Spatio-temporal hierarchical Bayesian analysis of wildfires with Stochastic Partial Differential Equations. A case study from Valencian Community (Spain)

Dr Pablo Juan Verdoy[a*]

[a]*Department of Mathematics, University Jaume I, Campus Riu Sec, 12071, Castellón, Spain*

**Corresponding author*

Telephone number: +34 964 728381

e-mail: juan@uji.es

ORCID number: http://orcid.org/0000-0002-2197-7502

Spatio-temporal hierarchical Bayesian analysis of wildfires with Stochastic Partial Differential Equations. A case study from Valencian Community (Spain)

Abstract

The spatio-temporal study of wildfires has two complex elements that are the computational efficiency and longtime processing. Modelling the spatial variability of a wildfire could be performed in different ways, and an important issue is the computational facilities that the new methodological techniques afford us. The Markov random fields methods have made possible to build risk maps, but for many forest managers, it is more advantageous to know the size of the fire and its location. In the first part of this work, Stochastic Partial Differential Equation with Integrated Nested Laplace Approximation is utilised to model the size of the forest fires observed in the Valencian Community (Spain) and so it does the inclusion of the time effect, and the study of the emergency calls. The most crucial element in this paper is the inclusion of the improved meshes for the spatial effect and the time, these are, 2d (locations) and 1d (time) respectively. The advantage of the use of spatio-temporal meshes is described with the inclusion of Bayesian methodology in all the scenarios.

Keywords: Bayesian Inference; INLA; SPDE; Spatio-temporal Mesh; wildfire.

1. Introduction

Modelling the incidence of wildfires and the fire size is necessary to understand how global warming and climate change may affect the landscape in the coming years, and to determine what factors are related to spatial incidence and the size of the burned area ([1]; [2]). Wildfires are associated with their spatial coordinates, the time effect and the corresponding covariates. Thus, even though methods such as Markov Random Fields may also be useful to respond to some scientific questions of interest, spatial point processes are the most appealing analytical tool to investigate the spatial and spatio-temporal distribution of forest fires ([3]; [4]).

Previous studies have solved the wildfires problem by producing risk maps or by calculating the probability of a starting wildfire at some location inside a study area D using statistical methods ([5]; [6]; [7]). These studies used statistical methods to produce wildfire risk maps included Markov Random Fields ([1]) and spatial point processes ([8]; [9]). Despite their usefulness, most of the studies have not considered the burned area caused by each wildfire as in this paper given that it is used INLA-SPDE.

To begin with, the first part of the work is devoted to the study of the best spatial mesh for the data. The models that are introduced in this work have been compared with the use of the Deviance Information Criterion (DIC) and the Watanabe-Akaike information criterion (WAIC) ([10]; [11]).

In addition, the most essential point of this work is to show the benefits of using Stochastic Partial Differential Equation (SPDE) with Integrated Nested Laplace Approximation (INLA) for spatio-temporal wildfire data ([12]; [13]; [14]; [15]; [16]), including the mesh for the temporal effect, one dimension (1d - time) and for the spatial effect (2d - locations). Nowadays, the problem is to choose the perfect mesh formed with the SPDE of each pattern in spatio-temporal processes. A prerequisite for creating mesh processes is the exploratory data analysis which is fundamental in this research.

Moreover, the noteworthy elements in the paper are the different data used to this methodology (INLA - SPDE). The outline applied to wildfires in Valencian Community includes two different data set. Firstly, are the real locations and temporal wildfires, and secondly the emergency calls about the corresponding wildfires.

Data set

The patterns produced by wildfire in the Valencian Community are analysed and its location is in the north-east coast of the Iberian Peninsula. The region is bordered by Catalonia to the north and the Iberian System range of mountains to the west. Furthermore, the region is delimited to the east by the Mediterranean Sea. It is a region with a surface area of 23,245 square kilometres, representing 4.6% of the Spanish national territory.

A total of 315 fires were recorded in the studied area in 2015 (Figure 1). In addition to the locations of the fire centroids in Cartesian coordinates (Mercator transversal projections, UTM, Datum ETRS89, zone 31-N), several covariates were also considered.

[Here Figure 1]

Data exploration steps used in such cases are:

1) Exploratory Data Analysis
2) Outliers
3) Collinearity
4) Relationships between the response variable (Y) and the covariates (X's)
5) Variance Inflation factors
6) Interactions
7) Zero inflation justification

The covariates that directly affect wildfires are typology, cause, causative, days last rain, maximum temperature, relative humidity, wind speed, wind direction, combined model (relation between elements in a wildfire), danger degree and a fire type. The relationships between the covariates are shown in Figure 2 (boxplots).

[Here Figure 2]

The next step is the Variance Inflation factors (GVIF) and seeks the optimal variables and using only values below 2 because these are informative (Table 1).

[Here Table 1]

The collinearity study of the covariates, the possible structure of outliers' problems and the possible pattern appear in Figure 3. The most likely option to obtain patterns is the relationship between each covariate and the response variable TOTAL (burned tree and not tree together), and as a result, in this illustration there is no pattern present.

[Here Figure 3]

In Figure 4 all the possible distances between points, extensive or short, are studied and this enables us to continue with the study.

[Here Figure 4]

Finally, before selecting the model, the number of zeroes determines if the variable data is high. The decision regarding the selected model depends on the number of zeroes found, and if there are many of them proximate zero or zero, the best likelihood family is Zero-inflated. In our case, there is 42.22% of zero data, and it is unavoidable the use of Zero-inflated model. (Figure 5).

[Here Figure 5]

On the other hand, the second data used which reinforces the analysis of the wildfires in the Valencian Community, is the use of the emergency calls in the same period for the incidences. In Figure 6 appear the locations of the emergencies calls. The main characteristic of this data set is the position of a real wildfire which is proximate to any population, and in consequence it is necessary an early intervention and resolution of the problem.

[Here Figure 6]

The rest of the paper is organised as follows. Section 2, Methodology, gives all the details needed to clarify the Bayesian methodology used and Spatial Point Process. Section 3 is devoted to the Data set, and Section 4 includes the models for burned area among different wildfires scenarios. Finally, Discussions and Conclusions are in Section 5.

2. Methodology

Integrated Nested Laplace Approximation (INLA)

This work offers the possibility of studying Spatial Point Processes by using integrated nested Laplace approximation (INLA) [29 and 30]. [16] develops the INLA methodology for approximate Bayesian inference as an alternative to traditional Markov chain Monte Carlo methods. INLA focuses on models that can be expressed as latent Gaussian Markov random fields (GMRF) for their computational properties, and the data applied in this case possesses such characteristics.

The data can be idealised as realisations of a stochastic process indexed by:

$$Y(\cdot) = \{y(s_i, t_i) \in R^2 \times R\} \qquad (1)$$

where s_i represents spatial and t_i is for temporal, with both of them $Y(\cdot)$ is a spatio-temporal subset of R^2xR.

The advantages of using INLA over other methods, such as basic statistical methods or more complex ones (like Markov Chain Monte Carlo (MCMC) ([17])), are the following:

- It works with reasonable computational times, thereby allowing the user to work with complex models quickly and efficiently.
- It allows the integration of as many covariates as desired, and also the incorporation of new covariates in the model in later steps.
- It allows the level of significance of covariates to be analysed.
- It does not require working with normal distributions exclusively, since its base is on Bayesian inference.

The data can be presented by a collection of observations $y = \{y_1, \ldots, y_n\}$ ([18]; [19]; [20]; [29] and [30]). In statistical analysis, to estimate a general model it is useful to shape the mean for the additive linear predictor, defined on a suitable scale:

$$\eta_i = \beta_0 + \sum_{m=1}^{M} \beta_m z_{mi} + \sum_{i=1}^{L} f_l(v_{li}) \qquad (2)$$

Where β_0 is a scalar, which represents the intercept, $\beta = (\beta_1, \ldots, \beta_M)$ are the coefficients of the linear effects of the covariates $z = (z_1, \ldots, z_M)$ on the response, and $f = \{f_1(.), \ldots f_L(.)\}$ is a collection of functions defined in terms of a set of other covariates represented as $v = v(v_1, \ldots, v_L)$, different from the previous covariates. The first step in defining the structure of the data $y = \{y_1, \ldots, y_n\}$. A very general approach consists in specifying a distribution for $y = \{y_1, \ldots, y_n\}$ characterised by a parameter ϕ_i (usually the mean $E(y_i)$) defined as a function of a structured additive predictor η_i

through a link function $g(.)$, such as $g(\phi_i)=\eta_i$. The additive linear predictor η_i is defined as follows ([21]):

$$\eta_i = \beta_0 + \sum_i \beta_i z_i \qquad (3)$$

Where β_i represents the coefficient that quantifies the effect of the covariates in the response z_i. This statistical analysis can be carried out with the freeware statistical package R, version 3.4.3 ([22]) and the R-INLA package 2017 ([15]).

