 January 2011; final version received 16 November 2011*)

Most forest fires in Korea are spatially concentrated in certain areas and are highly related to human activities. These site-specific characteristics of forest fires are analyzed by spatial regression analysis using the R-module generalized linear mixed model (GLMM), which can consider spatial autocorrelation. We examined the quantitative effect of topology, human accessibility, and forest cover without and with spatial autocorrelation. Under the assumption that slope, elevation, aspect, population density, distance from road, and forest cover are related to forest fire occurrence, the explanatory variables of each of these factors were prepared using a Geographic Information System-based process. First, we tried to test the influence of fixed effects on the occurrence of forest fires using a generalized linear model (GLM) with Poisson distribution. In addition, the overdispersion of the response data was also detected, and variogram analysis was performed using the standardized residuals of GLM. Second, GLMM was applied to consider the obvious residual autocorrelation structure. The fitted models were validated and compared using the multiple correlation and root mean square error (RMSE). Results showed that slope, elevation, aspect index, population density, and distance from road were significant factors capable of explaining the forest fire occurrence. Positive spatial autocorrelation was estimated up to a distance of 32 km. The kriging predictions based on GLMM were smoother than those of the GLM. Finally, a forest fire occurrence map was prepared using the results from both models. The fire risk decreases with increasing distance to areas with high population densities, and increasing elevation showed a suppressing effect on fire occurrence. Both variables are in accordance with the significance tests.

**Keywords:** word; forest fire; spatial statistics; variogram; GLMM

## 1. Introduction

Forest fires are a major disaster, damaging both the forest ecosystem and the human society. Fire threat in Korea is very strong because about 65% of the Korean peninsula is covered by forest. All Korean forest fires are caused by humans as there are no lightening-induced

*Corresponding author. Email: leewk@korea.ac.kr

ISSN 1365-8816 print/ISSN 1362-3087 online

http://dx.doi.org/10.1080/13658816.2011.642799
http://www.tandfonline.com

natural fires because of the condition of the Korean meteorology. Because all forest fires in South Korea are closely related to human activities, fires tend to occur repeatedly in environments closer to the metropolitan areas according to the pattern of population distribution. The most frequent reason for fires is the accidental fires started by humans, followed by fires by field incineration (Kwak *et al.* 2009). To prevent forest fires caused by human activities, it is important to determine the related factors and the extent of their influence on the forest fire occurrence.

Many factors influence forest fire occurrence. Previous studies have reported weather, vegetation types, and topography (elevation, slope, and aspect) to be significant factors in the occurrence of forest fires (Kushla and Ripple 1997, Diaz-Avalos *et al.* 2001, Latham and Williams 2001, An *et al.* 2004, Wotton and Martell 2005). The moisture content of fuel was reported to influence forest fire occurrence (Renkin 1992, Chuvieco *et al.* 2002, Wotton *et al*. 2003). Forest fires have also been reported to be related to human residence and population density (Veblen *et al.* 1999, Guyette *et al*. 2002, Bergeron *et al.* 2004, Hessburg *et al*. 2005). Human accessibility is also a main cause for forest fires (Cardille *et al*. 2001, Prestemon and Butry 2005). Because almost all forest fires in Korea are started by human activities (Won *et al*. 2006, Kwak *et al.* 2010), the factor of human accessibility should be included in the analysis of forest fire occurrence.

Spatial count data arise in many situations in epidemiology, ecology, and agriculture (Zhang 2002). Typical methods of statistics are limited in their ability to detect specific relationships, including spatial autocorrelations. Therefore, additional techniques are required, and handling count data with spatial autocorrelation is becoming important for ecology and forestry. To analyze spatial autocorrelation, geostatistical methods are well known for point-based data, and hybrid methods such as regression-kriging, which is a combined approach with ordinary least squares regression and kriging, have recently been adopted (Oliver 1990, Stein 1999, Hengl *et al.* 2004, 2007). In statistical research, several generalized linear model (GLM) and generalized linear mixed model (GLMM) based on different assumptions have been proposed to account for spatial autocorrelation (Diggle *et al.* 1998, Christensen and Waagepetersen 2002, Venables and Ripley 2002, Zhang 2002). These methods are used in a variety of research fields, not only forest fires but also epidemiology, social geography, and remote sensing (Curran 1988, Anselin 1992, Rezaeian *et al.* 2007).

## 2. Study site

The whole of South Korea was considered as the study site. The eastern part of the Korean peninsula is highly elevated with many steep mountains, dense forest stands, and high concentrations of *Pinus densiflora* stands that are vulnerable to forest fires. Large forest fires have recently occurred in this region. The human population density in these mountainous regions is relatively low. On the other hand, most mega cities with high population densities are located in the western and southern parts of the peninsula. In this urbanized region, there are few large forest fires, but the frequency of forest fire occurrence appears higher than that of the mountainous eastern area (Figure 1)

## 3. Material and methods

### *3.1. Forest fire data*

The forest fire occurrence history from 1991 to 2008 was collected by the Korea Forest Service (KFS). This daily fire information contains time and location of fire occurrences.

