

# **Metodologías para el análisis espacial de retención y disipación de herbicidas en suelos**

*Tesis presentada para optar al título de Doctor de la Universidad de Buenos Aires,  
Área Ciencias Agropecuarias*

**Franca, Giannini Kurina**

Ingeniera Agrónoma- Universidad Nacional de Córdoba - 2015

Lugar de trabajo: Facultad de Ciencias Agropecuarias, UNC



**FAUBA**

Escuela para Graduados Ing. Agr. Alberto Soriano  
Facultad de Agronomía – Universidad de Buenos Aires



**COMITÉ CONSEJERO**

Directora

**Mónica Graciela, Balzarini**

Ingeniera Agrónoma (Universidad Nacional de Córdoba)  
Magister Scientiae en Biometría (Universidad de Buenos Aires)  
Philosophy Doctor (Louisiana State University)

Codirectora

**Susana Beatriz, Hang**

Ingeniera Agrónoma (Universidad Nacional de Córdoba)  
Magister en Ciencias del Suelo (Universidad Nacional del Sur. Bahía Blanca)  
Doctora en Ciencias Agropecuarias (Universidad Nacional de Córdoba)

Consejero de Estudios

**Gervasio, Piñeiro Guerra**

Ingeniero Agrónomo (Universidad de la República)  
Doctor en Ciencias Agropecuarias (Universidad de Buenos Aires)

**JURADO DE TESIS**

JURADO

**Antonio, Paz González**

Licenciado en Biología (Universidad de Santiago de Compostela)  
Magister en Ciencias del Suelo (Universidad de Santiago de Compostela)  
Philosophy Doctor (Universidad de Santiago de Compostela)

JURADO

**María Valeria, Amé**

Licenciada en Bioquímica (Universidad Nacional de Córdoba)  
Doctora en Ciencias Químicas (Universidad Nacional de Córdoba)

JURADO

**Susana Beatriz, Perelman**

Ingeniera Agrónoma (Universidad de Buenos Aires)  
Magister Scientiae en Biometría (Universidad de Buenos Aires)

Fecha de defensa de la tesis: 27 de marzo de 2020

*Dedicatoria*

*A quienes piensan el ambiente como una matriz diversa compleja y fascinante, cuyos componentes nos damos el lujo de llamar recursos.*

## *Agradecimientos*

*A quienes confiaron en la potencialidad de mi aporte brindando el suyo.  
Al sistema científico-educativo público y nacional, por ser el conjunto de  
instituciones motoras de este proceso. Especialmente a quienes lo  
construyen, sostienen y disputan cada día.  
A las grandes brujas en esta causa.*

## Declaración

Declaro que el material incluido en esta tesis es, a mi mejor saber y entender, original producto de mi propio trabajo (salvo en la medida en que se identifique explícitamente las contribuciones de otros), y que este material no lo he presentado, en forma parcial o total, como una tesis en esta u otra institución.

## Publicaciones derivadas de la tesis

Giannini Kurina, F., Hang, S., Cordoba, M.A., Negro, G.J. and Balzarini, M.G. (2018) Enhancing edaphoclimatic zoning by adding multivariate spatial statistics to regional data. *Geoderma*, 310, 170–177. (Capítulo II)

Giannini Kurina, F., Hang, S., Macchiavelli, R., and Balzarini, M. (2019). Spatial predictive modelling essential to assess the environmental impacts of herbicides. *Geoderma*, 354, 113874. (Capítulo I y III)

Giannini Kurina, F., Balzarini, M., Rampoldi, A. y Hang, S. (2019) Site-specific data on herbicide soil retention and ancillary environmental variables. *Data in Brief*, 104754.

# Índice de contenidos

Introducción general.....	2
Parámetros de la dinámica de herbicidas en suelos .....	2
Desafíos metodológicos.....	5
Objetivo general .....	9
Objetivos específicos.....	9
Capítulo I: Análisis de la variabilidad y predicción espacial.....	11
Introducción.....	11
Construcción de un modelo predictivo.....	13
Marcos teóricos para la predicción espacial .....	17
Modelación de la autocorrelación espacial.....	18
Modelos de regresión para predicción espacial .....	24
Enfoque frecuentista .....	24
Enfoque bayesiano.....	26
Enfoques desde el aprendizaje automático .....	27
Comparación de predicciones obtenidas desde distintos modelos .....	29
Capítulo II: Zonificación según variabilidad edafoclimática.....	33
Introducción.....	33
Materiales y Métodos .....	36
Área de estudio .....	36
Base de datos .....	37
Preprocesamiento de datos .....	38
Protocolo estadístico.....	39

Paso 1: Obtención de variables sintéticas que capturan correlación espacial .....	39
Paso 2: Conglomerados de sitios .....	40
Validación del protocolo de zonificación .....	40
Resultados.....	41
Discusión .....	47
Metodología para la zonificación regional .....	48
Zonificación edafoclimática de la provincia de Córdoba .....	50
Conclusiones.....	53
Capítulo III: Predicción de retención de herbicidas en suelo.....	55
Introducción.....	55
Materiales y Métodos .....	60
Base de datos .....	60
Protocolo estadístico.....	62
Selección de variables predictoras .....	62
Ajuste de un modelo de regresión espacial para el coeficiente de adsorción.....	63
Mapeo de retención de glifosato y atrazina.....	64
Resultados y Discusión.....	65
Selección de variables predictoras .....	65
Predicción espacial y mapeo de retención .....	68
Distribución espacial del proceso de retención e implicancias ambientales .....	71
Conclusiones.....	76
Capítulo IV: Modelación de la disipación y su covariabilidad con la retención de herbicidas en suelo .....	77
Introducción.....	77
Materiales y Métodos .....	84
Base de datos .....	84

Selección de sitios e información auxiliar .....	84
Determinaciones de disipación .....	85
Protocolo estadístico.....	86
Estimación de la curva de disipación sitio-específica y cálculo de vida media .....	86
Modelación de la variabilidad espacial de la vida media.....	88
Variabilidad conjunta de retención y disipación.....	88
<b>Resultados y Discusión.....</b>	<b>89</b>
Curvas de disipación y cálculo de vida media .....	89
Variabilidad espacial de la vida media de atrazina en Córdoba.....	94
Variabilidad conjunta de retención y disipación a escala regional .....	99
Peligrosidad ambiental por el uso de atrazina en Córdoba .....	103
Conclusiones.....	107
<b>Síntesis .....</b>	<b>109</b>
Protocolos estadísticos para el mapeo regional de retención y disipación de herbicidas .....	111
Implicancias ambientales de la variabilidad en retención y disipación de herbicidas .....	114
Relevancia de las contribuciones.....	115
Futuras líneas de investigación.....	116
<b>Bibliografía .....</b>	<b>119</b>
<b>Anexo .....</b>	<b>133</b>
Anexo I: Descripción de variables edafoclimáticas.....	133

# Índice de cuadros

Cuadro 1. Comparación del desempeño estadístico de distintas estrategias metodológicas para construir modelos de predicción sitio-específica del coeficiente de adsorción de un herbicida en suelo a partir de características del suelo.-----	31
Cuadro 2. Variabilidad edafoclimática en la Provincia de Córdoba (n=355 sitios) -----	42
Cuadro 3. Autovalores de los análisis PCA clásico y PCA espacial.-----	43
Cuadro 4. Contribución de variables originales a las componentes principales -----	44
Cuadro 5. Índices de validación de agrupamiento (cantidad de clústers).-----	46
Cuadro 6. Diferencias entre las zonas delimitadas para variables externas.-----	46
Cuadro 7. Caracterización de zonas edafoclimáticas delimitadas en provincia de Córdoba -----	47
Cuadro 8. Estimación y validación del modelo de retención de herbicidas -----	70
Cuadro 9. Estimaciones de modelos exponencial de disipación de atrazina en suelos -	90
Cuadro 10. Modelo de disipación según uso del suelo -----	94
Cuadro 11. Modelo de disipación en función de variables de sitio-----	96
Cuadro 12. Retención y disipación de atrazina por zona edafoclimática -----	103
Cuadro 13. Distribución de sitios según retención y persistencia de atrazina-----	103

## Índice de figuras

Figura 1. Semivariograma. Parámetros rango, umbral y nugget .....	19
Figura 2. Área de estudio y sitios de muestreo. ....	38
Figura 3. Protocolo de zonificación regional .....	41
Figura 4. Zonificación edafoclimática de Córdoba: a) obtenida por PCA clásico b) obtenida por PCA espacial .....	45
Figura 5. Importancia de variables edáficas y climáticas para explicar la variabilidad del coeficiente de retención de atrazina según el método BRT y el índice DIC.....	66
Figura 6. Importancia de variables edáficas y climáticas para explicar la variabilidad del coeficiente de retención de glifosato según el método BRT y el índice DIC .....	66
Figura 7. Distribución espacial de retención de glifosato en Córdoba e incertidumbre asociada expresada como desvío estándar (abajo, izquierda) y como amplitud de intervalos de credibilidad (abajo derecha) .....	74
Figura 8. Distribución espacial de retención de atrazina en Córdoba e incertidumbre asociada expresada como desvío estándar (abajo, izquierda) y como amplitud de intervalos de credibilidad (abajo derecha) .....	75
Figura 9. Curvas de disipación de atrazina en suelos. Curvas sitio-específicas (arriba). Curvas medias según uso de suelo (abajo).....	91
Figura 10. Vida media de atrazina en suelos de la provincia de Córdoba, Argentina. ...	92
Figura 11. Variabilidad espacial en valores predichos de vida media de atrazina según características del sitio en la provincia de Córdoba (arriba). Mapas de incertidumbre de	

la predicción (abajo): desvío estándar (izquierda) y amplitud de intervalos de credibilidad (derecha).....	98
Figura 12. Correlación entre retención medida por el parámetro <i>Koc</i> y persistencia de atrazina en suelos de Córdoba. En colores se representan las distintas categorías de propensión a la lixiviación del herbicida según el índice GUS .....	100
Figura 13. Propensión a la lixiviación de atrazina según GUS.....	101
Figura 14. Mapa de correlaciones cruzadas espaciales entre retención medida por el parámetro <i>Koc</i> y persistencia de atrazina en Córdoba .....	102
Figura 15: Diseño metodológico para explicar variabilidad espacial marginal y conjunta de los procesos de retención y disipación a escala regional .....	113

## Abreviaturas

AIC: criterio de información de Akaike, del inglés “*Akaike Information Criteria*”

Al (Ox): óxidos de Aluminio

BIC: criterio bayesiano de Schwarz, del inglés “*Bayesian Information Criteria*”

BLUP: mejor predictor lineal insesgado, del inglés “*Best Linear Unbiased Predictor*”

BRT: árboles de regresión mejorados por remuestreo, del inglés *Boosting Regression Tree*

CART: árboles de clasificación y regresión, del inglés “*Classification And Regression Trees*”

CC: Capacidad de Campo

CE: Conductividad Eléctrica

cHLS: Muestreo hipercubo latino condicionado, del inglés “*conditioned Latin Hypercube Sampling*”

CIC: Capacidad de Intercambio Catiónico

COT: Carbono Orgánico Total

CV: Coeficiente de Variación

DIC: criterio de información por deviance, del inglés “*Deviance Information Criterion*”

Fe (Ox): óxidos de hierro

GBR: árboles de regresión generalizados, del inglés “*Generalized Boosting Regression*”

GUS: índice de propensión a la contaminación por lixiviación, del inglés “*Grounwater Ubiquity Score*”

HPLC: cromatografía líquida de alta precisión, del inglés, “*High Performance Liquid Chromatography*”

INLA: aproximación anidada integrada de Laplace, del inglés, “*Integrated Nested Laplace Aproximation*”

$k$ : constante de velocidad específica

$K_d$ : coeficiente de adsorción

$K_{oc}$ : coeficiente de adsorción por unidad de carbono orgánico de suelo

LRT: cociente de verosimilitud, del inglés "*Likelihood Ratio Test*"

MCMC: métodos de simulación por cadenas de Markov Monte Carlo, del inglés "*Markov Chain Monte Carlo*"

ML: máxima verosimilitud, del inglés "*Maximum Likelihood*"

MLM: Modelos Lineales Mixtos

MOS: Materia Orgánica de Suelo

NT: Nitrógeno Total

PDA: Detector de arreglos de Diodo

PLSR: Regresión por mínimos cuadrados parciales, del inglés "*Partial Least Square Regression*"

PP: precipitación

REML: Máxima verosimilitud restringida, del inglés "*REstricted Maximum Likelihood*"

RFR: regresión por bosques aleatorios, del inglés "*Random Forest Regression*"

RLM: Regresión Lineal Múltiple

RMSE: raíz cuadrada del error cuadrático medio, del inglés "*Root Mean Square Error*"

RMSPE: raíz cuadrada del error cuadrático medio de predicción, del inglés "*Root Mean Square Prediction Error*"

RSV: varianza estructural relativa, por el término en inglés "*Relative Structural Variance*"

sPC: componente principal espacial, del inglés "*spatial Principal Component*"

SPDE: ecuaciones diferenciales parciales estocásticas, del inglés "*Stochastic Partial Differential Equations*"

$t_{1/2}$ : vida media del herbicida en suelo

$T_m$ : temperatura media anual

$T_{vsPP}$ : indicador del balance hídrico (cociente entre pluviometría anual acumulada y temperatura media)

WLS: mínimos cuadrados ponderados, del inglés "*Weighted Least Squares*"

# Metodologías para el análisis espacial de retención y disipación de herbicidas en suelos

## Resumen

Los procesos de retención y disipación que gobiernan la dinámica de herbicidas en suelo han sido estudiados en contextos experimentales puntuales que dificultan la inferencia a escalas mayores. El objetivo de esta tesis es generar modelos de predicción espacial para mapear variabilidad y covariabilidad espacial, a escala regional, de índices que parametrizan la dinámica de herbicidas en suelos. Se desarrolló un protocolo para zonificación regional, acoplado interpolación geoestadística de características de sitio y Análisis de Componentes Principales Espacial, identificando cuatro zonas edafoclimáticas en la provincia de Córdoba para la interpretación ambiental de los procesos de retención y disipaciones. Usando muestras de suelos distribuidas sistemáticamente en el territorio provincial, se determinaron en laboratorio coeficientes de adsorción ( $Kd$ ) de atrazina y de glifosato. Se construyeron curvas de disipación exponencial sitio-específica (modelos no lineales mixtos) desde las que se derivó la vida media ( $t_{1/2}$ ) de atrazina en cada suelo. Se diseñó una metodología de 3 pasos para mapeo digital de  $Kd$  y  $t_{1/2}$ : 1) Selección de variables de sitio con capacidad predictiva mediante aprendizaje automático y minimización de errores de predicción, 2) ajuste de un modelo de regresión bayesiano mediante aproximación integrada por Laplace (R-INLA) incluyendo efecto sitio aleatorio valuado mediante ecuaciones diferenciales estocásticas, y 3) mapeo de  $Kd$  y  $t_{1/2}$  a escala regional. Para atrazina, la variabilidad del  $Kd$  se explicó como función de balance hídrico, carbono orgánico total y arcilla, mientras que la retención de glifosato se explicó con óxidos de aluminio, pH y arcilla. La  $t_{1/2}$  atrazina fue menor en suelos con historial de cultivos de gramíneas. Se obtuvieron mapas, con medidas de incertidumbre, del comportamiento regional de  $Kd$  y  $t_{1/2}$  que fueron usados para aproximar la peligrosidad ambiental en función de la correlación espacial entre

retención y persistencia. Los protocolos metodológicos propuestos para el análisis estadístico podrían ser aplicados para mapear otras moléculas con potencial contaminante.

**Palabras claves:** Zonificación regional, Predicción espacial, Datos espaciales, Covariación, Atrazina, Glifosato.

# Methodological approaches for spatial analysis of soils herbicides retention and dissipation

## Abstract

Retention and dissipation are the main processes that leads the dynamics of herbicides in soil. These have been studied in pointwise experimental contexts that could limit the inference. The objective of this work is to build spatial prediction models able to map the spatial (co)variability of indices that parameterize the herbicide dynamics in soils. A protocol for regional zoning has been developed, coupling geostatistical interpolation o and Principal Component Analysis. Four edaphoclimatic zones were identified in Córdoba province, subsequently used for the environmental understanding of the retention and dissipation processes. The adsorption coefficient ( $Kd$ ) of atrazine and glyphosate was determined. Site-specific atrazine dissipation exponential curves were modeled by nonlinear mixed models to obtain half-life ( $t_{1/2}$ ). A 3-step workflow has been designed for digital mapping of  $Kd$  and  $t_{1/2}$ : 1) selection of explanatory variables with predictive ability by coupling through machine learning techniques and predictive criteria, 2) Bayesian regression model using an Integrated Nested Laplace Approximation (R-INLA), including a random site effect calculated by stochastic differential equations, and 3) mapping of  $Kd$  and  $t_{1/2}$  at a finer grid in a regional domain. The variability of atrazine  $Kd$  was explained as a function of two climatic variables, total organic carbon and clay, while glyphosate  $Kd$  was explained by aluminum oxides, pH and clay. Atrazine dissipation was higher in soils with a history of grass growing. Retention and dissipation maps were obtained and used to delimitate areas based on the spatial correlation between retention and persistence. The proposed methodological protocols for statistical analyses could be applied to map other molecules with polluting potential.

**Keywords:** Regional zoning, Predictive models, Spatial data, Covariation, Glyphosate, Atrazine.

# Introducción general

## Parámetros de la dinámica de herbicidas en suelos

La contaminación difusa o no puntual por el uso de plaguicidas es una problemática asociada a los cambios introducidos por la intensificación agrícola (Holland, 2004). Con el aumento de la superficie bajo sistemas agrícolas basados en la reducción de las prácticas de labranzas y el incremento de la productividad por unidad de superficie, se ha registrado un crecimiento en el uso de productos plaguicidas. Según estadísticas de la FAO, Argentina en el año 2014 llegó a posicionarse en el tercer lugar a nivel mundial de países consumidores de herbicidas, siendo los herbicidas el 93,5 % del total de consumo de plaguicidas. Dentro de los herbicidas de mayor uso se encuentran en primer lugar Glifosato (N-fosfometilglicina) representando el 65% de las aplicaciones de plaguicidas, seguido por 2,4-D (ácido 2,4-diclorofenoxiacético) y atrazina (6-cloro-N2-etil-N4-isopropil-1,3,5-triazina-2,4-diamina) (Aparicio et al., 2015). La aplicación de herbicidas en agroecosistemas tiene aparejado un riesgo contaminante, cuya magnitud dependerá de la dosis y las características del ambiente de aplicación y propiedades intrínsecas de la sustancia en sí. Particularmente los herbicidas usados en pre siembra interactúan con el suelo y son las características fisicoquímicas y la posición en el paisaje las que regularán el destino y la potencialidad contaminante de la sustancia aplicada (Calvet, 2005; Mamy et al., 2015; Silva et al., 2019).

Entre los múltiples servicios ecosistémicos de los suelos se encuentra su capacidad amortiguadora y descontaminante basada en propiedades físicas, químicas y biológicas

que permiten que actúen como barrera protectora de otras matrices ambientales como aguas superficiales y subterráneas. Algunos autores consideran al suelo como un “reactor natural” ya que tiene la capacidad de transformar, filtrar, amortiguar y depurar sustancias (Targulian y Sokolova, 1996). En el suelo ocurren múltiples reacciones de naturaleza química (redox, precipitación, quelación, neutralización, fotólisis) fisicoquímicas (adsorción, atrapamiento físico), biológica (cometabólicas, biodegradación, conjugación, absorción) que causan la desaparición del herbicida. La potencialidad contaminante de un herbicida en un ambiente particular depende de la probabilidad que éste tenga de salir del sistema suelo y alcanzar otros compartimentos como la biosfera y la hidrósfera donde el riesgo de persistencia se intensifica con los consecuentes problemas hacia los seres humanos y en general la vida de todo tipo de organismos (Arias-Estévez et al., 2008).

La retención, la disipación y el transporte, son los tres procesos que gobiernan el comportamiento de herbicidas en el suelo (Calvet, 2005). Por ello, la dinámica de un herbicida en suelo se parametriza a través de los índices de retención ( $Kd$ ), vida media ( $t_{1/2}$ ) y la constante de velocidad específica ( $k$ ) asociados a cada uno de los tres procesos fundamentales (Cheng, 1990). El coeficiente  $Kd$  expresa la relación entre la concentración de la molécula retenida en la fase sólida y la concentración de la misma en fase acuosa (Bailey y Withe, 1970). Los valores de  $Kd$  determinados para cada herbicida no son constantes y varían en función de las propiedades edáficas particulares de cada suelo. Se expresan en unidades de volumen por masa; mientras mayor es el  $Kd$ , mayor es la adsorción del herbicida al suelo. En términos de riesgo ambiental la erosión con arrastre de sólidos incrementa la peligrosidad de los herbicidas en suelos con altos  $Kd$ , mientras que valores bajos de retención sugieren mayor potencial de pérdida por lixiviación y escurrimiento de la molécula de herbicida.

El parámetro denominado vida media ( $t_{1/2}$ ) caracteriza el proceso de disipación y representa el tiempo transcurrido hasta alcanzar la mitad la concentración de aplicada. La vida media indica la persistencia del compuesto en el suelo que está condicionada por procesos de fotólisis, hidrólisis, degradación microbiana y óxido reducción (Weed et al., 1995). Los sitios donde las moléculas del herbicida alcanzan mayor vida media, ie. son más persistentes, e implica a que la molécula esté más disponible para ser lixiviada (Sinkkonen y Paasivirta, 2000).

La caracterización de los procesos de retención y disipación, a través de los índices  $Kd$  y  $t_{1/2}$ , se ha realizado para usualmente a partir de suelos contrastantes seleccionados puntualmente y sin mediciones de covariables de sitio como para entender mejor la dinámica ambiental del potencial contaminante Si bien se avanzó en establecer rangos esperables para estos índices (Gustafson, 1989) se observa, en la práctica, gran amplitud de los valores de  $Kd$  y  $t_{1/2}$ , incluso a escala de lote. Es decir que, aunque se conoce cómo describir el comportamiento de una molécula de herbicida en un suelo, la alta variabilidad ambiental de los procesos, ligada no sólo al tipo de suelo, sino también al clima y al manejo, ha sido una limitante para el uso de estos índices en análisis globales del potencial de contaminación a escala regional. La mayoría de los estudios existentes analizan el comportamiento de herbicidas sobre adsorbentes particulares y en laboratorio (Calvet, 2005; Hang et al., 2010). Si bien estas aproximaciones son indispensables para profundizar el conocimiento de los procesos que regulan su dinámica, proveen información difícilmente extrapolable a escalas regionales. Numerosas covariables ambientales o características del sitio de aplicación del herbicida pueden explicar la variabilidad regional de los procesos de retención y disipación de las moléculas. Las más usuales son textura del suelo, contenido de materia orgánica y pH (Weber, Wilkerson y Reinhardt, 2004), aunque variables climáticas y relativas al uso del suelo también han

sido citadas como explicativas de la dinámica de la molécula (Hang, Barriuso y Houot, 2003; Rampoldi, Hang y Barriuso, 2008).

Es importante resaltar que estas covariables edafoclimáticas generalmente se encuentran espacialmente estructuradas e imprimen variabilidad espacial a los indicadores de la dinámica de los herbicidas en estudios a escala regional. La identificación y estimación de la variabilidad espacial de la respuesta a mayor escala podría contribuir a mejorar la eficiencia de los modelos de predicción de retención, disipación y transporte de un herbicida sobre un sitio específico (Petach, Wagenet y DeGloria, 1991; Montoya et al., 2003; Chaplain et al., 2011). Los procedimientos experimentales necesarios para cuantificar los parámetros de la dinámica de herbicidas son costosos por su complejidad analítica y por ello, no pueden ser realizados sobre todos los suelos de un territorio. En esta tesis, se proponen protocolos estadísticos para zonificar territorios en función de la variabilidad edafoclimática y orientar muestreos de sitios y para realizar predicciones sitio específicas, de indicadores de retención y persistencia de herbicidas en un dominio espacial continuo. Finalmente, se delimitan zonas según la covariabilidad espacial de los procesos de retención y disipación de herbicidas en relación con la variabilidad edafoclimática para realizar aproximaciones de la peligrosidad del uso de atrazina a escala regional.

## Desafíos metodológicos

Muchos de los procesos relacionados a contaminaciones de suelo varían de manera continua en el espacio. Por ello, resulta de interés predecir el comportamiento de una variable continua en referencia a su ubicación geográfica. Los datos que alimentan estos análisis se conocen como datos espaciales ya que no solo contienen el valor de la(s) variable(s) medidas en un sitio sino también la georreferencia o posicionamiento del sitio. Claro que la variabilidad espacial no se encuentra determinada sólo por la distribución en

el espacio de la variable de interés, sino por múltiples covariables que también varían espacialmente como hemos mencionado anteriormente para el caso de variables edafoclimáticas y que impactan la respuesta que se modela.

La modelación estadística de variables aleatorias, como son los índices que parametrizan la dinámica del herbicida en suelo, dependen de al menos dos tipos de componentes. Por un lado, un componente estructural, explicado por factores y covariables de sitio que se suponen conocidas y determinan el valor esperado del herbicida en suelo. Por el otro, un componente no estructural o aleatorio que puede estar a su vez regulado por un proceso de autocorrelación espacial. En el caso de una variable aleatoria con distribución normal estos componentes se relacionan de manera aditiva. Así, el modelo lineal para sus realizaciones contiene un componente determinístico y un componente aleatorio explicado, en parte, por el proceso de estructuración en el espacio y un remanente de variación no explicada o variación netamente residual.

La correlación espacial observada en una respuesta a una determinada escala puede tener orígenes múltiples y diversos, razón por la cual los métodos analíticos para modelar este tipo de variable incluyen numerosas alternativas. Independientemente de las alternativas que se elijan para modelar la correlación espacial en la componente aleatoria, las predicciones espaciales siempre serán resultado del ajuste de un modelo predictivo al que se incorporan las funciones que contempla la dependencia de las observaciones según la distancia entre los sitios de observación (Schabenberger y Gotway, 2005).

En la actualidad existen numerosas estrategias para modelar la variabilidad espacial de una respuesta continua. Entre éstas: 1) el modelado de la autocorrelación en la respuesta y la predicción en un sitio a partir de la estimación de sus valores en un vecindario (Schabenberger y Gotway, 2005; Cressie y Wikle, 2015), 2) el modelado de la estructura de medias con covariables estructuradas espacialmente (Prasad, Iverson y Liaw, 2006;

Grimm et al., 2008), 3) la incorporación de dependencias espaciales a través de efectos aleatorios de sitio vía un modelo lineal mixto ya sea abordado desde un enfoque frecuentista (Webster y Oliver, 2007) o desde uno bayesiano (Besag et al., 1995; Lindgren y Rue, 2015a), 4) la modelación de la estructura de correlación espacial en los errores del modelo a través de la identificación de estructuras en los residuos del modelo ya sea lineal (Hengl, Heuvelink y Rossiter, 2007) o no lineal y 5) la modelación de la variación espacial en los residuos de un modelo estimado por algoritmos de aprendizaje de máquina (Li et al., 2011; Hengl et al., 2018) una rama de la inteligencia artificial. No obstante, pocos estudios comparan la eficiencia de la predicción espacial lograda bajo las distintas metodologías.

La incorporación y modelación de las covariaciones o correlaciones espaciales de variables de sitio continuas que pueden ser abordadas a través de modelos geoestadísticos basados en la teoría de variables regionalizadas y predicción kriging (Schabenberger y Gotway, 2005) es una de las estrategias más usadas en la modelación de fenómenos espaciales continuos. Aun cuando los procesos involucrados en la dinámica del herbicida en suelo se describen por múltiples variables correlacionadas, la geoestadística clásica presenta mayor desarrollo en la construcción de modelos para una única variable; los mapas obtenidos por kriging permiten visualizar la variabilidad espacial de esa variable, pero no aportan información sobre interacciones con otras variables que condicionan la variación espacial conjunta. Estas interacciones pueden adquirir especial relevancia para predecir valores asociados a procesos complejos y costosos de determinar como la retención y disipación de un herbicida en el territorio.

Si bien se ha abordado la covariación de dos variables mediante covariogramas y modelos de correogionalización (Schabenberger y Gotway, 2005; Arrouays et al., 2011) su extrapolación a contextos de mayor dimensionalidad no es obvia y la implementación de

modelos de covariación espacial en estudios de dinámica de contaminantes es escasa. La inferencia sobre retención y disipación de moléculas de herbicidas en suelo, a escala regional, aún presenta desafíos metodológicos tanto en la selección de covariables ambientales como en la predicción espacial de los índices que describen el destino de los herbicidas en el ambiente.

En este trabajo de tesis se investigan metodologías para identificar métodos y modelos más apropiados para el tratamiento de geodatos relacionados a parámetros de la dinámica de herbicidas en suelo con el fin de generar mapas de variabilidad espacial y de riesgo de procesos dispersivos de contaminantes según las propiedades del sitio que operen a escala regional. Se trabajó con datos de retención y disipación obtenidos para glifosato y atrazina, dos herbicidas de amplio uso en Argentina. En una primera etapa se evaluaron las bondades y las inconveniencias de metodologías provenientes de distintos marcos teóricos (frecuentista, bayesiano y de inteligencia artificial) para la predicción espacial de los indicadores de la dinámica de estos herbicidas en suelos. Paralelamente, se trabajó en una zonificación edafoclimática de la provincia de Córdoba, Argentina para potenciar la interpretación ambiental de mapas de variabilidad espacial generados para ambos herbicidas.

A partir de la investigación metodológica se seleccionaron los métodos estadísticos de mejor desempeño para la predicción espacial y el mapeo y se desarrolló una metodología de 3 pasos orientada al mapeo digital basado en modelos que se utilizó para explicar la variabilidad espacial de los índices de retención de glifosato y atrazina en los suelos de Córdoba. Para obtener mapas digitales de persistencia de atrazina en el ambiente se obtuvieron valores de vida media para una muestra de sitios a partir de modelos no lineales sitio-específicos y se volvió a aplicar la metodología propuesta para el mapeo digital, en este caso de la vida media. Una vez descripta la variabilidad de la retención y

la persistencia de atrazina en suelos, se trabajó sobre la modelación de la correlación cruzada espacial entre ambos indicadores de proceso. Finalmente, sobre la base de estos insumos (mapas) y en complementación con información auxiliar disponible públicamente sobre variables biofísicas y de procesos del territorio en estudio, se realizaron aproximaciones al dimensionamiento de la peligrosidad ambiental por el uso de glifosato y atrazina en ambientes agrícolas de la provincia de Córdoba.

## Objetivo general

Caracterizar la variabilidad espacial, a escala regional, de índices de retención y disipación de herbicidas y predecir valores de estos índices en función de variables ambientales.

## Objetivos específicos

- i. Identificar, en el territorio provincial, áreas homogéneas respecto a variables edafoclimáticas potencialmente importantes para la dinámica de herbicidas, generando protocolos estadísticos para la delimitación de áreas homogéneas en contexto de abundante información respecto a características de sitio.
- ii. Determinar índices de retención y de disipación, de glifosato y atrazina, para suelos de la provincia de Córdoba.
- iii. Evaluar y proponer metodologías de ajuste de modelos estocásticos para predicción espacial de índices de retención y de disipación de herbicidas según variables de sitio.
- iv. Mapear la variabilidad espacial marginal de los índices de retención y de disipación de atrazina y glifosato en la provincia de Córdoba.
- v. Estimar la covariación espacial de la retención y de la disipación de atrazina.
- vi. Dimensionar la peligrosidad ambiental por el uso de herbicidas en el territorio.