The priors for the formulas (1) to (3), are an important element. The *fixed effects* and are typically normally distributed, centered on 0 and with a large variance, while v_{1j} are called *random effects(hyperparameters)* and are typically normally distributed with an exchangeable structure, i.e., $v_{1j} \sim$ Normal(0, σ_v^2). With this, a prior distribution needs to be specified on the regression parameters $\boldsymbol{\beta} = \{\beta_0,\ldots,\beta_M\}$ including the intercept, and on the variance σ^2 of the outcome. The choice of prior is:

$\beta_m \sim$ Normal(0,10^6), $m = 1,\ldots,M$

$\log(\tau) = \log(1/\sigma^2) \sim$ logGamma(1, 10^{-5}).

The aim is to perform the inferential process and to obtain the posterior distribution for $\boldsymbol{\beta}$ and σ^2.

If we are interested in changing the prior for the regression parameters, for instance, reducing the variability on the prior for β_0 and β_1, specifying $\beta_0 \sim$ Normal(0, 10000) and $\beta_1 \sim$ Normal(0, 1), we can achieve it in R- I NLA using the option control.fixed. It is also possible to modify the specification of the prior on the outcome precision (remember $\tau=1/\sigma^2$) using the option control.family of the inla command. By default, a noninformative logGamma prior is assumed on the logarithm of the precision, which is equivalent to assume a Gamma prior on the precision $\tau \sim$ Gamma(1, 10^{-5}).

In the case of the temporal correlation is considered here the commonly used *random walk* (RW). This random walk structure is characterized by a variance parameter, on which we need to specify a prior distribution, in our case logGamma(1, 10^{-5}). In R-INLA the default internal representation for the SPDE parameters is log(τ) = θ_1 and log(κ) = θ_2, with θ_1 and θ_2 being given a joint Normal prior distribution (by default independent Normal(0, 1) priors are used).

When the battery of competing models has been obtained, the DIC and the WAIC criterium can be obtained for each one of the models to select the most suitable one, those that occur to have a higher level of complexity and a greater goodness-of-fit. That is to say, models that show the lowest WAIC and DIC should be chosen ([10]; [11]):

$$DIC =' goodness\ of\ fit' + 'complexity' = D(\theta) + 2p_D \qquad (4)$$

Where $D(\theta)$ is the deviance evaluated at the posterior mean of the parameters and p_D denotes the ‘effective number of parameters’, which measures the complexity of the model ([10]). When the model is true, $D(\theta)$ should be approximately equal to the ‘effective degrees of freedom’, $n - p_D$.

Stochastic Partial Differential Equation (SPDE)

The Stochastic Partial Differential Equation (SPDE) approach is used for the study of spatial effect with the Matérn covariance function. The triangulation presented allows the spatio-temporal covariance function and the dense covariance matrix of a Gaussian Field (GF) to be replaced with a neighbourhood structure and a sparse precision matrix. This yields substantial computational advantages ([5]; [13]), and this approach makes possible to detect the risk factors effects in the spatial distribution of wildfire patterns ([23]).

In this case, the covariance structure of the Matérn type for the dispersion matrix Σ, that is, if $h_{ij}=||x_i-x_j||$ denotes the distance between two arbitrary wildfires within W the covariance of their fire sizes is given by

$$C(h_{ij}) = \frac{1}{\Gamma(\tau)2^{\tau-1}}(\kappa h_{ij})^{\tau} K_{\tau}(\kappa h_{ij}) \qquad (5)$$

Where K_τ denotes a Bessel function of second kind and order τ, which controls the smoothness of the process. The parameters τ and κ relate empirically to the nominal range of the spatial covariance ($r = \sqrt{8\tau}/\kappa$). The Matérn covariance is a general model that encompasses covariance models such as the Exponential and the Gaussian, commonly used in geostatistical analyses ([12]).

The structure and the basis functions used are defined on a triangulation of domain *D*:

$$X(s) = \sum_{l=1}^{n} \varphi_l(s)\, \omega_l \qquad (6)$$

where *n* is the total number of vertices in the triangulation; $\{\varphi_l(s)\}$ is the set of base functions and $\{\omega_l\}$ are the zero-mean Gaussian distributed weights.
These basis are presented as:

$$\varphi_l(s) = \begin{cases} 1 & at\ vertix \\ 0 & elsewhere \end{cases} \qquad (7)$$

The key is to calculate $\{\omega_l\}$, which reports the value of the spatial field at each vertex of the triangle. The values inside the triangle will be determined by linear interpolation ([5]; [24]).

The choice problem of the best mesh for the SPDE approximation could be solved using the correct elements of the mesh shown in Figure 7, where it is different for 1d mesh used for the temporal effect because it is easier for a one-dimension mesh.

[Here Figure 7]

3. Application. Burned area in wildfires

Spatial mesh with SPDE

A wildfire is any uncontrolled fire in combustible vegetation that occurs in the countryside or a wilderness area ([8]; [25]). A wildfire differs from other fires by its extensive size, and they are characterised in terms of the cause of ignition, their physical properties or the weather effect on the fire ([26]; [27]).

Wildfires are a natural element of the Mediterranean ecosystem, and their prevention and suppression is in the benefit of lowering the levels of risk and its vulnerability to values that are tolerable for society ([8]).

A wildfire is associated with its spatial coordinates, longitude and latitude of the centroid of the burned area or the place where it was detected, along with other variables such as size or cause of the forest fire. The spatial-temporal stochastic is the process by which controlling the moment in time when it was produced, all wildfires can be identified. Temporal clustering of wildfires, whether deriving from multiple ignition lightning events, arson ([28]), or other sources, combined with favourable fuel and weather conditions, can force suppression resource rationing across space. Spatial clustering can also indicate the presence of risk factors. The temporal effect is included as a covariate mesh (1d) in the model.

The steps for modelling the application that includes the spatial effect created with the mesh, are firstly creating the spatial locations, by Matérn covariance, and then, is created the spatial mesh structure ([13]). In the next steps, the covariates are included, and the SPDE spatial model is done and introduced as a function in the final model ([16]).

Following the possibilities studied in previous works, the meshes are shown in Figure 8, and in Figure 9, following the natural instructions ([12]), are selected the best meshes.

[Here Figure 8 and Figure 9]

The models applied for the real data, that is, wildfires in the Valencian Community in 2015, have both non-spatial and spatial effects and these affects the computing time (Table 2). Table 3 shows the value of real data parameters for each model. In Figure 10, the difference between the values of the parameters appears, and it is crucial for the final results.

[Here Table 2, Table 3, Figure 10]

Table 4 below shows the summary results related to goodness-of-fit for the battery of models. DIC and WAIC.

[Here Table 4]

With these results, it can be deduced that including a higher number of covariates improves any statistical model because DIC and WAIC become lower and provides a better prediction.

[Here Figure 11]

In this case, validation is performed by comparing residuals and the correlation between the real data and the model data. The relationship with distances can also be seen in Figure 11, above. Figure 12 bellow shows the spatial effect of the Valencian Community in two formats. The correlation in this case, $\rho = 0.8309389$, has a high value which suggests a good result.

[Here Figure 12]

Spatio-temporal meshes

The second part for the applied data is the analysis of emergency calls about the wildfires in 2015 in the Valencian Community.

In this case, for the inclusion of the time effect, the temporal 1d mesh is developed in 4 parts (blue lines in Figure 13). The black lines are the mouthparts of the data.

[Here Figure 13]

The next step is developing the spatial 2d mesh (left) and regionalised mesh (right) when the contour of the region is not used (Figure 14). The leftward mesh is used for the inclusion of the spatial effect in the models and this reduces the computing time.

[Here Figure 14]

With these data and with k= 4 (four temporal parts), the models tested are shown in Table 5 and the structure of their parameters in Figure 15. The inference to the parameters is another advantage of the use of Bayesian methodology. In Figure 16 the region models are presented without the use of region contour.

