Two-step procedure to model site-specific herbicide soil persistence

Giannini Kurina F.[1,2]*, Hang S.[1], Balzarini M.[1,2], Borello, J.[3]
[1]Facultad de Ciencias Agropecuarias, Universidad Nacional de Córdoba; [2]CONICET-UFYMA Unidad de Fitopatología y Modelización Agrícola; [3]CEPROCOR, Córdoba

Abstract

Soil herbicide persistence is the length of time the herbicide molecule remains active in soil and it is crucial to describe risks of diffuse contamination in agriculture. Persistence is characterized by "half-life", which is the time it takes to reach half of the initial concentration supplied to soil. Half-life is estimated as a function of the dissipation curve parameters. Analytic quantification is costly for obtaining dissipation curves at many sites. Methodological tools to predict half-life in a continuous spatial domain, from a sample of dissipation curves, become crucial in regional studies. Since herbicide persistence in the environment depends on sites variables, model-based predictions of half-life as function of environmental features, are pursuit. The objective of this work was to design a statistical workflow for digital modeling of soil herbicide persistence at regional scale. From a regional soil survey, a sample of sites was drawn using the cLHS method. Samples were fortified with the herbicide atrazine and incubated for 21 days. Herbicide concentrations were measured at days 0,3,7,14 and 21 on each soil by liquid chromatography coupled to tandem mass spectrometry (LC-MS / MS) using QuEChERs. A two-step procedure was proposed for digital mapping of herbicide persistence in the environment. First, an exponential model with a random site effect, associated to the decay rate, was fitted to derive atrazine half-life for each sampled soil. Second, a Bayesian regression with a site random effect relating the resulting half-life values with soil and land-use values was adjusted to predict the spatial distribution of atrazine persistence at un-sampled sites for mapping. The addition of a random effect on the decay rate produced a better fit than a fixed exponential model and allowed us to explore half-life variability among soils. Atrazine persistence was mainly explained by the agricultural use of land (sites with previous grass crops had higher decay rates than other land-uses). The two-step procedure made possible to accurate map the spatial variability of atrazine persistence in soil and enhanced its environmental understanding.

Keywords: Non-linear mixed models, Longitudinal data, Spatial prediction models, R-INLA

Introduction

Anticipate fate of pesticides released into the environment is necessary to minimize adverse impacts of agricultural activities. Mobility and persistence of a pesticide in soil are dynamical processes that explain the potential losses of a pesticide into environment. A measurement named as half-life characterizes herbicide soil persistence and represents the time necessary to reach half of the initial pesticide concentration supplied to soil. A variety of chemical, physical and biological soil properties affect the persistence of herbicides in soil. It can occur through abiotic and biotic mechanisms, being microbial degradation, the most relevant aspect contributing pesticides soil dissipation (Mamy et al., 2015). Such degradation depends on the acting microbial community structure, its functionality and its metabolic activity (Nannipieri et al., 2003) which are influenced by agriculture land uses. The existing microflora adaptation to mineralize some herbicides frequently used on grass crops has been widely documented. Therefore, the understanding of herbicide persistence through the analysis of half-life variability across sites in a region requires the knowledge of site-specific characteristics, mainly those related to soil chemical and physical properties, climate, and land-uses. The regression of half-life values on environmental features would provide knowledge and predictions to support decisions focused on mitigating impacts of pesticides released into soil.

Half-life quantification for a soil under lab conditions demands the modeling of dissipation curves from data obtained by incubation experiments where the concentration of an herbicide supplied to soil is repeatedly monitored in time. The exponential function is commonly used to describe the decay of the pesticide concentration in time (Gustafson and Holden, 1990). Once the dissipation curve is estimated the half-life is derived from a simple function of its parameters. Unfortunately, this procedure cannot be applied to a large collection of soil samples because the herbicide analytic quantification along time is costly. Therefore, methodological tools to predict persistence in a continuous spatial domain, from a sample of dissipation curves, become crucial in regional