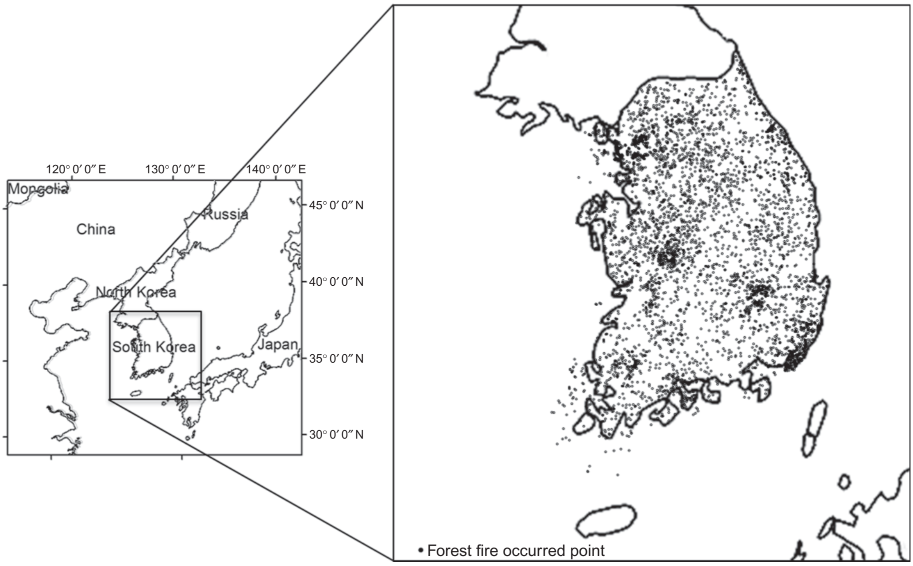


Figure 1. Study area showing the original points of forest fire occurrence in South Korea collected from the Korea Forest Research Institute.

The location was geo-coded with spatial coordinates using a land registry map from the Korea Cadastral Survey Corporation. The point data were converted to fire counts per 5 km quadrats (Figure 2). The quadrat size was fixed at 5 km to minimize the error arising from the conversion of point to quadrat counts because the average of the polygon area of the land registry map that was used for geo-coding was approximately 25 km$^2$. After that, the mean values of all external variables consisting of topographical factors, human accessibility, and vegetation types within the quadrats were also attached to each quadrat.

For the quadrat, forest area is a very important factor affecting forest fires. The ratio of forest area to quadrat size varies among quadrats. In Poisson regression, it is possible to consider such ratio data by either space or time unit, which is handled by an ‘offset’ variable (Christiansen and Morris 1997). Here, we used the forest area per quadrat as an offset term. The forest area map was determined by the forest-type map from KFS. Every quadrat had some forest area, the smallest of which was 796.261 m$^2$.

Totally, 4474 grids covered the whole study site of South Korea. For model estimation, half of the grids, 2237, were randomly selected. The others were used for model validation.

### *3.2. Explanatory variables*

We focused on the spatial, not temporal, characteristics of forest fire occurrence. We presumed that the important factors influencing forest fire occurrence are related to the following three major sources: topography, forest cover, and human accessibility.

Topographical factors such as elevation, slope, and aspect are known to influence the original ignition of a forest fire and its subsequent spread speed and direction (Rothermel 1972, Richards *et al*. 1999, Lee *et al.* 2004a). We presumed that these factors also can affect human accessibility. Topographical data were prepared using the digital maps with a scale of 1:25,000 issued from the National Geographic Information Institute (NGII) of Korea. The elevation from the DEM (Digital Elevation Model), aspect, and slope maps

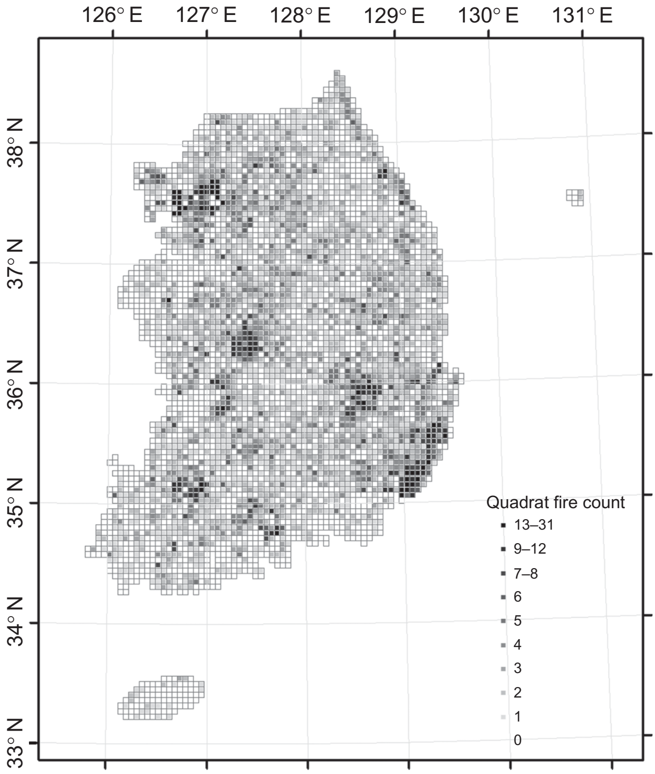


Figure 2. Forest fire distribution by 5 km quadrat counts in South Korea. The grey areas indicate increased fire occurrence.



were derived using the spatial analyst module of ESRI® ArcGIS™. Particularly, the aspect was converted to the aspect index (AI) with a value ranging from 0 to 2, using the equation below, where 0 means the southern and 2 the northern aspect (Lee *et al.* 2004b). In this research, we assumed that the aspect explains the solar insolation and the human accessibility. In South Korea, there are many tombs with southerly aspect as this is traditionally favored. Moreover, people access the tombs more in spring season (which is also the main fire season) to manage them. This situation increases the forest fire occurrence. Thus, we supposed that the division of north and south aspects is very important. Aspect variables were not classified by categorical variables in this research; instead, they were converted to a continuous index. Because the aspect data are originally continuous in the range of 0–360°, continuous characteristics can be conserved by using this index.

$$\text{Aspectindex} = 2\left|1 - \frac{(\text{Aspect})}{180}\right|$$

Originally, the flat areas were calculated as $-1$ in the ArcGIS™ module. However, this is not within the range of AI. So, all flat areas were replaced by 1 because the meaning of flat is between the extremes, northern and southern, in that AI.