[Here Table 5, Figure 15 and Figure 16]

Then, the introduction of the contour in the model defines the studied region. The 2d mesh is different, and the outside part of the region is not necessary. In Figure 17 the new mesh and regionalised zones are shown, and the red dots are the regions without points.

[Here Figure 17]

With these data and with k= 4 (four temporal parts), the models tested are (Table 6 and Figure 18):

[Here Table 6, Figure 18 and Figure 19]

Finally, in Figure 19 the models with the contour included is presented.

4. Discussions and Conclusions

In this work, the study of wildfires with Bayesian methodology, including SPDE for spatial and temporal effect, is done. The phases proposed for considering the best solution for each case are presented and modelled using latent Gaussian field which is extended to Gaussian Random Markov Fields.

A computationally efficient method for Bayesian inference, based on INLA and SPDE, was presented and the newest element is the inclusion of meshes in time (one dimension - 1d) and space (two dimensions - 2d). We looked at an application for modelling a wildfire data set in which the process is faster and more precise. This method is faster than the other ones proposed in previous works, and it a basic advantage for new research.

The advantage of INLA-SPDE is that it can predict the subsequent marginal distributions of the model parameters as well as the model responses without carrying out extensive simulations. This methodology can be potentially applied to mapping the spatial distribution of environmental variables or in all kinds of spatial point patterns in geostatistical issues, including covariates. The use of triangulated meshes in INLA-SPDE may lead to two possibilities, working with simple or complex databases.

As a conclusion, the wildfire data and emergency calls have a similar behaviour in space and time as it is shown in this study. Hence, it is relevant for following research since data set about real locations or emergency calls of wildfires depend on the accessibility of getting both. The obtained models in this investigation show that the time covariate is an essential element for the behaviour of the wildfires, and not only the other covariates as the elevation, human effects or climatological effect.

The results show that INLA-SPDE could also be a complementary tool in the wildlife biologist's analytical toolkit, where models are specified using a syntax that should be familiar to users of R, and where data are formatted straightforwardly with relatively few lines of codes, and that implies a leading advance in spatial statistics.

Acknowledgements.

The author is grateful to all professionals who have provided information and help, specially to the Forest Fire Prevention Service and Government of Valencian Community for providing wildfire data. Pablo Juan was partially funded by grant MTM2016-78917-R from the Spanish Ministry of Science and Education.

5. References

[1] Díaz-Avalos C, Peterson DL, Alvarado E, Ferguson SA, Besag JE (2001) Space-time modelling of lightning-caused ignitions in the Blue Mountains, Oregon. *Canadian Journal of Forest Research* 31:1579–1593.

[2] Díaz-Avalos, Carlos, Juan, Pablo and Serra-Saurina, Laura (2016) Modelling fire size of wildfires in Castellon (Spain), using spatiotemporal marked point processes. *Forest Ecology and Management* 381 (2016) 360–369.

[3] Mandallaz, D., Ye, R., 1997. Prediction of forest fires with Poisson models. *Can. J. For. Res*. 27, 1685–1694.

[4] Xu, H., Schoenberg, F.P., (2011). Point process modelling of wildfire hazard in Los Angeles County, California. *Ann. Appl. Stat.* 5 (2A), 684–704.

[5] Serra, L., Saez, M., Mateu, J., Varga, D., Juan, P., Diaz-Ávalos, C. and Rue, H., (2014a). Spatio- temporal log Gaussian Cox processes for modelling wildfire occurrence. The case of Catalonia, 1994–2008. *Environmental and Ecological Statistics,* 21 (3), 531–563.

[6] Serra, L., Saez, M., Juan, P., Varga, D., Mateu, J., (2014b). A spatio-temporal Poisson hurdle point process to model forest fires. . *Stochastic Environmental Research and Risk Assessment*, 28 (7), 1671–1684.

[7] Barros, A.M.G., Pereira, J.M.C., (2014). Wildfire selectivity for land cover type: Does size matter? *PlosOne* 9 (1), e84760. http://dx.doi.org/10.1371/journal. pone.0084760.

[8] Juan, P., Mateu, J. and, Saez, M., (2012). Pinpointing spatio-temporal interactions in wildfire patterns. *Stochastic Environmental Research and Risk Assessment*, 26 (8), 1131–1150.

[9] Møller, J., Díaz-Avalos, C., (2010). Structured spatio-temporal shot-noise Cox point process models, with a view to modelling forest fires. *Scand. J. Stat*. 37, 2–15.

[10] Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van der Linde (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B* 64 (4), 583–616.

[11] Watanabe, S. (2010). Asymptotic equivalence of Bayes cross-validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research* 11, 3571–3594.

[12] Krainski, E.t, Gómez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D., Simpson, D., Lindgren, F. and Rue, H. (2019). Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and INLA. CRC Press/Taylor and Francis Group. Boca Raton, FL. ISBN: 978-1-138-36985-6.

[13] Lindgren, F., Rue, H. and Lindstrom, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields the SPDE approach. J. *Roy. Stat. Soc., Ser. B*, Volumen 73 (4), pp. 423-498.

[14] Martins, T., G., Simpson, D., Lindgren, F. And Rue, H. (2013) Bayesian computing with INLA: New features. Computational Statistics & Data Analysis, 67; 68-

83.

[15] R-INLA R-INLA project. <http://www.r-inla.org/home> (accessed on January 15th, 2019).

[16] Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *J. Roy. Stat. Soc. B,* 71, 319-392.

[17] Tsanas, A., Xifara, A., (2012). Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. *Energy and Buildings*, 49, 560-567.

[18] Blangiardo, M., Cameletti, M., Baio, G. and Rue, H. (2013). Spatial and Spatio-temporal models with R-INLA. *Spatial and Spatio- temporal Epidemiology* 4:33–49.

[19] Cameletti, M., Lindgren, F., Simpson, D., Rue, H., 2013. Spatio-temporal modleling of particulate matter concentration through the SPDE approach. *Adv. Stat. Anal*. 97 (2), 109–131.

[20] Vlad, Iulian, Juan, Pablo and Mateu, Jorge. (2015). Bayesian spatio-temporal prediction of cancer dynamics. *Computers and Mathematics with Applications*, 70, 857-868.

[21] Blangiardo, M., & Cameletti, M. (2015). Spatial and Spatio-temporal Bayesian Models with R-INLA. John Wiley & Sons, Chichester, UK.

[22] R Core Team, 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

[23] Aragó, P., Juan, P., Díaz-Avalos, C., Salvador, P., (2016). Spatial point procesmodellingng applied to the assessment of risk factors associated with forest wildfires incidence in Castellón, Spain. *Eur. J. For. Res*. http://dx.doi.org/10.1007/s10342- 016-0945-z.

[24] Simpson, D., Illian, J.B., Lindgren, F., Sørbye, S. H. and Rue, H. (2016). Going off grid: computationally efficient inference for log-Gaussian Cox processes, *Biometrika*, Volume 103, Issue 1, 49–70.

[25] Cambridge Advanced Learner's Dictionary, Colin McIntosh,Third Edition. Cambridge: Cambridge University Press, 2008.

[26] National Interagency Fire Center. The Science of Wildland Fire (available in: http://www.nifc.gov/, accessed on February 14, 2011).

[27] Flannigan MD, Amiro BD, Logan KA, Stocks BJ, Wotton BM. Forest fires and climate change in the 21st century. *Mitigation and Adaptation Strategies for Global*

Change 2006; 11(4):847-859.

[28] Butry DT, Prestemon JP. Spatio-temporal wildland arson crime functions. *American Agricultural Economics Association Annual Meeting*, 26–29 July 2005, Providence, Rhode Island, USA (available in: http://ageconsearch.umn.edu/handle/19197, accessed on March 7, 2011).

[29] Braulio-Gonzalo, M., Juan, P., Bovea, M.D., and Ruá, M. J. (2016). Modelling energy efficiency performance of residential building stocks based on Bayesian statistical inference, *Environmental Modelling & Software*, 83, 198-211.

[30] Bovea, M.D., Ibáñez-Forés, V., Juan, P., Pérez-Belisa, V., and Braulio-Gonzalo, M. (2018), Variables that affect the environmental performance of small electrical and electronic equipment. Methodology and case study, *Journal of Cleaner Production*, 203, 1067-1084.