## Análisis de la variabilidad y predicción espacial

### Introducción

En las últimas décadas se ha impulsado el desarrollo y la utilización de nuevas tecnologías que permiten capturar datos espaciales, es decir, datos de una variable regionalizada o asociados a una localización en el espacio. La infraestructura de datos espaciales es cada vez mayor en tamaño y calidad, especialmente la asociada a la generación de datos que provienen del sensoramiento ya sea remoto o proximal. También existen datos georreferenciados que se han recopilado a través de muestreos espaciales conducidos a distintas escalas. Los volúmenes de datos espaciales no sólo son vastos y variados, sino que también, en la mayoría de los escenarios, son accesibles. Estos datos generan nuevas oportunidades para la investigación en Ciencias Ambientales y en Agricultura dado que la caracterización de la variabilidad espacial es crucial para comprender fenómenos biológicos en el territorio. Usualmente, esta variabilidad en los procesos aleatorios que generan datos espaciales se modela con herramientas de la Estadística Espacial (Webster y Oliver, 2007; Cressie y Wikle, 2015) y se representa gráficamente en mapas de variabilidad espacial donde se observa cómo cambian los valores de una o más variables aleatorias según su posición en el espacio.

Aún cuando se estudian dominios espaciales continuos, usualmente no existen observaciones de la variable de interés para todos las localizaciones o sitios del espacio analizado; así se hace necesario obtener predicciones espaciales, es decir, predecir el valor de la variable en sitios sin datos. Con grillas de predicción densa, es posible obtener mapas de contorno casi continuos espacialmente. Existen numerosos métodos para la interpolación espacial de una variable en un dominio espacial continuo, algunos determinísticos y otros geoestadísticos como el método Kriging (Cressie, 1990). Cuando existe abundante información sobre cada sitio podrían aplicarse métodos multivariados para su análisis. Aunque la disponibilidad de métodos multivariados que consideran la naturaleza espacial de las observaciones es escasa, existen técnicas de reducción de dimensión como el Análisis de Componentes Principales que incorporan restricciones espaciales en la búsqueda de soluciones óptimas para representar la variabilidad total y la variabilidad espacialmente estructurada en un conjunto de varias variables (Dray, Saïd y Débias, 2008).

En esta tesis se supone que cuando existen varias variables, pero una es interpretada como resultantes de un proceso espacial subyacente y otras como variables explicativas de esa variabilidad espacial, es posible obtener predicciones sitio específicas a partir de modelos para datos correlacionados espacialmente tanto desde un marco teórico frecuentista (Schabenberger y Gotway, 2005; Cressie y Wikle, 2015) como desde el marco teórico bayesiano (Correa Morales, Causil y Javier, 2018). También la predicción espacial puede realizarse desde algoritmos de aprendizaje automático que incorporan la espacialidad (Li et al., 2011).

Existe una plétora diversa de metodologías analíticas para generar predicciones espaciales y mapear variabilidad espacial uni y multivariada. Dado que provienen de marcos teóricos diferentes, su desempeño es raramente evaluado de manera simultánea y no resulta

simple, para quienes necesitan construir un mapa de variabilidad espacial en un contexto de abundante información sobre variables auxiliares, seleccionar uno u otro método de predicción espacial.

Las predicciones espaciales, en este contexto de disponibilidad de datos sobre variables auxiliares, provienen del ajuste de un modelo predictivo que relaciona la o las variables respuestas con las variables explicativas y que, a su vez, incorpora funciones que contemplan la dependencia de las observaciones en relación con la distancia que las separa en el espacio (Schabenberger y Gotway, 2005; Hengl et al., 2007). La modelación predictiva espacial es el proceso por el cual se construye un modelo estadístico capaz de generar una predicción, la cual se asocia a una o más componentes aleatorias relacionadas con variables latentes o no observables y con la variación no explicada por el modelo.

Construir un modelo predictivo implica no solo dilucidar los patrones estocásticos subyacentes a los datos observados, sino que requiere como en cualquier modelación predictiva hacer uso del criterio experto en la disciplina, dar un correcto tratamiento a los datos, identificar los predictores influyentes y validar correctamente el modelo (Kuhn y Johnson, 2013). En este capítulo se describen diferentes metodologías a partir de las cuales es posible construir modelos predictivos para datos espaciales.

## Construcción de un modelo predictivo

Inicialmente, se debe recurrir a una instancia de preprocesamiento de datos que incluye la depuración y acondicionamiento a los formatos necesarios para el estudio relacional. En el caso de datos espaciales en el preprocesamiento hay que considerar los diferentes formatos de la información espacial (raster y vectorial) y las particularidades del sistema de referencia debido a los diferentes sistemas de proyección geográfica. Diversos Sistemas de Información Geográfica (SIG) (Warner y Diab, 2002) ofrecen un amplio rango de funciones para crear, integrar, transformar, visualizar y analizar de manera

exploratoria datos espaciales (Longley, 2005). Los SIG más avanzados también disponen de funciones que generan interfase con software estadístico ampliando así la capacidad para la modelación y análisis más complejos (Bivand, 2014; Lovelace, Nowosad y Muenchow, 2019).

La etapa de selección de las variables es una etapa crucial en el ajuste de modelos predictivos. Otro tópico que es necesario tener en cuenta es el de variables explicativas correlacionadas (multicolinealidad) ya que necesita particular tratamientos en el ajuste de modelos lineales (Dormann et al., 2013). Las técnicas de selección de variables en el contexto de modelos lineales entre las variables predictoras y la variable respuesta no necesariamente arrojan los mismos resultados que otras técnicas de selección con capacidad de captar estructuras no lineales como las derivadas de métodos de inteligencia artificial (Elith et al., 2006; Kanevski et al., 2009).

Luego de la depuración de datos atípicos y la selección criteriosa de variables, basada en el conocimiento del problema y potenciada por métodos estadísticos de selección, prosigue la etapa de ajuste del modelo predictivo. El concepto de autocorrelación espacial es crucial en el ajuste de modelos para datos georreferenciados. La autocorrelación espacial es la correlación de los valores de una misma variable entre sitios del dominio espacial y se expresa como función de la distancia que existe entre los sitios (Matérn, 1986). En el tratamiento de datos espaciales se recurre a la modelación de este proceso para poder generar las predicciones en sitios donde el atributo no ha sido observado (Webster y Oliver, 2007).

Desde la Estadística Espacial se han realizado numerosos aportes para el ajuste de modelos de regresión lineal en datos con autocorrelación espacial tanto desde una perspectiva frecuentista (Schabenberger y Gotway, 2005) como desde la teoría Bayesiana (Cameletti et al., 2013; Lindgren y Rue, 2015a; Wang, Ryan y Faraway, 2018). Desde un

enfoque más computacional como es el Aprendizaje Automático también existen algoritmos capaces de ajustar modelos predictivos (Breiman, 1996, 2001) que, aunque menos explícitos en la forma relacional entre variables, permiten la predicción de valores de la variable de interés. Recientemente se ha incorporado a estas predicciones el concepto de la autocorrelación espacial (Li et al., 2011).

Una vez que el modelo predictivo espacial se pone en marcha es necesario evaluar las predicciones. La predicción implica asignar nuevos valores de las variables respuesta a contextos o escenarios que no corresponden al conjunto de escenarios medidos, es decir no se trata de aquellos sitios que utilizaron para ajustar el modelo de predicción espacial, por lo cual estimar el error de un modelo de predicción espacial es solo posible a través de la comparación de los valores predichos por un modelo y los valores observados en sitios cuya información no fue utilizada en la construcción del modelo.

Entre las alternativas para estimar la exactitud de la predicción están aquellas técnicas de validación cruzada o técnicas de partición del conjunto de datos en datos de calibración y datos de validación (Efron y Hastie, 2016). Es necesario identificar un grupo de observaciones sobre las que se ajustará el modelo, usualmente llamado datos de calibración, y otro grupo que se usará para validar, llamado datos de validación. Muchos trabajos discuten cuál es la forma más adecuada de muestrear para dividir el conjunto de datos relativo a los casos en estudio y los criterios de validación (Golub, Heath y Wahba, 1979; Efron y Tibshirani, 1997; Brenning, 2012). Como regla general se puede establecer que en cuanto menor es la proporción de datos que contiene el grupo de datos de validación menor capacidad de extrapolación del modelo, aunque existe un compromiso con la cantidad de información disponible para ajustar y lograr un buen modelo. Esto se relaja a medida que la base de datos es de mayor tamaño (Kuhn y Johnson, 2013).

Cuando se modelan variables continuas una métrica usual para validar el modelo predictivo ajustado es el Error Cuadrático Medio de Predicción (RMSPE por el término en inglés “Root Mean Square Prediction Error”) (Willmott, 1981). En primer lugar, se calculan los errores como la diferencia entre los valores predichos por el modelo y los observados sobre el grupo de validación. Entonces, el RMSPE se calcula tomando la raíz cuadrada de la media de los errores cuadráticos. Suele interpretarse como la distancia promedio de los valores observados y las predicciones del modelo y muchas veces es conveniente expresarla como una medida relativa a la media de la variable respuesta. El muestreo sobre la base de datos completa, para generar el grupo de entrenamiento y de validación, suele realizarse aleatoriamente y en múltiples ocasiones, es posible calcular también la varianza del RMSPE y varianza de los parámetros del modelo. Sobre la base de estas medidas puede realizarse un análisis de sensibilidad (Hamby, 1994). Existen otras medidas, utilizadas en estimación de modelos de regresión como medidas de bondad de ajuste, como el  $R^2$  o coeficiente de determinación de los modelos lineales y los criterios basados en verosimilitud penalizada como el criterio de información de Akaike y BIC que se pueden obtener en modelos lineales o no lineales. Estas métricas pueden adjuntarse también a los modelos predictivos ya que un modelo bueno como herramienta predictiva debería representar también un buen ajuste de los datos de entrenamiento. Sin embargo, existen riesgos de modelos de buen ajuste, pero sobreparametrizados que pueden derivar en una mala capacidad predictiva o complejizar inútilmente la interpretación. Para el caso de datos espaciales, no solo el error de predicción global es necesario, sino que también es crucial evaluar del error de la predicción en cada sitio específico para dimensionar el error puntual de la predicción espacial. Finalmente, es importante resaltar que siempre debe existir un compromiso entre el ajuste, la predicción y la interpretación a la hora de evaluar un modelo predictivo que sea pertinente y útil.

## Marcos teóricos para la predicción espacial

La modelación de procesos espaciales se desarrolla a través de la teoría de campos aleatorios (Besag, 1974; Cressie, 1990). Un campo aleatorio en el dominio espacial se suele modelar como la suma de un componente estructural o determinístico (tendencia o media), un proceso aleatorio de autocorrelación espacial y un proceso aleatorio de errores independientes (Burrough y McDonnell, 1998). Un proceso de autocorrelación espacial (correlación de un mismo atributo entre dos sitios) positivo, implica que los puntos cercanos en el espacio tienen valores más similares que los puntos distantes. Se asume comúnmente que los procesos de autocorrelación espacial son estacionarios de segundo orden, esto implica que existe una media constante y una función de covariancia que depende únicamente de la distancia entre observaciones. Además, en el contexto espacial se deben contemplar dos particularidades; una es la existencia de bordes o fronteras que delimitan el proceso de autocorrelación y otra es que la correlación no siempre se da en una sola dirección, sino que puede describir procesos anisotrópicos (autocorrelaciones diferentes de acuerdo con la dirección del dominio).

La variable aleatoria bajo estudio puede no resultar medible en todos los puntos del dominio espacial generando distintos tipos de datos espaciales. La clasificación más conocida de los tipos de datos espaciales es la propuesta por Cressie (1993), quien establece tres variantes de acuerdo con la definición del dominio del problema en estudio. Éstas son: datos del tipo geoestadísticos, datos regionales y patrones de puntos. Los datos del tipo geoestadísticos son datos de dominio continuo, es decir, supone que entre dos sitios pueden existir infinitos datos. No se refiere a continuidad en el atributo en estudio sino en la estructura espacial del proceso subyacente, los datos pueden ser del tipo continuo o discreto indistintamente. Los datos regionales o referidos a un área, son datos donde el dominio es fijo y discreto haciendo referencia a una superficie o polígono

(Schabenberger y Gotway, 2005). En la tercera categoría se definen los patrones de puntos donde la definición del dominio debe modificarse ya que provienen de un proceso puntual aleatorio conformado por los puntos o sitios donde ocurren los eventos.

### **Modelación de la autocorrelación espacial**

La teoría de variables regionalizadas ha sido ampliamente documentada por lo que en esta sección se describen las principales características establecidas por otros autores (Schabenberger y Gotway, 2005) La teoría establece que la observación de la variable aleatoria en cada sitio es una realización de un proceso aleatorio, por lo cual existe una distribución de probabilidad asociada al mismo. En este contexto la herramienta clásica para modelar autocorrelación espacial positiva en un proceso estacionario de segundo orden es el semivariograma o función de semivarianza (mitad de la varianza de las diferencias entre observaciones separadas en el espacio por una distancia o *lag h*).

Para estimar la semivarianza deben agruparse los sitios de acuerdo con un parámetro de distancia denominado *lag*. Entre los factores a tener en cuenta para ajustar el semivariograma se encuentra el tamaño de muestra con el que se estima cada semivarianza por lo que la distribución de los sitios en el plano determinará el *lag*. De esta manera se construye a partir de las observaciones lo que se conoce como semivariograma empírico denotado por la siguiente ecuación,

$$\hat{\gamma}(h) = \frac{1}{2m(h)} \sum_{i=1}^{m(h)} [Y(x_i) - Y(x_i + h)]^2$$

donde *h* es el *lag* o distancia entre los sitios  $x_i$  y  $(x_i + h)$  y  $m(h)$  es el número de observaciones contenidas en el *lag h*. La función del semivariograma empírico se puede describir a partir de la varianza *nugget*, la varianza estructural y el rango (Figura 1).

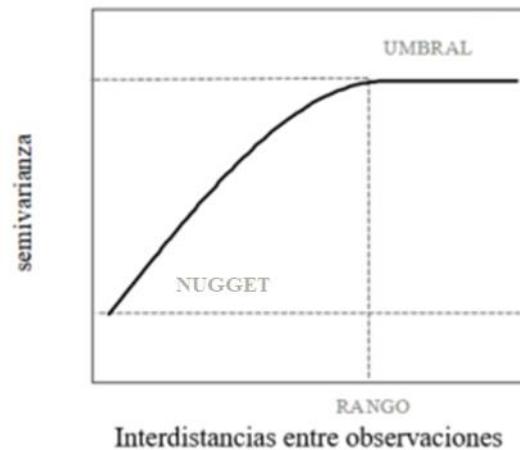


Figura 1. Semivariograma. Parámetros rango, umbral y nugget

La varianza *nugget* refiere a la ordenada al origen del semivariograma y representa la suma de errores aleatorios no espaciales, o errores asociados con la variabilidad espacial a escalas más finas de las medidas en la grilla de datos espaciales con la que se trabaja (Schlesinger et al., 1996). La varianza estructural representa la varianza de observaciones independientes, es decir observaciones que fueron tomadas a una distancia tal que la autocorrelación debida a la dependencia espacial es cero y se expresa como la diferencia entre el umbral o meseta y el *nugget* del semivariograma. Otro parámetro del semivariograma es el rango, que representa la distancia a la cual se alcanza el máximo de la función de semivarianza, es decir la distancia hasta la cual se registran correlaciones entre las observaciones. El parámetro rango del semivariograma es también entendido como la extensión del proceso de autocorrelación espacial. Cuando las funciones que describen la semivarianza no tienen un máximo absoluto a menudo se utiliza lo que se llama un rango práctico el cual representa la distancia a la cual se obtiene el 95% de la varianza estructural.

Una medida que se utiliza para resumir el grado de estructura espacial de una variable georreferenciada en un dominio continuo es la varianza estructural relativa (RSV por el término en inglés “Relative Structural Variance”), que es el cociente entre la varianza

estructural y la suma de la varianza *nugget* y la varianza estructural. Un valor alto de RSV indica que las predicciones geoestadísticas serán más eficientes que aquellas obtenidas con métodos de predicción que ignoran la información espacial. Zimback (2001) establece que el grado de dependencia en función del RSV entre muestras puede ser clasificado como:  $\leq 25\%$  bajo, entre 25% y 75% medio y  $\geq 75\%$  alto.

Al semivariograma empírico se le debe ajustar un modelo teórico el cual debe describirse por funciones definidas positivas, ya que la varianza puede tomar solo valores positivos no nulos. Entre las funciones más clásicas se encuentran una serie de modelos no lineales, entre los que se pueden citar el gaussiano y el exponencial. Estas funciones se derivan de una clase flexible de funciones provenientes de una representación espectral de la función de covarianza de un proceso isotrópico llamadas funciones Matérn (Matérn, 1986). Esta familia de funciones posee una serie de propiedades convenientes para el ajuste del semivariograma empírico (Schabenberger y Gotway, 2005). Otro modelo utilizado es el modelo esférico (Isaaks y Srivastava, 1989) que deriva de una familia de funciones indicadoras donde el valor indicador corresponde al diámetro de la esfera que es equivalente al rango es decir cuando la autocorrelación es exactamente cero. El modelo esférico tiene dos características principales, un comportamiento lineal cerca del origen y el hecho que al rango el semivariograma alcanza la meseta. El modelo de semivariograma exponencial, por el contrario, se aproxima a la meseta de manera asintótica. Para el mismo rango y meseta, de un modelo esférico, el modelo exponencial alcanza el rango más fácilmente, es decir, a menor distancia que el modelo esférico. Los modelos de semivariograma denominados *lineal* y *power* (Schabenberger y Gotway, 2005) tienen la particularidad de no ser estacionarios de segundo orden. Para contemplar dentro de estos modelos estructuras espaciales que no sean isotrópicas se puede recurrir a una transformación lineal del sistema de coordenadas cuando existe anisotropía geométrica

(Stroup, 2016). Otro tipo de proceso anisotrópico es el que se conoce como anisotropía zonal, que puede modelarse adicionando un término dependiente de la dirección del *lag* (Goovaerts, 1999).

Los modelos de semivariaograma son no lineales. A diferencia de los modelos lineales donde el método de estimación por mínimos cuadrados garantiza una solución óptima y estable, para los modelos no lineales la optimalidad del método de ajuste depende no sólo del modelo, sino también de las características de los datos que están siendo ajustados. Entre estos el método de mínimos cuadrados ponderados (WLS) suele ser el elegido para la estimación de funciones de semivariaograma. Se elige una función y valores iniciales de los parámetros basados en la observación del semivariogramas empírico.

Alternativamente, la estimación de los parámetros del semivariaograma puede hacerse simultáneamente a la estimación de tendencia para la media bajo la teoría de los modelos lineales mixtos (MLM) (West *et al.*, 2007) con métodos de estimación basados en la función de verosimilitud. El MLM usado cuando la respuesta esperada depende de otras variables de sitio es el modelo de regresión lineal múltiple (RLM) para datos correlacionados. La estimación de los parámetros de varianza y covarianza del MLM, que son también parámetros del semivariaograma, puede realizarse en este marco de trabajo de manera simultánea a la de aquellos parámetros relacionados a la estructura de media del proceso que reflejan tendencias a gran escala en una o más dimensiones. La estimación de parámetros se puede realizar por los métodos de máxima verosimilitud (ML) (Searle *et al.*, 1992) o por máxima verosimilitud restringida (REML) (Patterson y Thompson, 1971). Bajo este marco teórico se ajusta un modelo directamente sobre los datos y no sobre las semivarianzas como en la geoestadística clásica. La estimación REML de la estructura de covariación espacial se define considerando que la misma es función de la distancia entre la separación de las observaciones. Las más utilizadas son

las funciones espacial esférica, exponencial y gaussiana los que contienen parámetros relacionados a los del semivariograma del mismo nombre (Córdoba et al., 2015).

Al ajustar distintos MLM a un mismo conjunto de datos, es necesario utilizar criterios para la comparación de los ajustes, para ello se usan criterios de información. En el marco de los MLM, los criterios de información se basan en el logaritmo de la verosimilitud residual (log-reslikelihood) y aplican una función de penalización debida a la cantidad de parámetros del modelo ajustado. Un menor valor del criterio indica un "mejor" ajuste. Los criterios de información de Akaike (AIC) (Akaike, 1973) y el criterio bayesiano (BIC) (Schwarz, 1978) son los más usados en la selección de un MLM.

Otra de las herramientas estadísticas utilizadas para la selección de modelos es la prueba del cociente de verosimilitud (LRT, Likelihood Ratio Tests) (West *et al.*, 2007). Esta se basa en una prueba de hipótesis que se formula en el contexto de dos modelos anidados en sus parámetros. Utiliza el valor de la función de verosimilitud evaluada en las estimaciones ML o REML de los modelos comparados. El modelo más general o con más parámetros, abarca tanto la hipótesis nula como la alternativa, es denominado modelo de referencia. El segundo modelo, más simple, satisface la hipótesis nula (parámetros igual a cero) y se denomina modelo anidado. La única diferencia entre estos dos modelos es que el modelo de referencia contiene todos los parámetros, mientras que el modelo anidado no contiene aquellos que se suponen podrían ser iguales a cero. Si el estadístico LRT es suficientemente grande, hay evidencias para rechazar el modelo reducido y preferir el modelo de referencia o modelo más parametrizado. Si los valores de verosimilitud de los dos modelos están muy cerca, el estadístico LRT será pequeño sugiriendo evidencia a favor del modelo reducido.

Aún cuando la diferencia en las estimaciones logradas tanto con la aproximación basada en técnicas geoestadísticas como con los MLM puede ser poca, la utilización de MLM

presenta claras ventajas. Cuando se trabaja con técnicas geoestadísticas es necesario realizar en la etapa exploratoria de los datos el ajuste de tendencias a gran escala. En caso de que la tendencia fuese significativa, será necesario descontarla y trabajar con los residuos del modelo. Mientras que usando MLM, se puede modelar la correlación espacial y la tendencia a gran escala en un solo paso. En esta estrategia las coordenadas se incorporan en la estructura de medias del modelo, permitiendo que en el término de error aleatorio se elimine el sesgo producido por esa tendencia

Las predicción espacial, es decir la predicción de valores de la variable continua en sitios del campo espacial donde no existen observaciones, usualmente se hace por el método Kriging basándose en el semivariograma ajustado (Cressie, 1990). Kriging proporciona el mejor estimador lineal insesgado (BLUE por el término en inglés, “Best Linear Unbiased Estimator”) y un error de estimación conocido como varianza kriging para cada predicción. Esta varianza depende del modelo de semivariograma ajustado y de la ubicación en el espacio de los datos originales. Las interpolaciones geoestadísticas tienen ciertas ventajas respecto a interpolaciones espaciales determinísticas. Por un lado, las distancias sobre las que trabaja son distancias estadísticas en contraste con las distancias geométricas utilizadas en métodos no estadísticos. Además, evita muestras redundantes, ponderando de forma distinta muestras que estén muy cerca entre sí y procedan de la misma región respecto a muestras que estén, por ejemplo, en lados opuestos al sitio sobre el que se quiere asignar la predicción (Webster y Oliver, 2007). Los parámetros del semivariograma son los que gobiernan la asignación de los pesos a las observaciones que rodean el sitio al cual se le asignará la predicción. Particularmente, el *nugget* determinará como se reparten estos pesos, si la varianza del error es muy alta, todas las muestras tenderán a tener el mismo peso en la interpolación. Por el contrario, si la varianza del error es pequeña, los coeficientes de ponderación serán distintos para las diferentes

muestras, muchas de las cuales por encontrarse cercanas en el espacio serán redundantes. El rango del semivariograma determinará cuando el semivariograma se vuelve horizontal. Si el rango aumenta, cada punto tendrá mayor peso en la interpolación de otras observaciones. Luego, el método kriging se basa en el conocimiento del comportamiento de la variable en el espacio.

Entre los métodos de interpolación espacial que utilizan todos los datos simultáneamente se destacan los métodos de kriging ordinario, simple y universal. En el kriging ordinario la media de la variable es estimada localmente. En caso de conocer la media de la variable, hecho que raramente ocurre, se utiliza el kriging simple. En el kriging universal la media es estimada y se incluye también la influencia de una tendencia espacial de los datos. La predicción asignada a los puntos incógnita puede realizarse de manera puntual (kriging puntual) o definiendo bloques (kriging en bloques) (Webster y Oliver, 2007). La interpolación puntual es la estimación del valor de la variable en el punto incógnita, mientras que la interpolación por bloques estima la media de puntos de un área predeterminada que rodea al punto incógnita. La interpolación por bloques (que produce un “suavizado” de las estimaciones) suele correlacionar mejor con los valores verdaderos (Isaaks y Srivastava 1989). Este tipo de predicción Kriging resulta apropiado para las predicciones espaciales a escalas regionales.

### **Modelos de regresión para predicción espacial**

#### *Enfoque frecuentista*

En la estadística frecuentista los parámetros del modelo a ajustar se consideran constantes fijas que son desconocidas y deben ser estimadas a partir de los datos. Por ejemplo, para la predicción de la respuesta en un sitio en relación con el valor de otras variables del sitio, se puede usar un modelo de regresión lineal múltiple (RLM), donde los parámetros a estimar serán los coeficientes de regresión que relacionan cada variable independiente

con la respuesta y las varianzas y covarianzas subyacentes. Dado estos parámetros se asignan probabilidades a las observaciones las cuales se consideran aleatorias. La probabilidad asignada a la inferencia de los parámetros refiere a una frecuencia relativa (asumiendo un tamaño grande de muestra) y se evalúa en un número infinito de repeticiones hipotéticas del proceso generatriz de datos. Con datos estructurados en el espacio, se usarán modelos lineal mixtos (MLM) que asumen una relación lineal entre la respuesta y las variables explicativas, pero también permiten para considerar la autocorrelación espacial subyacente en los términos de error del modelo (Raudenbush y Bryk, 2002; West et al., 2014). En estos MLM, también llamados modelos de covarianza residual, no se trata a los errores como independientes, sino que se supone algún modelo para el proceso que determina las autocorrelaciones espaciales en la componente aleatoria. Es decir, las funciones de semivariaograma se ajustan a partir de los residuos de un modelo lineal que especifica vía la suma de efectos fijos el valor esperado de la respuesta. La estimación del modelo completo (efectos fijos o coeficientes de regresión y varianzas y covarianzas asociadas a los efectos aleatorios) para respuestas gaussianas puede hacerse a partir de mínimos cuadrados ponderados, máxima verosimilitud o máxima verosimilitud restringida (Schabenberger y Gotway, 2005).

Para modelar la estructura de medias en modelos lineales de regresión múltiple que involucran numerosas variables explicativas, potencialmente correlacionadas, también suele recurrirse a la regresión por mínimos cuadrados parciales (PLSR por el término en inglés “Partial Least Square Regression”) (Geladi y Kowalski, 1986). PLSR es un modelo que regresiona la variable respuesta con combinaciones lineales ortogonales generadas a partir de las variables explicativas. Aún cuando la técnica ha sido desarrollada para datos independientes y no para datos con correlación espacial, al igual que en el modelo de

RLM, en PLSR es posible considerar la estructura de autocorrelación en los términos de error.

### *Enfoque bayesiano*

En estadística bayesiana se considera que los parámetros del modelo son variables aleatorias y se calculan distribuciones de probabilidad para los parámetros de las cuales se deriva medidas de incertidumbre (Correa Morales et al., 2018). La información previa sobre los parámetros debe resumirse en distribuciones de probabilidad denominadas distribuciones *a priori*, a partir de las cuales se estima la distribución de probabilidad *a posteriori* dadas las observaciones. Estimaciones puntuales de los parámetros de interés se pueden obtener calculando medidas resumen de la distribución *a posteriori*, como la media o el modo, y se informan juntos a intervalos de credibilidad calculados desde percentiles de la distribución *a posteriori*. La credibilidad se interpreta como la probabilidad de que el valor estimado para el parámetro pertenezca al intervalo reportado, dado los datos observados.

Los métodos de simulación por cadenas de Markov Monte Carlo (MCMC) (Besag et al., 1995), han permitido resolver modelos complejos sin la necesidad de imponer estructuras que lo simplifiquen. Éstos han sido usados para la estimación de modelos con datos espaciales (Best, Richardson y Thomson, 2005; Reich y Fuentes, 2007). Sin embargo, el método MCMC presenta una alta demanda computacional. Rue et al (2009) propusieron una alternativa para aproximar la distribución *a posteriori* en contextos de datos espaciales a partir de aproximaciones basadas en el algoritmo INLA (del término en inglés “Integrated Nested Laplace Approximation”) lo que ha permitido simplificar las estimaciones (Bivand, Gomes-Rubio y Rue, 2015; Wang et al., 2018). Sobre la base de las aproximaciones por INLA y la implementación de la alternativa en el lenguaje de

programación R (R-INLA) se han popularizado las aplicaciones de la regresión bayesiana espacial y espaciotemporal (Cameletti et al., 2013).

Una particularidad de INLA en la estimación de estructuras espaciales es que permite modelar estructuras ralas, es decir con gran presencia de valores ceros, en la inversa de la matriz de variancias y covariancias (matriz de precisión). La estructura rala de la matriz de precisión puede provenir de la no dependencia de las variables aleatorias en la distribución multivariada conjunta (Rue y Held, 2005). En R-INLA particularmente se estiman matrices de precisión ralas utilizando aproximaciones por ecuaciones diferenciales parciales estocásticas (SPDE) (Lindgren, Rue y Lindström, 2011; Lindgren y Rue, 2015a). Bajo este enfoque la grilla de predicción espacial, propia de los métodos frecuentistas de la estadística espacial, se reemplaza por una malla de predicción que se construye a partir de triángulos que cubren el dominio entero, cada vértice de los triángulos representa un nodo sobre los que se predice por interpolación (Blangiardo y Cameletti, 2015). Además de las ventajas computacionales que el algoritmo ofrece, permite trabajar con dominios espaciales de límites y bordes complejos (Bakka et al., 2018) y generar mapas de variabilidad espacial a partir de estas grillas a los que será posible asignar medidas de incertidumbre de cada predicción puntual.

#### *Enfoques desde el aprendizaje automático*

El término aprendizaje de máquina o aprendizaje automático corresponde a una rama de la inteligencia artificial, por definición este tipo de métodos no asumen supuestos sobre la estructura de los datos, bajo el enfoque de la estadística computacional esa afirmación no es determinante, en este sentido el término minería de datos contempla una combinación de herramientas de estadística y computación (Witten et al., 2016). Se hace referencia a aquellos algoritmos usualmente basados en intenso proceso computacional que “aprenden” automáticamente de los datos intentando minimizar la intervención

humana. Algunos métodos de aprendizaje automático se basan en algoritmos heurísticos de particiones recurrentes de los datos y evaluaciones de éstas hasta identificar la mejor para explicar el comportamiento de la variable respuesta como es el caso de los árboles de clasificación y regresión (algoritmos CART) (Breiman, 2001) o algoritmos de redes neuronales, como máquinas de vectores de soporte (Kohonen, 1988) y mapas autoorganizados (Van Hulle, 2012). Estas herramientas se utilizan con fines predicción y clasificación y son particularmente útiles para interpretar relaciones lineales y no lineales en contextos de regresión múltiple con variables explicativas correlacionadas. Estos algoritmos pueden ser empoderados mediante métodos de remuestreo que obtienen muestras aleatorias a partir de los datos observados, derivan modelos para cada muestra y ensamblan los modelos resultantes para obtener predicciones con menor error de predicción a partir del modelo ensamblado. Entre los algoritmos de este tipo usados con fines predictivos, este trabajo revisa la oportunidad del uso de regresiones por bosques aleatorios (RFR por el término en inglés “Random Forest Regression”) y de árboles de regresión generalizados (GBR por el término en inglés “Generalized Boosting Regression”) (Efron y Hastie, 2016). Si bien estos algoritmos se han utilizado para datos espaciales (Kanevski et al., 2009), en escasas ocasiones se modela la estructura espacial. Una propuesta para incorporar la estructura espacial es utilizar las coordenadas o matrices de distancias generadas a partir de las mismas como una covariable más en la construcción del modelo (Hengl et al., 2017, 2018). Otra propuesta, es modelar el término remanente del ajuste del algoritmo de aprendizaje automático con una función de autocorrelación espacial (Li et al., 2011) de manera análoga a los modelos del enfoque frecuentistas. El objetivo de este Capítulo es comparar el desempeño de los métodos citados arriba para obtener modelos de regresión que puedan ser usados como modelos predictivos en un contexto de datos con correlación espacial,

## Comparación de predicciones obtenidas desde distintos modelos

Se realizó una investigación metodológica utilizando datos espaciales de variables de suelo y el coeficiente de adsorción de un herbicida como variable dependiente. Los datos de suelo provienen de un muestreo regional conducido en la provincial de Córdoba, Argentina. Involucran muestras de los 15 cm superficiales de suelo relevadas a campo sobre una grilla de 40×40 km (n=89 sitios) (Hang et al., 2015; Giannini Kurina et al., 2019a), mayor información sobre el muestreo y la base de datos edafoclimática se puede encontrar en el Capítulo III. Las variables de suelo medidas en cada sitio fueron pH, nitrógeno total (NT), carbono orgánico total (COT), Na, K, Ca, Mg, Zn, Mn, Cu, capacidad de intercambio catiónico (CIC), porcentaje de arena, limo y arcilla, capacidad de campo (CC), óxidos de hierro y aluminio (Ox. Al, Ox. Fe). El coeficiente de adsorción del herbicida en la matriz suelo fue determinados según la técnica de *batch-equilibrium* para la preparación de las suspensiones de suelo fortificadas y cuantificados por HPLC según se detalla en el Capítulo III de esta tesis. Los datos de variables de sitio fueron usados para predecir retención con modelos predictivos construidos a partir de los 89 sitios muestreados.