Human accessibility is also known to affect forest fire occurrence. Fire ecologists and historians have found that the fire regimes of many forest ecosystems are anthropogenic and shaped largely by human settlement and management (Yang *et al.* 2007). We presumed that indirect factors such as population density and distance from road are related to forest fire occurrence. Because measuring human accessibility quantitatively was very difficult, the factors of human accessibility were measured indirectly using the distance from roads and population density as proxy variables (Yang *et al.* 2007, Romero-Calcerrada *et al.* 2008).

The average population density from 1991 to 2008, the same period as that of the forest fire data, was used for the estimation. Population data were collected by the government in the form of polygon shapes with discontinuities at the edges. To overcome this problem, the discontinuous population density was smoothed by kernel density estimation (KDE) (Silverman 1981, Diggle 1985). The centroids of the polygons containing population density values were used to perform KDE. We used a fixed smoothing parameter (bandwidth) which was defined as 18,325 m, as determined by visual choice.

The road map was extracted from the digital map of NGII. The complete road network is very complex with very small spaces between the roads which made measuring from the main road to each grid cell excessively difficult. Therefore, only the expressways and city roads were used in the analysis. The distances from the roads were calculated by the Euclidian distance module of the ArcGIS™ spatial analyst with 30 m resolution raster. Finally, these raster values of distance from road within each 5 km quadrat were averaged in order to get one representative value for each quadrat as an independent variable.

Vegetation types were extracted from the fourth forest-type map of KFS. We rearranged the vegetation types into five dummy variables: needle leaf forest, broad leaf forest, mixed forest, grassland, and other. Grassland includes cultivated land, pasture, and fruit garden. In the winter, illegal field incineration is the cause for many fires in Korea. Other includes all miscellaneous types that cannot be assigned to one of the other four types. This class was usually distributed near cities and army communities.

### *3.3. Spatial regression model*

Spatial data, particularly those collected on a systematic grid as our fire count data, are usually assumed to be autocorrelated (Anselin 1988, Legendre 1993). Observations are often more similar to others located nearby than to more distant observations. This is usually assumed to be the effect of unobserved covariates. It is expressed in the so-called first law of geography: 'Everything is related to everything else, but near things are more related than distant things' (Tobler 1979). Therefore, a model aimed at the description of the effects of covariates on a regionalized dependent variable should take spatial autocorrelation into account.

Recent statistical approaches to the analysis of forest fire occurrence in Korea (An *et al.* 2004, Lee *et al.* 2004a) have not considered that phenomenon. This is also true for many other studies using multivariate regression analysis with logistic models to estimate the effects of drivers of forest fire risk (Martell *et al*. 1987, Garcia *et al.* 1995, Pew and Larsen 2001). Recently, spatial point pattern analysis with a log-linear regression model was used to explain the pattern of forest fire occurrence (Yang *et al.* 2007), where the dependent variables were smoothed using KDE. The study area is relatively small and spatial autocorrelation is not incorporated.

Although logistic regression is used in many case studies to predict the probability of fire events assuming binomially distributed binary data, the structure of the Korean fire count data suggests using a Poisson model approach. The number of fire events, $y_i$, counted in a quadrat $i$ is explained by the model

$$y_i = \exp(x_i\beta) + \varepsilon_i \qquad i = 1, \dots, n \tag{1}$$

where $n$ is the total number of quadrats, $x_i$ the vector of covariate values attached to quadrat $i$ (the first component equals 1 if an intercept is to be included in the linear predictor), $\beta$ the vector of regression coefficients, and $\varepsilon_i$ a random error term. The number of fire events $y_i$ is assumed to follow a Poisson distribution with local expectation $\mu_i = \exp(x_i\beta)$, variance $v(y_i) = \mu_i$, and the linear predictor $\eta_i = x_i\beta$. With the Korean fire count data, we used an additional offset factor $\log(f_i)$ to multiply the linear predictor, where $f_i$ is the forest area in quadrat $i$. If the covariance matrix, $C(y)$, equals $V_\mu$, which is an (n,n)-matrix with the variances $v(y_i)$ on the main diagonal and zeroes as off-diagonal elements, we have a classical log-linear Poisson model, a GLM with independent errors and inhomogeneous error variances. This can be used as a simplified approach, where we also consider overdispersion by an additional scale factor, and finally compare it to the according model with an autocorrelated error structure.

Spatial autocorrelation can be introduced into model (1) in two different ways (see e.g. Schabenberger and Gotway 2005), either by an additive random effect in the linear predictor $\eta_i$ leading to a GLMM, called G-side effect in SAS PROC GLIMMIX or by a modification of the covariance matrix of the residuals, called R-side effect in GLIMMIX. Using R($\theta$) as a correlation matrix, with $\theta$ being a low-dimensional parameter vector, $\sigma^2$ as an overdispersion parameter, and $c_0$ as a nugget effect parameter, the new covariance matrix is

$$C(y) = c_0 V_\mu + \sigma^2 V_\mu^{1/2} R(\theta)\, V_\mu^{1/2} \tag{2}$$

Thus, $R(\theta)$ is the correlation function of the Pearson residuals

$$r_{Pi} = \frac{y_i - \mu_i}{\sqrt{\mu_i}}$$

Since PROC GLIMMIX failed to estimate a plausible autocovariance structure close to the empirical variogram of the GLM residuals of the fire count data, we followed the advice of Dormann *et al.* (2007, appendix) and used the R-function glmmPQL {MASS} with a constant grouping variable that assigns all observations (quadrats) to only one group or subject (abbreviated GLMM in this article). It simultaneously estimates the regression coefficients of the linear predictor and the variogram parameters of nugget, sill, and range. Even in cases where a nugget effect can be excluded because of general considerations, it may still be helpful to include it into the model to avoid numerical instability (Dormann *et al.* 2007).