Figures:

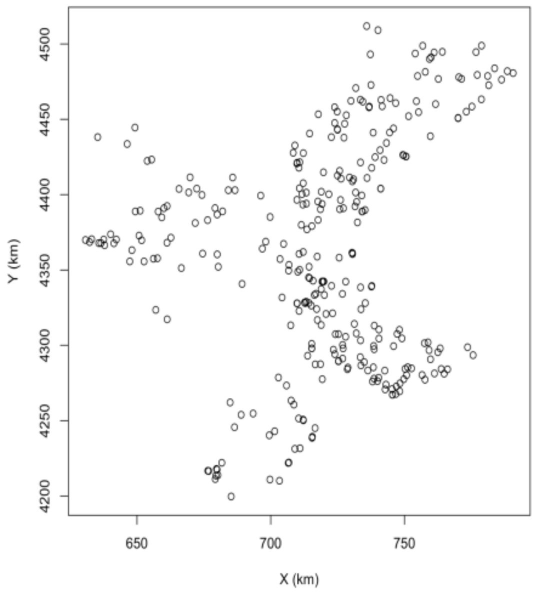

Figure 1: Distribution of forest fires in the Valencian Community in 2015.

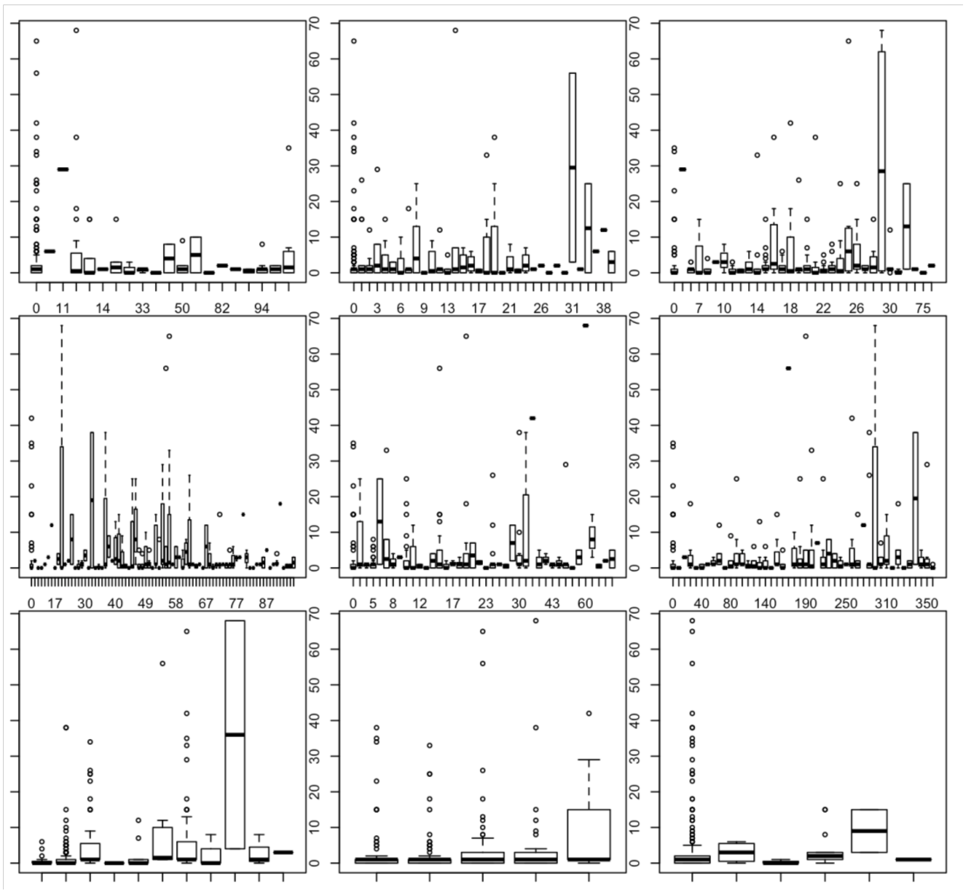

Figure 2: Boxplot covariates, in the first line are, cause, days last rain, temp_max, in the second line are Relative_H, wind speed, wind direction and in the third line are combined model, danger degree, a fire type.

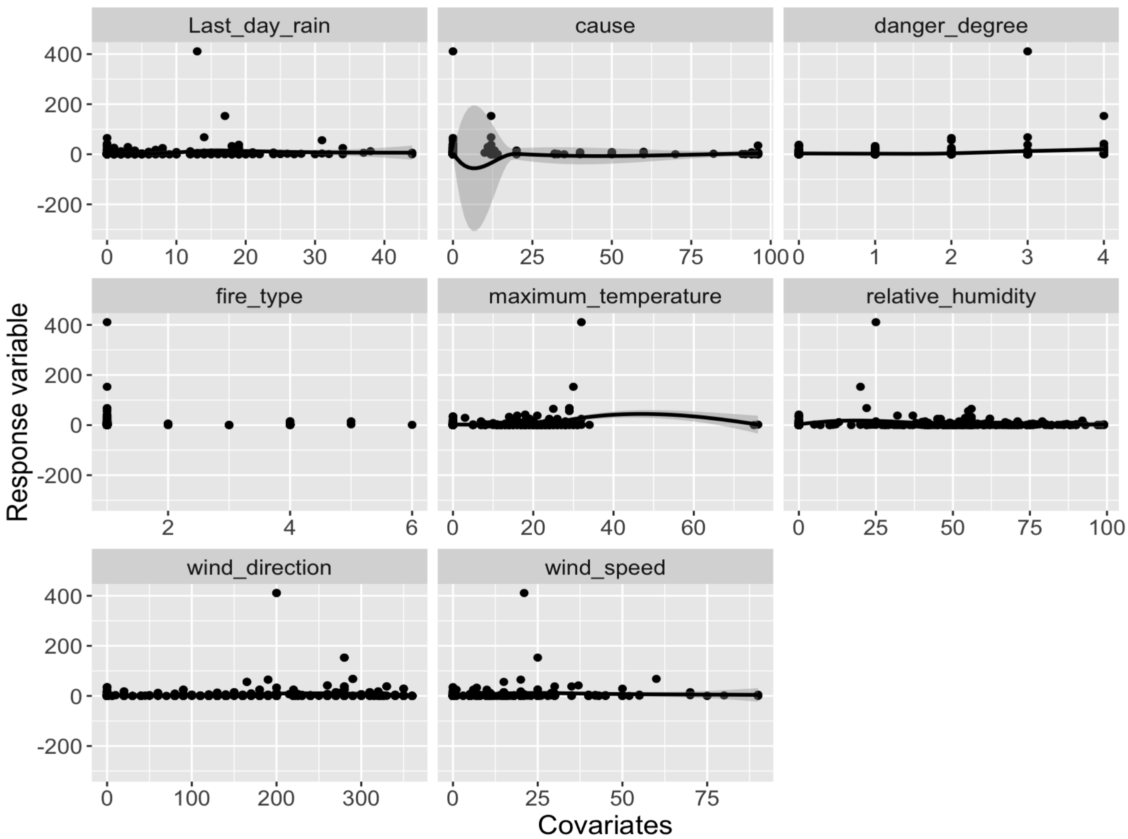

Figure 3: Response vs covariates.

.

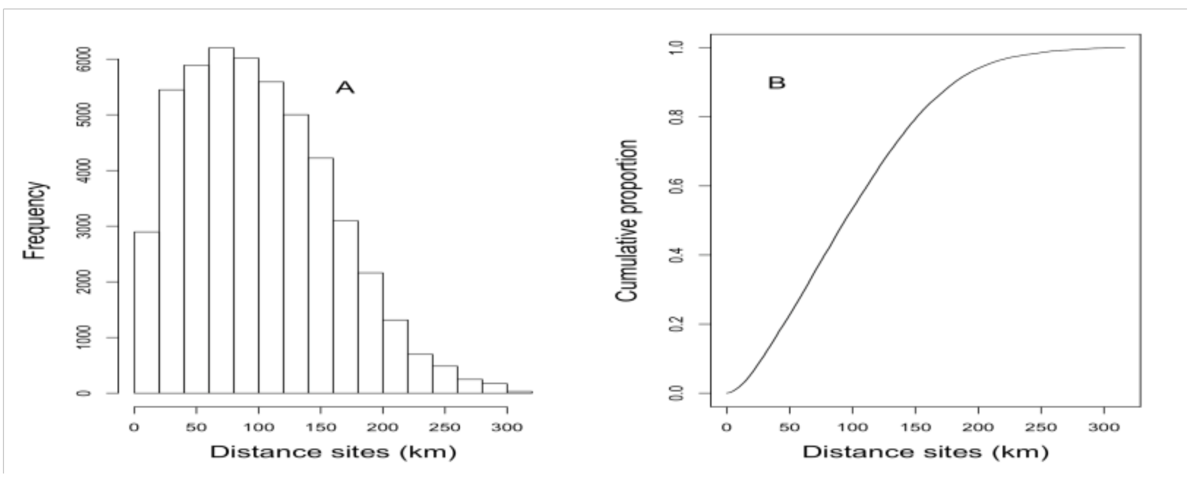

Figure 4: Frequency and cumulative proportion of distances between points.