Para la modelación estadística los valores del coeficiente de retención ( $K_d$ ) fueron transformados a escala logarítmica debido a su distribución ligeramente sesgada hacia la derecha. Debido a la alta correlación entre covariables de sitio, la selección de variables para construir el modelo predictivo se hizo con el algoritmo BRT (del inglés, *boosting regression tree*) (Elith et al., 2008) optimizando los parámetros de este procedimiento mediante validación cruzada para minimizar el error de predicción. Las funciones `gbm.step` y `gbm.simplify` del paquete R `gbm` fueron usadas en esta etapa de selección de variables. Se implementaron 11 estrategias diferentes para la construcción de modelos de

predicción espacial para predecir el  $Kd$  en sitios sin determinaciones de retención de herbicida.

Se usaron modelos de regresión múltiple del coeficiente de adsorción en función de las variables de sitio seleccionadas (Al (Ox), arcilla, pH). Los modelos fueron estimados en el marco frecuentista mediante RLM y por PLSR, todos se estimaron dos veces, una suponiendo errores independientes y otra, considerando errores correlacionados espacialmente. La regresión lineal con errores correlacionados fue estimada por máxima verosimilitud residual (REML). También se ajustó el modelo de predicción espacial mediante regresión bayesiana con R-INLA. Bajo este marco teórico también los datos se supusieron independientes o, alternativamente, correlacionados espacialmente en cuyo caso se usó un término aleatorio de sitio con semivariograma continuo aproximado por SPDE al modelo de Matérn. La modelación de la correlación espacial en el marco bayesiano se realizó previa optimización del parámetro de suavizado ( $\alpha$ ) de la malla que modela la autocorrelación espacial. También se usaron los algoritmos RFR y GBR sin considerar correlación espacial y restringidos espacialmente mediante Kriging sobre los residuos del modelo. Para disponer las predicciones se usó una grilla fina (2,5×2,5 km) generando mapas de variabilidad espacial a escala regional.

Para validar las predicciones se calcularon errores de predicción (RMSPE) global y sitio-específicos mediante el método *Jackknife* (Efron y Hastie, 2016). Los errores se expresaron relativos a los valores de coeficiente de adsorción. Finalmente, los valores observados y predichos fueron categorizados como bajos ( $Kd \leq 10 \text{Lkg}^{-1}$ ), medios ( $10 < Kd < 30 \text{Lkg}^{-1}$ ) y altos ( $Kd \geq 30 \text{Lkg}^{-1}$ ) (Rampazzo et al., 2013). Como complemento del RMSPE se construyó una table de clasificación cruzada entre las categorías observadas y predichas para  $Kd$  con el objetivo de calcular el porcentaje de clasificación

correcta. Todos los modelos se obtuvieron con y sin tendencia lineal en las coordenadas geográficas (Latitud y Longitud) en la estructura de medias.

Cuadro 1. Comparación del desempeño estadístico de distintas estrategias metodológicas para construir modelos de predicción sitio-específica del coeficiente de adsorción de un herbicida en suelo a partir de características del suelo.

	Correlación espacial	Coordenadas en la estructura de medias			
		Incluídas		No incluídas	
		Error de predicción <sup>†</sup>	Proporción de clasificación correcta	Error de predicción	Proporción de clasificación correcta
RLM	no	18,5	83	18,3	83
	si	19,1	81	18,7	81
REML	si	18,9	83	18,6	88
PLSR	no	19,5	78	19,0	82
	si	18,8	79	20,1	80
GBR	no	18,9	87	19,3	90
	si	20,0	88	20,2	85
RFR	no	19,3	85	19,9	84
	si	20,3	88	20,2	85
INLA	no	18,5	82	18,2	82
	si	18,2	81	<b>17,6</b>	81

<sup>†</sup>Error de predicción expresado como porcentaje de la media general. RLM, regresión lineal múltiple; REML, regresión lineal múltiple estimada por máxima verosimilitud; PLSR, regresión por mínimos cuadrados parciales; GBR, regresión por árboles de regresión mejorados por remuestreo; RFR, regresión por bosques aleatorios

Todos los modelos mostraron buen (y similar) desempeño predictivo. No obstante, las regresiones bayesianas con efectos aleatorios de sitio fueron las más competitivas sobre todo en situaciones donde no se modela la tendencia espacial en la estructura de media. Estos resultados coinciden con reportes realizados para este tipo de regresión por otros autores (Poggio et al., 2016; Huang et al., 2017). En adelante, en este trabajo de tesis, se ha utilizado el modelo de regresión bayesiana con efectos aleatorios de sitio en construcción de modelos de predicción sitio específicos. La disponibilidad de nuevas rutinas de software para el ajuste de este modelo (Lindgren y Rue, 2015a) ha sido, indudablemente, otro factor líder en la selección de la metodología a usar para desarrollar

el modelo. Además, en contraste con el enfoque frecuentista, el cual asume a los parámetros del modelo como fijos, el enfoque bayesiano los considera variables probabilísticas que cuentan con funciones de distribución y, consecuentemente, permite obtener medidas de incertidumbre para la predicción realizada en cada sitio. Las medidas de incertidumbre se derivan de la distribución *a posteriori* de los valores predichos para cada sitio (Gelman, 2004) Suelen informarse como mapas de desvíos estándares o mapas de intervalos de credibilidad del 95% de probabilidad (Huang et al., 2017; Giannini Kurina et al., 2019b). Las distribuciones a posteriori predichas en cada sitio pueden operarse por otras funciones(Correa Morales et al., 2018), lo que facilita la interpretación de las predicciones cuando se trabaja con variables respuesta transformadas.

# Zonificación según variabilidad edafoclimática

## Introducción

Una forma de identificar y caracterizar la interacción entre el clima y el suelo con variables biofísicas de interés en un territorio es a través de mapas de variabilidad espacial para cada variable ambiental (mapas biofísicos). Sin embargo, cuando las variables que caracterizan los sitios son numerosas puede ser conveniente previo a la construcción de modelos, zonificar el dominio espacial en estudio, es decir, reconocer áreas con características comunes entre los sitios de una misma zona. Así, existirá más información para la interpretación ambiental de las predicciones de los parámetros de la dinámica del herbicida.

Numerosas técnicas geoestadísticas univariadas han sido usadas para modelar la variabilidad espacial de una variable e identificar gradientes en sus valores (Cressie y Chan, 1989; Lark, 2000). Para identificar y describir áreas geográficas homogéneas respecto a tipo de suelo y capacidad de uso se han utilizado con éxito las cartas de suelo (Imbellone y Teruggi, 1993; Jarsún et al., 2006). La clasificación de suelos se realiza comúnmente usando modelos umbrales que llevan a una secuencia de particiones binarias de las variables que son tratadas independientemente (Burrough y McDonnell, 1998;

Burrough et al., 2015). Si bien éstas son herramientas útiles no permiten reconocer zonificaciones multivariadas en un continuo espacial en sentido. La metodología clásica, no tiene en cuenta la correlación y covariación o variación conjunta espacial entre variables. Cuando se registra más de una variable por sitio, la covariabilidad espacial entre todas o algunas de ellas demanda análisis geoestadísticos multivariados (Schabenberger y Gotway, 2005).

Se ha demostrado que el análisis de la variabilidad espacial de una variable puede ser mejorada a partir de la estructura de covarianza de esa variable con respecto a una variable anexa (Wu et al., 2003; Hengl, Heuvelink y Stein, 2004). Así, el estudio del patrón de variación espacial no sólo de cada variable independientemente sino también de las correlaciones espaciales entre variables podría proveer conocimiento útil para zonificaciones en sentido multidimensional, es decir contemplando simultáneamente una serie o conjunto de atributos o variables de sitio. La variabilidad conjunta de dos variables georreferenciadas ha permitido el análisis exitoso de otros procesos edáficos (Cosby et al., 1984) y se supone puede contribuir en la explicación de la dinámica de herbicidas en suelo.

En la actualidad se dispone de técnicas de análisis de datos que no solo tienen en cuenta la naturaleza multivariada del conjunto de datos, sino también su espacialidad cuando estos están georreferenciados (Wackernagel, 2013). El método MULTISPATI-PCA (Dray et al., 2008; Arrouays et al., 2011) es una extensión del análisis multivariado de componentes principales que permite tener en cuenta la covariabilidad espacial entre las variables en estudio. El método provee combinaciones lineales de las variables originales que maximizan su autocorrelación espacial. Sin embargo, su potencialidad en zonificaciones a escala regional ha sido poco explorado.

Las técnicas de interpolación espacial (Cressie, 1990; Webster y Oliver, 2007) suelen ser usadas para realizar re-escalamiento y así disponer datos correspondientes a diferentes variables sobre una grilla común para facilitar el estudio de correlaciones (Long, 1998). Para clasificación de sitios suelen usarse distintos métodos de conglomerados (Anderberg, 2014), ya sea implementándose sobre el conjunto de variables originales o sobre las variables sintéticas obtenidas como combinaciones lineales desde las variables originales. Los métodos de conglomeración difusa (Bezdek et al., 1981) resultan particularmente útiles para clasificar sitios en áreas con variación espacial de dominio continuo. Un método de conglomerados difusos clasificará cada sitio en uno u otro grupo luego de evaluar probabilísticamente la similitud multivariada del sitio con cada uno de los conglomerados posibles, pero asume que la clasificación no es exacta, sino que el sitio podría pertenecer a uno u otro grupo (pero a alguno con mayor probabilidad). Este tipo de método de conglomeración puede clasificar sitios en áreas de transición de mejor manera que los métodos que asignan los objetos a clasificar a uno u otro clúster sin considerar la probabilidad de pertenencia de cada objeto al conjunto de clústeres. Sin embargo, la mayoría de los algoritmos de clúster, incluso los de clasificación difusa, no han sido desarrollados para contemplar la covariabilidad espacial entre las variables que caracterizan los sitios del dominio espacial en estudio. Una combinación de MULTISPATI PCA y conglomerado difuso ha sido usada para obtener clúster de sitios en un dominio espacial y definir zonas homogéneas a escala fina en aplicaciones de agricultura de precisión (Córdoba et al., 2016), pero no se ha evaluado su desempeño a escala regional.

Un problema crítico en un protocolo de zonificación, además de la selección de la escala, la densidad de sitios con observaciones y el método de agrupamiento elegido, es la determinación del número de conglomerados existentes una vez que se ha implementado

una clasificación no supervisada de sitios ya que cualquiera de estos métodos se aplica sin conocer cuál es el agrupamiento subyacente y por tanto no se conoce cuál es el número de grupos que define la estructura de aglomeración. A pesar de la existencia de numerosos índices (Hennig, 2007) para identificar un número de grupo óptimos para representar la estructura de agrupamiento subyacente, no existen claros consensos sobre cuál de éstos usar en cada conjunto de datos según el propósito de conglomeración. Charrad et al. (2014) propuso calcular más de un índice y aplicar la regla de la mayoría, es decir seleccionar el número de conglomerados recomendado por la mayoría de los índices. Últimamente, con el incremento de la capacidad computacional, ha proliferado la cantidad de índices sugeridos para identificar la cantidad de clústers que definen una estructura. Entre éstos se encuentran los índices de estabilidad que resultan de comparar los conglomerados obtenidos a partir de la remoción aleatoria sucesiva de parte de los datos y que son usados como medidas de validación obtenidas desde los mismos datos (validación interna). También es posible evaluar los resultados con la validación externa o con información no usada en el proceso de conglomeración (Theodoridis y Koutroumbas, 2009).

Este capítulo tiene como objetivo proponer un protocolo de zonificación basado en un análisis de conglomerados difuso de sitios de un territorio que incluya la covariabilidad espacial existente entre las variables de sitio y permita identificar zonas edafoclimáticas homogéneas a escala regional.

## Materiales y Métodos

### **Área de estudio**

El área bajo estudio corresponde a la provincia de Córdoba, Argentina (Figura 2) limitada por los paralelos de 29° y 35° de latitud sur y los meridianos de 61° y 65° de longitud

oeste. El paisaje, está constituido predominantemente por planicies aproximadamente en un ~60% y el resto por cordones montañosos con dirección norte sur hacia el oeste del territorio. La elevación varía entre 79 a 2884 msnm. El territorio es atravesado por las isohietas de 700 mm y de 500 mm determinando un gradiente de humedad en dirección Este-Oeste pasando por: húmedo, subhúmedo, semiárido y árido. La precipitación media anual varía entre 900 y 400 mm y la temperatura media anual oscila entre 10 °C y 24 °C. El balance hidrológico arroja valores de deficiencia hídrica anual que oscilan entre -80 mm y -480 mm. Los suelos son de origen loésico y en base al *Soil Taxonomy* se los agrupa en Molisoles (61%), Entisoles (13%), Alfisoles (7%) y Aridisoles (5%) (Jarsún et al., 2006).

### **Base de datos**

Se construyó una base de información ambiental georreferenciada para la provincia de Córdoba que incluye datos de propiedades edáficas proveniente de un muestreo de estos atributos realizado a escala regional previo al trabajo de tesis (Hang et al., 2015). El muestreo sistemático llevado a cabo en ese estudio cubre el territorio de la provincia de Córdoba con una grilla regular de resolución 20 x 20 km<sup>2</sup> (tamaño de celda adaptado en función de la presencia y el estado de los caminos). El tamaño muestral n=354 sitios reproducen las proporciones de los distintos ordenes de suelos reportados para la provincia (Hang et al., 2015). Los sitios muestreados (Figura 2) correspondieron a Molisoles (72 %), seguido por Entisoles (13 %) mientras que los Aridisoles y Alfisoles fueron representados con un 5% cada uno. En cada uno de los 354 sitios se tomó una muestra de los primeros 15 cm de profundidad compuesta por 20 submuestras extraídas en un radio de 100 m según protocolo recomendado (Gregorich y Carter, 2007). Cada sitio fue caracterizado con las variables edáficas descriptos en detalle en el Anexo 1. El muestreo fue principalmente en suelo bajo agricultura (72 %) y el porcentaje restante se

repartió entre vegetación natural (pastizales y monte) y pasturas implantadas. En la presente tesis, se extendió la caracterización de cada sitio con datos climáticos (precipitaciones, temperaturas medias y extremas), topográficos y usos del suelo, también descritos en el Anexo I. En los análisis de Capítulos posteriores, que demandan datos sobre características de sitio, se utilizó la misma base de datos.

Para este capítulo se usaron 10 variables edáficas (pH, NT, COT, Na, K, CIC, Cu, arena, arcilla, CC) y las variables climáticas de la base de datos.

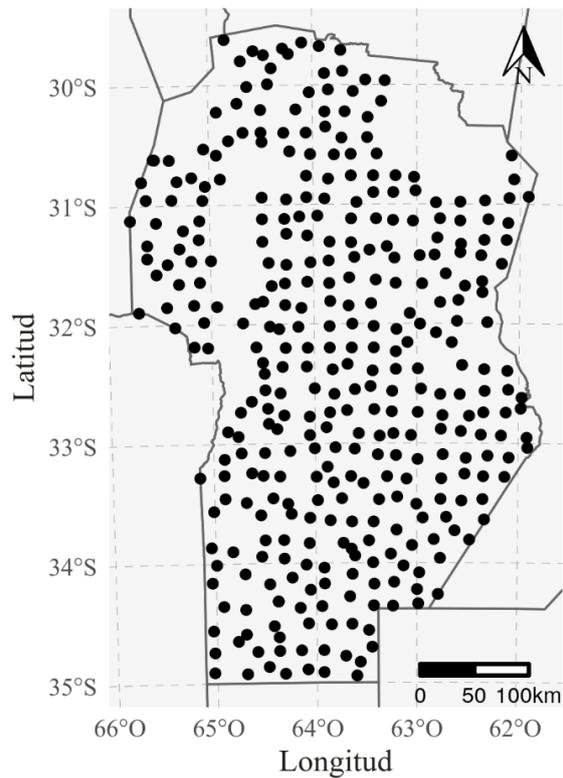


Figura 2. Área de estudio y sitios de muestreo.

### Preprocesamiento de datos

Previo a la aplicación del protocolo de zonificación cada atributo de suelo fue modelado espacialmente y predicho en una grilla de mayor resolución ( $2,5 \times 2,5$  km). Las predicciones de cada variable se realizaron a través de un modelo de regresión tomando

la variable pendiente (derivada del MDE) como regresora y asumiendo una estructura espacial en los residuos modelada a partir del ajuste sobre un semivariograma empírico construido a partir del estimador robusto de Cressie (Cressie, 1993). En la estructura de media del modelo ajustado también se tuvo en cuenta tendencias espaciales en sentido de la latitud y longitud (Schabenberger y Gotway, 2005). Los modelos de autocorrelación espacial se estimaron con el método mínimos cuadrados ponderados (WLS, por su sigla en inglés, Weighted Least Squares). La estimación de la variabilidad espacial de cada variable se hizo utilizando la librería “gstat” (Pebesma, 2004) en lenguaje de programación R (R Development Core Team, 2011).

### **Protocolo estadístico**

#### *Paso 1: Obtención de variables sintéticas que capturan correlación espacial*

Se realizó un análisis de componentes principales con restricciones espaciales MULTISPATI-PCA (Dray et al., 2008) tomando como entrada las variables de suelo, elevación y clima asociadas a los sitios de la grilla fina obtenida luego del preprocesamiento de los datos. Para calcular autocorrelación espacial, se usó una red de vecindarios construida a partir de la inversa de distancia, usando 50.000 m como distancia máxima. Este valor fue determinado considerando el tamaño de celda de la grilla a clasificar y los rangos que presentaron los semivariogramas ajustados de cada variable. Se seleccionaron las componentes principales espaciales (sPC) que permitieron explicar una proporción de variabilidad total acumulada de al menos 80 % de la variabilidad total. La no inclusión de algunas componentes se realizó con la finalidad de eliminar variabilidad residual, es decir, variabilidad no asociada a patrones espaciales repetibles. El procedimiento MULTISPATI-PCA se implementó con las librerías “ade4” (Chessel, Dufour y Thioulouse, 2004) y “rgdal” (Bivand et al., 2014) en lenguaje de programación R (R Development Core Team, 2011).

### *Paso 2: Conglomerados de sitios*

Las sPC resultantes del MULTISPATI-PCA se utilizaron como inputs del análisis de conglomerado difuso *Fuzzy Kmeans* (Bezdek et al., 1981). El algoritmo de clasificación *Kmeans* agrupa objetos en  $k$  grupos haciendo máxima la variación entre conglomerados y minimizando la variación dentro del conglomerado. Comienza con un agrupamiento inicial o con un grupo de puntos semilla (centroides) que formarán los centros de los grupos. Luego asigna a cada objeto al grupo que tiene el centroide (media) más cercano. En *Fuzzy Kmeans* se determina además el grado de semejanza de un objeto con un clúster para decidir su pertenencia al clúster. Se utilizó la distancia euclidiana al cuadrado y el coeficiente difuso  $k = 2$  con el paquete "e1071" (Meyer et al., 2014).

El número de conglomerados usados para la zonificación fue determinado en función de la disminución de la suma de cuadrados de las distancias dentro de los grupos (Krzanowski y Lai, 1988) y mediante la regla de la mayoría aplicada sobre 17 índices propuestos para determinar número óptimo de conglomerados (librería "Nbclust" en R). Los índices utilizados y sus correspondientes referencias bibliográficas se indican en el Cuadro 5 (sección Resultados). Estos índices combinan información sobre la compactación de los agrupamientos y la separación entre agrupamientos, es decir, la variabilidad multidimensional dentro y entre grupos. La formulación detallada de cada uno de estos 17 índices se pueden encontrar en Charrad et al. (2014).

### **Validación del protocolo de zonificación**

Para evaluar la secuencia metodológica propuesta (Figura 3), el algoritmo de agrupamiento se aplicó tanto a las componentes principales restringidas espacialmente derivadas del MULTISPATI-PCA, como a las componentes principales derivadas de un Análisis de Componentes Principales clásico. Para cada tipo de análisis se evaluaron mediante validación cruzada ( $k = 10$ ) las distintas estratificaciones generadas. La

pertenencia de cada sitio no utilizado en la clasificación se estableció en función de la distancia euclídea entre el sitio a clasificar y el centroide de cada conglomerado. Los sitios fueron asignados al conglomerado con centroide más próximo a su perfil de variables y el porcentaje de clasificación correcta fue calculado comparando la asignación a un conglomerado dada por el protocolo con la verdadera localización del punto dada por sus coordenadas. La zonificación también fue validada externamente mediante la comparación de medias de zonas para un conjunto de variables (Zn, Fe, Mn, P, CE) no utilizadas en la implementación del protocolo de zonificación. Tal comparación se realizó mediante análisis multivariado de varianza (Johnson, 1998).

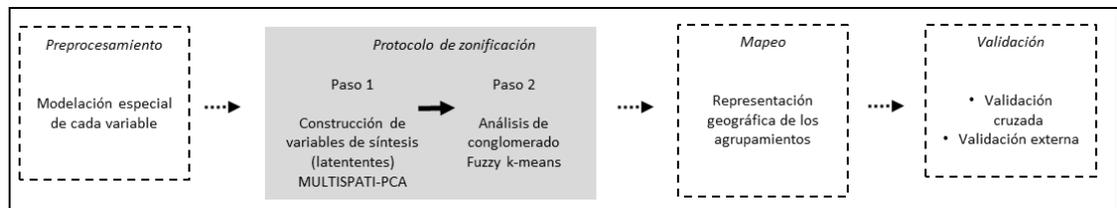


Figura 3. Protocolo de zonificación regional

## Resultados

En el Cuadro 2, se presentan las variables edáficas (pH, NT, COT, Na, K, Cu, CIC, CC, arena y arcilla) así como las variables topográficas (elevación) y las variables climáticas (pp y Tm) junto con los parámetros de los modelos seleccionados de la estructura espacial de cada variable edáfica. Todas las variables presentaron alta variabilidad en el área de estudio con coeficientes de variación entre 10 y 273 % (pH y Na, respectivamente). Las variables vinculadas con la composición de la fracción sólida (arena, arcilla, CIC y CC), presentaron un coeficiente de variación entre 33 y 54%. Como era de esperar COT y NT, presentaron un porcentaje de variación similar. Elevación presentó un 76 % de variación asociada con la topográfica serrana de parte del territorio. Las variables climáticas

registraron intervalo de variación con amplitud de 547 mm para pp y 7 °C para Tm (Cuadro 2).

Los parámetros de las funciones de los semivariogramas ajustados para las variables edáficas mostraron que la extensión de la autocorrelación espacial (rango) varió entre 38 y 181 km. Los valores más bajos se registraron para arena, arcilla, Cu y Na y especialmente CC, mientras que los más altos se observan para las variables vinculadas a la materia orgánica como COT CIC, NT, K y pH. Se observó también que un alto porcentaje de la variabilidad total se encuentra espacialmente estructurada (entre el 40 y 68 %).

Cuadro 2. Variabilidad edafoclimática en la Provincia de Cordoba (n=355 sitios)

Variable	Unidad	Media	Min	Max	CV (%)	Parámetros del semivariograma				Error <sup>††</sup> (%)	
						Mejor Modelo <sup>†</sup>	nugget	Sill Parcial	Rango (km)		RSV
pH		6,8	5,31	10	10	Sph	0,2	0,2	102	48	8,5
NT	%	0,13	0,04	0,52	45	Sph	0,0	0,0	149	45	30,8
COT	g kg <sup>-1</sup>	13,8	2,8	59,3	45	Exp	13,7	7,3	180	35	33,3
Na	Cmol kg <sup>-1</sup>	1,44	0,02	53,7	273	Gau	0,3	0,5	113	64	37,6
K	Cmol kg <sup>-1</sup>	1,86	0,4	4,8	31	Sph	0,2	0,1	58	40	24,7
CIC	Cmol kg <sup>-1</sup>	17,7	5,3	35,8	33	Sph	7,8	11,4	101	59	4,4
Cu	mg kg <sup>-1</sup>	1,63	0,1	5,9	61	Exp	0,3	0,7	104	68	23,9
arena	%	42	0,6	94,8	57	Gau	95,8	171,5	91	64	30,8
arcilla	%	18,6	0,1	44,6	43	Gau	20,1	16,4	86	45	32,3
CC	% p:p	17,7	5	34	33	Exp	8,3	9,6	50	54	22,8
Elevación	m	312	80	1421	76						
pp	mm	738	461	908	14						
Tm <sup>2</sup>	°C	17,2	13,8	20,8	7						

<sup>†</sup>Sph: Modelo esférico, Exp: Modelo exponencial, Gau: Modelo Gaussiano modelos seleccionados por validación cruzada y optimización del porcentaje de varianza total que se encuentra espacialmente estructurado (RSV); CV coeficiente de variación; <sup>††</sup>Error de predicción explicado expresado como porcentaje de la media de la variable.

La implementación de MULTISPATI-PCA sobre la grilla preprocesar indico que eran necesarias tres sPC para resumir la información edafoclimática explicando más del 80 % de la variabilidad total. En el Cuadro 3, se presentan las varianzas y los coeficientes de autocorrelación de cada una de las sPC generadas a partir de PCA clásico y MULTISPATI-PCA tanto para los datos de la grilla original como los datos obtenidos por interpolación espacial (Cuadro 3).

Cuadro 3. Autovalores de los análisis PCA clásico y PCA espacial.

Eje	MULTISPATI-PCA				PCA			
	Autovalor	Porcentaje	Porcentaje acumulado	Índice de Moran	Autovalor	Porcentaje	Porcentaje acumulado	Índice de Moran
1	4,46	0,38	0,38	0,90	5,12	0,39	0,39	0,82
2	2,92	0,3	0,68	0,75	3,91	0,3	0,69	0,78
3	1,46	0,16	0,85	0,69	2,01	0,15	0,85	0,77

El índice de autocorrelación espacial fue menor en la primer PC que en la primer sPC. Los autovectores asociados a las variables sintéticas de ambos métodos (MULTISPATI-PCA y PCA Clásico) se muestran en el Cuadro 4. Los elementos en estos vectores representan los pesos de cada variable original en la combinación lineal que representa las nuevas variables (variables sintéticas). Cuanto mayor sea el peso (expresado como valor absoluto), mayor contribución de la variable para explicar la variabilidad de la variable sintética. Para la primer componente espacial, la variabilidad fue principalmente explicada por arena, arcilla, CC y CIC. Las variables más importantes en la segunda sPC fueron pH, K, Tm y pp. Para la tercer sPC las variables de mayor envergadura fueron Elevación y Na (Cuadro 4). Las principales diferencias entre sPC y PC se observaron en el rol asignado a Tm y Elevación. MULTISPATI-PCA pondera en la primer sPC aquellas variables con mayor estructura espacial.

Cuadro 4. Contribución de variables originales a las componentes principales

Variable	MULTISPATI-PCA			PCA		
	sPC 1	sPC 2	sPC 3	PC 1	PC 2	PC 3
pH	0,04	-0,43	0,3	0,1	-0,32	0,4
NT	-0,13	-0,31	-0,19	0,32	-0,3	-0,09
COT	-0,17	-0,33	-0,18	0,29	-0,32	-0,11
Na	-0,04	-0,05	0,56	4,00E-03	0,06	0,57
K	-0,26	0,38	0,01	0,1	0,45	-0,06
CIC	-0,39	-0,14	0,05	0,42	0,01	0,07
Cu	-0,25	-0,32	0,09	0,32	-0,15	0,2
arena	0,45	-0,01	-0,06	-0,38	-0,21	-0,08
arcilla	-0,43	0,03	-0,05	0,39	0,19	-0,05
CC	-0,44	-0,04	0,06	0,41	0,15	0,08
Elevación	-0,02	0,1	0,61	0,09	-0,44	-0,15
pp	-0,29	0,36	-0,27	0,19	0,35	-0,33
Tm	0,07	-0,44	-0,25	-0,08	0,22	0,55

Como resultado del análisis de agrupamiento *Fuzzy Kmeans* identificamos cuatro agrupamientos de sitios, independientemente de qué tipo de variable sintética fue usada como variable de entrada en el algoritmo de agrupamiento. Los valores de todos los índices para 2, 3, 4, 5 y 6 grupos se indican en el Cuadro 5. La representación espacial de los cuatro agrupamientos permite visualizar las zonas delineadas por el protocolo (Figura 4).

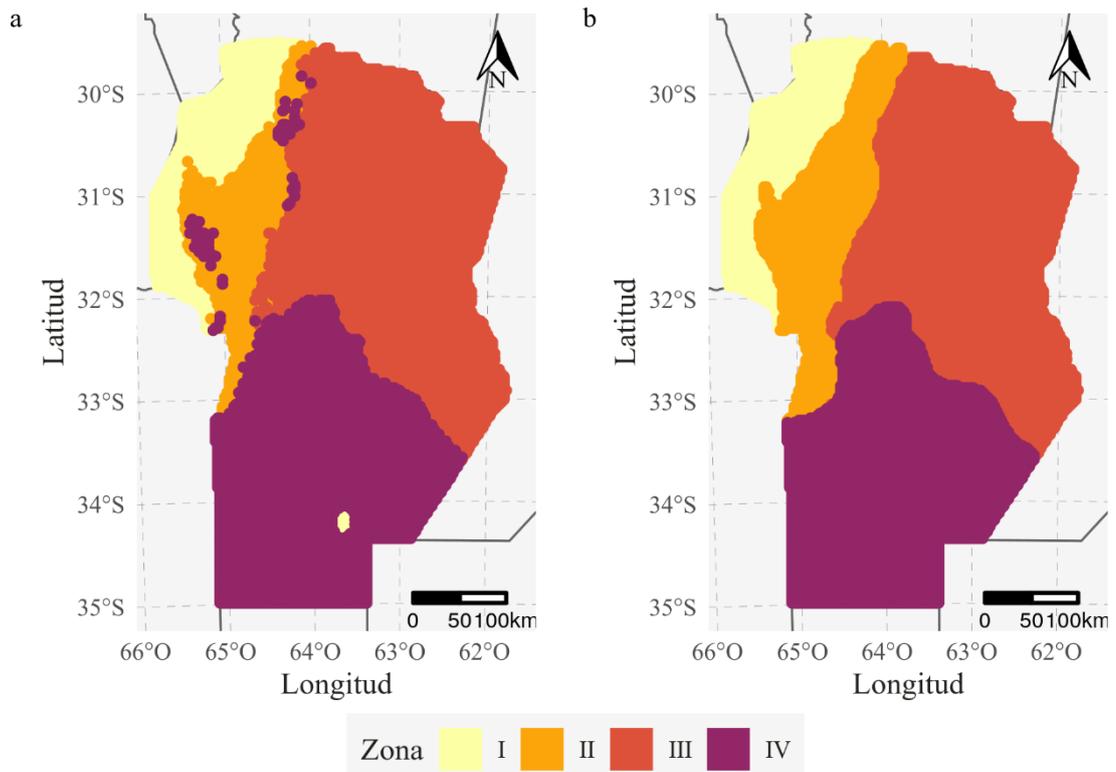


Figura 4. Zonificación edafoclimática de Córdoba: a) obtenida por PCA clásico b) obtenida por PCA espacial

La validación de la secuencia metodológica completa realizada por validación cruzada arrojó una clasificación correcta del 80 % de los sitios al usar MULTISPATI-PCA; ese valor se redujo a 60 % con el uso del PCA convencional, es decir sin considerar las covariaciones espaciales entre las variables originales. La comparación de medias entre zonas para las variables externas de validación (Zn, Fe, Mn, P, y CE) indicaron diferencias estadísticas ( $p < 0,05$ ) entre las cuatro zonas delimitadas (Cuadro 6).