The empirical Pearson residuals

$$\hat{r}_{Pi} = \frac{y_i - \exp(x_i\hat{\beta})}{\sqrt{\exp(x_i\hat{\beta})}}$$

of GLM can be used to estimate an empirical variogram

$$\hat{\gamma}(h) = \frac{1}{2N_h}\sum_{j=1}^{N_h}(\hat{r}(s_j) - \hat{r}(s_j + h))^2$$

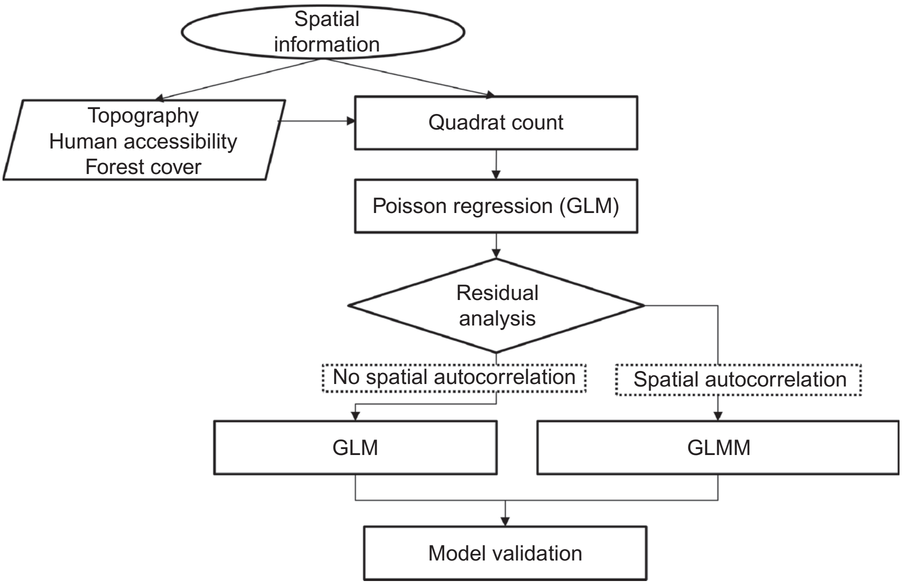


Figure 3. Process for predicting forest fire occurrence with GIS-based quadrat count data using GLM and GLMM.

and fit a variogram model in order to initialize the GLMM estimation procedure and to compare it to the covariance parameters estimated by the R-module as a plausibility control (Figure 3).

Here, the linear predictor $\eta_i = \log(\mu_i)$ comprises three groups of covariates

$$\eta_i = \beta_0 + \beta_p' \text{ topography} + \beta_q' \text{ accessability} + \beta_r' \text{ forest cover}$$

in our application to fire count data, with $p = 3$ topographical (*mean AI, elevation, and slope)*, $q = 2$ human accessibility (*population density and distance from road*) and $r = 4$ forest cover covariates (*broad leaf and mixed forest, grassland, and other; needle leaf forest serves as a reference category*) and a constant at the beginning.

## 4. Results and discussion

### *4.1. Parameter estimation in GLM and GLMM*

The coefficients of GLM and GLMM were estimated as shown in Table 1. The insignificant variables such as grassland and other were dropped in the final model. The population density and elevation were the best explanatory variables with the lowest standard error and *P*-value. Population density, distance from road, elevation, and AI showed similar results as in previous research (Pew and Larsen 2001, Yang *et al.* 2007). The fact that most of the forest fires occurrences in South Korea are a result of human impacts explains the positive relationship between population density and fire occurrence. The distance from the road was also highly significant. These results demonstrated the strong influence of human accessibility on forest fire occurrence.

Slope showed a positive sign. Many previous studies have reported that slope affects fire spread and occurrence (Chuvieco and Salas 1996, Yang *et al.* 2007, Beaty and Taylor

Table 1. Estimated coefficients, standard error, and significance level of independent variables using GLM and GLMM to predict forest fire occurrence in South Korea.

| | GLM | | | GLMM | | |
|---|---|---|---|---|---|---|
| Variables | Coefficient | Std. error | *P*-value | Coefficient | Std. error | *P*-value |
| (Intercept) | −1.206 | 0.1104 | $<2\times10^{-16}$ | −1.417 | 0.1297 | $<2\times10^{-16}$ |
| Population density (KDE) | 1.888 | 0.06736 | $<2\times10^{-16}$ | 1.7579 | 0.1223 | $<2\times10^{-16}$ |
| Distance from road | −0.0296 | 0.01109 | 0.008 | −0.0409 | 0.01428 | $4\times10^{-3}$ |
| Elevation | −0.002731 | 0.0002139 | $<2\times10^{-16}$ | −0.002093 | 0.0003165 | $<2\times10^{-16}$ |
| Slope | 0.02415 | 0.004979 | $1\times10^{-6}$ | 0.0232 | 0.005255 | $1\times10^{-5}$ |
| Aspect index | −0.3367 | 0.07633 | $1\times10^{-5}$ | −0.3186 | 0.07575 | $3\times10^{-5}$ |
| Needle leaf forest | 0 (N/A) | – | – | 0 (N/A) | – | – |
| Broad leaf forest | −0.1656 | 0.06504 | 0.01 | −0.1960 | 0.0868 | 0.01 |
| Mixed forest | −0.07796 | 0.0884 | 0.38 | −0.20387 | 0.1024 | 0.05 |

2008). High slope increases the risk of ignition. When the degree of slope is too high, fire risk may be decreased due to low human accessibility. In the study by Yang *et al.* (2007) about human-caused forest fire, slope showed a positive relationship with risk on gentle slopes from 10° to 25°. In our research in South Korea, most values of slope are also in a range up to 20°. Therefore, our positive sign coincides with other previous research results.