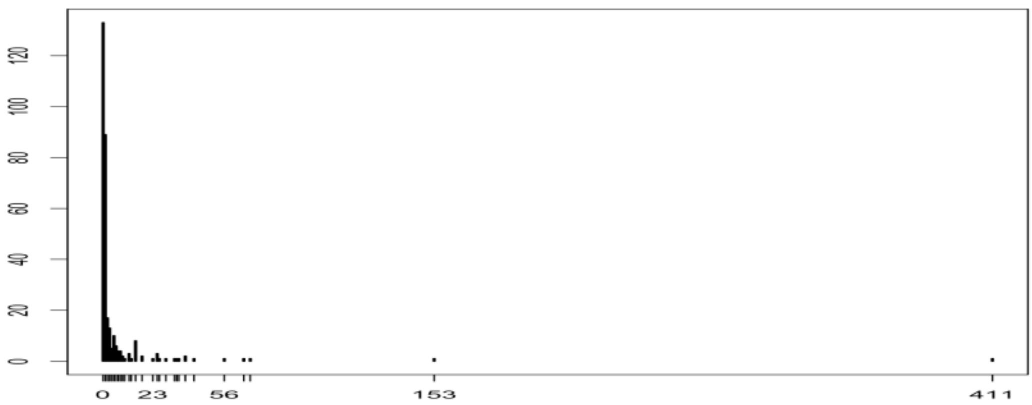

Figure 5: Frequency of the response variable (total burned area). The highest values are zeroes.

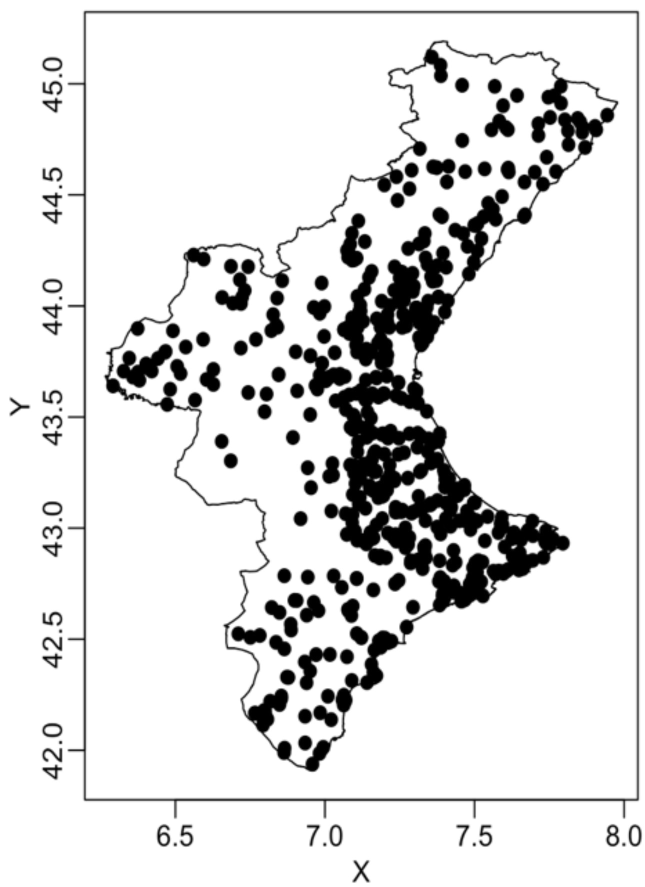

Figure 6: Locations of emergency calls about wildfires.

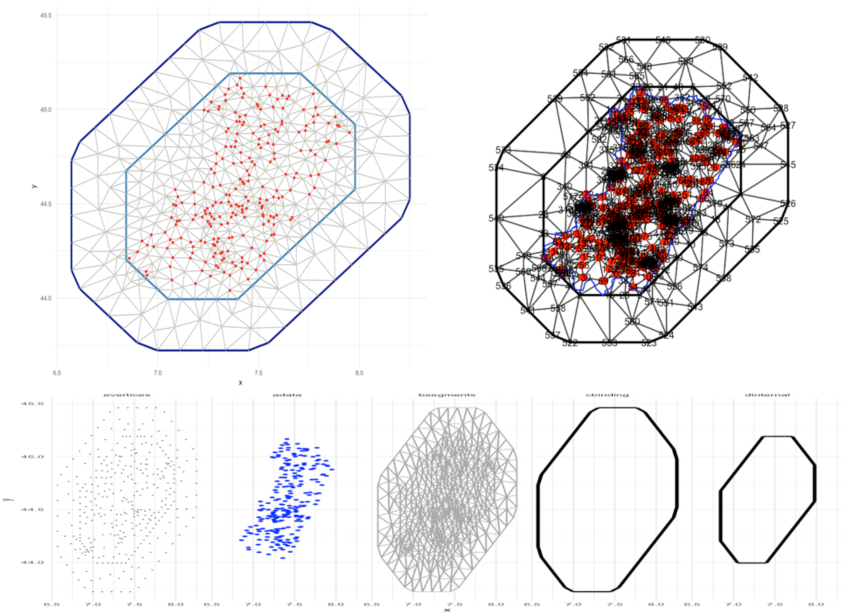

Figure 7: Mesh elements. in the first line on the left the mesh with the points and on the right the number of vertices. In the second line are all the elements of the mesh.

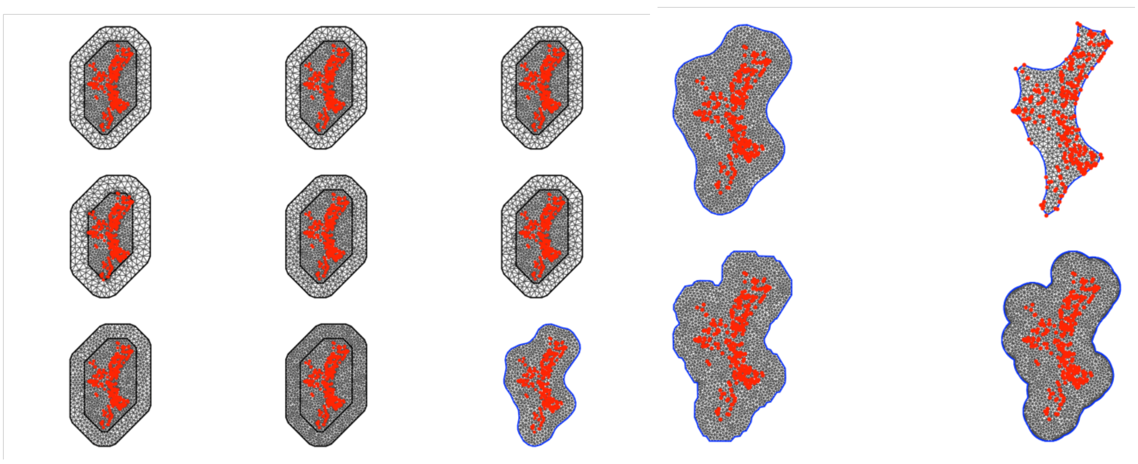

Figure 8: The 13 meshes, with respective numbers of vertices: 1612, 1430, 1612, 2021, 1732, 1356, 1592, 2071, 794, 1778, 600, 1836, 2801.