studies. Assuming that a sample of different sites is available, we propose to address model fitting with non-linear mixed models (Davidian, 2017) including random effects associated with one or more of the exponential model parameters to account for serial correlations among herbicide concentrations coming from the same site. Mixed models would broaden the possibilities of inference, since it would be possible to infer the average population soil dissipation but also on the site-specific dissipation. A collection of half-life values will be obtained after the site-specific dissipation curves are obtained. This way it would be possible to quantify the variability of the dissipation process among soils.
Since herbicide persistence in the environment depends on sites variables, model-based predictions of half-life as function of environmental features, will be pursuit. The spatial variability of the dissipation process can be studied through the characterization of the spatial variability of site-specific half-life using Bayesian regressions with random site effects. Bayesian framework allows the implementation of algorithms to speed the fitting of spatial regression models. The computation objectives are the marginal posterior distributions for each fixed model parameter (regression coefficients) and for each hyper-parameter (the variance component of a random site effect, the variance component of the model error and the parameters of the function modeling the spatial dependence). Rue et al. (2009) proposed deriving the posteriori distribution of such spatial regression model by using an Integrated Nested Laplace Approximation. Based on the implementation of this approach in the language of R programming (R-INLA), applications of Bayesian regression have become popular for spatial modeling (Blangiardo and Cameletti, 2015). R-INLA handles latent Gaussian models in which fixed effects and structured and unstructured Gaussian random effects are combined linearly. The sparse precision matrices such us the covariance matrix of spatially correlated data are computational achieved using Stochastic Partial Differential Equations (SPDE) (Lindgren and Rue, 2015; Krainski et al., 2018).
The aim of this work was to couple sampling estimation of site-specific half-life with Bayesian spatial regression for digital mapping of herbicide persistence in environment at a regional scale.

Materials and Methods

Data & Proposed Statistical Workflow

The dissipation experiments were made for the herbicide atrazine on a sample of soils obtained from a regional soil survey of Cordoba, Argentina that counts with a broad ancillary environmental information (edaphic, climatic and agronomic) for a total of 355 geo-referenced sites. A samples of n=60 sites were selected by conditionate Latin Hypercube method (cLHS, Minasny and McBratney, 2006) to capture the underlying edaphoclimatic variability in the study region (Brus, 2019; Minasny and McBratney, 2006). Dissipation data were obtained fortifying with atrazine and incubating for 21 days at 28ºC. Each soil was fortified with a concentration of atrazine equivalent to an application of 3 L ha^{-1}, and the amount of water to reach 80% of water holding capacity. Atrazine concentration were measured at days 0,3,7,14 and 21 by liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) using QuEChERs procedure to extract and clean up.
A two-step procedure was performed for data analysis: First, we estimated the decay of atrazine along time with a mixed nonlinear model including a random site effect on the decay rate; and second, a site-specific statistical modeling of half-life as function of soil, climatic and land-use covariates by a Bayesian spatial regressions.

Step 1. Half-life derived from site-specific dissipation curves

For each soil sampled, atrazine concentration along the dissipation experiment was modeled as follows:

$$C_i(t) = C_0\, e^{-(k+u_i)\, t} \qquad u \sim N(0, \sigma_u^2)$$

where $C_i(t)$ is the percentage of atrazine concentration (expressed as relative to the observed concentration at time 0) remaining at the time t (expressed as days from the incubation beginning) for the soil i, k is the decay rate of the population of soils, u_i is the random effect for the soil i expressed as the deviation from de population decay rate. C_0 is the intercept representing the concentration at t=0. The random effect u is assumed normally distributed with zero mean and variance σ_u^2 and assumed to be independent of the error term. Direct likelihood was used to estimate the curves with the *nlme* package in R (Pinheiro et al., 2017). The maximum likelihood estimates for the decay rate in the population of sites and the best linear unbiased prediction (BLUP, u_i) of the random site effect. Significance tests for σ_u^2 was evaluated by comparing the nonlinear mixed model with the analogous nonlinear fixed effect model using a likelihood ratio test (LRT) with 0.5 degrees of freedom (Molenberghs & Verbeke, 2007). AIC and BIC criteria were also used for model selection. Site-specific half-life ($t_{1/2_i}$) was calculated from the dissipation curve at each soil:

$$t_{1/2_i} = \frac{ln(0.5)}{-(k + u_i)}$$

Step 2. Modeling site-specific half-life and spatial prediction

A selection of explanatory variables was made based on deviance information criteria (DIC, (Huang et al., 2017)) to fit a Bayesian regression including random site effects spatially correlated. The distribution of log half-life for the i[th] soil is:

$$\ln(t_{1/2_i}) \sim N(\eta_i, \sigma_e^2) \qquad \eta_i = \beta_0 + \sum_{j=1} x_{ij}\beta_j + \xi(s_i)$$

where β_0 is the intercept; β_j is the fixed regression coefficient for the explanatory variable x_j; x_{ij} is the value of x_j at site i; and $\xi(s_i)$ is a random site effect that is assumed to be the realization of a latent Gaussian field $GF\xi(s_i) \sim MVN(0, \Sigma)$, with Σ being the covariance matrix of site effects defined by the Matérn spatial covariance function. Within R-INLA, the estimation of Σ inverse (precision matrix) is solved by SPDE (Krainski et al., 2018). Predictive ability was performed with a Jackknife validation to obtain a global measurement of error (root mean square prediction error, RMSPE) and a pointwise site-specific error (site-specific residual expressed as percentage of the site mean, SSE). Using the fitted model, atrazine half-life was obtained for a prediction grid of 2.5×2.5 km resolution. The random spatial effect estimated by SPDE was projected in the prediction domain following Krainski et al., 2018. The spatial prediction was re-expressed in the original scale and uncertainty of predictions were obtained from the 95% credibility intervals of the prediction at each site.

Results and Discussion

Dissipation curve estimation

The estimation of the dissipation curve, with and without a random site effect on the decay rate, are shown in Table 1. The AIC and BIC information criteria decreased under the random effect model, and the LRT test confirms that the model considering a random effect on the decay rate has a better fit. Random effect on other parameters of the exponential model did not improves the fit (data not showed). There is high variability among soils in the dissipation process (Fig 1). The population average curve describes a global behavior of atrazine persistence of Córdoba soils where half-life lasts 10 days ranged between 3 to 100 days with 75% of the soils associated with atrazine half-life lesser than 19 days. From an environmental point of view this is an encouraging result since it implies a low herbicide persistence in most soils. The fitted model includes a high serial correlation among atrazine concentrations measured in successive days of incubation of the same soil sample. The random site effect on the decay rate of the exponential function, improves the estimation of the dissipation curve and consequently the accuracy of half-life extending the possibilities of inference, since it allows inferring an average dissipation curve (in the population of soils samples) as well as the site-specific behavior (Fig 1).

Table 1. Atrazine dissipation curve

	Fixed-effect model	Random-effect model
Model parameters		
C_0	98.11 ± 2.12	100.39 ± 0.88
k	-0.056 ± 0.003	-0.072 ± 0.007
σ_u^2	-	3.9×10^{-5}
Goodness of fit		
AIC	2748.46	2426.62
BIC	2759.66	2441.53
Loglikelihood	-1371.23	-1209.31
LRT		161.92 (p-value<0,001)

C_0, intercept; k, decay rate; σ_u^2 variance of random effects; AIC, Akaike information criterion; BIC, Bayesian information criterion; LRT, likelihood ratio test.

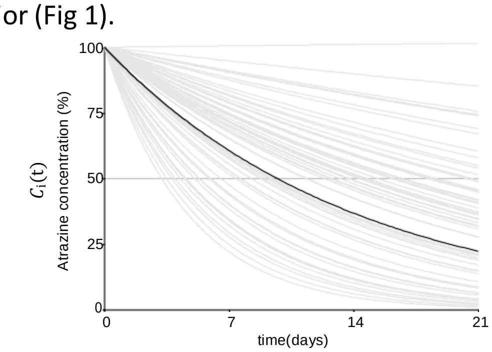

Figure 1: Atrazine dissipation curves, population curve (in bold) and site-specific curves (in gray).

Model fitted for digital mapping of half-life

DIC prioritized the contribution of TvsPP (hydrological balance), potassium (K), clay content, electrical conductivity (EC) and land-use of soil. Table 2 shows the fitted regression model; the positive regression coefficient for other land-uses with respect to grass crops indicate a lower atrazine half-life under grasses were the herbicide is historically applied (reference category). Soils covered by forest showed the highest values of atrazine half-life. Higher EC also increases atrazine half-life. On the contrary higher K and clay decreases the persistence of the molecule in soil. The global prediction error was 22.7%. Figure 2 shows the spatial model-based prediction of atrazine half-life within the spatial domain (Cordoba, Argentina 165.321 km^2) and its standard deviation.