Cuadro 5. Índices de validación de agrupamiento (cantidad de clústers).

Índices	Números de conglomerados					Mejor Partición	Autor
	2	3	4	5	6		
KL <sup>†</sup>	2,9	0,6	4,9	3,0	1,0	4	(Krzanowski y Lai 1988)
CH <sup>†</sup>	165	139	149	133	119	2	(Calinski y Harabasz 1974)
CCC <sup>1</sup>	-0,3	0,0	4,4	4,8	4,6	5	(Sarle 1983)
Silhouette <sup>†</sup>	0,3	0,4	0,4	0,3	0,2	4	(Rousseuw 1987)
Ratkowsky <sup>†</sup>	0,2	0,2	0,3	0,2	0,2	4	(Ratkowsky y Lance 1978)
Ptbiserial <sup>†</sup>	0,5	0,5	0,6	0,5	0,5	4	(Milligan 1980, 1981)
McClain <sup>†</sup>	0,7	0,9	1,1	1,7	2,3	2	(McClain y Rao 1975)
Dunn <sup>†</sup>	0,05	0,05	0,06	0,04	0,04	4	(Dunn 1974)
Cindex <sup>††</sup>	0,2	0,3	0,2	0,2	0,2	4	(Hubert y Levin 1976)
DB <sup>††</sup>	1,4	1,3	1,2	1,2	1,4	4	(Davies y Bouldin 1979)
SDindex <sup>††</sup>	1,0	0,8	0,8	1,0	1,2	4	(Halkidi et al. 2000)
SDbw <sup>††</sup>	1,1	0,6	0,6	1,0	0,7	4	(Halkidi y Vazirgiannis 2001)
Hartigan <sup>†††</sup>	74	89	36	23	20	4	(Hartigan 1975)
Scott <sup>†††</sup>	426	736	1167	1310	1422	4	(Scott y Symons 1971)
TrCovW <sup>†††</sup>	$2,2 \times 10^{+5}$	$1,2 \times 10^{+5}$	$7,9 \times 10^{+4}$	$6,0 \times 10^{+4}$	$5,3 \times 10^{+4}$	3	(Milligan y Cooper 1985)
Friedman <sup>†††</sup>	2,9	4,3	8,2	9,9	12,2	4	(Friedman y Rubin 1967)
Ball <sup>†††</sup>	1180	634	369	264	205	3	(Ball y Hall 1965)

<sup>†</sup> Valor del índice debe ser maximizado, <sup>††</sup> Valor del índice debe ser minimizado, <sup>†††</sup> diferencias entre niveles consecutivos del índice debe ser maximizado

Cuadro 6. Diferencias entre las zonas delimitadas para variables externas.

Zona	Zn	Fe	Mn	P	CE 1:2.5 (s:w) (dSm <sup>-1</sup> )	Prueba Hotelling <sup>†</sup>
	(mg kg <sup>-1</sup> )					
I	1,2	154,6	34,7	41,3	2,4	a
II	1,8	142,4	25,2	82,8	0,2	b
III	1,7	188,3	55,6	124,9	0,5	c
IV	0,7	94,9	36,9	135,9	0,6	d

<sup>†</sup> Letras diferentes indican diferencias estadísticas entre zonas ( $p < 0.05$ ).

La caracterización de cada una de las 4 zonas edafoclimáticas delimitadas se muestra en el Cuadro 7. La Zona I, ubicada al noroeste de la provincia, mostró la precipitación media anual más baja (531 mm) y pH medio más alto (7,5). Las zonas I y IV presentaron los valores más bajos de COT (I = 11,3 g kg<sup>-1</sup> y IV = 9,9 g kg<sup>-1</sup>), NT (I = 0,11% y IV = 0,10%), arcilla (I = 11,7% y IV = 13,4%) y CIC (I = 12,2 cmol<sub>c</sub> kg<sup>-1</sup> y IV = 10,9 cmol<sub>c</sub> kg<sup>-1</sup>). A su vez, la zona IV localizada al sur de la provincia, presentó un bajo pH (6,5) y alto milimetraje acumulado (pp = 744 mm); esta zona, además se asocia con mayor contenido de arena (63,7 %) y el menor de Cu y (0,9 mg kg<sup>-1</sup>). La zona II corresponde al cordón serrano central de la provincia y su piedemonte, presenta la media de elevación

más alta, 761 m.s.n.m., y la media de Tm más baja (16 °C). En esta zona las variables: COT, NT y Cu tuvieron los valores más elevadas (22,3 g kg<sup>-1</sup>, 0,21% y 2,2 mg kg<sup>-1</sup>, respectivamente). La zona III cubre una gran área al este del territorio provincial; las precipitaciones anuales medias fueron las más altas comparativamente (807 mm), y la menor elevación, 200 m.s.n.m., también mostró los valores de arcilla, K, CIC y CC más altos (2,3 cmol<sub>c</sub> kg<sup>-1</sup>, 19,8 cmol<sub>c</sub> kg<sup>-1</sup> y 22,7%, respectivamente).

Cuadro 7. Caracterización de zonas edafoclimáticas delimitadas en provincia de Córdoba

Zona	pH 1:2.5 (s:w)	NT (%)	COT (g kg <sup>-1</sup> )	Na (cmol <sub>c</sub> kg <sup>-1</sup> )	K (cmol <sub>c</sub> kg <sup>-1</sup> )	CIC (cmol <sub>c</sub> kg <sup>-1</sup> )	Cu (mg kg <sup>-1</sup> )	arena (%)	arcilla (%)	CC (%)	Elevación (m)	pp (mm)	Tm (°C)
I	7,5 (8) <sup>†</sup>	0,11 (41)	11,3 (51)	4,9 (237)	1,3 (31)	12,2 (26)	1,9 (28)	57,8 (19)	11,7 (63)	14,3 (25)	332 (43)	531 (12)	19,5 (5)
II	7,2 (7)	0,21 (43)	22,3 (39)	1,2 (38)	1,4 (38)	18,1 (24)	2,2 (47)	41,3 (41)	20,0 (39)	19,0 (26)	761 (26)	673 (8)	16,0 (7)
III	6,8 (8)	0,15 (24)	15,3 (26)	1,1 (118)	2,3 (20)	19,8 (25)	2,0 (48)	18,9 (54)	24,5 (23)	22,7 (16)	200 (60)	807 (7)	17,3 (4)
IV	6,5 (9)	0,10 (30)	9,9 (33)	0,9 (106)	1,7 (24)	10,9 (22)	0,9 (65)	63,7 (21)	13,4 (36)	12,6 (27)	258 (66)	744 (9)	16,8 (2)

<sup>†</sup>Valores entre paréntesis corresponden a coeficientes de variación.

## Discusión

El área de estudio abarca un amplio rango de variación latitudinal (6 grados) y longitudinal (4 grados), además existen importantes diferencias en topografía debido a la existencia de un cordón de elevación central y su pie de sierra con elevados coeficientes de variación de la mayoría de las variables edáficas, topográficas y climáticas. La cobertura del muestreo sistemático realizado resultó útil para caracterizar la variabilidad espacial en el territorio. Considerando el contenido de arena, una de las variables de mayor varianza, una muestra de 354 sitios permite estimar con un 95% de confianza la media de la variable con un intervalo de amplitud igual al 12% de la media. Los muestreos

sistemáticos sobre grillas regulares han sido recomendados para capturar la varianza estructurada espacialmente (Gili, 2013). Dado que las muestras serán utilizadas para construir modelos estadísticos donde las variables explicativas contienen o se relacionan con la variabilidad espacialmente estructurada, el tamaño muestral para la construcción de modelos debería ser juzgado desde la varianza del error puro, la cual es más pequeña que la varianza total y por tanto la demanda del tamaño muestral será menor a la requerida para estimar una media poblacional (Hengl, Rossiter y Stein, 2003; Li, 2010).

Para las variables estudiadas una proporción importante de esta variabilidad se encontró espacialmente estructurada (Cuadro 2). Es decir, se trata de un territorio con suficiente heterogeneidad espacial para justificar y soportar una zonificación edafoclimática. Se registró variabilidad en la magnitud de la dependencia espacial (rango) de las funciones de semivarianza ajustadas para cada variable. Algunas variables tales como CC y Cu, mostraron variaciones de carácter local mientras que otras presentaron variaciones de carácter regional tal como pH y COT. La magnitud del rango resulta alta para variables de reconocida heterogeneidad a escalas finas, como es el caso de pH y Na, (Odeh, Chittleborough y McBratney, 1991), no obstante, en el contexto regional, aportaron a la zonificación.

### **Metodología para la zonificación regional**

El protocolo propuesto, basado en el agrupamiento de sitios mediante el algoritmo *Fuzzy Kmeans* aplicados a sPCs, constituye una herramienta para explorar variabilidad espacial entre las variables usadas para la zonificación. Permite resumir no solo el comportamiento individual de las variables utilizadas sino también su covariación espacial. La propuesta generó clústeres de sitios (que denominamos zonas) con coherencia espacial, es decir, los sitios vecinos mayoritariamente pertenecen al mismo clúster.

El preprocesamiento de los datos a través de la modelación de la estructura espacial de cada variable permitió redistribuir y adaptar la información de diferentes fuentes y formatos. De esta manera, de un conjunto de observaciones muestrales se llega a predicciones regularmente espaciadas a mayor resolución.

La comparación entre los resultados de PCA clásico y MULTISPATI-PCA mostró ventajas del enfoque espacial. El método que incluye las restricciones espaciales generó variables sintéticas en donde se maximizó la autocorrelación espacial, resultados consistentes con lo publicado por otros autores (Arrouays et al., 2011). La validación cruzada de la clasificación con sPC mostró un error de clasificación inferior que el logrado con las PC (20 % vs. 32 %). MULTISPATI-PCA acoplado a *Fuzzy Kmeans* permitió mapear la variabilidad espacial en el dominio continuo con mejor nitidez que cuando se trabaja con las variables derivadas del PCA no espacial a escala regional re-escalando las variables originales.

Un aspecto importante en la delimitación de zonas a partir de métodos de clasificación no supervisada es la definición del número óptimo de clúster o zonas que compondrán la zonificación. Diferentes estrategias fueron usadas para evaluar la pertinencia de agrupamientos alternativos. Algunos de ellos se basan en la comparación de los conglomerados con otras variables, mientras que otros utilizan la información empleada en el mismo proceso de aglomeración (criterio de validación interna) (Theodoridis y Koutroumbas, 2009). El protocolo propuesto incluye una serie de índices de validación interna y, además, sugiere un proceso de validación a partir de variables no utilizadas en el proceso de aglomeración. El protocolo estadístico realiza una zonificación basada en los datos y en métodos multivariados donde todas las variables, así como sus interdependencias, contribuyen a la identificación de zonas homogéneas sin la necesidad de indicar relaciones de tipo causa efecto entre ellas.

Los modelos digitales de elevación e imágenes del sensoramiento remoto son fuentes de datos que pueden proveer gran volumen de datos para la clasificación multivariada. La disponibilidad de este gran volumen de información ha estimulado la implementación de técnicas multivariadas no supervisadas y de técnicas de clasificación mediante aprendizaje automático. Además de los métodos de clúster clásicos, se están utilizando algoritmos de bosques aleatorios (Breiman, 2001), máquinas de soporte vectorial y mapas autoorganizados (Kohonen, 1982) para obtener clasificaciones automáticas basadas en información multidimensional. Sin embargo, estos métodos proporcionan menos información sobre la importancia de las variables originales y su correlación espacial.

### **Zonificación edafoclimática de la provincia de Córdoba**

Los principales rasgos del territorio de la provincia de Córdoba fueron captados por la zonificación realizada. Es decir, las áreas generadas contribuyen en la descripción y comprensión ambiental del territorio en estudio desde una perspectiva multivariada, aun así, existe cierta heterogeneidad dentro de cada área que se puede constatar al analizar la variabilidad de cada atributo ya sea edáfico o climático marginalmente. El territorio presenta gradientes abruptos de tres de los cinco factores formadores de suelo, ellos son el clima, el material originario, y el relieve. El loess pampeano presenta una gradación granulométrica decreciente en dirección oeste-este asociado a la dirección de los vientos que lo transportaron (Rocca, Redolfi y Terzariol, 2006). Además, los niveles de  $\text{CaCO}_3$  del loess pampeano oscilan entre 5 y 10 % (Gorgas y Tassile, 2003), y su profundidad en el perfil de suelo se corresponde con el gradiente de precipitaciones que crece en sentido oeste-este (Iriando y García, 1993; Manzur, 1997) . Así, las zonas de mayor superficie III y IV, que cubren en conjunto el 72,7% del total de sitios, corresponden al área de planicies de la provincia y las dos zonas restantes más pequeñas (I, II) corresponden al área serrana y piedemonte. Las características generales del territorio que comprende las zonas I y II

incluyen una alta heterogeneidad espacial debido a los cambios abruptos en los parches de pampas de altura (Gorgas y Tassile, 2002), inclusive presenta  $\text{CaCO}_3$  cerca de la superficie y una región de salinas en el extremo noroeste. Otro aspecto particular de esta área es que tiene la mayor proporción de vegetación nativa remanente (Cabido y Zak, 1999; Cabido et al., 2003). Existen otras zonificaciones precedentes del territorio estudiado generadas para describir procesos geomorfológicos (Carignano, 1999) y agroeconómicos (Ghida Daza et al., 2009), que responden a la estructuración espacial aquí observada. Por ejemplo Ghida Daza et al., 2009 reconocen en la provincia de Córdoba 5 áreas y 15 zonas agroeconómicas homogéneas en las que las zonas I y II se corresponde con el área descrita como *Área ganadera del NO*, la zona II con las *Áreas agrícola ganadera del Centro* y la *Zona núcleo agrícola del SE* y la zona IV con el *Área mixta del S de Córdoba*. A su vez también reconocen cierta heterogeneidad dentro de estas grandes áreas por lo cual recurren a una estratificación aún más detallada por zonas. La clasificación realizada en esta tesis proporciona menos zonas de mayor tamaño ya que se basa sólo en aquellas variables con potencialidad de impactar la retención y disipación de herbicidas en suelo y tiene como objetivo identificar grandes áreas para la interpretación de estos procesos que se realiza al final del trabajo de tesis. La zona I presentó concordancia con un área denominada Bolsón chaqueño de fuerte heterogeneidad topográfica (pie de monte y planicie), climáticamente se destaca el balance hidrológico negativo y la marcada amplitud térmica (Gorgas y Tassile, 2002). Es la zona con el pH más alto, que se explica en parte por el origen de la alcalinidad asociada con la presencia de  $\text{CaCO}_3$  en la superficie (Manzur, 1997), corresponde a un área salina, que se originó a partir de una falla geológica en sentido noreste-suroeste que dejó a la intemperie un antiguo lecho marino (Gorgas y Tassile, 2002). También se caracteriza por la abundancia de cloruro de sodio, sulfato de sodio y otros minerales acumulados por los

ciclos de sedimentación y evaporación de aguas superficiales (Bertolino et al., 2000; Gorgas y Tassile, 2002). La zona II abarca la región montañosa y parte del pie de sierra este y el oeste, siendo el área de mayor altitud de las cuatro zonas definidas. Edáficamente mostró mayor COT y NT. La CIC fue la mayor de las dos zonas del oeste (I, II), pero menor que la encontrada en zona II con menor contenido de COT. La relación positiva entre COT y CIC se vincula a los coloides o materia orgánica humificada (Parfitt et al., 1995), sugiriendo que la MOS de la zona II presentaría abundancia de materia orgánica fresca o poco humificadas. Otro indicador que refuerza esta suposición es que el contenido de Cu es el más alto de las cuatro zonas teniendo en cuenta que las formas extraíbles de este elemento se ven favorecidas por el aumento del pH. Sin embargo, la dinámica del Cu se relacionada con la MOS debido a la formación de complejos (Mortensen, 1963; McGrath, Sanders y Shalaby, 1988; Alvarez y Lavado, 1998) debe ser profundizada.

La Zona III, fue una de las más extensas, y geomorfológicamente se corresponde en gran parte con las llamadas “planicies/llanuras loessicas” (alta, plana, de altos de morteros) (Gorgas y Tassile, 2002). Los rasgos distintivos de la zona III fueron numerosos dado que corresponde a la zona de menor altitud como también porque presentó los valores más altos de pp, arcilla, K, CC, CIC. Este conjunto de propiedades define una región con aptitud agrícola y, de hecho, este es el principal uso (Cabido et al., 2003; Jarsún et al., 2006). Finalmente, la zona IV es la de mayor superficie (40,3 % de los sitios) y dado que el contenido de arena es el rasgo más característico, el tamaño del grano fue claramente el factor que discriminó esta porción del territorio de las zonas restantes. Corresponde a un grupo de regiones definidas anteriormente en la literatura por su tamaño de grano como Pampas caracterizadas por dunas y Pampas arenosas altas y planas (Gorgas y Tassile, 2002). Los contenidos medios más bajos de COT, NT y CIC y varios elementos como

Zn, Mn y Cu fueron en promedio consistentes con los rasgos de tamaño de grano de la zona IV. Estos resultados sugieren que, a pesar de sus indicadores edafoclimáticos de potencial agrícola, esta zona también presenta características de ambientes frágiles.

## Conclusiones

Modelar el comportamiento espacial de cada variable original permitió fusionar información de diferentes fuentes y escalas.

Contemplar la covariabilidad espacial entre variables mediante PCA con restricción espacial (MULTISPATI-PCA) produjo mejores resultados que usar PCA clásico y resultó útil para entender las correlaciones entre variables que dan lugar a la zonificación.

El protocolo propuesto permite diferenciar zonas edafoclimáticas coherentes a escala regional.

Para la provincia de Córdoba se identificaron cuatro zonas edafoclimáticas que podrían contribuir en la interpretación de la variabilidad ambiental de procesos característicos de la dinámica de herbicidas en suelos.



# Predicción de retención de herbicidas en suelo

## Introducción

El proceso de retención condiciona el resto de las interacciones entre la matriz de suelo y la molécula herbicida ya que retarda e inhibe el desplazamiento de las moléculas en el perfil (Bailey y White, 1970; Mamy et al., 2015). La matriz de suelo adsorbente es heterogénea, porosa, con distintos niveles de cristalinidad. Este conjunto de factores dan lugar a un efecto “matriz” que define la interacción herbicida-suelo (Wauchope y Myers, 1985). En la mayoría de los suelos los herbicidas son adsorbidos en mayor o menor medida por los coloides ya sean de origen orgánico o inorgánico y la magnitud con la que participa uno u otro es determinada por características específicas del herbicida y de los coloides (Khan, 1980). Pasado un tiempo (dependiente de la intensidad del proceso) la retención conduce a una estabilización de los compuestos convirtiéndolos en formas menos disponibles y biodegradables, denominadas residuos no extractables o residuos ligados (Koskinen y Clay, 1997). Dado que la mayor fracción del herbicida que se degrada proviene de la fracción que se encuentra en la fase líquida, la retención impacta en las tasas de los procesos subsiguientes como la disipación y mineralización.

La partición del producto incorporado que es retenido en la fracción sólida del suelo (fracción adsorbida) puede variar ampliamente, dependiendo de las propiedades edáficas y para dimensionarla se utiliza el índice de adsorción o  $Kd$ . La determinación del  $Kd$  para cada herbicida y ambiente edáfico se realiza en experimentos que se basan en la agitación de una disolución acuosa conteniendo el herbicida, con una cantidad determinada de suelo. Luego, se calcula la relación entre la concentración del herbicida adsorbido al suelo y la concentración del herbicida en la fase acuosa a través de la siguiente función:

$$x/m = Kd C_{eq}$$

donde  $x/m$  es la concentración adsorbida,  $Kd$  el coeficiente de adsorción de una molécula a una matriz particular y  $C_{eq}$  es la concentración de equilibrio.

Los valores de  $Kd$  varían, entre otras dependencias, en función de las propiedades de los suelos por lo cual es frecuente expresarlos con relación al contenido de carbono orgánico del suelo, obteniéndose así la constante de adsorción relativa al carbono orgánico  $Koc$  (Hodson y Williams, 1988; Calvet, 2005). Ambas métricas se expresan en unidades de volumen por masa. Mientras mayor es  $Kd$  o  $Koc$ , mayor potencialidad de adsorción del herbicida a la matriz de suelo.

Los coeficientes  $Kd$  y  $Koc$ , junto con la vida media del herbicida en suelo pueden ser utilizados en la construcción de indicadores de potencial de pérdida por flujo de agua. Diversos modelos de riesgo ambiental los requieren como dato de entrada. Los valores de  $Kd$  más bajos están relacionados a mayor potencial de pérdida por lixiviación o escurrimiento. El  $Koc$  es especialmente útil para analizar el destino de herbicidas no ionizables ya que éstos se adsorben directamente sobre la materia orgánica del suelo. El potencial de pérdida debido al proceso de erosión del suelo se asocia a valores altos de  $Kd$ , ya que la molécula se pierde junto con la partícula a la cual se encuentra adherida (Lammoglia et al., 2018).

El Glifosato (N-fosfometilglicina) es un herbicida no selectivo usado en cultivos tolerantes tanto en etapas de preemergencia como en postemergencia. El uso de glifosato en la agricultura se ha incrementado notablemente con la adopción de cultivos modificados genéticamente para su resistencia en sistemas de no labranza donde las malezas se controlan químicamente. Su uso, además, se ha generalizado a diversos sectores como el ámbito domiciliario, parques, jardines e incluso el control de malezas acuáticas. Pertenece a una clase de agentes fuertemente quelantes con alta adsorción que la hace una molécula persistente en la matriz suelo. No obstante, el  $K_d$  de glifosato ha demostrado tener un amplio rango de variación a través de distintos suelos; entre 3 y 1100  $Lkg^{-1}$  (Eberbach, 1998; Grünwald y Walraven, 2001; Mamy, Barriuso y Gabrielle, 2005). Esta variabilidad se explica, en parte, por los múltiples mecanismos a través de los cuales se puede retener la molécula en la matriz de suelo. El glifosato forma complejos superficiales con goetita, caolinita, illita, montmorillonita y suelos con similar superficie máxima de adsorción. Según De Jonge et al., 2001 mayores  $K_d$  podrían provenir de un aumento de aluminio amorfo reactivo, óxidos de hierro y pH. Algunos otros procesos que impactan la variabilidad del  $K_d$  de glifosato son: la formación de enlaces de coordinación con cationes polivalentes ( $Al^{3+}$ ,  $Fe^{3+}$ ) adsorbidos a los planos basales de los minerales, cationes adsorbidos en los espacios interlaminares de arcillas expandibles, cationes que forman parte de la estructura superficial de los minerales (McBride y Kung, 1989) mediante quelación con metales en solución (McBride y Kung, 1989; Gimsing y dos Santos, 2005) y enlaces puente hidrógeno con la fracción humificada de la materia orgánica del suelo (MOS) (Piccolo, Celano y Conte, 1996; Albers et al., 2009).

Otro herbicida, ampliamente usado es la Atrazina (6-cloro-N<sub>2</sub>-etil-N<sub>4</sub>-isopropil-1,3,5-triazina-2,4-diamina) especialmente para el control de malezas en cultivos de gramíneas como el maíz. (*Zea mays* L.) y sorgo (*Sorghum bicolor*). Atrazina, es parte de la familia

de las triazinas, se utiliza como herbicida de preemergencia y postemergencia. La bibliografía informa que el contenido de materia orgánica del suelo y su composición afectan la retención de atrazina en el suelo (Bollag y Myers, 1992; Barriuso y Houot, 1996), pero también el pH y el porcentaje de arcilla (Hang, Bocco y Sereno, 2000; Houot et al., 2000). La fracción de MOS gruesa ( $> 50 \mu\text{m}$ ) tiene mayor capacidad para ligar residuos de atrazina que las fracciones más finas (Barriuso et al., 1994). Es importante destacar que los cambios en la concentración relativa de las fracciones de MOS dependen principalmente de las prácticas de manejo del suelo (Yakovchenko, Sikora y Millner, 1998; Skjemstad et al., 2002).

Los  $K_d$  fueron estudiados en suelos diversos, pero bajo condiciones fijadas experimentalmente. Por ejemplo, diferentes condiciones texturales, niveles de materia orgánica del suelo, condiciones de acidez y manejos del suelo (sistemas de labranza, niveles de fertilización, rotaciones de cultivo) (Rampoldi, Hang y Barriuso, 2011; Rampazzo et al., 2013; Aslam et al., 2014). Si bien estos estudios han permitido comprender la naturaleza fisicoquímica del proceso de adsorción a la matriz suelo, no son suficientes para comprender la variabilidad a través de ambientes en un dominio espacial continuo a escala de territorial. Como, la retención es un proceso que se describe a partir de los fenómenos de superficie, la variabilidad del  $K_d$  se encuentra condicionada por las estructuras de variación de las propiedades que definen la afinidad de la molécula a una superficie dada. Entonces, estudiar el proceso de retención del herbicida en un suelo, aislado de las propiedades edafoclimáticas y del manejo del sitio de aplicación, limita la posibilidad de interpretar la información a nivel regional. En esta tesis se propone el mapeo del  $K_d$  basado en modelos que expresan el proceso de retención del herbicida en un sitio como función de variables del sitio. El mapeo, a escala regional, de la variabilidad de este proceso de la dinámica del herbicida en suelo permitirá contextualizar y mejorar

las interpretaciones de los coeficientes de adsorción necesarias para diagnosticar el grado de fragilidad de diferentes ambientes y actuar en consecuencia.

Dado que la mayoría de las variables edáficas que explican la variabilidad del proceso de retención se encuentran espacialmente estructuradas, la modelación de la variabilidad de  $K_d$  deberá contemplar de alguna manera los procesos de autocorrelación espacial subyacentes. Bajo un paradigma frecuentistas son usuales las aplicaciones de regresiones lineales que contemplan la autocorrelación a partir de funciones no-lineales de los errores y de la distancia entre los sitios donde se generaron las observaciones. La estimación de estos modelos vía REML podría implicar costos computacionales importantes en estudios con muchos datos espaciales. Una alternativa en creciente auge debido a los avances computacionales es la regresión bayesiana estimada con procedimientos de cómputo optimizados. En los modelos de regresión bayesiana el objetivo es computar la distribución marginal *a posteriori* para cada parámetro coeficiente de regresión y cada hiperparámetro de la estructura de varianzas y covarianzas del modelo (componente de varianza del efecto espacial aleatorio, componente de varianza del error puro y el parámetro de la función que modela la dependencia espacial). Rue, Martino y Chopin (2009) propusieron derivar la distribución *a posteriori* de una variable estructurada en el espacio con aproximaciones de la dependencia espacial realizadas a partir de integraciones anidadas por Laplace (INLA) y desarrollaron la implementación sobre lenguaje R (R-INLA, Lindgren y Rue, 2015) que facilita su implementación aún con grandes bases de datos. Las aplicaciones de esta alternativa se ha vuelto popular en la modelación de datos espaciales en contextos ambientales (Blangiardo y Cameletti, 2015) y ha mostrado, en el Capítulo I de esta tesis ser competitiva en términos de minimizar el error de predicción respecto a otras estrategias existentes para ajustar regresiones con datos espaciales.

R-INLA trabaja con modelos gaussianos latentes donde los efectos fijos y los efectos aleatorios estructurados o netamente residuales se combinan linealmente (Bakka et al., 2018). La conveniencia computacional radica en la construcción de matrices de precisión ralas de autocorrelación espacial, siendo la alternativa más usada la función de Matérn resuelta por ecuaciones estocásticas diferenciales parciales (Lindgren et al., 2011; Lindgren y Rue, 2015b). La grilla de predicción se construye a través de triángulos que cubren el dominio en  $R^2$ , cada vértice representa un nodo donde se predice la variable respuesta como una combinación lineal de las covariables, esto le permite al algoritmo trabajar bien aún con límites y bordes complejos (Bakka et al., 2018). El marco bayesiano, seleccionado en esta tesis para la construcción de modelos de predicción espacial, ofrece una alternativa viable para el mapeo digital de propiedades de suelos y facilita el mapeo de la incertidumbre de la predicción sitio específica (Poggio et al., 2016).

La hipótesis de este capítulo sostiene que el proceso de retención depende de variables edafoclimáticas que se encuentran estructuradas espacialmente que contienen información útil en la generación de modelos de predicción espacial de  $Kd$  para derivar mapeos digitales a escala regional. Las propiedades del suelo con potencial para predecir los procesos de retención dependen de la interacción matriz-molécula por lo que, en este capítulo, se abordará el caso particular de los herbicidas atrazina y glifosato, modelando la variabilidad espacial de cada uno.

## Materiales y Métodos

### **Base de datos**

Se modelaron los  $Kd$  de glifosato y atrazina a partir de la base de datos descrita en el Capítulo II. Cada sitio cuenta con información de múltiples variables edáficas, topográficas, climáticas y de manejo. Para este capítulo las variables edáficas utilizadas

fueron: pH, NT, COT, Na, K, Ca, Mg, Zn, Mn, Cu, CIC, CC, óxidos de hierro y aluminio totales, contenido de arena, limo y arcilla. Además, se utilizaron las variables elevación, pendiente, media diaria anual de temperatura ( $T_m$ ), precipitaciones medias anuales ( $pp$ ), balance hidrológico como el índice entre  $T_m$  y  $pp$  ( $TvsPP$ ).

Los  $K_d$  no se midieron en la totalidad de los sitios de la base de datos, sino sobre una muestra de sitios (glifosato  $n=89$  y atrazina  $n=156$ ). Los sitios se seleccionaron utilizando la técnica de muestreo hipercubo latino condicionado (cHLS, por el término en inglés *conditioned Hypercube Latin sampling*, Minasny y McBratney, 2006). El muestreo cHLS es usado para obtener muestras en situaciones donde existe abundante información auxiliar sobre cada sitio del espacio a muestrear. Dados  $N$  sitios con información multivariada, cHLS selecciona una muestra de tamaño  $n$  en la que las variables auxiliares se encuentran bien representadas, es decir existen sitios con valores de estas variables en todos los intervalos de clase de sus distribuciones marginales (Minasny y McBratney, 2006; Brus, 2019). Para implementar la técnica se utilizó el paquete *chls* (Roudier, 2011) del software R.