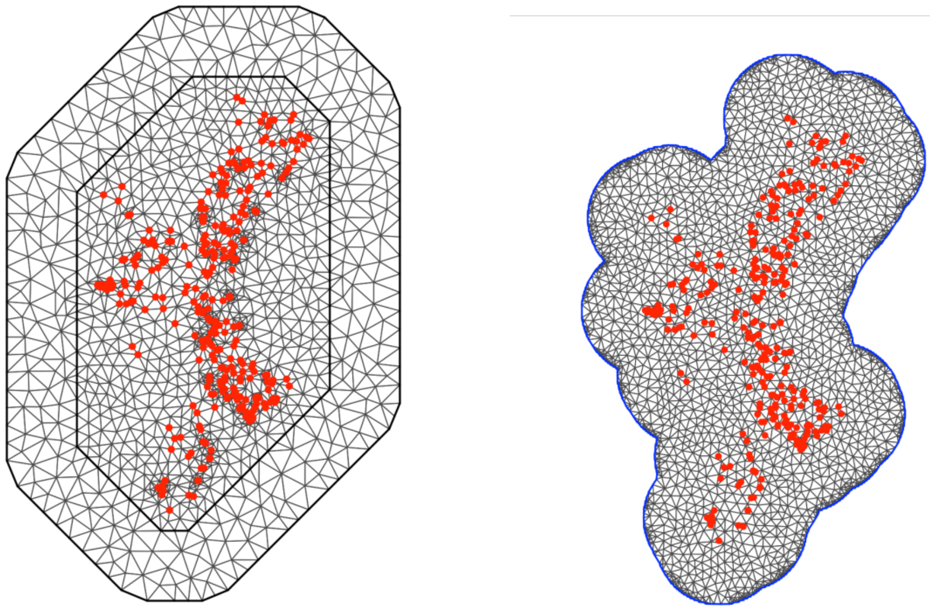

Figure 9: The meshes selected, with the previous steps.

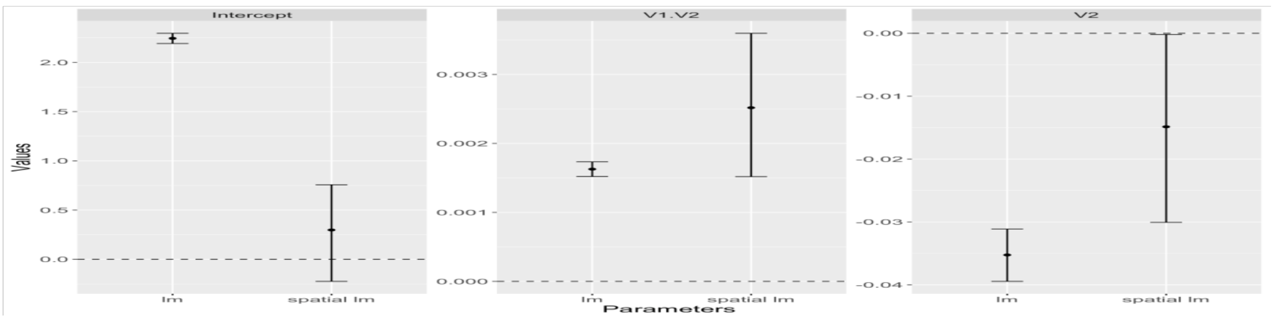

Figure 10: Comparison of parameters with and without spatial effect.

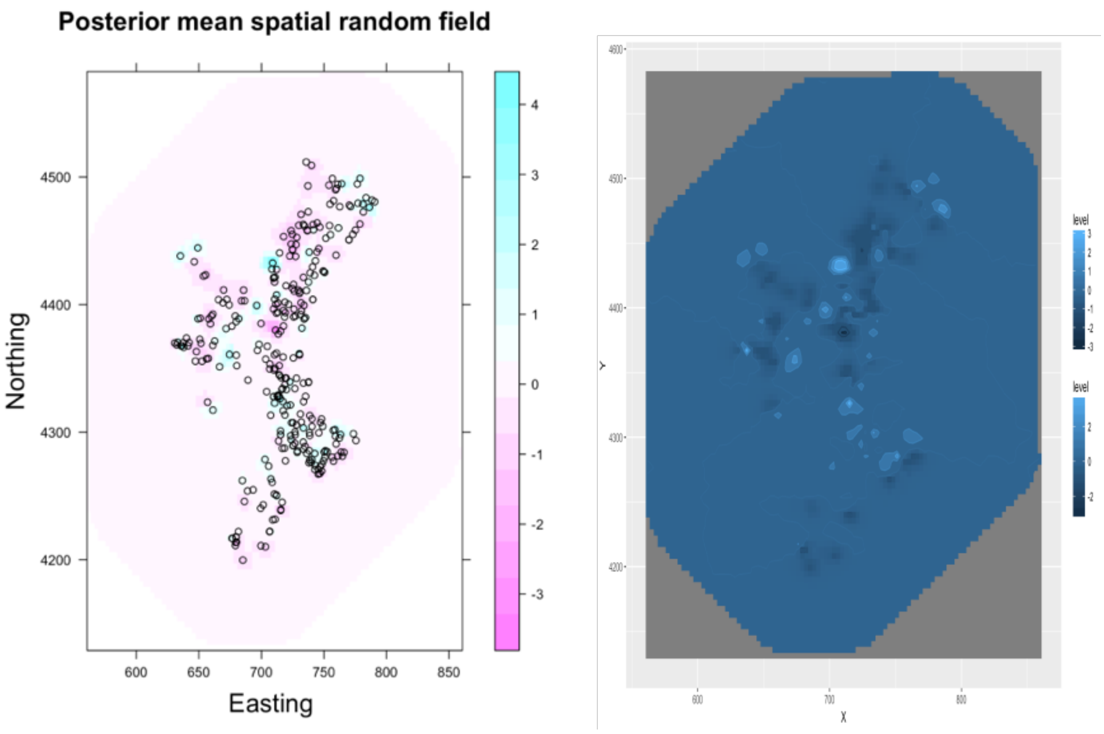

Figure 11: Spatial effect.

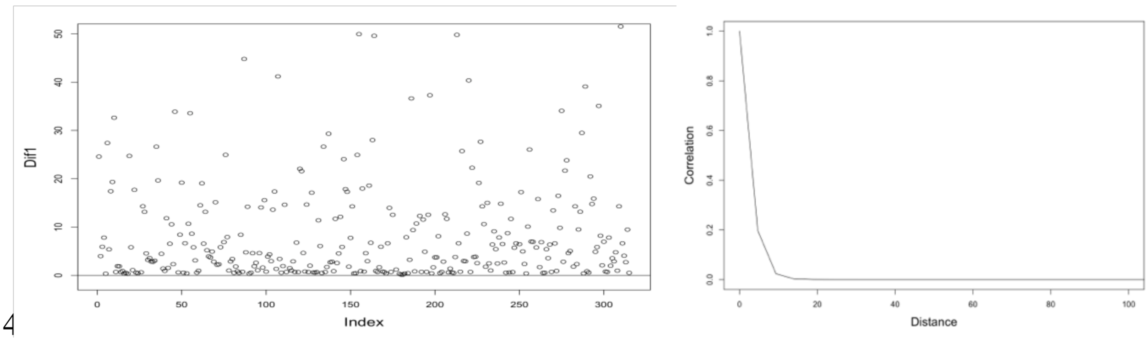

Figure 12: Differences between real and model data (left), and Correlation–distance (right).

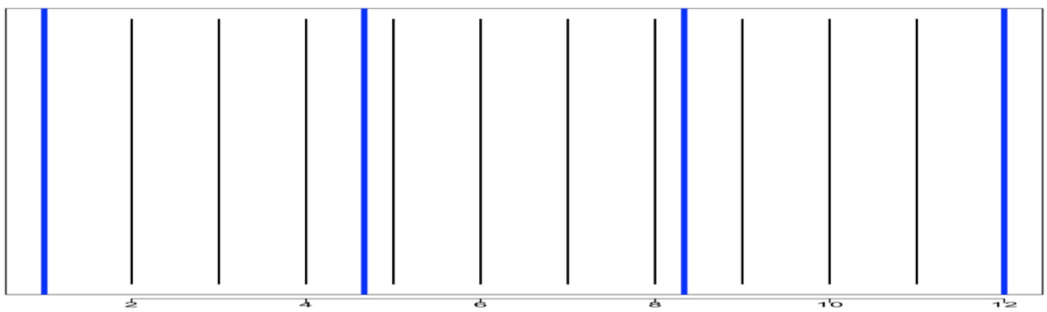

Figure 13: 1d mesh for the emergency calls of wildfires.