Table 2. Spatial regression model for atrazine hal-life on environmental variables

Model	Mean ± SD
Regression coefficients	
Intercept	4,31±0,75
TvsPP [C°mm^{-1}]	-5,82±2,84
K [ppm]	-0,0007±0,0004
Caly [%]	-0,02±0,01
CE [dSm^{-1}]	0,27±0,12
Crop without grass†	0,26±0,15
Pasture†	0,70±0,27
Forest †	1,35±0,26
Hyperparameters ††	
Range	24076.08
Sill	0.38
Cross-validation	
RMSPE relative to the mean	22,7%
Sites with RMSPE <25%	70,0%

† Difference of means in terms of $ln(t_{1/2})$ between grass crops and the other land uses. ††Hyperparameters of Matérn function for the spatial autocorrelation function in the random effect. SD, standard deviation derived from the marginal posterior distribution of each parameter. RMSPE, root mean square prediction error relative to the mean.

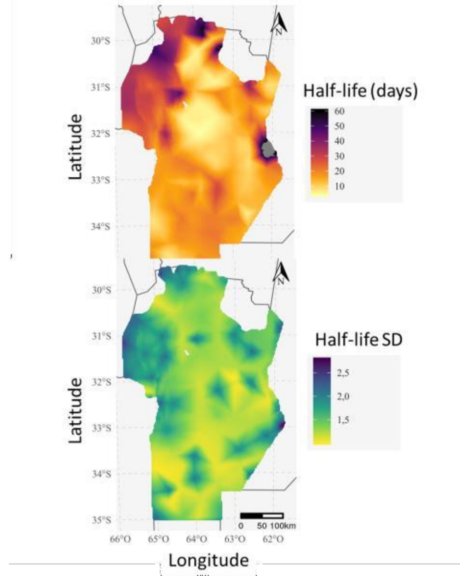

Figure 2: Atrazine half-life in soil of Cordoba, Argentina

Conclusion

The addition of a random effect in the exponential model of herbicide dissipation in soil provided a useful tool to explore persistence (herbicide half-live) across a region. Bayesian spatial regression of half-life on environmental variables allowed to obtain predictions of half-live in a continuous domain. The two-step modeling workflow made possible to accurate map the spatial variability of atrazine persistence in soil and enhanced its environmental understanding.

References

Blangiardo, M. and Cameletti, M. (2015) *Spatial and Spatio-Temporal Bayesian Models with R-INLA*. John Wiley & Sons.

Davidian, M. (2017) *Nonlinear Models for Repeated Measurement Data*. Routledge.

Gustafson, D.I. and Holden, L.R. (1990) Nonlinear pesticide dissipation in soil: a new model based on spatial variability. *Environmental Science & Technology*, **24**, 1032–1038.

Huang, J., Malone, B.P., Minasny, B., McBratney, A.B. and Triantafilis, J. (2017) Evaluating a Bayesian modelling approach (INLA-SPDE) for environmental mapping. *Science of the Total Environment*, **609**, 621–632.

Krainski, E.T., Gómez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D., Simpson, D., et al. (2018) *Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and INLA*. Chapman and Hall/CRC.

Lindgren, F. and Rue, H. (2015) Bayesian spatial modelling with R-INLA. *Journal of Statistical Software*, **63**, 1–25.

Mamy, L., Patureau, D., Barriuso, E., Bedos, C., Bessac, F., Louchart, X., et al. (2015) Prediction of the fate of organic compounds in the environment from their molecular properties: A review. *Critical Reviews in Environmental Science and Technology*, **45**, 1277–1377.

Minasny, B. and McBratney, A.B. (2006) A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Computers and Geosciences*, **32**, 1378–1388.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., Heisterkamp, S., Van Willigen, B., et al. (2017) Package 'nlme.' *Linear and Nonlinear Mixed Effects Models, version*, 1–3.

Rue, H., Martino, S. and Chopin, N. (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)*, **71**, 319–392.