AI was estimated as negative sign. That means that the south face of a slope is more vulnerable for forest fire occurrence. As we assumed, the southern part of the slope would be drier and human accessibility for managing ancestor's tombs, which are placed on the south face of slopes, is also increased in the fire season because of the traditional custom.

Distance from road and elevation had a negative effect on forest fire occurrence, meaning that any increase in their values reduces human accessibility. We additionally analyzed the model substituting distance from (primary) road by distance from primary and secondary road. Yet the new distance variable was no longer significant, because the number of extremely short distances to the 25 km$^2$ increased remarkably leading to a much lower variability of that distance.

The ranking of vegetation type classes in terms of fire occurrence probability was needle leaf forest + grassland + other > mixed forest > broad leaf forest. The negative coefficients of broad leaf and mixed forest (Table 1) show that these vegetation type classes have a lower fire risk than grassland (including crops and pasture), other and needle leaf forest. Park proved that the conifer have more possibility of forest fire throughout combustion experiment in South Korea (2009). The result of our research showed the coincidence.

In GLMM, the standard errors of coefficients were increased in almost all cases compared to the GLM approach, a well-known effect of the spatial autocorrelation. Quite large differences of coefficients between GLM and GLMM existed in the mixed forest cover factor, the factor with the highest *P*-value.

### 4.2. Empirical variogram

An empirical variogram of the standardized Pearson residuals is shown in Figure 4. It is fitted with a spherical model with the parameters presented in Table 2. In variogram analysis, the sill is the value at which the variogram becomes flat, which indicates the variance of the two separated points of spatial data. The nugget relates to the variance between pairs of points separated by very small distances, random measurement errors, or both of them (Western *et al.* 1998). The range is the distance where the model first flattens out. A partial sill means the width between the nugget and sill. The variogram analysis revealed a correlation range of about 33 km, which is clearly beyond the size of the quadrats. The parameters estimated by GLMM were quite similar to those fitted to the empirical variogram. The sill (nugget plus partial sill) of the GLMM spatial structure was not estimated because it is set to one in the spatial structure in R (Pinheiro and Bates 2006).

### 4.3. Spatial prediction and validation of models

The prediction maps of forest fire occurrence (Figure 5b and c) were derived from the estimated fire counts of the two GLMs. In the GLM approach, we used the estimated model parameters to estimate fire counts for all quadrats. In the GLMM approach, we simply added the kriged residuals to the fixed part of the GLMM regression model. First, we conducted the kriging interpolation using the variogram parameters of the standardized Pearson residuals in GLMM. We then transformed these kriging predictions to the raw residuals, which can be reasonably added to the log-linear model. Although only an ad

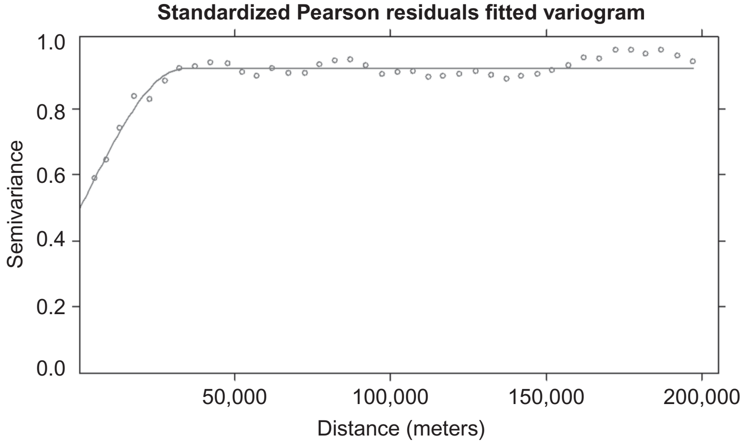


Figure 4. Variogram of standardized Pearson residuals of GLM. The $x$-axis indicates the distance and the $y$-axis the semivariance. The solid line is a spherical model which is computed empirically.

Table 2. Estimated coefficients of the random effects (spherical spatial structure).

| | Primitive estimation by standardized residuals of GLM | | Estimated GLMM spatial structure | |
|---|---|---|---|---|
| Model | Parameter | Value | Parameter | Value |
| Spherical | Range | 34,184 | Range | 32, 301 |
| Sill | 0.93 | Sill | 1 | |
| Nugget | 0.48 | Nugget | 0.47 | |

The spatial model of GLM is computed using residuals of GLM and that of GLMM is integrated with the model.

hoc method at the moment, it will be justified later with the results of the validation. This method was also discussed on the basis of results from more sophisticated approaches in the study of Dormann *et al.* (2007).