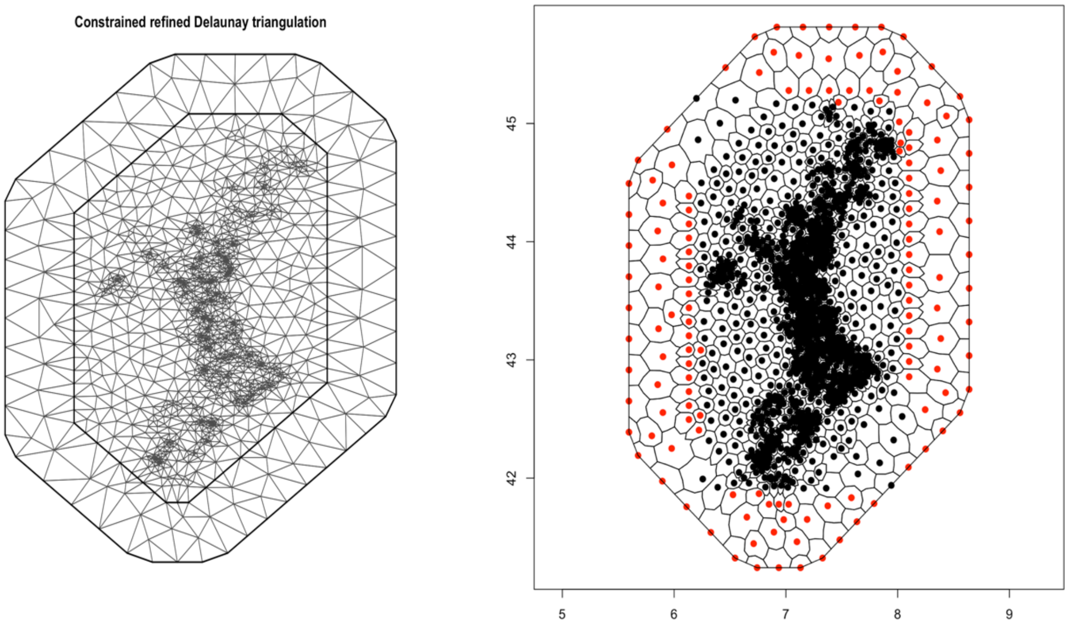

Figure 14: Spatial 2d mesh (left) regionalised mesh (right) without contour.

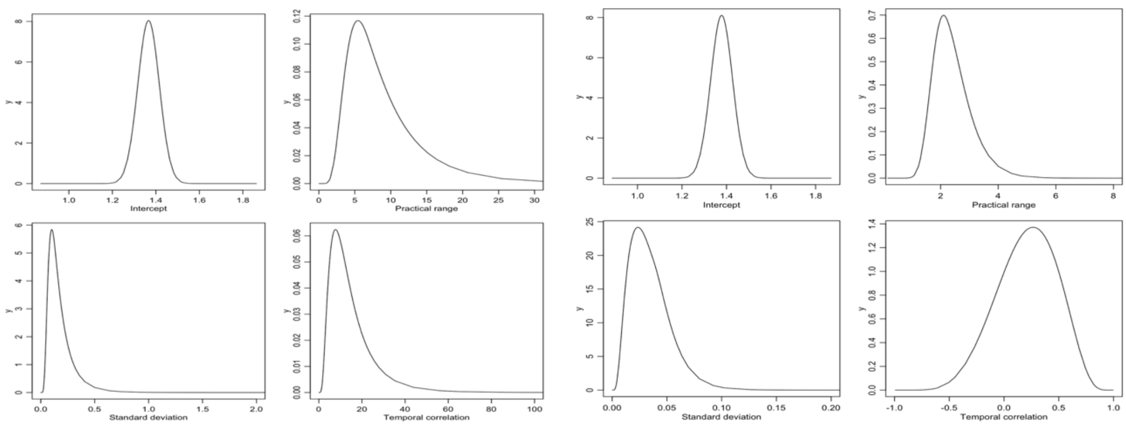

Figure 15: Parameters of M1 and M2.

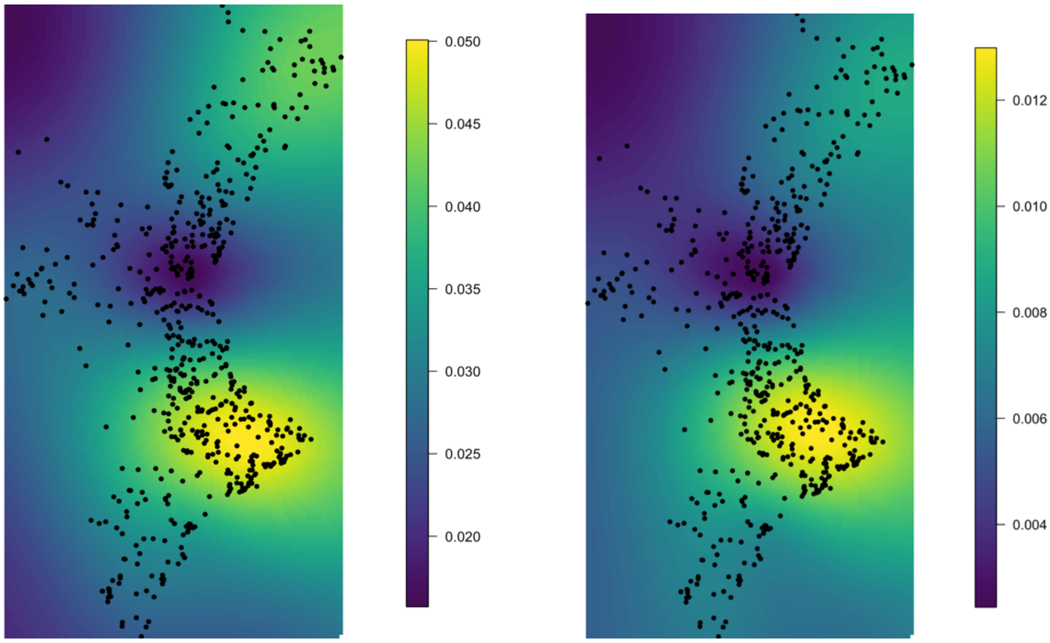

Figure 16: Modelled total burned area with the model M1 (left) and model M2 (right).

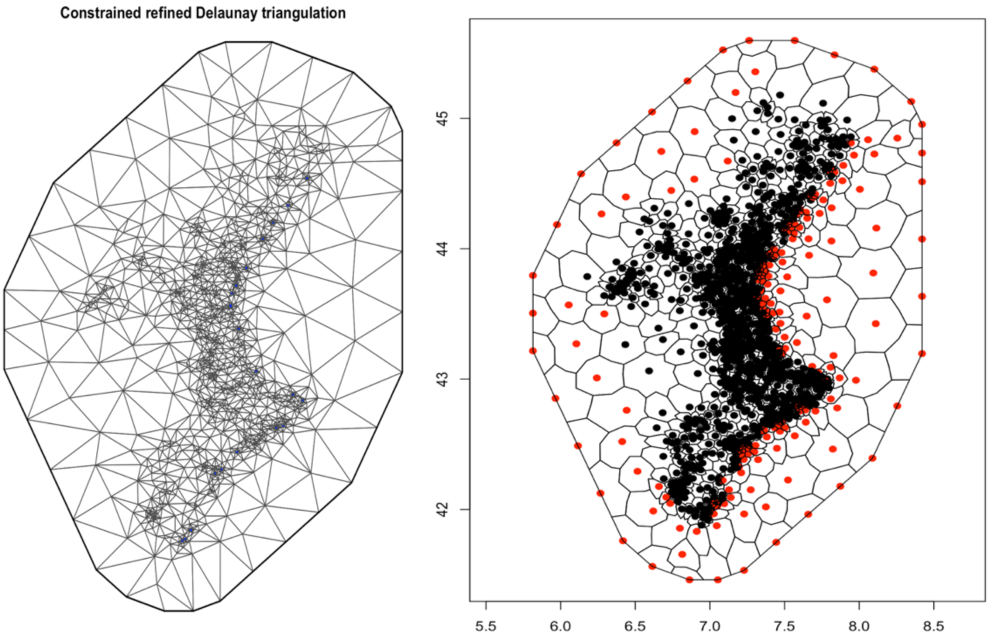

Figure 17: Spatial 2d mesh (left) regionalised mesh (right).