Los índices de  $K_d$  se determinaron según la técnica de *Batch-equilibrium* mediante la preparación de suspensiones suelo: solución de glifosato o atrazina en relación 1:5 por duplicado. Las soluciones de trabajo se prepararon a partir de soluciones madre de glifosato ( $10 \text{ mg L}^{-1}$ ) y atrazina ( $20 \text{ mg L}^{-1}$ ). En tubos de centrífuga se pesaron 2 g de cada suelo y se agregaron 10 mL de soluciones de distintas concentraciones ( $C_i$ ) de cada herbicida; se agitaron durante 24 h en oscuridad y a  $25 \pm 1 \text{ }^\circ\text{C}$ . Se centrifugaron los tubos y los sobrenadantes se filtraron por filtros de celulosa de  $0,45 \text{ }\mu\text{m}$  en viales de 1,5 mL de volumen. Se cuantificó la concentración de cada herbicida ( $C_{eq}$ ) mediante cromatografía líquida de alta presión (HPLC) según Marek y Koskinen, (2014) para glifosato y Steinheimer, (1993) para atrazina. Para atrazina se usó un detector de arreglos de Diodo

(PDA) utilizando una columna con fase estacionaria de octadecil-silano (XBridge C18 3,5  $\mu\text{m}$  – 2,1 x 100 mm – Part N° 186003022 (WATERS)). Para Glifosato se utilizó una columna de intercambio iónico, se realizó una derivatización post columna para la posterior detección por fluorescencia. La concentración adsorbida ( $C_{\text{ad}}$ , x/m,  $\mu\text{g/g}$ ) se calculó como diferencia entre la concentración inicial y la de equilibrio. El  $Kd$  ( $\text{L kg}^{-1}$ ) se obtuvo como el cociente  $C_{\text{ad}}/C_{\text{eq}}$ . Las determinaciones se realizaron en el Laboratorio de Calidad de Alimentos Pampeanos (LabCAP)

### **Protocolo estadístico**

#### *Selección de variables predictoras*

Con la finalidad de simplificar el análisis de regresión bayesiano, se seleccionó un subconjunto de variables con potencial predictiva. utilizando dos métricas: 1) La importancia relativa de cada variable obtenida a partir de árboles de regresión mejorados por remuestreo (por el término en inglés *Boosting Regression Tree*, BRT Elith et al., 2008) y 2) el criterio de deviance (por el término en inglés *Deviance Information Criterion*, DIC, Huang et al., 2017). Las funciones `gbm.step` y `gbm.simplify` del paquete `gbm` en lenguaje R se utilizaron para el análisis BRT optimizando los parámetros del algoritmo (*bag.fraction*, *tree complexity* y *learning rate*) siguiendo el criterio de minimizar el error de predicción calculado por validación cruzada con 10 grupos de entrenamiento y validación (*k fold* con  $k=10$ ). El criterio DIC se obtuvo a partir del cálculo sucesivo de diferencia de deviance entre un modelo de regresión Bayesiana saturado (con todas las variables) y un modelo con una variable menos, con la finalidad de cuantificar la contribución de esa variable (Rue et al., 2009). Las variables ambientales con una contribución relativamente baja sugerida tanto por BRT como por DIC fueron eliminadas de los análisis subsecuentes.

*Ajuste de un modelo de regresión espacial para el coeficiente de adsorción*

Se ajustó un modelo de regresión bayesiana para datos espaciales para el  $Kd$  de cada herbicida en función de variables edafoclimáticas. El ajuste se efectuó a partir de la variable transformada en escala logarítmica debido a la distribución asimétrica que presentaron ambos coeficientes. El modelo de regresión bayesiana para el  $\ln(Kd)$  asume la siguiente distribución para la  $i^{\text{ésima}}$  observación:

$$\ln(Kd_i) \sim N(\eta_i, \sigma_e^2)$$

$$\eta_i = \beta_0 + \sum_{j=1}^p x_{ij}\beta_j + \xi(s_i)$$

donde  $\beta_0$  es el intercepto;  $\beta_j$  es el coeficiente de regresión asociado a la variable explicativa  $x_j$ ;  $x_{ij}$  la valuación de  $x_j$  en el sitio  $i$  y  $\xi(s_i)$  el efecto aleatorio de sitio que se asume una realización de un proceso gaussiano latente  $\xi(s_i) \sim \text{MVN}(0, \Sigma)$ , siendo  $\Sigma$  la matriz de varianza y covarianza de los efectos de sitio definidos por la función de covariación espacial de Matérn (Matérn, 1986).

La estimación de la inversa de  $\Sigma$  (matriz de precisión) se resolvió por SPDE usando R-INLA. El parámetro de suavizado de la función de covarianza ( $\alpha$ ), se optimizó minimizando el error de predicción (RMSPE) por validación cruzada con 10 grupos de entrenamiento y validación (*k fold* con  $k=10$ ) (Krainski y Lindgren, 2013).

Para evaluar la capacidad predictiva de los modelos se utilizó un error de predicción global y un error de predicción puntual o sitio específico calculados a partir del procedimiento de validación *Jackknife* o *leave-one-out* (Efron y Hastie, 2016). Los errores de predicción se calcularon como la diferencia entre los valores observados y la media de la distribución *a posteriori* del valor predicho para cada sitio. Estos errores se utilizaron para calcular el RMSPE global con el fin de evaluar el ajuste espacial. A su vez, las medidas de error puntual o sitio específico se calcularon expresando los errores

de predicción para cada sitio en relación con su valor observado de  $Kd$ . Finalmente, para dimensionar el impacto del error en la interpretación del proceso de retención los valores de  $Kd$  observados y predichos para cada herbicida fueron categorizados como bajo, medio y alto. Bajo  $Kd \leq 10 \text{ Lkg}^{-1}$  y  $\leq 1,5 \text{ Lkg}^{-1}$ , Medio  $Kd: 10-30 \text{ Lkg}^{-1}$ ;  $1,5-2$  y Alto  $Kd: \geq 30 \text{ Lkg}^{-1}$ ;  $\geq 2,5 \text{ Lkg}^{-1}$ , para glifosato y atrazina, respectivamente. La asociación entre las categorías de  $Kd$  observado y predichos fue evaluada por una matriz de confusión desde la cual se derivó la tasa de clasificación correcta. Los valores usados para construir las categorías de  $Kd$  de ambos herbicidas fueron definidos de acuerdo con la información existente de estas moléculas en la bibliografía (Rampazzo et al., 2013; Becerra, Hang y Díaz Zorita, 2015)

#### *Mapeo de retención de glifosato y atrazina*

La predicción espacial de  $Kd$  para ambos herbicidas se realizó en una grilla densa de  $2,5 \times 2,5$  km. Las predicciones se expresaron en la escala original aplicando la función exponencial a los valores predichos para cada sitio por el modelo de regresión bayesiana (Stow, Reckhow y Qian, 2006; Correa Morales et al., 2018). La malla utilizada para el ajuste de la matriz de precisión por SPDE se escogió entre seis opciones (combinatoria entre tres valores alternativos del largo máximo de los triángulos y dos alternativas para el parámetro de suavizado) bajo el criterio de maximizar la representación espacial en el  $\ln(Kd)$  medida como el grado de la varianza espacialmente estructurada RSV (Robertson et al., 1993) derivada de los parámetros  $\theta$  según Blangiardo y Cameletti, (2015). La incertidumbre respecto a las predicciones se informó a través de mapas de desvíos estándares de la distribución *a posteriori* predicha en cada sitio y con mapas de intervalos de credibilidad del 95% de probabilidad (Huang et al., 2017; Giannini Kurina et al., 2018).

## Resultados y Discusión

El protocolo de análisis propuesto incluye: la selección de variables, la construcción de un modelo de variabilidad y la predicción espacial sobre una grilla a escala regional. El diseño metodológico es innovador y ofrece la posibilidad de obtener mapas de la variabilidad espacial del proceso de retención de un herbicida en suelo, desde muestras convenientemente seleccionadas. Dado que la retención se predice desde variables de sitio, el protocolo facilita la interpretación del proceso de retención en diferentes ambientes. La matriz de datos construida a partir de determinaciones de sitios muestrales, y de determinaciones de algunos  $Kd$  permitió la construcción de un modelo con capacidad de predicción sitio-específica. Los resultados se presentan acorde a los pasos del protocolo de análisis propuesto.

### **Selección de variables predictoras**

La selección de variables predictoras es un paso crítico, más aún cuando existen múltiples variables correlacionadas (Kuhn y Johnson, 2013). Las técnicas de selección de variables y el criterio agronómico deben acoplarse en esta etapa. En este trabajo se usó el algoritmo BRT por su capacidad de manejar escenarios con variables colineales y relaciones no lineales. Además se usó DIC como es usual en el ajuste de modelos bayesianos (Rue y Held, 2005).

El conjunto de variables ambientales usadas en la construcción de modelos de retención de cada herbicida, según los criterios BRT y DIC (Figura 5), fue diferente. Para el  $Kd$  de atrazina ambos métodos priorizaron variables climáticas relacionadas a la disponibilidad hídrica ( $TvsPP$ ,  $pp$ ) y el contenido de COT, y arcillas de los suelos. La selección por DIC permitió identificar también una contribución del pH y del contenido de Cu del suelo superficial que la selección por BRT no destacó. Todas estas variables fueron

consideradas en el ajuste del modelo predictivo. Para el  $K_d$  de glifosato ambos métodos destacan la contribución de arcilla y pH y el BRT sugiere la consideración del contenido de óxido de aluminio. Sin embargo, las variables climáticas, posiblemente determinantes de las mencionadas, no aportan información marginal al modelo.

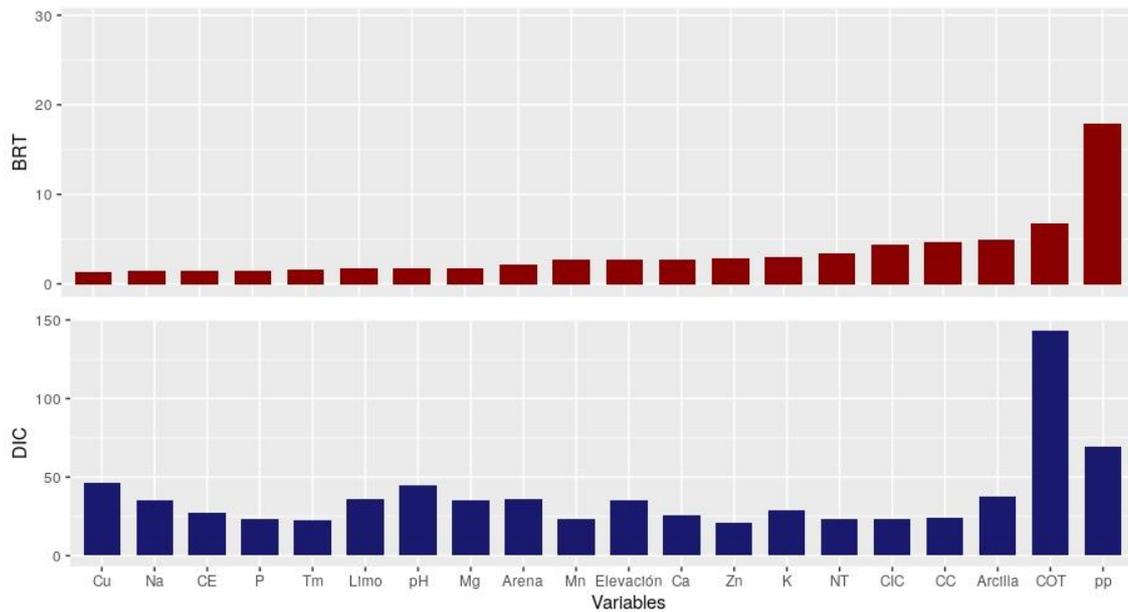


Figura 5. Importancia de variables edáficas y climáticas para explicar la variabilidad del coeficiente de retención de atrazina según el método BRT y el índice DIC.

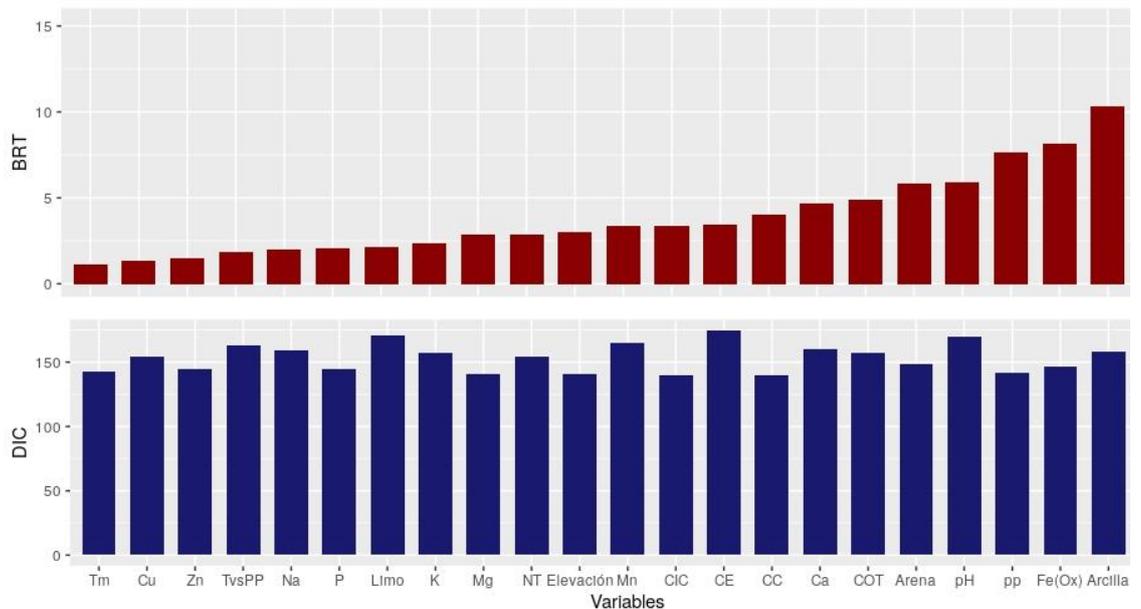


Figura 6. Importancia de variables edáficas y climáticas para explicar la variabilidad del coeficiente de retención de glifosato según el método BRT y el índice DIC

Los procesos de adsorción de moléculas orgánicas al suelo se encuentran condicionados a fenómenos que ocurren en el complejo de intercambio donde el contenido de arcilla, la acidez y el COT, resultan variables influyentes dado su naturaleza coloidal (Moreale y Van Bladel, 1979; Celis et al., 1998). Por lo tanto, incluir estas variables resulta consistente desde un punto de vista agronómico.

El COT y las arcillas participan en la generación de cargas y en sentido general, favorecen la retención. El incremento de pH favorece el incremento de cargas negativas y ha sido documentado como una variable que disminuye la adsorción de moléculas orgánicas a la matriz de suelo (Calvet, 2005). La contribución del Cu a la variabilidad del proceso de retención no ha sido documentada pero podría deberse a que este elemento se encuentra fuertemente relacionado con el tipo de materia orgánica debido a la formación de complejos del suelo (Alvarez y Lavado, 1998). Las variables climáticas pp y TvsPP que representan la pluviometría anual acumulada y su relación con la temperatura media, i.e. un proxy de balance hídrico, son frecuentemente usadas para resumir la acción del clima como factor de formación de suelo (Jenny, 1994; McBratney, Mendonça Santos y Minasny, 2003) y además condicionan el uso agrícola del mismo. Por eso, fueron seleccionadas como potenciales variables explicativas, particularmente pp fue señalada por ambos criterios usados en la selección de variables como potencial predictora (Figuras 5 y 6).

En glifosato, se ha reportado la formación de enlaces de coordinación con cationes polivalentes ( $\text{Al}^{3+}$ ,  $\text{Fe}^{3+}$ ) adsorbidos a los planos basales de los minerales y los espacios interlaminares de arcillas expandibles (Hance, 1976; Meng et al., 2003), hallazgo congruente con la alta participación de los óxidos de hierro y aluminio en la explicación de la variabilidad del  $K_d$  de glifosato (Figura 6). El DIC selecciona variables con relaciones lineales con la respuesta, por lo que resulta apropiada su complementación con

otro criterio que permita explorar relaciones no lineales como BRT (Elith et al., 2008). Los óxidos de Al y Fe podrían manifestar relaciones más complejas con la retención del herbicida en suelo.

### **Predicción espacial y mapeo de retención**

Los modelos para mapeo digital de glifosato y atrazina se construyeron a partir de las variables seleccionadas, incluyendo términos de segundo orden para modelar posibles relaciones cuadráticas. En el Cuadro 8 se informan los valores del parámetro de suavizado para resolver la ecuación de Matérn por SPDE que arrojaron los menores RMSPE. Además, se informan las estimaciones de los modelos de predicción tanto para atrazina como para glifosato. La combinación de variables que mejor explica la variabilidad del proceso de retención de glifosato incluyó las variables óxidos de aluminio, pH y contenido de arcilla. El modelo permitió explicar más del 85% de la variabilidad en el  $\ln(Kd)$  de glifosato (Cuadro 8). A mayores valores de óxidos de aluminio se incrementó la retención de glifosato mientras que el aumento de pH la disminuyó.

En el caso de atrazina, las variables pp, TvsPP, COT y arcilla incrementaron la retención de la molécula a las partículas de suelo. Estudios puntuales con suelos contrastantes sobre el mecanismo de retención de atrazina en suelo también identificaron a la materia orgánica del suelo y el porcentaje de arcillas como variables explicativas del  $Kd$  de atrazina (McGlamery y Slife, 1966; Weber et al., 2004). Esos trabajos también destacan la participación de la acidez suelo (McGlamery y Slife, 1966; Linn et al., 1993), aunque para los suelos y la escala aquí estudiada la relación entre la adsorción de atrazina y el pH de suelo no resulta evidente. Los resultados obtenidos a escala regional corroboran relaciones entre el proceso de retención y condiciones edáficas estudiadas en laboratorio con condiciones controladas, pero amplían la información sobre el modelo relacional que

permite mapear la variabilidad territorial del proceso e interpretarlo en relación al ambiente.

Los modelos para  $Kd$  propuestos en la bibliografía se han construido a partir de regresiones lineales múltiples donde los efectos de las variables ambientales se tratan como fijos y las observaciones como independientes (Becerra et al., 2015; Rampazzo et al., 2013; Spurlock et al., 2006; Weber et al., 2004). En este trabajo se modeló la variabilidad de los  $Kd$  considerando que parte de esta variabilidad se encuentra espacialmente estructurada, como sugieren los índices de autocorrelación espacial para el  $\ln(Kd)$  de ambas moléculas (Cuadro 8). La regresión múltiple con efecto aleatorio de sitio, bajo un marco bayesiano, resultó una buena aproximación para describir el continuo espacial del proceso de retención. A partir de este modelo es posible generar predicciones espaciales para el territorio en estudio en una rápida ejecución computacional (<3min en un computador con un procesador de 2 núcleos y 8 GB en memoria RAM). La modelación del efecto espacial en el término aleatorio permite dimensionar de mejor manera los efectos correspondientes a las variables predictoras de la estructura de medias del modelo ya que ésta se estima junto a la estructura de varianzas y covarianzas del modelo pudiendo distinguirse del efecto inducido por la correlación espacial subyacente (Schabenberger y Gotway, 2005; Krainski y Lindgren, 2013; Bakka et al., 2018).

Cuadro 8. Estimación y validación del modelo de retención de herbicidas

Parámetros y ajuste	Atrazina		Glifosato
Coeficientes de regresión			
Intercepto	-4,332±1,772	Intercepto	4,11±1,09
TvsPP[C°mm <sup>-1</sup> ]	3,577±1,101	arcilla [%]	0,05±0,01
pp[mm]	0,005±0,001	Al(Ox) [%]	21,26±6,89
COT [(g kg <sup>-1</sup> )]	0,018±0,008	Al(Ox) <sup>2</sup> [%]	-51,07±24,73
COT <sup>2</sup> [(g kg <sup>-1</sup> )]	0,0004±0,0002	pH	-0,56±0,12
arcilla [%]	0,006±0,003		
Hiperparámetros†			
$\theta_1$	1,21±0,14	$\theta_1$	1,52±0,26
$\theta_2$	-13,97±1,24	$\theta_2$	-12,55±0,85
$\alpha$	1	$\alpha$	1,085
Validación cruzada y bondad de ajuste			
RMSPE relativo a la media	30,49	RMSPE relativo a la media	12,96
IM ln(Kda)	0,59 (p <0,001)	IM ln(Kdg)	0,39 (p <0,001)
IM residual	-0,05 (p=0,976)	IM residual	-0,05 (p=0,788)
Tasa de clasificación correcta	87,62%	Tasa de clasificación correcta	88,06%
Sitios con ESE <25%	74%	Sitios con ESE <25%	67%

† hiperparámetros de la función de Matérn que modela el proceso de autocorrelación espacial de los efectos aleatorios de sitio. DE, desvío estándar de la distribución *a posteriori* del parámetro; RMSPE, raíz cuadrada del error cuadrático medio de predicción espacial obtenido con una malla de vecinos con parámetro de suavizado  $\alpha=2$ ; ESE, error sitio específico

La evaluación de un modelo a partir del error global de predicción puede resultar insuficiente y resulta conveniente complementar este criterio a partir de errores sitio-específicos, más aún cuando se trabaja sobre modelos de predicción espacial. Por otro lado, incorporar una medida de evaluación que permita dimensionar el impacto de los errores de exactitud en la interpretabilidad del problema permitiría mejorar las conclusiones.

Respecto a la proporción de sitios con error de predicción sitio específico menor al 25 %, en la implementación realizada en esta tesis se observó una proporción de 74 % y 67 % para glifosato y atrazina, respectivamente (Cuadro 8). Los errores de predicción sitio específicos obtenidos tienen bajo impacto respecto a la interpretación ambiental ya que

las altas tasas de clasificación correcta (Cuadro 8) fueron altas 88 % para glifosato y 87 % para atrazina.

### **Distribución espacial del proceso de retención e implicancias ambientales**

La predicción espacial del  $K_d$  de glifosato y atrazina en Córdoba se muestra en las Figuras 7 y 8. El mapa de la distribución espacial de glifosato muestra que los menores valores de  $K_d$  se encuentran al oeste de la provincia y los valores altos en el centro-norte y noreste. Los mayores valores coinciden con zonas de alta intensificación agrícola donde los riesgos de pérdidas por lavado y escurrimiento superficial son menores, habilitando prácticas de manejo conservacionistas como sistematización de suelo, siembras contra pendiente, cultivos en franja, entre otras. En general, estas prácticas tienden a minimizar las pérdidas de material por erosión de suelo ya sea de origen hidrológico o eólica y disminuir el transporte de la molécula a destinos no deseados. Las zonas sur y sureste de la provincia donde se concentran los suelos arenosos, según se observó en la zonificación edafoclimática realizada en el Capítulo II, registra bajos valores de retención de herbicidas en suelo por lo que las pérdidas por lavado o percolación pueden representar un problema, no obstante, la acidez de suelo podría contrarrestar los efectos del bajo contenido de arcilla (Hance, 1976; Rampazzo et al., 2013). La situación se agrava en el extremo noroeste donde además de contener bajos contenidos de arcilla aumenta el pH del suelo (Jarsún et al., 2006) es decir son suelos que se predisponen una menor retención. La retención de atrazina al suelo persigue un patrón similar en cuanto a tendencia oeste-este, registrando los mayores valores al este de la provincia. El hecho de que la adsorción de atrazina esté fuertemente asociada con la variabilidad del COT, que es una variable que depende en gran medida de las prácticas de manejo (West y Post, 2002; Sun, Zhou y Zhao, 2003), implica que prácticas que tiendan a estabilizar la materia orgánica del suelo, como la siembra directa y la rotación de cultivos, podrían reducir la peligrosidad de

contaminación del sitio por pérdidas de la molécula del herbicida. Bajo condiciones de altas retenciones, los riesgos de pérdida de las moléculas de herbicidas estarían más asociadas al desplazamiento de partículas de suelo por transporte o escorrentía y consecuentemente, las prácticas de conservación para mitigar erosión evitarían la contaminación.

En referencia al diseño metodológico propuesto, se observa que la regresión bayesiana estimada con R-INLA resultó adecuada para la interpolación espacial de los valores de  $Kd$  a escala regional proveyendo tanto un modelo explícito para comprender la retención de herbicidas al suelo como para mapear su distribución espacial. Una ventaja adicional, respecto a otros modelos para datos espaciales, es la simplicidad con la que la predicción espacial realizada en una escala transformada, por ejemplo logarítmica, se puede expresar fácilmente en la escala original (Minasny y McBratney, 2016). En contraste con el enfoque frecuentista el cual asume como fijos los parámetros del modelo, en el marco bayesiano estos se consideran variables probabilísticas que cuentan con funciones de distribución que pueden resultar útiles para obtener intervalos de credibilidad e incorporar información externa (Gelman, 2004). La predicción *a posteriori* de la variable respuesta a su vez ofrece una herramienta potente para dimensionar la incertidumbre (Diggle, 2013; Huang et al., 2017). En la Figura 7 y 8 Se ilustran las medias de la distribución *a posteriori*, pero también la desviación estándar derivada. Puede notarse que la mayor incertidumbre se da en zonas con menor densidad de observaciones. Las amplitudes de los intervalos de credibilidad son mayores para mayores  $Kd$  porque la desviación estándar se encuentra en la escala de la medición. No obstante, la interpretación del patrón de variabilidad subyacente es independiente de la escala. En el marco teórico frecuentista, la incertidumbre de predicciones espaciales (Goovaerts, 2001; Diggle, 2013) se obtienen por simulaciones a partir de modelos geoestadísticos, focalizándose en la variable

respuesta y no en el resto de los parámetros del modelo. En el marco bayesiano hemos descrito la incertidumbre a partir de la distribución *a posteriori* generada para los valores predichos en cada sitio. El modelo bayesiano permite calcular un intervalo de credibilidad para la predicción de cada sitio, es decir que se informa un rango específico dentro del cual pueden estar las predicciones (Banerjee, Carlin y Gelfand, 2014; Blangiardo y Cameletti, 2015). Es importante destacar que el *Kd* caracteriza un proceso multidimensional y que a pesar de que el enfoque analiza el proceso fisicoquímico en condiciones experimentales, el mapeo digital basado en modelos de la retención según características de sitio espacialmente estructuradas permite abordar la variabilidad ambiental en un continuo.

Aunque no se muestran medidas de incertidumbre para las predicciones espaciales, los modelos del tipo regresión kriging (Hengl et al., 2007) han sido usados para el mapeo digital de propiedades de suelo (Hengl et al., 2004; Vaysse y Lagacherie, 2015). La aplicación de cualquiera de estos marcos teóricos para el mapeo digital de procesos de la dinámica de las moléculas de herbicida en suelo es novedoso. Aquí, un modelo de regresión, abordado desde una perspectiva Bayesiana para datos espaciales, ha permitido realizar el mapeo digital de la retención de los herbicidas en el suelo. Es importante destacar que, si bien la retención es un parámetro relevante en la dinámica de una molécula en suelo, no es el único (Linn et al., 1993; Mamy et al., 2015; Lammoglia et al., 2018). La comprensión más acabada de la dinámica requiere de otros estudios que aborden las estructuras de covariación respecto a los procesos de persistencia en suelo para inferir el posible destino ambiental de moléculas con potencial contaminante.

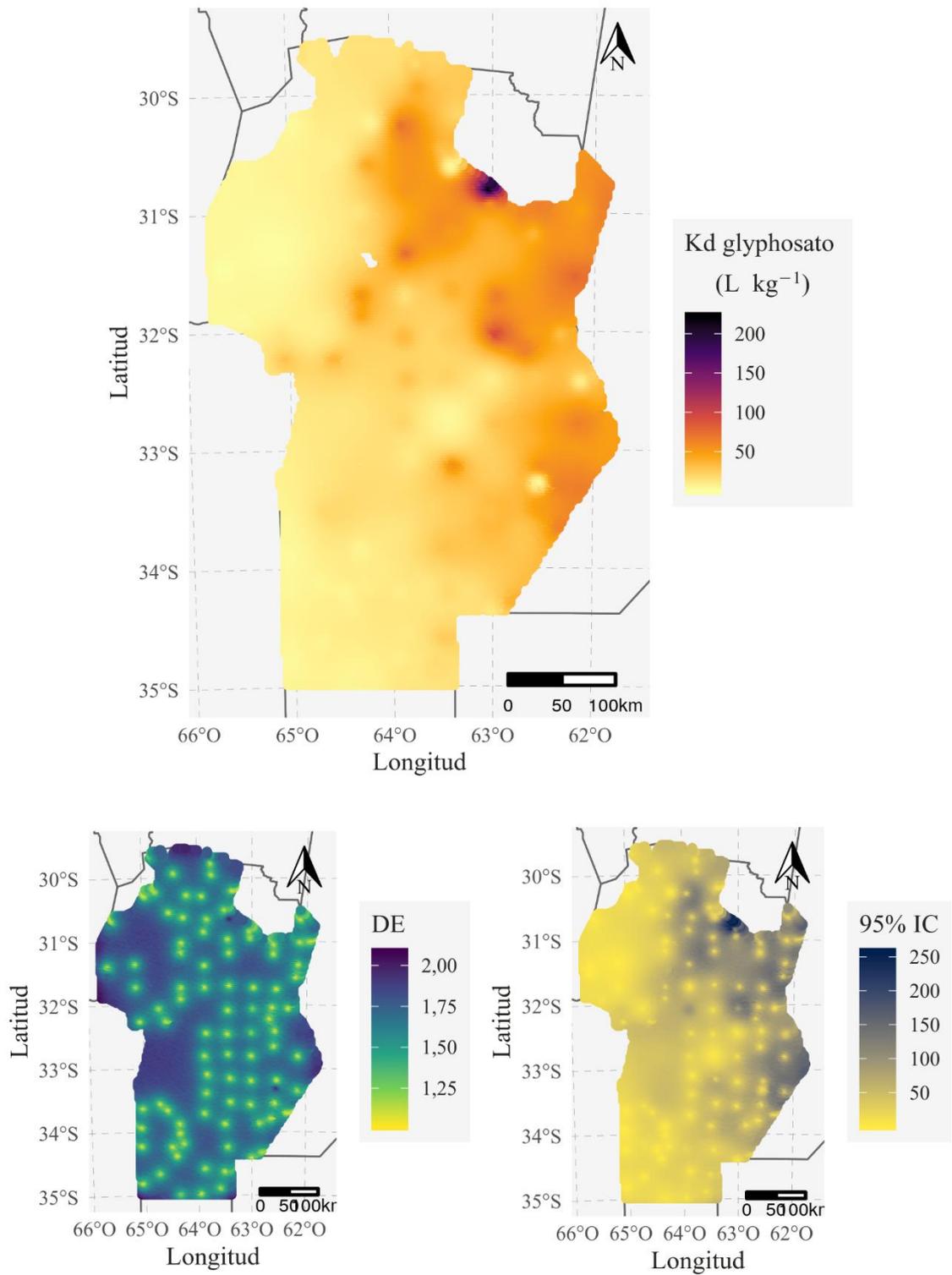


Figura 7. Distribución espacial de retención de glifosato en Córdoba e incertidumbre asociada expresada como desvío estándar (abajo, izquierda) y como amplitud de intervalos de credibilidad (abajo derecha)

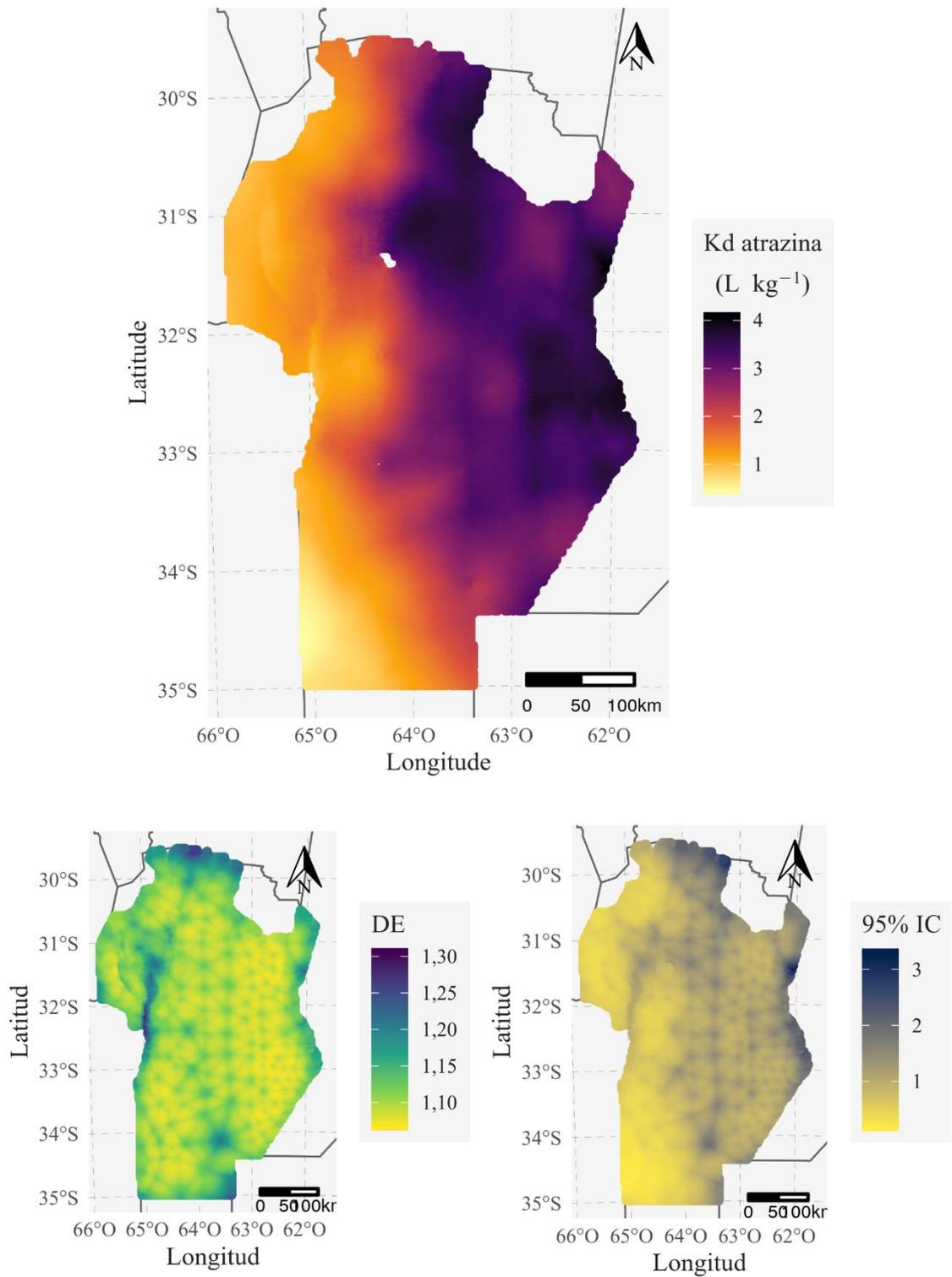


Figura 8. Distribución espacial de retención de atrazina en Córdoba e incertidumbre asociada expresada como desvío estándar (abajo, izquierda) y como amplitud de intervalos de credibilidad (abajo derecha)

## Conclusiones

La retención de glifosato depende de las variables óxidos de aluminio, pH y arcilla mientras que la retención de atrazina se ve condicionada a COT, arcilla y a indicadores climáticos relacionados al déficit hídrico. De esta manera, a escala regional se ha encontrado que la estructura de relación e interacción de las propiedades que afectan el proceso de retención es consistente, a nivel de variables de suelo, con la publicada desde experimentos puntuales y controlados. No obstante, con el cambio de escala, el comportamiento ambiental en un continuo se ha podido plasmar en mapas que cuentan con medidas de incertidumbre de predicciones. Así, el mapeo digital basado en modelos a partir de la regresión Bayesiana espacial no sólo permitió detectar los factores incidentes en el proceso retención, sino también describir su distribución espacial para identificar situaciones de peligrosidad ambiental por la salida de la molécula del sistema suelo.