The spatial characteristics of forest fire occurrence which are expressed in the prediction map revealed again the strong relationship between forest fires and the population factor, as already shown by the model. The cities, such as Seoul (1), Daejeon (2), Daegu (3), and Busan (4), have high forest fire occurrence, whereas the western coastal part of the peninsula, with low forest coverage, showed relatively low fire occurrence. The highly elevated area through the Taebaek and Sobaek Mountains in the eastern part had a low level of fire risk. These results reflect the topographical factors very well and reveal the high similarity to the observed fire distribution.

The GLM map shows an irregular scattering of high occurrence regions (Figure 5b). On the other hand, the predicted counts by GLMM show a more clustered pattern which is similar to the real distribution of forest fire occurrence (Figure 5a). The tendency to be concentrated at hot spots was increased on GLMM, indicating that GLMM reflects the spatial correlation quite well.

This technique is very important to establish the management plan and strategy for forest fire prevention. It can be applied as an enhanced technique to predict 'hot spots' of fire ignition. The risk map (expected fire counts) can be useful for the positioning of fire-observing agents and for zoning restricted areas in the mountains. Thereby, a fire prevention

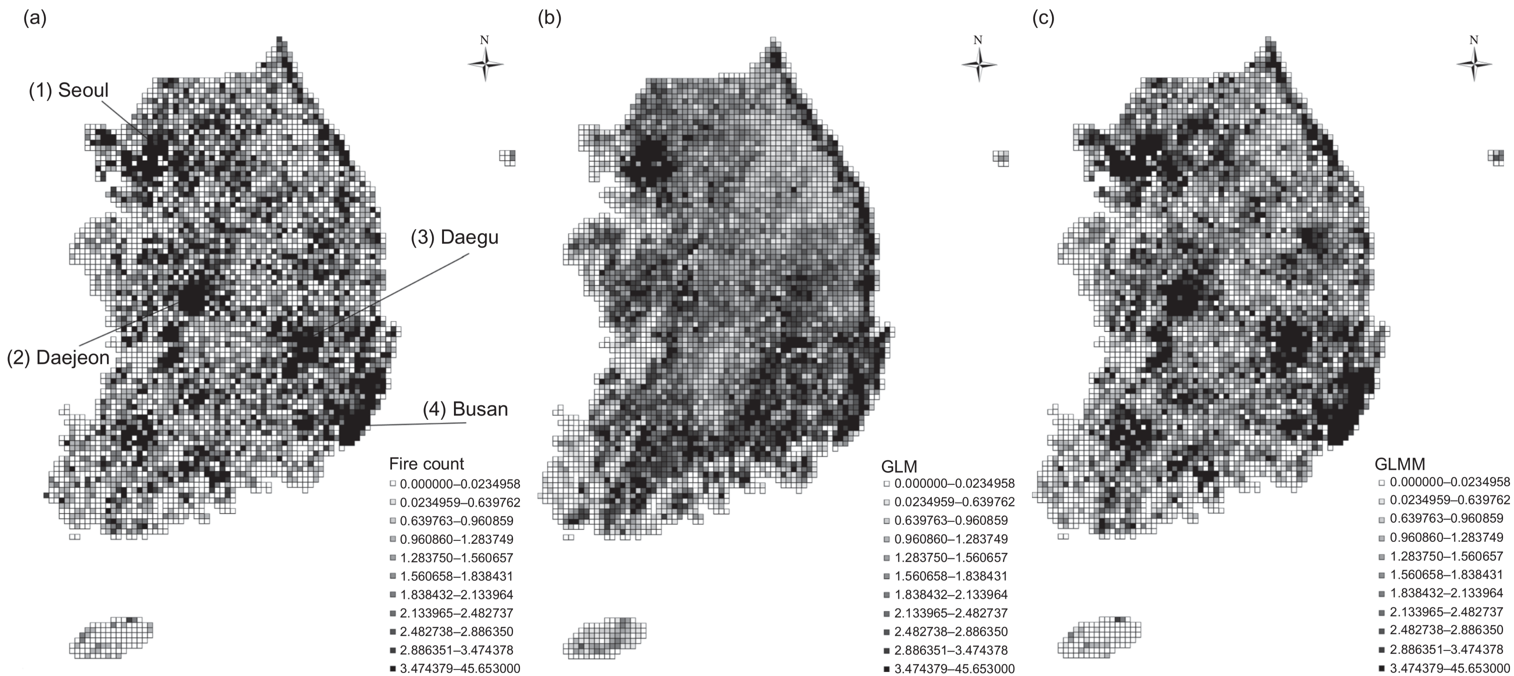


Figure 5. Prediction maps of GLM and GLMM: (a) observed fire count, (b) GLM, and (c) GLMM. The numbered areas are metropolitan areas.

Table 3. The result of RMSE and Pearson correlation for model validation.

| | RMSE | Pearson correlation |
|---|---|---|
| GLM | 1.4398 | 0.4088 |
| GLMM | 1.2595 | 0.6267 |

The lowest value is the best in RMSE. Pearson correlation is improved with increasing proximity to 1.