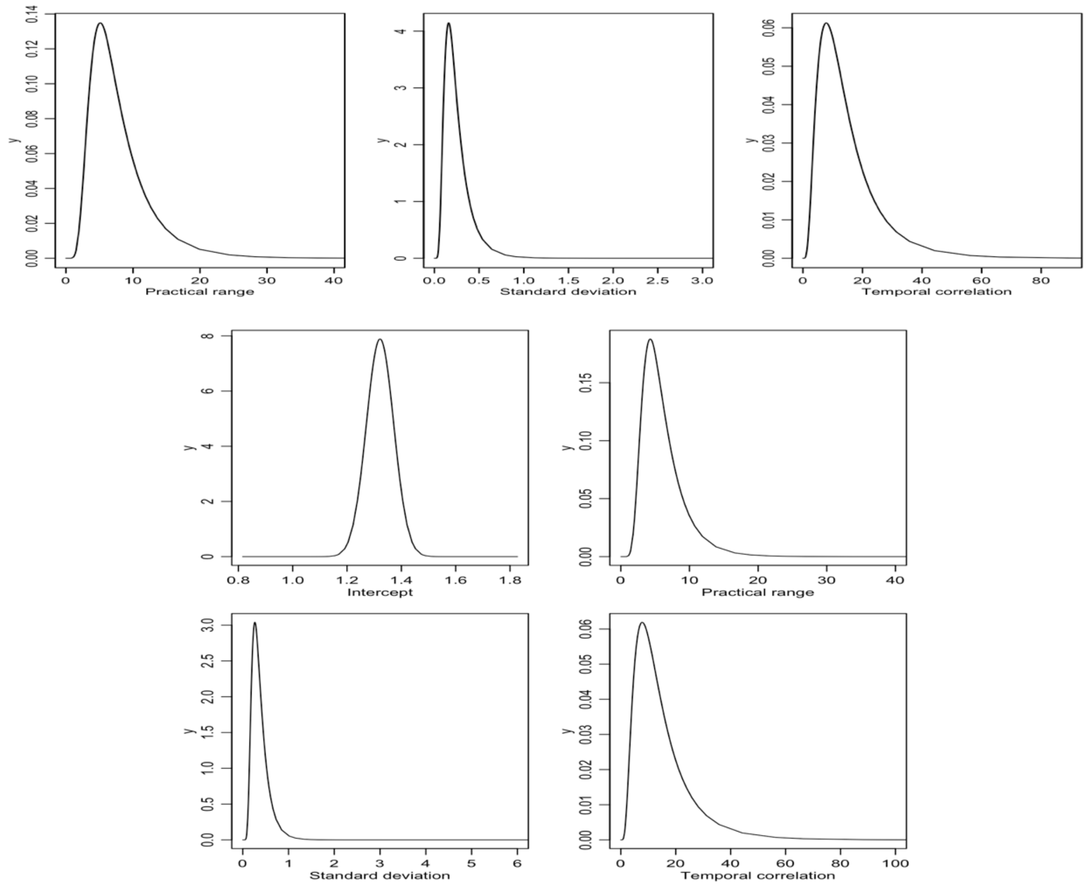

Figure 18: Parameters of M3 (up) and 4 (down).

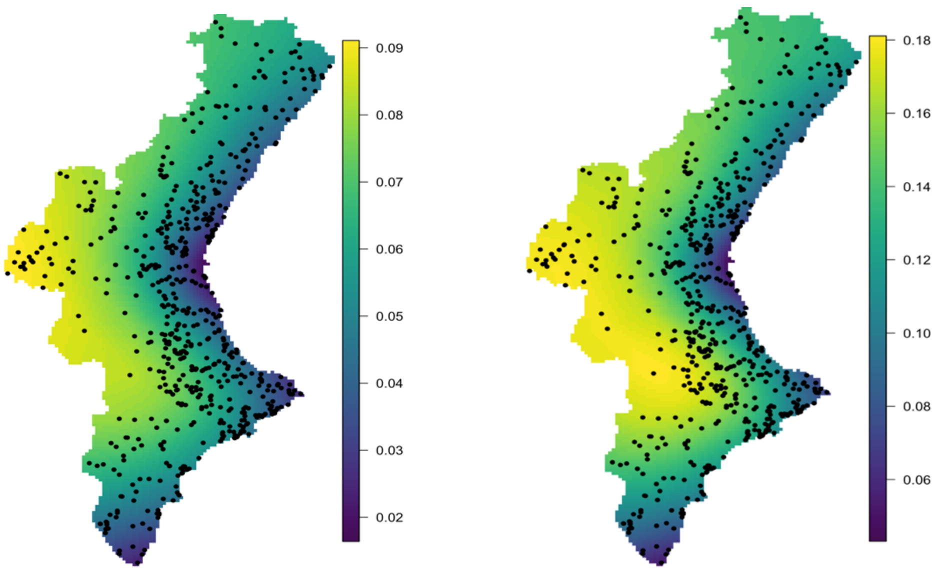

Figure 19: Modelled total burned area with the model M3 (left) and model M4 (right).

Tables:

	GVIF - Step 1	GVIF - Step 2
typology	1.060767	1.060414
group typology	2.441872	
cause	1.211273	1.080824
causative	2.444984	
days last rain	1.256075	1.221050
temp_max	1.520976	1.490312
Relative_H	1.412322	1.385338
wind speed	1.435472	1.435213
wind direction	1.690188	1.676258
combined model	1.182671	1.171493
danger degree	1.412806	1.393972
fire type	1.103011	1.100342
total tree	1.114397	1.107647
total no tree	1.199396	1.194232

Table 1: Values of GVIF in the two steps.

	Model 1	**Model 2**
Non-spatial effect	1.7078	1.3112
Spatial effect	14.7019	23.9632

Table 2: Computational Time for each model in seconds.

	β_0 Mean [0.025 quant, 0.975 quant]	β_1 Mean [0.025 quant, 0.975 quant]	β_2 Mean [0.025 quant, 0.975 quant]
M1 - Non-Spatial effect	2.2439 [2.1916, 2.2955]	-0.0352 [-0.0394, -0.0311]	0.0016 [0.00152, 0.0017]
M1 - Spatial Effect	0.2974 [-0.2243, 0.7566]	-0.0148 [-0.0300, -0.0002]	0.0025 [0.0015, 0.0036]
M2 - Non-Spatial effect	2.2439 [2.1917, 2.2955]	-0.0352 [-0.0394, -0.0311]	0.0016 [0.0015, 0.0017]
M2 - Spatial Effect	0.2384 [-0.3764, 0.7848]	-0.0163 [-0.029, -0.0039]	0.0033 [0.0022, 0.0045]
	κ = 0.5166625	σ_u = 2.467049	r = 5.694205

Table 3: Parameters of the model.

	DIC	**WAIC**
M1 - Non-spatial effect	5183.299	4574.274
M1- Spatial effect	5184.681	4573.108
M2 - Non-spatial effect	1104.704	1131.896
M2- Spatial effect	-Inf	-Inf

Table 4: DIC and WAIC.

	β_0 Mean [0.025 quant, 0.975 quant]	**β_1 (temporal covariate)** Mean [0.025 quant, 0.975 quant]	**DIC**
M1: y ~ β_0 + f(s, model = spde)	1.366 [1.269, 1.464]	14.861 [3.240, 43.698]	-1878.22
M2: y ~ β_0 + f(s, model = spde, group = s.group, control.group = list(model = 'ar1', hyper = list(theta = pcrho)))+ f(covariate, model = "rw2")	1.378 [1.282, 1.475]	14.734 [3.219, 43.328]	-1865.49

Table 5: Parameters of the models M1 and M2

	β_0 Mean [0.025 quant, 0.975 quant]	**β_1 (temporal covariate)** Mean [0.025 quant, 0.975 quant]	**DIC**
M3: y ~ f(s, model = spde)	-	15.054 [3.258, 44.191]	-1943.04
M4: y ~ β_0+ f(s, model = spde, group = s.group, control.group = list(model = 'ar1', hyper = list(theta = pcrho)))+ f(covariate, model = "rw2")	1.322 [1.222, 1.421]	15.007 [3.263, 44.248]	-1987.28

Table 6: Parameters of the models M3 and M3.