# Modelación de la disipación y su covariabilidad con la retención de herbicidas en suelo

## Introducción

Para explicar la dinámica de un herbicida en suelo, además del estudio del proceso de retención, resulta fundamental caracterizar la persistencia de la molécula del producto activo en dicha matriz. Para estudios de persistencia es usual analizar la vida media ( $t_{1/2}$ ) de la reacción química subyacente, medida que se refiere al tiempo necesario para que la concentración en solución de reactante, en este caso el herbicida, llegue a la mitad de su nivel inicial. Propiedades químicas, físicas y biológicas del suelo afectan la persistencia de los herbicidas en suelo (Koskinen y Clay, 1997). Los mecanismos que hacen a la disipación de la molécula del suelo comprenden volatilización, fotólisis, lixiviación potencial, sorción y degradación. Bajo condiciones controladas como las incubaciones en laboratorio, las pérdidas por volatilización, fotólisis y lixiviación pueden considerarse despreciables, la sorción se subsana mediante mediciones a tiempo cero y en sucesivos tiempos se evalúa la desaparición por degradación de la molécula. Ésta puede darse tanto por mecanismos abióticos (hidrólisis, reacciones óxido reducción) como bióticos

(biodegradación y cometabolismo), siendo la degradación microbiana la vía más importante en suelos (Goring et al., 1975; Spurlock et al., 2006). La degradación microbiana depende de la accesibilidad a la molécula; se supone que moléculas con mayor coeficiente de adsorción ( $K_d$ ) tendrían menor degradación por encontrarse menos disponibles en la solución. Moléculas con menor  $K_d$  se encuentran fácilmente disponibles para el ataque microbiano (Hang et al., 2003; Krutz et al., 2008).

Además de la biodisponibilidad y las condiciones ambientales, otro factor preponderante en la degradación microbiana es la estructura de poblaciones microbianas presentes y su actividad (Soulas y Lagacherie, 2001; Nannipieri et al., 2003). Debido a sucesivas aplicaciones de un herbicida en suelo se puede generar una adaptación microbiana permitiendo una mineralización acelerada del compuesto, lo que se ha observado tanto en condiciones controladas como a campo. La mineralización acelerada relacionada con el uso repetido de herbicidas ha sido tempranamente documentada (Leistra y Green, 1990). Para atrazina en suelos se cuentan con múltiples antecedentes, (Stolpe y Shea, 1995; Topp et al., 1995; Jenks et al., 1998; Houot et al., 2000; Hang et al., 2003; Hang, Houot y Barriuso, 2007). Como es un herbicida utilizado en el cultivo de gramíneas, es esperable que ciertos microorganismos de los suelos agrícolas cultivados con estas especies vegetales, favorezca la degradación de atrazina en tiempos acotados (Barriuso y Houot, 1996; Hang et al., 2007, 2011). En sitios del territorio en estudio incluso se ha alcanzado a mineralizar hasta el 50% de la cantidad aplicada inicialmente después de 28 días de incubación (Hang et al., 2003; Hang, Barriuso y Houot, 2005). Dado que la disipación está condicionada por procesos microbianos a su vez dependiente de las características que definen el ambiente del suelo, se supone que la distribución espacial de  $t_{1/2}$  debiera estar también condicionada por el uso del suelo. A diferencia de las variables de sitio que explican el proceso de retención, la presencia de determinados cultivos (principalmente

gramíneas) en el suelo es una variable para considerar en los modelos predictivos de vida media.

Para ajustar un modelo que explique la persistencia del herbicida en función de ésta y otras características de sitio, es necesario obtener mediciones de la variable respuesta, i.e. cuantificar  $t_{1/2}$ . Para el cálculo de vida media es necesario modelar el decaimiento temporal de la concentración de la molécula en cada suelo particular. Para ello, se pueden utilizar experimentos de incubación donde se fortifica el suelo con la molécula y se monitorea su concentración en distintos momentos de la incubación. Un modelo estadístico usado para representar curvas de disipación es el modelo exponencial (Gustafson, 1989; Fernandes et al., 2006) de dos parámetros que incluye la ordenada al origen  $C_0$  y la pendiente  $k$  del decaimiento en el tiempo de la concentración inicial. Éste es un modelo de regresión de efectos fijos donde la concentración del herbicida se expresa en función del tiempo de incubación desde la fortificación del suelo con el compuesto. Estimada la curva de disipación, la  $t_{1/2}$  se calcula como función de la estimación de  $k$  y representa el tiempo al cual la disipación ha alcanzado la mitad de la concentración inicial. Los estudios de disipación de herbicidas usualmente ajustan curvas de disipación para cada suelo o curvas promedio de una población de muestras de suelo, pero siempre en el marco teórico de un modelo no lineal de efecto fijo. Consecuentemente, el modelo es estimado asumiendo no solo la existencia de una única componente de varianza (la residual o asociada a la variabilidad no explicada por el modelo exponencial) sino también la independencia de los datos usados para el ajuste de la curva, los cuales se corresponden con mediciones sucesivas de una misma muestra de suelo. Sin embargo, debe considerarse que las observaciones en días sucesivos de la incubación podrían estar temporalmente correlacionadas debido a que son mediciones repetidas en el tiempo sobre la misma unidad de análisis (datos longitudinales). La correlación entre lecturas sucesivas de

concentraciones de herbicidas podría ser alta no solo por provenir de una misma muestra de suelo (sitio) con propiedades fisicoquímicas particulares, sino también por la presencia de un mismo pool de microorganismos. Luego, es el mismo efecto suelo o sitio el que imprime correlaciones entre las medidas repetidas. La falta de consideración de este patrón de correlación temporal en la modelación de curvas de disipación puede conducir a conclusiones erróneas respecto a las pendientes del proceso y consecuentemente a la estimación de la  $t_{1/2}$ .

La modelación no lineal con mediciones repetidas puede ser abordado a través del uso de modelos no lineales mixtos (Verbeke y Molenberghs, 2009) mediante la incorporación de efectos aleatorios asociados a uno o más de los parámetros de efectos fijos del modelo. En esta tesis proponemos adicionar un efecto suelo aleatorio al término de la pendiente de la función exponencial que modela de disipación ya que este efecto aleatorio no sólo permite cuantificar la variabilidad entre suelos, sino que también induce correlaciones entre las observaciones realizadas sobre un mismo suelo. Se hipotetiza que un modelo mixto con efecto fijo del tiempo de incubación y efecto aleatorio de sitio podría mejorar la estimación de la curva de disipación y consecuentemente, los cálculos de vida media. Más aún, los modelos mixtos amplían las posibilidades de inferencia y será posible inferir tanto respecto a la disipación promedio poblacional (conjunto de suelos) como también sobre la disipación sitio-específica. Aún cuando los modelos no lineales mixtos (Davidian, 2017) ofrecen una herramienta potente para la estimación de curvas de disipación y el cálculo de la  $t_{1/2}$ , éstos no han sido usados en estudios de persistencia de herbicidas.

Al igual que lo que sucede con el proceso de retención, si bien se conoce que la variabilidad de la  $t_{1/2}$  depende no solo de la molécula del herbicida sino también de las características del sitio donde se aplica (clima, suelo y uso) (Fernandes et al., 2006; Krutz

et al., 2008; Sarmah, Close y Mason, 2009; Bento et al., 2019), la disipación ha sido estudiada hasta el presente en contextos experimentales y a través de comparaciones de condiciones puntuales. Para extender el estudio de la variabilidad de  $t_{1/2}$  a escala regional como función de variables edafoclimáticas y de uso de suelo de los distintos sitios de un territorio, se propone ajustar un modelo lineal de regresión múltiple para la variable dependiente vida media que habilite a la predicción de la persistencia en un dominio espacial continuo.

En el capítulo anterior se modeló la variabilidad espacial del procesos de retención mediante regresión bayesiana para datos espaciales (Blangiardo y Cameletti, 2015) y se concluyó que esta metodología ofrece una alternativa eficiente para la predicción espacial de una variable en un dominio continuo a escala regional . El mismo procedimiento será usado para predecir  $t_{1/2}$  de un herbicida en función de variables de sitio incluyendo el uso del suelo.

Los procesos de retención y disipación determinan la propensión de un suelo a la desaparición de la molécula del herbicida, es decir la propensión a que el herbicida desaparezca del ambiente en forma definitiva por mineralización. Entender el comportamiento conjunto o covariación espacial de ambos procesos permite inferir los posibles destinos que puede tener una molécula de herbicida aplicada al suelo. Estos procesos no se dan de manera aislada o independiente, sino que se relacionan y esta relación puede manifestar un patrón espacial dado que la mayoría de las variables ambientales que intervienen en la explicación de estos procesos covarían espacialmente. Las técnicas propuestas para monitoreos ambientales de plaguicidas suelen basarse en valores umbrales de propiedades de las moléculas o en modelos matemáticos que explican el proceso de lixiviación. Un ejemplo, es el índice GUS (del inglés, Grounwater Ubiquity Score (Gustafson, 1989)), que ha sido ampliamente usado porque relaciona los procesos

de retención y disipación, para dimensionar el potencial contaminante de un plaguicida sobre las reservas acuíferas. El índice se basa en el análisis gráfico de la correlación entre vida media en suelo y el coeficiente de retención expresados en relación con las unidades de COT. La ventaja de expresar el  $K_d$  de esta manera, en adelante denominado  $K_{oc}$ , es que éste representa una medida de movilidad independiente del suelo. Además, la variabilidad del  $K_{oc}$  es usualmente menor que la variabilidad del  $K_d$ , por lo que  $K_{oc}$  es aceptado universalmente como un indicador para comparar herbicidas en términos de movilidad de suelo. Si bien, hay grandes diferencias en las magnitudes en las cuales los herbicidas exhiben propensión al lixiviado y varias propiedades físicas tales como solubilidad en agua y volatilidad impactan en la lixiviación, el análisis simultáneo de la retención y disipación resulta útil para reconocer situaciones con potencialidad contaminante (Weber et al., 2004).

Tanto los coeficientes de retención como la vida media de los herbicidas muestran amplios rangos de variabilidad en un territorio. El índice GUS es presentado como un sistema de clasificación para determinar en qué situaciones puede ocurrir lixiviación de pesticida, pero los procesos que impactan la potencialidad contaminante son varios y no resulta suficiente el estudio solo de la lixiviación hacia aguas subterráneas. Otra debilidad del índice GUS, es que ha sido desarrollado para clasificar moléculas respecto a la propensión de lixivarse con valores de  $K_d$  y  $t_{1/2}$  referidos a un suelo particular y no es fácil su extensión hacia otros suelos.

También existen modelos que permiten evaluar riesgos ambientales asociados a la actividad agropecuaria bajo diversos escenarios de uso del suelo y condiciones ambientales, como los modelos PRZM, (Pesticide Root Zone Model), LEACHM (Leaching Estimation and Chemistry Model) y CMLS (Chemical Movments in Layered Soils) (Carsel, 1984; Jury, Focht y Farmer, 1987; Marín-Benito et al., 2014). Algunos de

éstos alimentan modelos hidrológicos de amplio uso, como es el caso de SWAT, (Soil and Water Assessment Tool) (Klute, 1986; Neitsch, Arnold y Srinivasan, 2002; G. Vazquez-Amabile, B. A. Engel y D. C. Flanagan, 2006; Srinivasan, Zhang y Arnold, 2010) que permite analizar fenómenos de contaminación a nivel de cuenca. Estos modelos requieren abundante información para caracterizar las unidades o sistemas hidrológicos en estudio (aforos, dimensionamiento de curva número, erodabilidad del sistema, entre otras) y una serie de esfuerzos en la calibración local para su implementación y validación (Behrends Kraemer et al., 2011). Consecuentemente, la implementación de éstos se realiza a escalas locales más que regionales, donde es difícil conseguir monitoreos intensos y sistemáticos de las cuencas hidrológicas (Havrylenko et al., 2016; Romagnoli et al., 2017).

Una estrategia distinta a la del uso de índices y modelos de riesgos ambientales es el mapeo digital de la variabilidad y covariabilidad entre procesos claves de la dinámica de un herbicida en suelo. El mapeo basado en modelos estadísticos abre nuevas oportunidades para realizar aproximaciones a la peligrosidad a escala regional. Los modelos de correlación espacial entre dos variables comprenden tanto la autocorrelación espacial como la correlación cruzada espacial. Si bien, la teoría de autocorrelación espacial ha sido bien desarrollada proveyendo medidas de autocorrelación a través de estadísticos como el índice de Moran o el coeficiente de Geary y mediante modelos geoestadísticos (Cressie y Chan, 1989), menor avance metodológico existe para explorar la correlación cruzada espacial o covariación espacial entre dos variables espacializadas. (Chen, 2013), presentó una metodología para el análisis de la correlación cruzada espacial que es una modificación de una propuesta anterior para estudiar grado de asociación espacial entre dos variables (Anselin, 1995). Por analogía con el índice de Moran se derivó un marco teórico para modelos espaciales de correlaciones cruzadas. Se definieron

coeficientes de correlaciones cruzadas espacial (global y local) y se propusieron gráficos de dispersión para visualizar patrones subyacentes a los procesos espaciales. En base al coeficiente de correlación cruzada global la correlación se descompone como directa (correlación parcial) e indirecta (correlación cruzada). Se hipotetiza que estos coeficientes de correlación espacial cruzada pueden mapearse en el territorio para visualizar la covariación espacial de los índices de procesos que interesan para estudiar la dinámica de un herbicida en suelo.

Los objetivos del presente capítulo son: 1) la estimación de curvas de disipación y cálculo la  $t_{1/2}$  de un herbicida en distintos suelos, 2) la caracterización espacial de la persistencia en función de características de sitio, y 3) la variación espacial conjunta de los parámetros que caracterizan la retención y la disipación de un herbicida en suelo para dimensionar la peligrosidad ambiental por el uso del mismo. Se trabaja con el caso particular de la variación de  $t_{1/2}$  de atrazina a escala de la provincia de Córdoba.

## Materiales y Métodos

### Base de datos

#### *Selección de sitios e información auxiliar*

Desde el sistema de información territorial edafoclimática de la provincia de Córdoba descrito en el Capítulo II, se seleccionaron, para realizar determinaciones de disipación, un total de 61 sitios utilizando la técnica de muestreo hipercubo latino condicionado (cHLS, Minasny y McBratney, 2006). Además de las variables edáficas y climáticas que caracterizan a cada sitio, se incorporó una variable de uso derivada de información registrada durante el levantamiento de muestras sobre el uso del suelo en el sitio muestreado (al momento de la toma de la muestra y del cultivo antecesor). Las múltiples situaciones de uso del suelo registradas se agruparon en 4 categorías contrastantes desde

el punto de vista del microbioma que se espera encontrar. Estas categorías fueron: Cultivo con presencia de gramíneas, Cultivos sin presencia de gramínea, Pastizales y pasturas, Monte y Arbustal.

#### *Determinaciones de disipación*

Las curvas de disipación se construyeron incubando los 61 suelos previamente seleccionados por 21 días y cuantificando la concentración de atrazina a los días 0,3,7,14 y 21 después de suministrarles el herbicida. Se pesaron 2 g de suelo en tubos de centrifuga de 50 ml. Se fortificó cada suelo con una concentración de  $4,9 \mu\text{g g}^{-1}$  (equivalente a una aplicación de  $3 \text{ L ha}^{-1}$ ), y la cantidad de agua para llegar a un 80 % de capacidad de campo. Para la solución de fortificado se trabajó con estándar de atrazina, (6-cloro-N2etil-N4-isopropil-1,3,5-triazina-2,4-diamina), provistos por *Chem Service* y de pureza certificada superior al 99 %. Se trabajó en cámaras de cultivo a  $28 \text{ }^\circ\text{C}$ , en oscuridad. A los 0, 3, 7, 14 y 28 días se midieron 2 réplicas de cada suelo. Las determinaciones se realizaron en el Laboratorio del Programa de Agroquímicos del Centro de Excelencia en Productos y Procesos Córdoba (CEPROCOR), dependiente del Ministerio de Ciencia y Tecnología de la provincia. La extracción de atrazina y el proceso de limpieza se ejecutó mediante una adaptación del procedimiento QuEChERS (Pang, Wang y Hu, 2016) respetando la siguiente secuencia:

#### *Extracción:*

Agregado de 10 ml de agua Milli-Q, agitado manual durante 1 min.

Agregado de 10 ml de acetonitrilo y 0,5 g de acetato de sodio, agitado manual durante 1 min.

Agregado de 2 g de sulfato de magnesio, agitado manual durante 1 min.

Centrifugado durante 5 min a 3000 rpm

#### *Limpieza:*

Se toman 5 ml de sobrenadante y se agregan en un tubo con 0.3 g de sulfato de magnesio, 0.1 g de PSA (por el término en inglés, “*Primary/Secondary Amine*”) y 0,1 g de C18, se agita manualmente 1 min.

Centrifugado durante 5 min a 3000 rpm.

Se toma 1ml de sobrenadante y se coloca en vial para su lectura.

La cuantificación se realizó por cromatografía líquida acoplada a espectrometría de masas tándem (LC-MS/MS).

### **Protocolo estadístico**

#### *Estimación de la curva de disipación sitio-específica y cálculo de vida media*

La modelación del decaimiento de la concentración de herbicida en función del tiempo se realizó con un modelo no lineal mixto de la forma exponencial con efecto aleatorio asociado a la pendiente y variando entre suelos. La ecuación del modelo de disipación ajustado es:

$$C_i(t) = C_0 e^{-(k+u_i) t}$$

donde  $C_i(t)$  es el porcentaje de recuperación de la molécula de atrazina en el tiempo  $t$  para el suelo  $i$  (concentración relativa a la concentración recuperada al inicio de la incubación),  $k$  es la pendiente de la curva de disipación poblacional,  $u_i$  es el efecto aleatorio del sitio o suelo  $i$ , expresado como una desviación de la pendiente del sitio  $i$  respecto a la pendiente poblacional,  $t$  es el momento de tiempo (expresado en días desde el inicio de la incubación) en el que se realiza la medición y  $C_0$  es la ordenada al origen. El efecto aleatorio de sitio se modeló con distribución normal de media cero y varianza  $\sigma_u^2$  y se supuso no correlacionado con la componente aleatoria asociado al término de error, es decir:

$$u \sim N(0, \sigma_u^2)$$

La significancia de la componente de varianza asociada al término aleatorio fue evaluada a través de la prueba del cociente de verosimilitud (LRT) comparando un modelo exponencial sin efecto aleatorio de sitio con un modelo que incluye este efecto aleatorio de sitio asociado a la pendiente. Se utilizó una distribución Chi-cuadrado con 0,5 grados de libertad dado que se compara un modelo de efecto fijo con otro que incluye un nuevo parámetro asociado a un efecto aleatorio (Molenberghs, Verbeke y Demétrio, 2007). De esta manera, si la prueba LRT resulta significativa el modelo de mayor número de parámetros (modelo con efectos fijos y aleatorios) resulta más apropiado que el modelo más simple (modelo clásico o modelo de efectos fijos). El modelo fue estimado por aproximación directa a la verosimilitud usando el paquete nlme (Pinheiro et al., 2017) de R a través de la interfase de InfoStat (Di Rienzo et al., 2019), obteniendo el estimador máximo verosímil de la pendiente promedio poblacional y el mejor predictor lineal insesgado (BLUP) del efecto aleatorio de sitio. Además de la prueba LRT se utilizaron los criterios AIC y BIC (Schwarz, 1978; Sakamoto, Ishiguro y Kitagawa, 1986). Para estos criterios de ajuste se elige el modelo con valores más bajos de AIC y BIC.

Para evaluar falta de correlación entre los errores del modelo propuesto se utilizó la prueba Durbin Watson (Durbin, 1970). La prueba estadística es  $d = 2(1 - r)$ , donde  $r$  es la autocorrelación de los residuos del modelo. Entonces, si los errores son independientes,  $d$  obtendrá valores cercanos a 2, por el contrario, si los errores están fuertemente auto correlacionados,  $d$  estará cercano a cero.

Luego de estimar  $k$  y predecir  $u_i$ , el efecto aleatorio de sitio sobre la pendiente de la curva de disipación para cada suelo, la  $t_{1/2}$  para el suelo  $i$ , se calculó con la siguiente función:

$$t_{1/2} \text{ }_i = \frac{\ln(0,5)}{-(k + u_i)}$$

La incidencia de los usos de suelos en la disipación de atrazina se abordó a partir de un modelo lineal para explicar  $\ln(t_{1/2})$  en función de categorías de uso, con varianzas

heterogéneas entre usos. A partir de la distribución de los BLUPs de los efectos aleatorio de sitio, se obtuvo la desviación promedio de las desviaciones de la pendiente de cada sitio respecto al pendiente promedio poblacional ( $k$ ) para cada tipo de uso. Este promedio de BLUPs de efectos de sitio según tipo de uso puede ser usado para obtener predicciones de  $t_{1/2}$  condicionadas al uso existente.

#### *Modelación de la variabilidad espacial de la vida media*

Para modelar la variabilidad espacial de la  $t_{1/2}$  en función de variables edafoclimáticas climáticas se usó regresión bayesiana incluyendo un efecto aleatorio de sitio para modelar el patrón espacial con una función de Matérn resuelta por ecuaciones diferenciales estocásticas parciales (SPDE). El ajuste se realizó con R-INLA (Lindgren y Rue, 2015a) siguiendo la metodología propuesta en el Capítulo III de esta tesis. La preselección de variables se realizó combinando métodos de aprendizaje automático (BRT, Elith, Leathwick y Hastie, 2008) con la información derivada de la comparación de deviance de modelos alternativos (DIC, Huang et al., 2017). La capacidad predictiva del modelo ajustado se evaluó a través de validación cruzada del tipo *Jackknife* (Efron y Tibshirani, 1997) a partir de la cual se calculó un error de predicción global (RMSPE) así como errores sitio-específicos. Para obtener la distribución *a posteriori* de valores predichos de vida media, se usó el modelo de regresión ajustado y una grilla de predicción densa (2,5×2,5 km). Se construyó un mapa con la media de la distribución *a posteriori* de  $t_{1/2}$  para cada sitio de la grilla de predicción. También se generaron mapas de incertidumbre de la predicción usando la amplitud de los intervalos de credibilidad conformado con los percentiles 2,5 y 97,5 de la distribución *a posteriori* de  $t_{1/2}$  (Poggio et al., 2016; Huang et al., 2017).

#### *Variabilidad conjunta de retención y disipación*

La relación entre  $Koc$  y  $t_{1/2}$  se exploró a través del índice de GUS:

$$\text{GUS} = \log_{10}(t_{1/2}) \times (4 - \log_{10}(K_{oc}))$$

Cada sitio fue clasificado de acuerdo con su propensión a lixiviación (Gustafson, 1989). En un diagrama de dispersión del logaritmo decimal del  $K_{oc}$  y el logaritmo de la vida media se representaron las 61 muestras originales de suelo, donde se midió retención y disipación. Además, el índice GUS se calculó para cada uno de los sitios predichos para lograr un mapeo de la propensión a la pérdida de atrazina por lixiviación.

Posteriormente, se obtuvieron los coeficientes de correlación cruzada espacial para las dos variables de la dinámica de atrazina en suelo obtenida a partir de los mapas de variabilidad espacial de  $K_d$  y  $t_{1/2}$ . Se clasificaron los coeficientes de correlación según los valores de  $K_{oc}$  y  $t_{1/2}$  en las siguientes categorías: 1) Alto  $K_d$  y Alta vida media, 2) Alto  $K_d$  y baja vida media, 3) Baja  $K_d$  y Baja vida media, 4) Bajo  $K_d$  y Alta vida media y 5) No significativa. Se describió la correlación condicional a la localización geográfica (correlación dada su ubicación) entre  $K_d$  y  $t_{1/2}$  para cada zona edafoclimática homogénea definida en el Capítulo I.

## Resultados y Discusión

### **Curvas de disipación y cálculo de vida media**

La estimación de los parámetros del modelo de disipación exponencial para atrazina, con y sin efecto aleatorio de sitio sobre la pendiente, se muestran en el Cuadro 9. Los criterios de información AIC y BIC disminuyen para el modelo de efectos aleatorios y la prueba LRT confirma que el modelo con efectos aleatorios de sitio tiene un mejor ajuste que el modelo exponencial comúnmente usado. Luego, la variabilidad existente entre suelos es estadísticamente significativa y tiene efecto sobre la velocidad de disipación modificando la forma de las curvas provenientes de distintos suelos (Figura 9).

En la Figura 9 se observa alta variabilidad entre suelos respecto a la pendiente de la curva de disipación; por encima de la curva promedio poblacional (línea negra) se observan curvas provenientes de suelos que han experimentado una disipación con comportamiento prácticamente lineal en el tiempo, mientras que por debajo de la curva promedio poblacional se observan curvas con mayor velocidad de disipación existiendo suelos en los que se ha alcanzado a disipar el 100% de la concentración inicial de atrazina, en el periodo de seguimiento de 21 días. La curva promedio poblacional (línea negra) permite comprender globalmente el proceso de disipación de atrazina en la provincia de Córdoba y sugiere una  $t_{1/2}$  de aproximadamente 10 días. Según la literatura, este valor indica una persistencia baja de atrazina en el suelo. Sin embargo, la variabilidad relativamente alta de los efectos de sitio indica que  $t_{1/2}$  presenta un alto rango de variación entre 3 y 100 días según las características de sitio (Figura 10). En la Figura 10 se presenta la distribución de frecuencia de  $t_{1/2}$ , la cual es asimétrica izquierda con el 75 % de los suelos asociados a  $t_{1/2}$  de atrazina menor o igual a 19 días ( $P(75) = 18,7$ ). Desde un punto de vista ambiental este resultado puede ser positivo ya que representa baja persistencia del herbicida en la mayoría de los sitios.

Cuadro 9. Estimaciones de modelos exponencial de disipación de atrazina en suelos

	Modelo de efecto fijo	Modelo mixto
Parámetros		
$C_0$	98,11 ± 2,12	100,39 ± 0,88
$k$	-0,056 ± 0,003	-0,072 ± 0,007
$\sigma_u^2$		3,9×10 <sup>-5</sup>
Función de parámetros		
$t_{1/2}$	12,01	9,96
Criterios de ajuste		
AIC	2748,46	2426,62
BIC	2759,66	2441,53
Durbin Watson	1,73	1,82
Log Verosimilitud Residual	-1371,23	-1209,31
LRT		161,92 (p-valor<0,001)

$C_0$ , ordenada al origen;  $k$ , pendiente del modelo de disipación exponencial;  $\sigma_u^2$ , varianza de los efectos aleatorios de sitio sobre la pendiente del modelo exponencial; AIC, criterio de información de Akaike; BIC, criterio de información bayesiano; LRT, prueba de cociente de verosimilitud

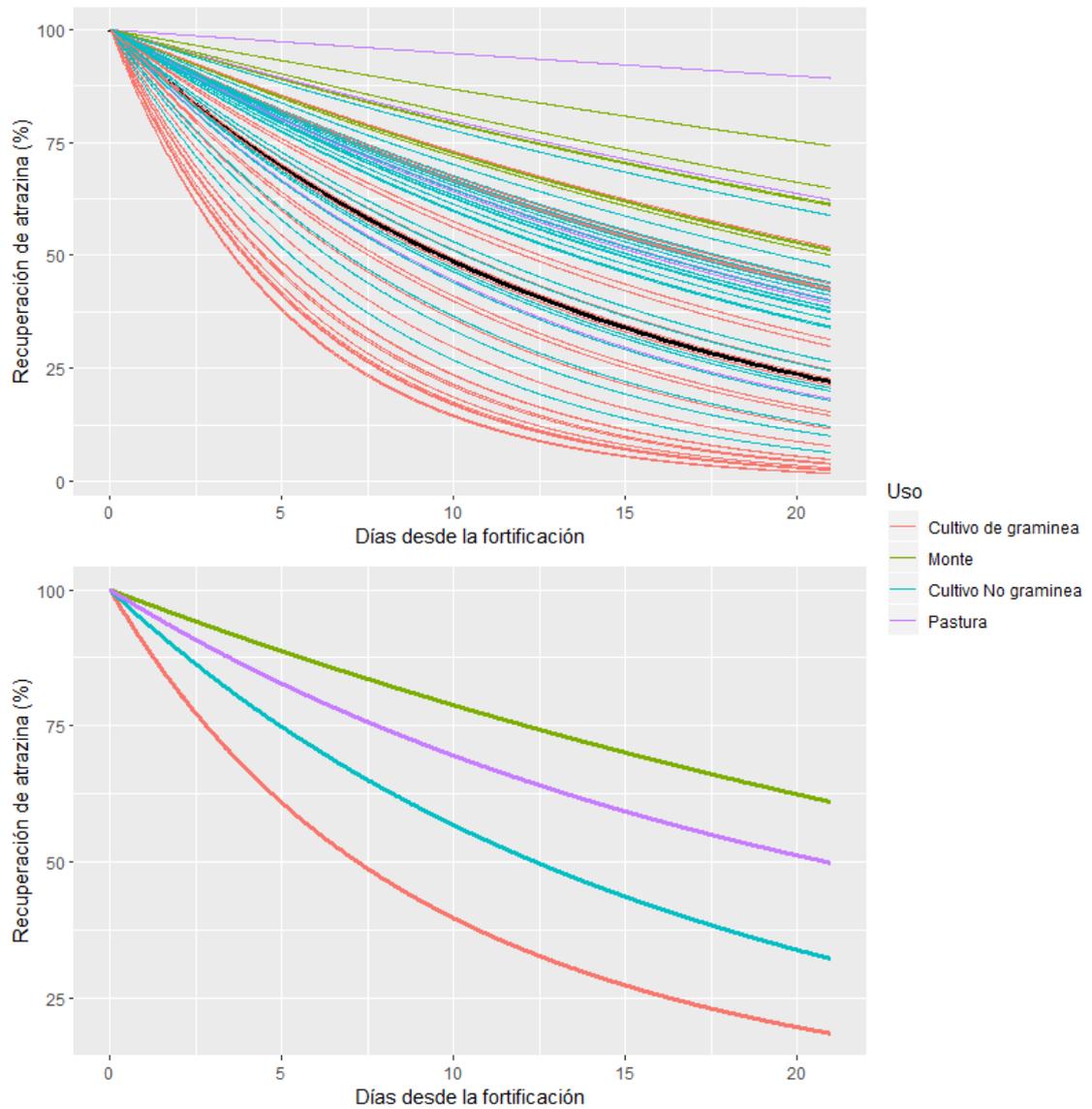


Figura 9. Curvas de disipación de atrazina en suelos. Curvas sitio-específicas (arriba). Curvas medias segun uso de suelo (abajo)

La disipación puede ocurrir por mineralización, transformación en compuestos intermedios o metabolitos o por retención. Es decir que el conjunto de mecanismos que pueden desencadenarse en el suelo actúa sobre las moléculas orgánicas de distintas maneras: 1) eliminándolas completamente como dióxido de carbono, iones y agua, 2) transformándolas en compuestos intermedios, 3) reteniéndolas a la matriz del suelo por diferente tipo de enlaces químicos y físicos (Fenner et al., 2013). Esta capacidad del suelo de funcionar inactivando y eliminando sustancias potencialmente tóxicas, existe por la

heterogeneidad física, química y biológica del sistema suelo, sus múltiples fases y la interacción de procesos y componentes (Kah, Beulke y Brown, 2007). Es importante destacar que las curvas de disipación que se obtuvieron para la molécula de atrazina en este trabajo no consideran los metabolitos derivados de la molécula que también tienen potencialidad contaminante. Encontrar varios metabolitos diferentes es indicador de una intensa transformación microbiana en el suelo (Hang et al., 2003).