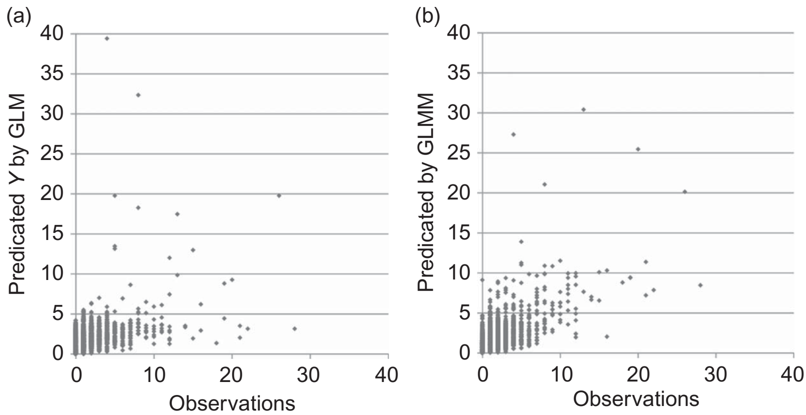


Figure 6. Scatter plots of observed and predicted fire counts using (a) GLM and (b) GLMM. A point cluster arranged from lower left to upper right indicates a positive relationship between the predictions and observations.

unit could efficiently make use of limited resources. Currently, the Korea Forest Research Institute is developing a forest fire risk map using a logistic regression model with static factors such as topographical variance and population density (Lee 2011). Compared to that risk map, which does not consider spatial autocorrelation, the result of our research gives more accurate predictions.

Validation was carried out for the other half of the data. The GLMM predictions for the validation quadrats showed a remarkably lower root mean square error (RMSE) than the estimates using GLM and also a higher correlation between predictions and observed values (Table 3). The resulting scatter plots of observed vs. predicted fire counts (Figure 6) show that GLMM had less extreme outliers than GLM and produced a scatter which was more tightly concentrated around the diagonal.

## 5. Conclusion

This study aimed to predict the forest fire occurrence in Korea using spatial data. The accuracy and precision of the prediction were clearly improved by using the GLMM approach, compared to the GLM approach. GLMM accounts for the spatial dependence of residuals, whereas GLM assumes stochastically independent residuals. Spatially predicted counts of fire remarkably increased after adding the interpolated residuals by kriging, whereas each estimated parameter and $P$-value were similar in both the approaches. In both models, the population density and elevation showed the clearest effect on fire occurrence according to their $P$-values, followed by slope, AI, and distance from the road. Two of the forest cover classes, broad leaf and mixed forest, were also significant compared to the reference class comprising needle leaf forest, grassland, and others.

Spatial autocorrelation of the standardized Pearson residuals of GLM was obvious in the empirical variogram. Fitting a spherical variogram model yielded nugget and range values quite similar to the nugget and range estimated by the GLMM.

Spatial autocorrelation is an important factor for estimation and prediction in regression modeling. Since most fires are caused by humans in South Korea, spatial autocorrelation may be understood as an indicator of other unknown factors of accessibility.

## References

An, S.H., *et al*., 2004. Developing the forest fire occurrence probability model using GIS and mapping forest fire risk. *Journal of the Korean Association of Geographic Information Studies*, 7 (4), 57–64.

Anselin, L., 1988. *Spatial econometrics: methods and models*. Boston, MA: Kluwer Academic Publishers.

Anselin, L., 1992. *Spatial data analysis with GIS: an introduction to application in the social sciences*. Santa Barbara, CA: National Center for Geographic Information Analysis.

Beaty, R.M. and Taylor, A.H., 2008. Fire history and the structure and dynamics of a mixed conifer forest landscape in the northern Sierra Nevada, lake Tahoe basin, California, USA. *Forest Ecology and Management*, 255 (3–4), 707–719.

Bergeron, Y., *et al*., 2004. Fire regimes at the transition between mixedwood and coniferous boreal forest in northwestern Quebec. *Ecology*, 85 (7), 1916–1932.

Cardille, J., Ventura, S., and Turner, M., 2001. Environmental and social factors influencing wildfires in the upper Midwest, United States. *Ecological Applications*, 11 (1), 111–127.

Christensen, O. and Waagepetersen, R., 2002. Bayesian prediction of spatial count data using generalized linear mixed models. *Biometrics*, 58 (2), 280–286.

Christiansen, C. and Morris, C., 1997. Hierarchical Poisson regression modeling. *Journal of the American Statistical Association*, 92 (438), 618–632.

Chuvieco, E. and Salas, J., 1996. Mapping the spatial distribution of forest fire danger using GIS. *International Journal of Geographical Information Science*, 10 (3), 333–345.

Chuvieco, E., *et al*., 2002. Estimation of fuel moisture content from multitemporal analysis of landsat thematic mapper reflectance data: applications in fire danger assessment. *International Journal of Remote Sensing*, 23 (11), 2145–2162.

Curran, P., 1988. The semivariogram in remote sensing: an introduction. *Remote Sensing of Environment*, 24 (3), 493–507.

Diaz-Avalos, C., *et al*., 2001. Space time modelling of lightning-caused ignitions in the Blue Mountains, Oregon. *Canadian Journal of Forest Research*, 31 (9), 1579–1593.

Diggle, P., 1985. A kernel method for smoothing point process data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 34 (2), 138–147.

Diggle, P., Tawn, J., and Moyeed, R., 1998. Model based geostatistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47 (3), 299–350.

Dormann, C., *et al.*, 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: A review. *Ecography*, 30 (5), 609–628.

Garcia, V., *et al*., 1995. A logit model for predicting the daily occurrence of human caused forest fires. *International Journal of Wildland Fire*, 5 (2), 101–111.

Guyette, R., Muzika, R., and Dey, D., 2002. Dynamics of an anthropogenic fire regime. *Ecosystems*, 5 (5), 472–486.

Hengl, T., Heuvelink, G., and Rossiter, D., 2007. About regression-kriging: from equations to case studies. *Computers and Geosciences*, 33 (10), 1301–1315.

Hengl, T., Heuvelink, G., and Stein, A., 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma*, 120 (1–2), 75–93.