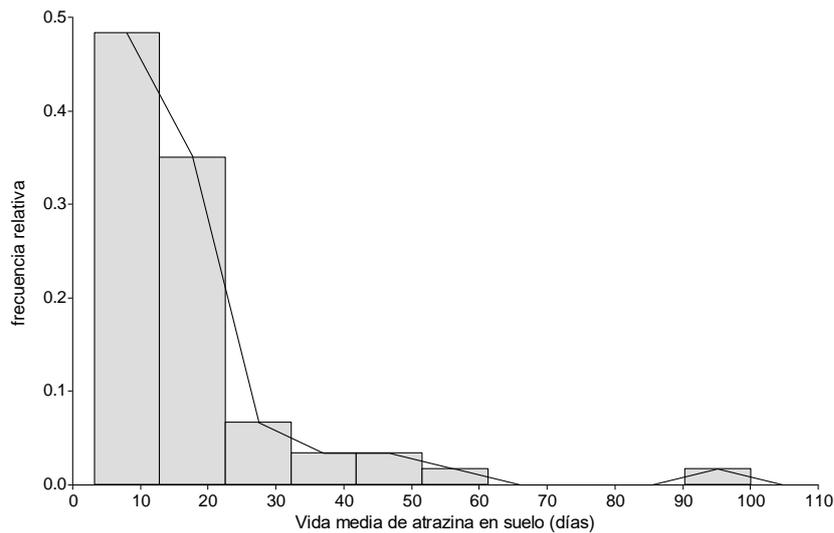


Figura 10. Vida media de atrazina en suelos de la provincia de Córdoba, Argentina.

La atrazina es un herbicida usado principalmente para el tratamiento del cultivo de maíz (*Zea mays* L.) por lo que la frecuencia de maíz en las sucesiones de cultivos de un lote agrícola determina la frecuencia de uso del herbicida. En la provincia de Córdoba, la rotación de cultivos más frecuente es la de soja (*Glicine max* L.) con maíz en años intercalados, siendo la aplicación de atrazina año de por medio suficiente para desarrollar microorganismos adaptados que generen la mineralización del anillo triazínico (Hang et al 2003). La posible adaptación de la microflora del suelo para la degradación de la atrazina ha sido demostrada y es la causal de mineralización acelerada con porcentajes de mineralización que pueden alcanzar el 60 % de la atrazina aplicada (Barriuso y Houot, 1996). En suelos de la región central de Córdoba con monocultivo de sorgo por más de

20 años se obtuvieron porcentajes de mineralización de atrazina de 80 % en 21 días (Hang et al., 2011).

Luego, se puede decir que la magnitud de la degradación de atrazina se encuentra condicionada a la adaptación microbiana que a su vez depende de la presión que generan sucesivas aplicaciones del herbicida en el suelo (Barriuso y Houot, 1996; Krutz et al., 2008). Hang et. al. (2007), trabajando con suelos colindantes, con y sin aplicación de atrazina, determinaron porcentajes de mineralización de atrazina 70 % y 6 %, respectivamente. Este amplio rango de variación fue explicado por la historia de uso de atrazina de los lotes.

En el Cuadro 10 se muestra la  $t_{1/2}$  de atrazina en la provincia de Córdoba según el uso del suelo. Suelos con montes o pasturas, donde la probabilidad de aplicación de atrazina es menor, registran valores significativamente mayores de  $t_{1/2}$  que aquellos con uso de gramíneas sobre los cuales la aplicación de la molécula es habitual. La persistencia en suelos fortificados con atrazina que provenían de monte fue, en promedio, 10 veces mayor a la registrada en suelos con uso de gramíneas (92,44 vs. 9,24 días). El uso de este herbicida en suelos sin historia de cultivo de gramíneas, por ejemplo, de desmonte reciente, conlleva altas persistencias del potencial contaminante. En otros estudios locales se detectaron después de 60 días de incubación porcentajes cercanos al 10 % de mineralización de atrazina, en suelos de monte o agrícolas sin uso de atrazina (Hang et al., 2007). Resultados que confirman que la mineralización acelerada de atrazina depende principalmente de la historia de uso del herbicida en suelos.

Las vidas medias provienen de curvas de disipación que fueron estadísticamente diferentes para los distintos usos; las pendientes del decaimiento de la concentración de atrazina respecto a la inicial fueron de -0,11 a -0,02 % día<sup>-1</sup> según el suelo tuviese antecedentes de cultivos con gramínea o no (monte). Estas pendientes se obtuvieron

adicionando a la pendiente de la curva de disipación promedio poblacional (-0,072 % día<sup>-1</sup>) el promedio de los efectos de sitio para la categoría de uso; por ejemplo, en los suelos con antecedente de cultivo de gramínea la velocidad de disipación aumento 0,03 unidades y en suelos provenientes de montes la pendiente disminuyó en 0,05 unidades (Cuadro 10).

Cuadro 10. Modelo de disipación según uso del suelo

Categorías de usos	n	$(\bar{u})$	$(k + \bar{u})$	$t_{1/2}$				Significancia
				Media	DE	Mín	Máx	
G	26	-0,034	-0,11	9,24	6,01	3,22	24,22	a
NG	21	0,01	-0,06	13,68	4,73	6,01	24,93	b
P	5	0,03	-0,04	34,24	164,25	18,98	529,00	c
M	9	0,05	-0,02	92,44	37,60	9,41	100,11	c

G, cultivo de gramíneas; NG, cultivo no gramíneas; P, pasturas; M, monte;  $\bar{u}$ , promedio de las desviaciones aleatorias de sitio respecto a la pendiente promedio poblacional;  $(k + \bar{u})$ , promedio de pendientes de las curvas de disipación;  $\bar{t}_{1/2}$ , vida media. Significancia, letras distintas indican diferencias estadísticas (p-valor < 0,05)

### Variabilidad espacial de la vida media de atrazina en Córdoba

El algoritmo BRT priorizó como variable explicativa de la variabilidad en vida media al uso de suelo (72 % vs 28 % de importancia relativa del uso respecto a la importancia promedio de las restantes variables potencialmente explicativas de  $t_{1/2}$ ). El criterio de información basado en la deviance de la regresión bayesiana espacial (DIC) indicó la importancia relativamente alta de TvsPP, un indicador de balance hídrico. De las variables edáficas del sitio las que mostraron mayor contribución para explicar la distribución de  $t_{1/2}$  fueron los contenidos de potasio (K), de arcilla y la conductividad eléctrica (CE), en ese orden.

En el Cuadro 11 se muestra el modelo de regresión lineal mixto ajustado para explicar  $t_{1/2}$  como función de efectos fijos de las variables de sitio seleccionadas y de un efecto aleatorio de sitio que refleja el proceso de autocorrelación espacial subyacente en los valores de  $t_{1/2}$ . El valor positivo de los coeficientes de regresión para las clases de uso indicadas en el Cuadro 11 sugiere una mayor  $t_{1/2}$  de atrazina bajo cualquiera de estos usos

que la registrada en suelos que provienen de cultivos con gramíneas. Mayores valores de CE también aumentan la  $t_{1/2}$  la vida media de atrazina mientras que mayores valores de contenido de K y arcilla disminuyen la persistencia de la molécula en suelo. La textura, en el caso de los suelos de Córdoba representado por el porcentaje de arcilla ha sido reportada como una variable que condiciona la  $t_{1/2}$  de atrazina en suelos bajo condiciones experimentales. Existen antecedentes que documentan la disminución de la velocidad de degradación en condiciones de salinidad elevada, la cual podría estar asociada a la contribución de la CE en el modelo ajustado. Sin embargo, menor atención ha recibido en la literatura el aporte de cationes, en particular del ion K, en el proceso de disipación. Futuros estudios relacionados al aporte de cationes en la degradación de la molécula de herbicida en el suelo podrían explicar la relación encontrada en este modelo empírico. El indicador climático usado en el modelo ajustado presenta una estructura espacial fuerte relacionada con la topografía de la provincia y resume múltiples procesos concomitantes. Por lo tanto, la contribución de esta variable en el modelo puede ser más indirecta que directa. La degradación microbiana en el ambiente está evidentemente influenciada por las temperaturas y precipitaciones que impactan sobre el sistema. Las temperaturas y el contenido de agua, así como el tipo de suelo, han sido documentados como factores determinantes de la vida media de atrazina en suelos (Stolpe y Shea, 1995; Topp et al., 1995; Houot et al., 2000). No obstante, debido a las condiciones experimentales en las que se realizó el estudio de disipación en este trabajo (incubación a temperatura y humedad controlada) el impacto de la variable  $T_{vsPP}$  se considera más asociado a un efecto espacial más que climático.

Cuadro 11. Modelo de disipación en función de variables de sitio

Parámetros y ajuste	Medias $\pm$ DE
<i>Coefficientes de regresión</i>	
Intercepto	4,31 $\pm$ 0,75
TvsPP [C°mm <sup>-1</sup> ]	-5,82 $\pm$ 2,84
K [ppm]	-0,0007 $\pm$ 0,0004
Arcilla [%]	-0,02 $\pm$ 0,01
CE [dSm <sup>-1</sup> ]	0,27 $\pm$ 0,12
Cultivo sin gramínea <sup>†</sup>	0,26 $\pm$ 0,15
Pastura <sup>†</sup>	0,70 $\pm$ 0,27
Monte <sup>†</sup>	1,35 $\pm$ 0,26
<i>Hiperparámetros<sup>††</sup></i>	
$\theta_1$	5,694 $\pm$ 0,95
$\theta_2$	-7,83 $\pm$ 0,50
Validación cruzada	
RMSPE relativo a la media	22,7%
Sitios con RMSPE <25%	70,0%

<sup>†</sup> El valor indica diferencia de medias entre  $\ln(t_{1/2})$  del uso respecto al uso con cultivo de gramíneas. <sup>††</sup> hiperparámetros de la función de Matérn que modela el proceso de autocorrelación espacial de los efectos aleatorios de sitio. DE, desvío estándar de la distribución *a posteriori* del parámetro; RMSPE, raíz cuadrada del error cuadrático medio de predicción espacial obtenido con una malla de vecinos con parámetro de suavizado  $\alpha = 2$ .

La media de la distribución *a posteriori* de los hiperparámetros ( $\theta_1, \theta_2$ ) asociados a los efectos aleatorios de sitio es relativamente alta comparada con el desvío estándar (Cuadro 11), sugiriendo la presencia de un proceso de autocorrelación espacial en la componente aleatoria distinta al error puro del modelo. Siguiendo (Krainski et al., 2018) a partir de los parámetros  $\theta_1$  y  $\theta_2$  se llega a un valor de 24076,8 m para el rango y una varianza espacialmente estructurada (Sill) de 0,38. El modelo estimado constituye una herramienta para predecir la  $t_{1/2}$  en función de las características de sitio sin necesidad de estimar curvas de disipación en laboratorio. Debido a que el proceso de disipación se encuentra estructurado espacialmente la  $t_{1/2}$  en sitios del vecindario, del sitio para el cual se requiere una predicción, es informativa y el conjunto de predicciones espaciales sobre una grilla apropiada de predicción provee valores de vida media para el mapeo a una escala regional. El modelo ajustado en este trabajo presentó un error global de predicción fue del 22,7 %

de la vida media promedio y bajos errores de predicción sitio específicos ya que en el 70 % de los sitios correspondientes a suelos donde las determinaciones de disipación fueron realizadas los errores de predicción relativos a la  $t_{1/2}$  observada fueron menores al 25% (Cuadro 11). En la Figura 11 se presenta el mapa de variabilidad espacial (Predicción y medida de incertidumbre) de la vida media de la provincia de Córdoba. Se observa que la vida media en el noroeste de la provincia donde hay una menor historia de cultivos agrícolas es mayor. Sin embargo, más del 80 % del territorio provincial presenta suelos donde es de esperar baja  $t_{1/2}$  para atrazina.

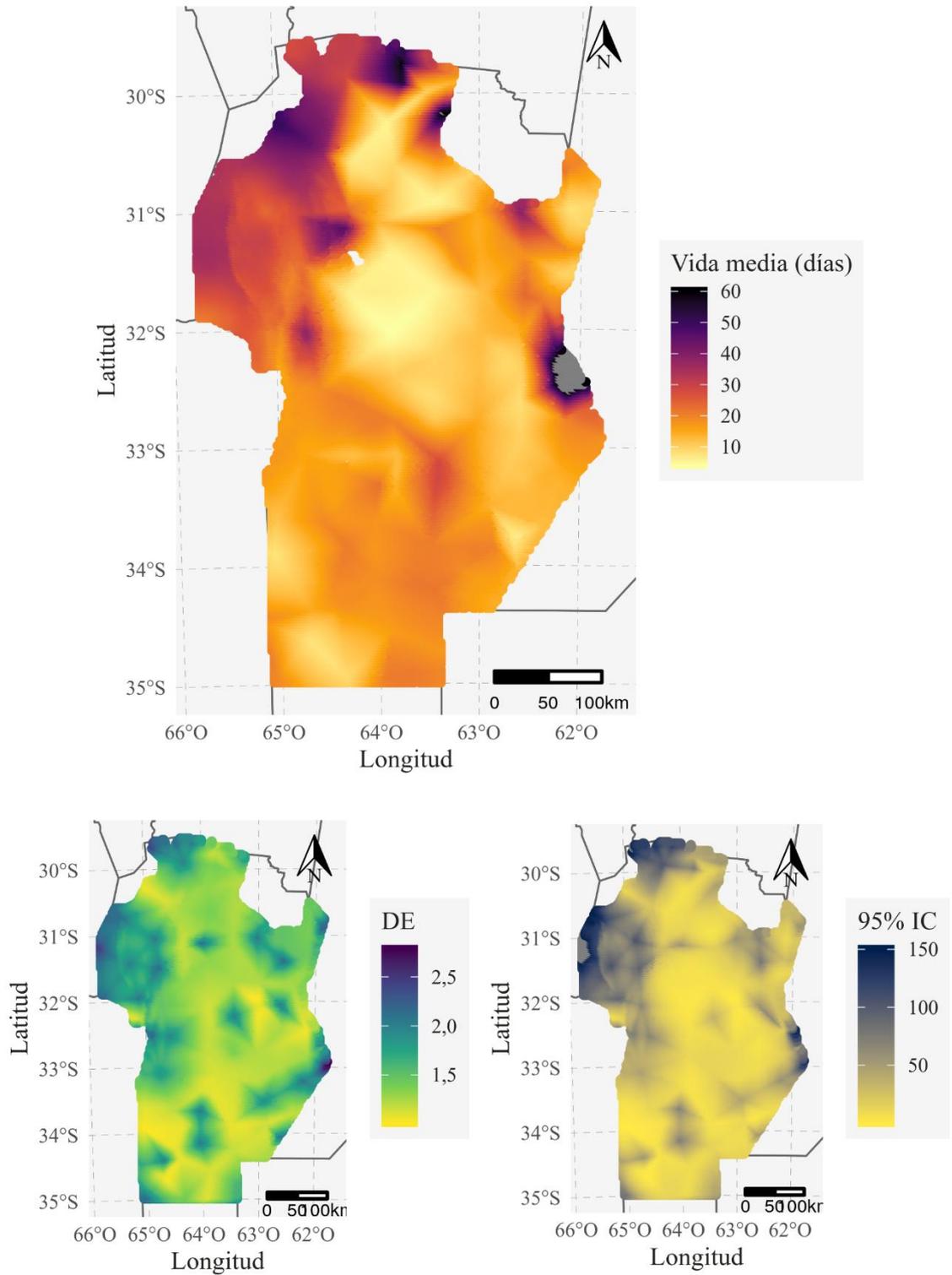


Figura 11. Variabilidad espacial en valores predichos de vida media de atrazina según características del sitio en la provincia de Córdoba (arriba). Mapas de incertidumbre de la predicción (abajo): desvío estándar (izquierda) y amplitud de intervalos de credibilidad (derecha)

### **Variabilidad conjunta de retención y disipación a escala regional**

La relación entre retención y vida media para los 61 suelos donde se determinaron estas mediciones experimentalmente se muestra en la Figura 12. Si bien la clasificación de GUS tiene 6 categorías (Gustafson, 1989) los resultados obtenidos muestran que estos suelos que representan la variabilidad edafoclimática de la provincia de Córdoba comprenden 4 de estas 6 categorías, ya que no se ven representadas las categorías de propensión a lixiviación *Extremadamente baja* ni *Muy alta*. Adicionalmente se trazaron dos líneas de referencia sobre la gráfica señalando umbrales de peligrosidad para  $K_{oc}$  y  $t_{1/2}$  citados en la literatura para plaguicidas en general (Gustafson, 1989; Pfeiffer, 2010). Se observa que la mayor proporción de muestras de suelo se ubica en el cuadrante IV baja  $t_{1/2}$  y altos  $K_{OC}$ , es decir alta retención con baja vida media. Desde una perspectiva ambiental esto implica bajos riesgos de pérdida de atrazina en fase acuosa en sentido vertical u horizontal.

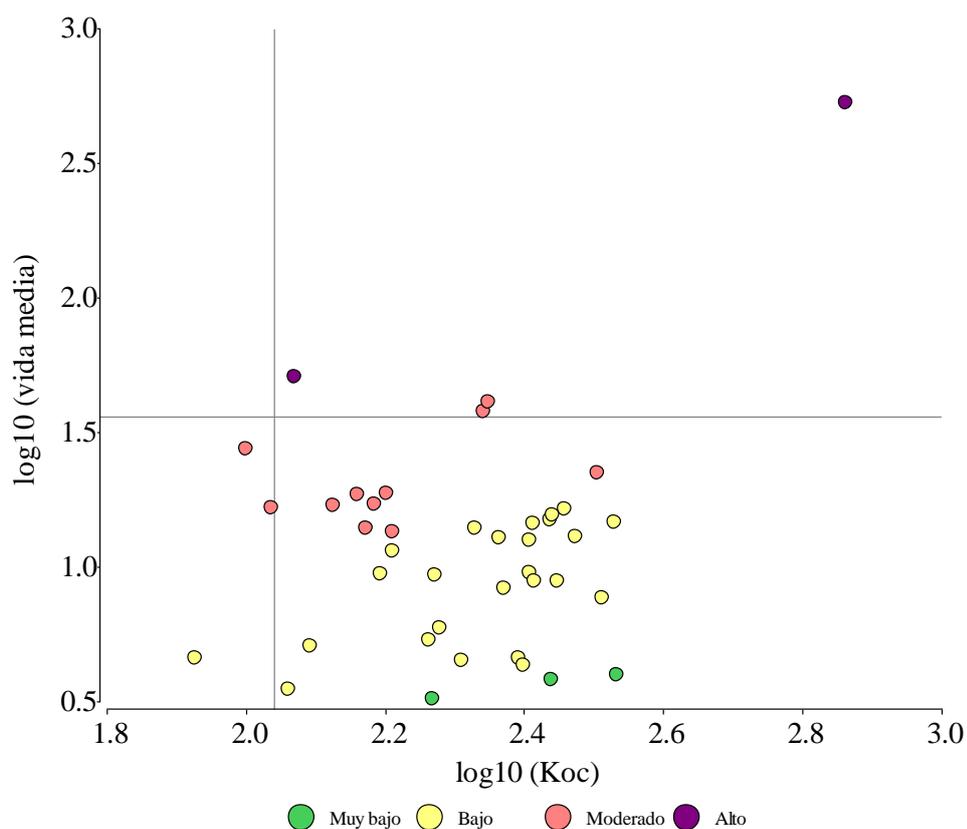


Figura 12. Correlación entre retención medida por el parámetro *Koc* y persistencia de atrazina en suelos de Córdoba. En colores se representan las distintas categorías de propensión a la lixiviación del herbicida según el índice GUS

Para complementar este análisis y visualizar este comportamiento en el territorio se mapeo y territorio según las categorías del índice GUS (Figura 13). Los índices GUS indican riesgo alto de pérdidas de lixiviación para el noroeste de la provincia y riesgo bajo para el centro-este.

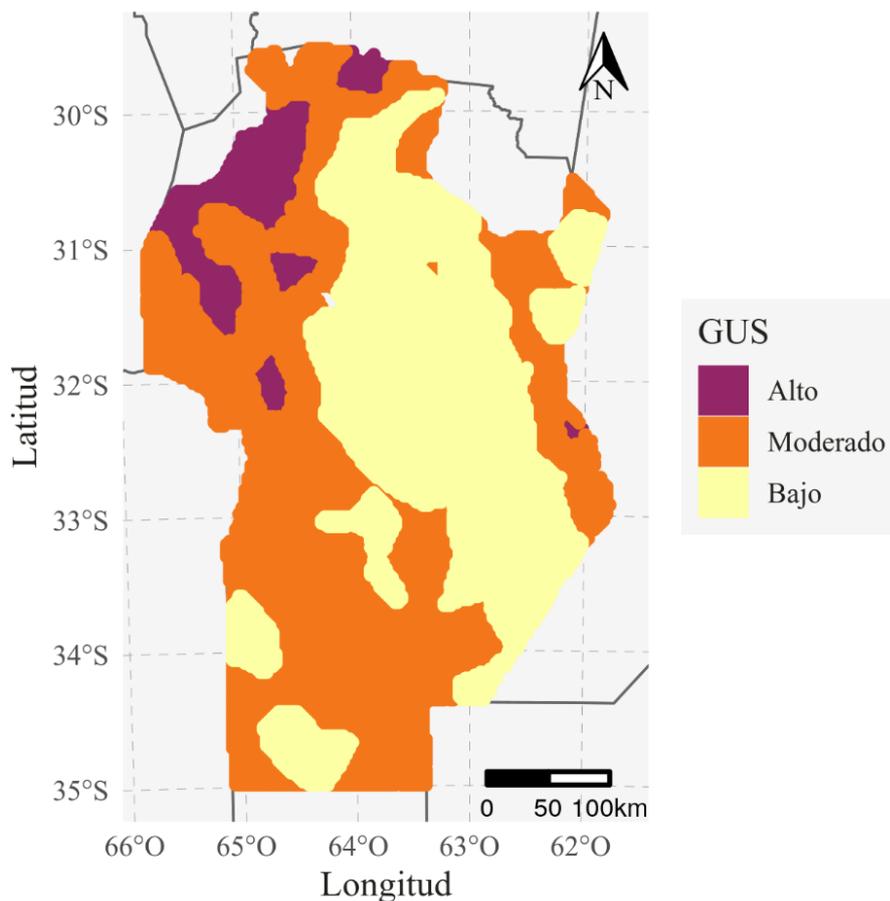


Figura 13. Propensión a la lixiviación de atrazina según GUS

El coeficiente de correlación de Pearson de estas dos variables, calculado sin tener en cuenta la espacialidad, no es estadísticamente significativo ( $p > 0,05$ ). Sin embargo, si se contempla la estructura espacial de la correlación se observa que las correlaciones cambian de magnitud y sentido dependiendo el lugar del territorio de estudio (Figura 14). Se observa que áreas con mayor  $t_{1/2}$  en el noroeste de la provincia donde el  $Kd$  es bajo y la  $t_{1/2}$  es alta, mientras que al sur de la provincia se aprecian  $Kd$  bajos y  $t_{1/2}$  bajas.

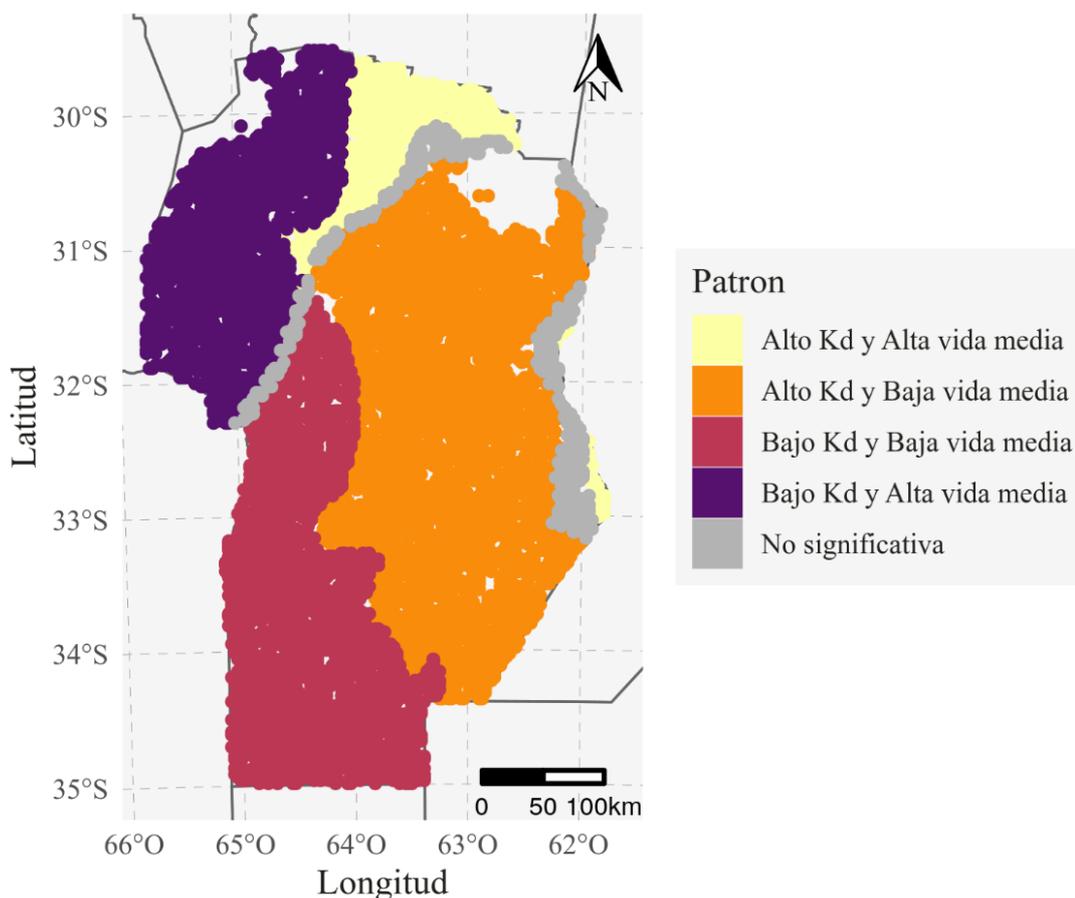


Figura 14. Mapa de correlaciones cruzadas espaciales entre retención medida por el parámetro *K<sub>oc</sub>* y persistencia de atrazina en Córdoba

El cambio de sentido de la correlación entre vida media y retención de atrazina dada la posición en el territorio puede evidenciarse en el Cuadro 12 donde se muestra la correlación condicional a la localización geográfica según las zonas edafoclimáticas homogéneas de la provincia (Giannini Kurina et al., 2018). En la zona I y II mientras mayor es la retención menor es la vida media y en las zonas II y IV lo contrario. Para complementar en el Cuadro 13 se presenta una tabla de contingencia entre las zonas edafoclimáticas homogéneas y la categorización sugerida por el análisis de correlación cruzada espacial. Este análisis complementa la interpretación ya que no solo indica el cambio de sentido de la correlación, sino que identifica las condiciones relativas en las

magnitudes de las dos variables en estudio, lo cual nos provee herramientas a la hora de la interpretación.

Cuadro 12. Retención y disipación de atrazina por zona edafoclimática

Zona	<i>Kd</i>			$t_{1/2}$			Corr( <i>Kd</i> , $t_{1/2}$ ) <sup>††</sup>
I	1,25	±0,21	a <sup>†</sup>	33,37	±6,27	a <sup>†</sup>	-0,59
II	1,89	±0,57	b	23,8	±8,84	b	0,15
III	3,27	±0,43	c	15,35	±10,6	c	-0,11
IV	2,00	±0,84	b	16,46	±3,86	d	0,48

<sup>†</sup>Letras diferentes indican diferencias estadísticas entre zonas ( $p < 0.05$ ). <sup>††</sup>Correlación condicional a la localización geográfica Corr(*Kd*,  $t_{1/2}$ )

Cuadro 13. Distribución de sitios según retención y persistencia de atrazina

Zona	Clasificación de sitios según <i>Kd</i> y $t_{1/2}$				
	Alto <i>Kd</i> y Alta $t_{1/2}$	Alto <i>Kd</i> y Baja $t_{1/2}$	Bajo <i>Kd</i> y Baja $t_{1/2}$	Bajo <i>Kd</i> y Alta $t_{1/2}$	No significativa
I	0,00	0,00	0,00	1,00	0,00
II	0,13	0,04	0,22	0,55	0,06
III	0,14	0,71	0,05	0,00	0,11
IV	0,00	0,39	0,61	0,00	0,00

Poder identificar situaciones contrastantes respecto al patrón que relaciona *Kd* y  $t_{1/2}$ , permite inferir distintos comportamientos de atrazina en suelo. Estos comportamientos diferenciales tienen una estructura espacial que se corresponde con la zonificación propuesta en el Capítulo I. Esto permite no solo aproximar los riesgos de contaminación debido a la covariación de estos índices sino interpretarlos respecto a su situación edafoclimática sitio-específica.

### **Peligrosidad ambiental por el uso de atrazina en Córdoba**

La erosión de suelo en sus diferentes formas y magnitudes constituye la principal amenaza frente a la contaminación por aplicación de herbicidas. La erosión extrae moléculas, agua, y partículas sólidas del suelo, deteriorando las funciones del mismo entre ellas la acción depuradora de moléculas orgánicas que en su lugar resultan trasladados a otros medios

donde la capacidad de disipación se reduce y aumenta su potencialidad contaminante (Lammoglia et al., 2018). En nuestro país se estima que alrededor de 50 millones de hectáreas están afectadas por erosión hídrica o eólica en grado moderado o grave (Gaitan et al., 2017). El problema de la erosión hídrica y eólica como fuente de contaminantes ha sido ampliamente descrito (Gevao, Semple y Jones, 2000), tanto para la dinámica de herbicidas en suelo como para otros problemas de contaminación asociados a la actividad agropecuaria como la eutrofización y la presencia de otros fitosanitarios en cursos de agua han sido reportada recurrentemente (Sasal, Wilson y NA, 2008; Sasal et al., 2015; De Gerónimo et al., 2014). Debido a esto, en este trabajo se entiende como peligrosa a aquellas situaciones que impliquen la salida de la molécula del ambiente suelo ya sea debido procesos de erosión hídrica o eólica.