Hessburg, P., Agee, J., and Franklin, J., 2005. Dry forests and wildland fires of the inland northwest USA: contrasting the landscape ecology of the pre-settlement and modern eras. *Forest Ecology and Management*, 211 (1–2), 117–139.

Kushla, J. and Ripple, W., 1997. The role of terrain in a fire mosaic of a temperate coniferous forest. *Forest Ecology and Management*, 95 (2), 97–107.

Kwak, H., *et al*., 2009. Spatial and temporal pattern of the human-caused forest fire occurrences in Korea. *Proceeding of the ESRI user conference*, 13–17 July, San Diego, CA, Unpaginated CD-ROM.

Kwak, H., *et al*., 2010. Cause-specific spatial point pattern analysis of forest fire in Korea. *Journal of Korean Forestry Society*, 99 (3), 259–266.
Latham, D. and Williams, E., 2001. Lightning and forest fires. *In*: E.D. Johnson and K. Miyanishi, eds. *Forest fires: behavior and ecological effects*, 376–418 (New York: Academic Press).
Lee, S.Y., *et al*., 2004a. Developing of forest fire occurrence probability model by using the meteorological characteristics in Korea. *Korean Journal of Agricultural and Forest Meteorology*, 6 (4), 242–249.
Lee, W.K., *et al*., 2004b. DBH growth model for *Pinus densiflora* and *Quercus variabilis* mixed forests in central Korea. *Ecological Modelling*, 176 (1–2), 187–200.
Lee, B., 2011. Prevention of forest fire by mapping forest fire risk. *Forest*, 2011 (2), 86–87.
Legendre, P., 1993. Spatial autocorrelation: trouble or new paradigm? *Ecology*, 74 (6), 1659–1673.
Martell, D., Otukol, S., and Stocks, B., 1987. A logistic model for predicting daily people-caused forest fire occurrence in Ontario. *Canadian Journal of Forest Research*, 17 (5), 394–401.
Oliver, M.A., 1990. Kriging: a method of interpolation for geographical information systems. *International Journal of Geographical Information Science*, 4 (3), 313–332.
Park, J., 2009. *A study on forest fire risk in youngdong areas through combustional characteristics of forest fuel.* Dissertation (PhD). Kangwon National University.
Pew, K. and Larsen, C., 2001. GIS analysis of spatial and temporal patterns of human-caused wildfires in the temperate rain forest of Vancouver Island, Canada. *Forest Ecology and Management*, 140 (1), 1–18.
Pinheiro, J. and Bates, D., 2006. Linear and nonlinear mixed effects models. *R package version*, 1–74.
Prestemon, J. and Butry, D., 2005. Time to burn: modeling wildland arson as an autoregressive crime function. *American Journal of Agricultural Economics*, 87 (3), 756–770.
Renkin, R.A., 1992. Fuel moisture, forest type, and lightning-caused fire in Yellowstone National Park. *Canadian Journal of Forest Research*, 22 (1), 37–45.
Rezaeian, M., *et al*., 2007. Geographical epidemiology, spatial analysis and geographical information systems: a multidisciplinary glossary. *Journal of Epidemiology & Community Health*, 61 (2), 98–102.
Richards, S., Possingham, H., and Tizard, J., 1999. Optimal fire management for maintaining community diversity. *Ecological Applications*, 9 (3), 880–892.
Romero-Calcerrada, R., *et al*., 2008. GIS analysis of spatial patterns of human-caused wildfire ignition risk in the SW of Madrid (Central Spain). *Landscape Ecology*, 23 (3), 341–354.
Rothermel, R., 1972. A mathematical model for predicting fire spread in wildland fuels. INT-115, U.S. Department of Agriculture, Forest Service, Ogden, UT.
Schabenberger, O. and Gotway, C.A., 2005. *Statistical methods for spatial data analysis*. Boca Raton, FL: Chapman & Hall /CRC.
Silverman, B., 1981. Using kernel density estimates to investigate multimodality. *Journal of the Royal Statistical Society: Series B (Methodological)*, 43 (1), 97–99.
Stein, M., 1999. *Interpolation of spatial data: some theory for kriging*. New York: Springer.
Tobler, W., 1979. Smooth pycnophylactic interpolation for geographical regions, *Journal of the American Statistical Association*, 74 (367), 519–530.
Veblen, T., *et al*., 1999. Fire history in Northern Patagonia: the roles of humans and climatic variation. *Ecological Monographs*, 69 (1), 47–67.
Venables, W. and Ripley, B., 2002. *Modern applied statistics with S*. 4th ed. New York: Springer.
Western, A., Blöschl, G., and Grayson, R., 1998. Geostatistical characterisation of soil moisture patterns in the Tarrawarra catchment. *Journal of Hydrology*, 205 (1–2), 20–37.
Won, M.S., Koo, K.S., and Lee, M.B., 2006. An analysis of forest fire occurrence hazards by changing temperature and humidity of ten-day intervals for 30 years in spring. *Korean Journal of Agricultural and Forest Meteorology*, 8 (4), 250–259.
Wotton, B. and Martell, D., 2005. A lightning fire occurrence model for Ontario. *Canadian Journal of Forest Research*, 35 (6), 1389–1401.
Wotton, B., Martell, D., and Logan, K., 2003. Climate change and people-caused forest fire occurrence in Ontario. *Climatic Change*, 60 (3), 275–295.
Yang, J., *et al*., 2007. Spatial patterns of modern period human-caused fire occurrence in the Missouri Ozark Highlands. *Forest Science*, 53 (1), 1–15.
Zhang, H., 2002. On estimation and prediction for spatial generalized linear mixed models. *Biometrics*, 58 (1), 129–136.