Existen al menos tres escenarios o posibilidades de salida de la molécula del suelo (Montgomery, 2007). Por un lado, las pérdidas por desplazamiento horizontal estando la molécula adherida a la partícula sólida, escenario que puede ocurrir ya sea por acción del agua (arrastre por escorrentía), o por acción del aire (arrastre por acción del viento) (Morgan, 2009; Zachar, 2011). El desplazamiento horizontal también puede ocurrir escorrentía cuando la molécula se encuentra disuelta en agua. El tercer escenario se refiere a pérdidas por desplazamiento vertical que ocurren cuando la molécula disuelta en agua experimenta movimiento descendente por transporte vertical, es decir lixiviación. Para la provincia de Córdoba existen descripciones y dimensionamientos de la la de erosión hídrica potencial (Gaitan et al., 2017) y de erosión eólica potencial (Observatorio Nacional de Degradación de Tierras y Desertificación (ONDTyD), 2019). Esta información de los procesos erosivos se yuxtapone con la zonificación edafoclimática y la caracterización de la covariación de retención y disipación de atrazina en suelo para

describir diferentes condiciones de potencial peligrosidad sitio-específica por el uso de este herbicida:

*Condiciones de Alto Kd y Baja  $t_{1/2}$ , zona III*

La región centro este de la provincia de Córdoba (zona III) presento el 71% de los sitios en esta condición (Cuadro 13). La zona corresponde a la llanura y piedemonte oriental de la provincia con intenso uso agrícola y predominio de los sistemas de siembra directa (Jarsún et al., 2006). Si bien se identifica a los sistemas de no labranza o labranza reducida como sistemas conservacionistas (Díaz-Zorita, Duarte y Grove, 2002; Fabrizzi et al., 2005), existen problemas de erosión hídrica de distinta magnitud y causas (Gaitan et al., 2017). Por un lado, el área de piedemonte con prolongadas pendientes ubicadas al centro norte, donde los procesos de arrastre a través de largas pendientes pueden transportar partículas por largas distancias. Cuenta con sistemas hidrológicos con abundante presencia de cursos de agua (Carignano, 1999) lo que implica que viajando cortas distancia la partícula tiene altas chances de llegar a cursos de agua. Al centro sur sobre el sistema hidrográfico del Carcarañá parte del territorio posee problemas de erosión hídrica asociado a limitantes de suelo por drenaje (Jarsún et al., 2006). Los problemas de erosión de esta zona se manifiestan en la época de primavera, previo al inicio del cultivo de verano, con cobertura limitada ya sea por suelo descubierto o en emergencia temprana de plántulas. Si bien la elevada retención de atrazina reduce el riesgo de su transporte por escorrentía, no minimiza el riesgo por arrastre de partículas. La vida media baja reduce la presencia de atrazina en la solución del suelo, el periodo de mayor peligrosidad a su desplazamiento del suelo se concentra en el mes previo y durante la siembra, que concuerda con el momento de su aplicación. Siendo la principal causa de salida de atrazina del sistema suelo el arrastre por adherencia a la partícula de suelo, las prácticas de conservación orientadas al control de la erosión, como sistematizaciones y siembras

contra pendiente, podrían reducir las posibilidades contaminantes por la aplicación de atrazina. Los antecedentes muestran que aún la provincia de Córdoba tiene reducido control de la erosión hídrica (100.000 ha en toda la provincia, Panigatti, 2015). La zona III es la que presenta menores posibilidades de contaminación potencial por sumar ambos mecanismos retención y disipación a que permanezca la molécula en el suelo.

#### *Bajo Kd y Alta t<sub>1/2</sub>, zonas I y II*

La región del oeste correspondiente a la zonas I y II se caracteriza por pampas de altura donde la actividad agrícola se realiza, valles poblados y numerosos cursos de agua (Cabido y Zak, 1999), especialmente en el territorio que comprende la zona I ya que en la zona II existe menor desarrollo de la actividad agrícola (Ghida Daza et al., 2009). Las características edáficas y climáticas de la zona I describen contenidos de arcilla bajos y pH ligeramente altos lo que contribuye a menores velocidades de disipación y retenciones. Es decir, las características de los suelos indican que tiene baja capacidad amortiguadora para evitar contaminación por atrazina ya que la molécula perecerá durante periodos largos en la solución con pocas posibilidades de retenerse. Estas particularidades sugieren que atrazina tendería a trasladarse vertical u horizontalmente en fase líquida, pudiendo llegar a aguas subterráneas en el sistema freático o a cursos superficiales. Incluso el mapeo de la propensión a pérdidas por lixiviación de GUS reconoce en esta zona áreas de alto riesgo (Figura 13). Por esto desde un punto de vista ambiental el arco noroeste representa la porción del territorio que mayor propensión a la contaminación por atrazina presenta ya que la molécula se encontrará durante periodos prolongados en la fase líquida del suelo siendo susceptible a perderse trasladada ya sea por procesos de lixiviación o de escurrimiento. Esto se intensifica debido a la topografía de altas pendientes (Carignano et al., 2014) y la concentración de las precipitaciones en la época estival (por tratarse de una zona ubicada en la región semiárida (Jarsún et al., 2006)).

### *Bajo $K_d$ y Baja $t_{1/2}$ , zona IV*

En el área centro sur correspondiente a la zona IV la combinación de  $K_d$  y vida media implicaría una zona de bajo potencial de contaminación por transporte de atrazina ya que queda poco retenida por lo tanto no puede perderse asociada al transporte de la partícula. No obstante, se trata de una zona donde la textura es predominantemente arenosa, situación que se intensifica al oeste de esta y donde el proceso de retención de atrazina es muy bajo (Figura 8). Esta zona se corresponde con un área de alto impacto de la erosión eólica (Panigatti, 2015) por su condición de suelos sueltos, lo cual por presentar bajas retenciones no representaría un problema directo respecto a la posibilidad de transporte por arrastre. La vida media indica que no es un área donde se prolonga la presencia de atrazina en solución (Cuadro 12), por lo que, tratándose de una región semiárida, la condición crítica para contaminación se asocia a la época de lluvias y con transporte vertical, más aun teniendo en cuenta que una de las limitantes principales de los suelos del área es el excesivo drenaje (Jarsún et al., 2006). Esta posibilidad de movimiento descendente indica que debiera plantearse monitoreos de aguas subsuperficiales por presencia de atrazina. Respecto a esto, existen antecedentes en la provincia de La Pampa que indican que atrazina en escenario de alto contenido de arena y precipitaciones de elevada intensidad (alta pluviometría en tiempo acotado) puede translocarse en el perfil (Montoya et al., 2006).

## Conclusiones

La variabilidad en la velocidad de disipación de un herbicida puede modelarse a partir de curvas de disipación sitio-específicas derivada del ajuste de modelos no lineales mixtos con efecto aleatorio de sitio asociado a la pendiente de la ecuación de disipación. La adición de un efecto aleatorio sobre la pendiente no solo puede producir mejores ajustes,

sino que también provee una herramienta para explorar la variabilidad de la disipación entre suelos de una región.

El uso de suelo resultó crucial en la explicación los tiempos de vida media de atrazina en la provincia de Córdoba, confirmando a escala regional que manejos que implican historia de uso de atrazina inducen una alta adaptación microbiana que acelera las tasas de degradación.

La regresión bayesiana para datos de  $t_{1/2}$  georreferenciados en un dominio de interés permitió explicar y mapear la persistencia de atrazina en suelos en función de características del sitio (suelo, clima y uso).

El análisis de covariación espacial de la retención y disipación de un herbicida en suelo permite visualizar la interdependencia de indicadores de la dinámica en un espacio geográfico que facilita interpretar la condición resultante en función de características de los sitios de las zonas.

Se identificaron cuatro zonas dentro del territorio de la provincia de Córdoba en función de las combinaciones de retención y persistencia que proveyeron información útil para dimensionar la peligrosidad por el uso de atrazina.

## Síntesis

El problema que contextualiza este trabajo es la contaminación difusa por el uso de herbicidas en la agricultura. Dado que los herbicidas son el grupo de fitosanitarios más utilizados, se trabajó con dos herbicidas de propiedades contrastantes y amplio uso (glifosato y atrazina). La dinámica de herbicidas en suelo depende de los procesos de retención y disipación que son caracterizados por el coeficiente de adsorción ( $Kd$ ) y la vida media ( $t_{1/2}$ ) como métrica de la persistencia de la molécula en suelo. La variabilidad de estos índices ha sido rigurosa y extensamente descrita en la literatura científica asociada a distintos tipos, usos y manejos de suelos. Si bien, estos estudios han permitido comprender y profundizar la naturaleza fisicoquímica de los procesos de retención y disipación, no son suficientes para describir la variabilidad espacial de procesos de contaminación difusa a escala regional. Mientras la disponibilidad de información ambiental con potencialidad explicativa de los procesos de retención y disipación es amplia y abundante espacialmente, las mediciones  $Kd$  y  $t_{1/2}$  son usualmente escasas para sus estudios a mayor escala. Las determinaciones de laboratorio necesarias para procesar las muestras de suelos en búsqueda de estos indicadores son costosas, laboriosas y demandan tiempo. Por esta razón, su uso es aún limitado ya que no se ha logrado disponer de estos índices a escala regional. En este trabajo, se integraron mediciones de variables ambientales con mediciones de  $Kd$  y  $t_{1/2}$  para generar modelos estadísticos que proveen predicciones espaciales de la dinámica de los herbicidas en un dominio espacial continuo.

El principal aporte de la tesis es el desarrollo de una batería de herramientas metodológico-estadísticas para la predicción espacial, a escala regional, de los índices de procesos (retención y disipación) a partir de un muestreo sistemático de suelos. En la metodología se integra información sitio-específica de distinta naturaleza para la construcción de modelos que permiten la predicción en un dominio espacial continuo de las variables que caracterizan la dinámica de herbicidas en suelos. Si bien, las dinámicas de glifosato y atrazina en suelo han sido estudiadas, la metodología generalmente usada en esos estudios es diferente ya que los procesos se caracterizan en condiciones controladas y para muestras de suelo puntuales. El diseño metodológico propuesto en esta tesis es novedoso y se sustenta en la construcción de modelos estadísticos para predecir los principales procesos de la dinámica en distintos sitios de un territorio a partir de variables de sitio. El enfoque metodológico de la predicción basada en modelos, aplicado al muestreo de suelos disponible para la provincia de Córdoba, resultó apto para la obtención de mapas de variabilidad espacial de los procesos de retención y disipación de una molécula de herbicida en suelo. El dimensionamiento de la retención y la persistencia de los herbicidas estudiados a una escala regional permitió abordar el problema de la contaminación difusa en relación con condiciones de sitio del territorio en estudio. El mapeo digital de retención y disipación a partir de características del ambiente permite estudiar la dinámica del herbicida en suelo con información del contexto resultando superador a los estudios realizados sobre muestras de suelos aisladas para realizar aproximaciones de la peligrosidad por el uso del potencial contaminante en distintos ambientes.

## Protocolos estadísticos para el mapeo regional de retención y disipación de herbicidas

Se diseñó un protocolo para zonificación edafoclimática regional basado en interpolación geoestadística de variables de sitio y métodos multivariados con restricción espacial (Análisis de Componentes Principales espacial) para la obtención de variables sintéticas. Éstas fueron usadas para la clasificación no supervisada de sitios con el algoritmo de clasificación difusa *Fuzzy Kmeans*. De esta manera se identificaron cuatro zonas edafoclimáticas homogéneas coherentes y contiguas en el área de estudio (provincia de Córdoba). Para un subconjunto de muestras de suelos, de las distintas zonas, se determinó el coeficiente de adsorción para atrazina y para glifosato. Así mismo se obtuvieron los tiempos de vida media de atrazina modelando curvas de disipación desde datos generados en incubaciones. Las curvas sitio-específicas se estimaron a partir del ajuste de un modelo no lineales mixto exponencial con efecto aleatorio de sitio sobre la pendiente, y desde cada una se derivó el tiempo de vida media de cada suelo. Tradicionalmente, las curvas de disipación son estimadas con un modelo exponencial de efectos fijos. La incorporación de un efecto aleatorio de sitio permitió obtener información sobre la variabilidad de la vida media y mapear la persistencia del herbicida en el territorio.

Para caracterizar la variabilidad espacial de los indicadores retención y disipación, se diseñó una secuencia metodológica orientada a la obtención de modelos predictivos espaciales para el mapeo digital. La metodología desarrollada involucra tres etapas: 1) selección de variables edafoclimáticas con potencialidad explicativa mediante aprendizaje automático y optimización de errores de predicción, 2) ajuste de un modelo de regresión donde se expresa el índice del proceso como dependiente de las variables seleccionadas. El modelo se construyó en el marco teórico bayesiano mediante

aproximación integrada por Laplace (R-INLA) e incluyendo un efecto aleatorio de sitio valuado mediante ecuaciones diferenciales parciales estocásticas. La elección del enfoque se basó en una investigación metodológica implementada para comparar modelos de predicción espacial construidos desde diferentes marcos teóricos (bayesiano, frecuentista y aprendizaje automático), y 3) mapeo digital de los índices de interés a mayor resolución y escala regional. La implementación de la metodología, basada en los modelos de regresión bayesianos, permitió evaluar el comportamiento ambiental de los procesos de retención y vida media a través de mapas que cuentan con medidas de incertidumbre de las predicciones. Finalmente, se abordó la covariación espacial de los procesos a través de correlaciones condicionales a la localización geográfica y correlaciones cruzadas espaciales. Tales correlaciones espaciales permitieron poner en evidencia la existencia de zonas sobre el territorio en estudio con distintos patrones de retención y disipación. El protocolo que integra estos desarrollos se presenta en la Figura 15. Si bien fue utilizado para el mapeo digital de retención y persistencia de atrazina y glifosato podría ser usada en otros contextos donde se demande el mapeo regional de procesos asociados a la dinámica de moléculas con potencialidad contaminante.

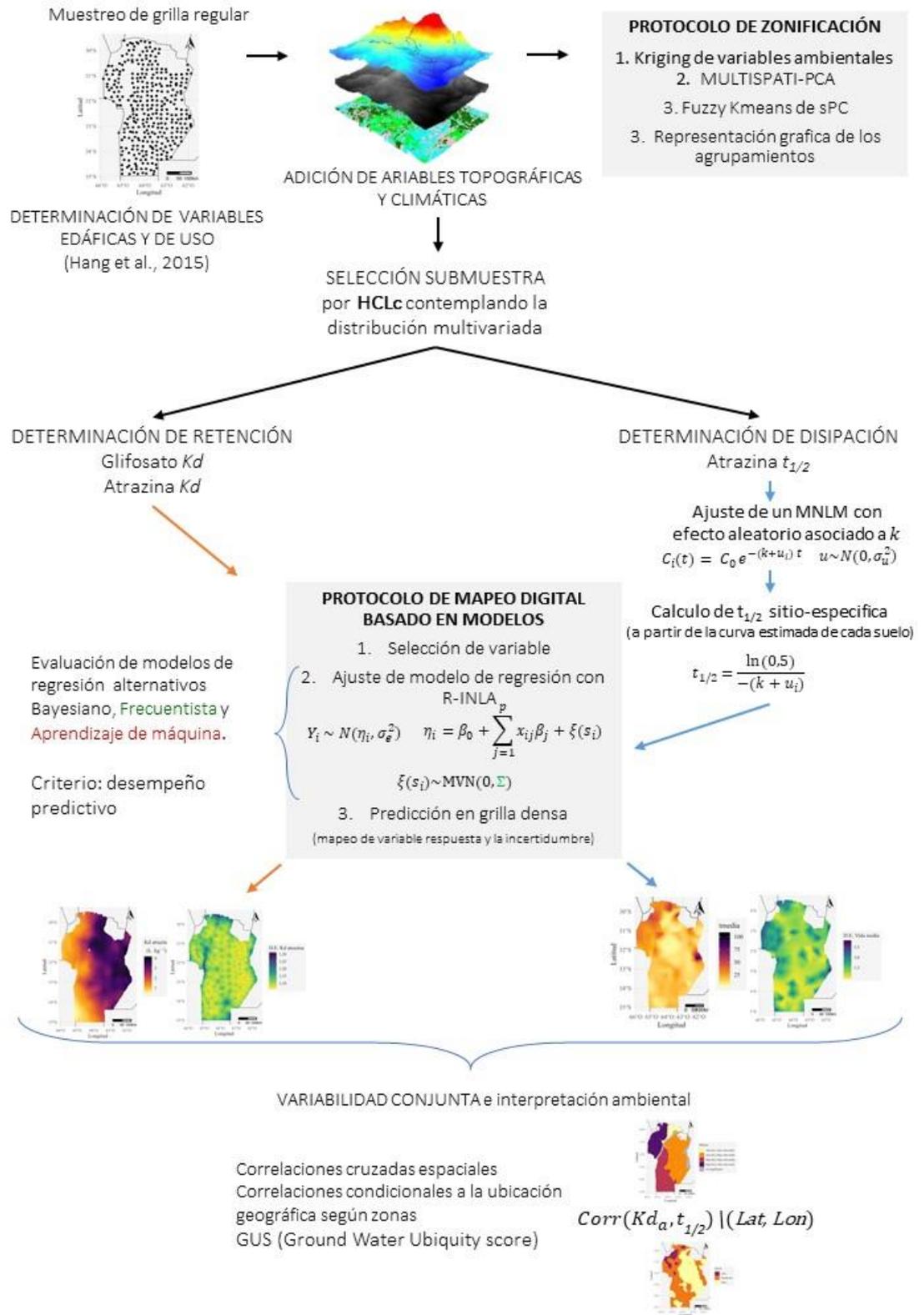


Figura 15: Diseño metodológico para explicar variabilidad espacial marginal y conjunta de los procesos de retención y disipación a escala regional

## Implicancias ambientales de la variabilidad en retención y disipación de herbicidas

Se confirmó que a escala regional la estructura de propiedades que afectan los procesos de retención y disipación de atrazina y glifosato se corresponde con la documentada por estudios previos realizados a otras escalas.

El rol descontaminante del suelo depende de la permanencia del contaminante en el sistema. Por ello, se identificaron situaciones con potencialidad contaminante en función a los diferentes desplazamientos que podría experimentar la molécula del herbicida en el suelo. Se relacionaron los mapas de erosión hídrica y eólica potencial, topografía y limitantes por drenaje de la provincia de Córdoba con los mapas generados a partir del estudio de covariación de la retención y disipación. El principal resultado de este análisis permitió comprender la variabilidad de la peligrosidad ambiental debida al uso de herbicidas a través del territorio.

Para el caso de atrazina se identificó que en el centro-este existe mayor propensión a la contaminación por transporte de la molécula adherida a partículas de suelo. En la región suroeste hay mayor riesgo debido a pérdidas por lixiviación, mientras que, en el noroeste de la provincia de Córdoba, la molécula puede salir del sistema disuelta o adsorbida a partículas de suelo. Por otra parte, los mapas generados para atrazina fueron consistentes con el conocimiento existente sobre el impacto de la historia de uso del herbicida sobre la persistencia. Es sabido que los suelos que reciben recurrentemente atrazina desarrollan una microflora adaptada causante de la mineralización acelerada, es decir el herbicida pierde efectividad por dilución y no por resistencia de las malezas como erróneamente se podría interpretar. Para el caso de glifosato los mapas resultantes mostraron que las características edáficas tienen fuerte dominio sobre el comportamiento y potencial

contaminante, más aún debido a que la degradación responde a mecanismos cometabólicos, es decir a una microflora inespecífica. Las zonas identificadas por su mayor retención, al este de la provincia, se correspondieron con las de mayor vulnerabilidad frente a la erosión hídrica donde glifosato podría ser transportado junto con suelo erosionado. Mientras que, al oeste, con baja retención, se incrementa el riesgo de salida en solución de la molécula de glifosato ya sea en dirección horizontal o vertical. De los resultados se desprende que las situaciones que presentan potencialidad contaminante se deben a diversas causas, pero todas confluyen en escenarios que implican la salida de la molécula del ambiente suelo, ya sea debido a procesos de erosión hídrica (adsorbido o en solución) o eólica. Es decir, que prácticas de manejo de suelo conocidas y ampliamente difundidas como la sistematización de terrenos, sistemas de siembra directa acoplados a rotaciones coherentes y las siembras contra pendiente constituyen una gran herramienta para reducir la potencialidad contaminante de un sitio. Otra estrategia, sería evitar la coincidencia temporal entre las aplicaciones y la ocurrencia de lluvias, especialmente en aquellas zonas donde la retención es baja justamente coincidente con suelos sueltos.

## Relevancia de las contribuciones

El principal valor de este trabajo radica en el desarrollo de herramientas para abordar multidimensionalmente el problema de la contaminación difusa a escala regional. La necesidad de comprender este proceso a dicha escala resulta crucial por la creciente demanda social hacia el sector agropecuario debido a las consecuencias ambientales asociadas a la aplicación de agroquímicos. Dadas las grandes superficies destinadas a la agricultura extensiva que demandan uso de herbicidas, el mapeo de la variabilidad espacial de los procesos que explican su destino solo puede interpretarse a escala regional y con información auxiliar del contexto ambiental de su aplicación. Por ello, la base de

datos edafoclimáticos generada para la provincia de Córdoba es relevante no solo para esta tesis. La existencia de esta nueva base de datos ambientales y de procesos asociados a la contaminación enriquece la infraestructura de datos espaciales de la Provincia y estimula futuros estudios ambientales. La identificación de escenarios contrastantes respecto a la peligrosidad de contaminación provee insumos para el diseño de acciones políticas territoriales tanto para monitorear como para mitigar el impacto de la aplicación de herbicidas al suelo.

Los aportes metodológicos realizados en la tesis establecen un protocolo para el manejo y tratamiento de datos espaciales cuando se pretende estudiar procesos de contaminación asociados a la matriz suelo a escalas regionales. Distintas herramientas estadísticas, provenientes del análisis multivariado, el aprendizaje de máquina, modelos mixtos bayesianos y frecuentistas fueron ordenadas en una secuencia lógica que permite resolver el mapeo a gran escala de la retención y la persistencia desde un muestreo relativamente pequeño de suelos para la determinación de los índices que caracterizan los procesos. Las distintas alternativas hilvanadas para contemplar la multidimensionalidad de fenómenos estructurados espacialmente representan el estado del arte en estadística espacial y potencian el protocolo diseñado para abordar mapeos regionales en estudios ambientales territoriales.

## Futuras líneas de investigación

La propuesta de líneas de trabajo futuras se presenta desde dos enfoques que se retroalimentan entre sí. Por un lado, aquellas necesarias para responder a interrogantes ambientales y por otro, las relacionadas al desarrollo y adaptación de herramientas metodológicas estadísticas que dan soporte a las respuestas a estos interrogantes.

Aun cuando el mapeo digital de los procesos que protagonizan la dinámica de un herbicida en suelo abre nuevas oportunidades para identificar zonas de distinta

peligrosidad por contaminación difusa, no cuantifica riesgos. Es necesario, a futuro, incorporar los insumos generados a modelos de riesgo y evaluar peligros ambientales asociados a la actividad agropecuaria bajo diversos escenarios de uso del suelo y condiciones ambientales. Los modelos hidrológicos que trabajan a escala de cuenca permiten comprender el proceso de transporte de contaminantes, pero no contemplan la variabilidad espacial de los procesos de retención y disipación. Un futuro desafío es alimentarlos con los insumos generados para llegar a describir la dinámica completa de los herbicidas aplicados al suelo. Otra línea podría considerar la incorporación, a los modelos desarrollados en este trabajo, de otras fuentes de información como las derivadas del sensoramiento remoto satelital cada vez más precisas y disponibles. Además, es importante destacar que la generación de datos del tipo de  $Kd$  y  $t_{1/2}$  requiere de compromisos político-institucionales para enriquecerse, sin duda, el aumento y actualización de la información que da origen a esta investigación permitirá robustecer los resultados aquí presentados. Otro aspecto para profundizar sería el diseño de una metodología que permita identificar sitios y momentos convenientes y relevantes (*hot-spot/ hot moments*) para monitorear moléculas de herbicidas a lo largo del año agrícola basados en los ciclos de los cultivos involucrados.

Dentro de los desafíos metodológicos vigentes es necesario abordar la variabilidad conjunta entre los procesos de retención y disipación desde otros enfoques como modelos de regresión bivariados. Éstos permiten predecir la distribución completa que explica la variabilidad y covariabilidad de los procesos para generar modelos de riesgo y realizar análisis de sensibilidad bajo distintos escenarios. Es posible aún generar productos derivados como, por ejemplo, mapas que describan la probabilidad de ocurrencia de una situación peligrosa, que pueden construirse volviendo binaria a la variable respuesta de acuerdo con el riesgo aparejado. En este sentido, los modelos bayesianos ofrecen una

serie de ventajas a la hora de describir la variabilidad espacial de datos binarios es decir en el contexto de modelos generalizados mixtos por la manera en la que construyen la matriz de varianzas y covarianzas. Esta misma particularidad que ofrecen los modelos bayesianos en el tratamiento de efectos aleatorios aplica para la modelación de procesos espaciotemporales, otro de los aspectos interesantes para abordar desde lo metodológico. Es decir, avanzar en el reconocimiento no solo los patrones espaciales sino temporales para poder modelar cómo cambian los patrones espaciales en distintos periodos de tiempo. Asimismo, es aún necesario profundizar el conocimiento respecto al dimensionamiento de la incertidumbre de las predicciones espaciales generadas por regresiones bayesianas y por métodos de aprendizaje automático. Todo esfuerzo orientado a cuantificar riesgos de contaminación y vulnerabilidades ambientales sitio-específicas a escalas regionales será de alto valor social.

## Bibliografía

- Albers, C.N., Banta, G.T., Hansen, P.E. y Jacobsen, O.S. (2009) The influence of organic matter on sorption and fate of glyphosate in soil—Comparing different soils and humic substances. *Environmental pollution*, **157**, 2865–2870.
- Alvarez, R. y Lavado, R.S. (1998) Climate, organic matter and clay content relationships in the Pampa and Chaco soils, Argentina. *Geoderma*, **83**, 127–141.
- Anderberg, M.R. (2014) *Cluster analysis for applications: probability and mathematical statistics: a series of monographs and textbooks*. Academic press.
- Anselin, L. (1995) Local indicators of spatial association—LISA. *Geographical analysis*, **27**, 93–115.
- Aparicio, V., De Gerónimo, E., Guijarro, K.H., Perez, D., Portocarrero, R. y Vidal, C. (2015) Los plaguicidas agregados al suelo y su destino en el ambiente. *INTA Ediciones. Imprenta El Vikingo. Balcarce, Buenos Aires, Argentina*.
- Arias-Estévez, M., López-Periago, E., Martínez-Carballo, E., Simal-Gándara, J., Mejuto, J.C. y García-Río, L. (2008) The mobility and degradation of pesticides in soils and the pollution of groundwater resources. *Agriculture, Ecosystems and Environment*, **123**, 247–260.
- Arrouays, D., Saby, N.P.A., Thioulouse, J., Jolivet, C., Boulonne, L. y Ratié, C. (2011) Large trends in French topsoil characteristics are revealed by spatially constrained multivariate analysis. *Geoderma*, **161**, 107–114.
- Aslam, S., Benoit, P., Chabauty, F., Bergheaud, V., Geng, C., Vieublé-Gonod, L., et al. (2014) Modelling the impacts of maize decomposition on glyphosate dynamics in mulch. *European Journal of Soil Science*, **65**, 231–247.
- Bailey, G.W. y White, J.L. (1970) Factors influencing the adsorption, desorption, and movement of pesticides in soil. *Single Pesticide Volume: The Triazine Herbicides* pp. 29–92. Springer.
- Bakka, H., Rue, H., Fuglstad, G.A., Riebler, A., Bolin, D., Illian, J., et al. (2018) Spatial modeling with R-INLA: A review. *Wiley Interdisciplinary Reviews: Computational Statistics*, 1–24.
- Banerjee, S., Carlin, B.P. y Gelfand, A.E. (2014) *Hierarchical modeling and analysis for spatial data*. Chapman and Hall/CRC.
- Barriuso, E. y Houot, S. (1996) Rapid meneralization of the s-triazine ring of atrazine in soils in relation to soil management. *Soil Biology and Biochemistry*, **28**, 1341–1348.
- Barriuso, E., Laird, D.A., Koskinen, W.C. y Dowdy, R.H. (1994) Atrazine desorption

- from smectites. *Soil Science Society of America Journal*, **58**, 1632–1638.
- Becerra, M.A., Hang, S.B. y Díaz Zorita, M. (2015) Análisis geoespacial del índice de adsorción [Kd] de atrazina calculado según cartas de suelos y muestreo en grilla. Geospatial analysis of atrazine adsorption index [Kd] as calculated by soil maps and grid sampling. *Ciencia del suelo. revista de la Asociación Argentina de la Ciencia del Suelo.*, **33**.
- Behrends Kraemer, F., Chagas, C.I., Vázquez Amábile, G., Palacín, E.A., Santanatoglia, O.J., Castiglioni, M.G., et al. (2011) Aplicación del Modelo Hidrológico-Swat-en una microcuenca agrícola de la Pampa Ondulada. *Ciencia del suelo*, **29**, 75–82.
- Bento, C.P.M., van der Hoeven, S., Yang, X., Riksen, M.M.J.P.M., Mol, H.G.J., Ritsema, C.J., et al. (2019) Dynamics of glyphosate and AMPA in the soil surface layer of glyphosate-resistant crop cultivations in the loess Pampas of Argentina. *Environmental Pollution*, **244**, 323–331.
- Besag, J. (1974) Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 192–236.
- Besag, J., Green, P., Higdon, D. y Mengersen, K. (1995) Bayesian computation and stochastic systems. *Statistical science*, 3–41.
- Best, N., Richardson, S. y Thomson, A. (2005) A comparison of Bayesian spatial models for disease mapping. *Statistical methods in medical research*, **14**, 35–59.
- Bezdek, J.C., Coray, C., Gunderson, R. y Watson, J. (1981) Detection and characterization of cluster substructure i. linear structure: Fuzzy c-lines. *SIAM Journal on Applied Mathematics*, **40**, 339–357.
- Bivand, R.S. (2014) GeoComputation and Open-Source Software. *GeoComputation* p. 329. CRC Press.
- Bivand, R.S., Gomes-Rubio, V. y Rue, H. (2015) Spatial Data Analysis with R - INLA with Some Extensions. *Journal of Statistical Software*, **63**, 1–31.
- Bivand, R., Keitt, T., Rowlingson, B. y Pebesma, E. (2014) rgdal: Bindings for the geospatial data abstraction library. *R package version 0.8-16*.
- Blangiardo, M. y Cameletti, M. (2015) *Spatial and spatio-temporal Bayesian models with R-INLA*. John Wiley & Sons.
- Bollag, J.-M. y Myers, C. (1992) Detoxification of aquatic and terrestrial sites through binding of pollutants to humic substances. *Science of the Total Environment*, **117**, 357–366.
- Booth, T.H., Nix, H.A., Busby, J.R. y Hutchinson, M.F. (2014) BIOCLIM: the first species distribution modelling package, its early applications and relevance to most current MAXENT studies. *Diversity and Distributions*, **20**, 1–9.
- Breiman, L. (1996) Bagging predictors. *Machine learning*, **24**, 123–140.
- Breiman, L. (2001) Random forests. *Machine learning*, **45**, 5–32.
- Brenning, A. (2012) Spatial cross-validation and bootstrap for the assessment of prediction rules in remote sensing: The R package sperrorest. *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International* pp. 5372–5375.

IEEE.

- Brus, D.J. (2019) Sampling for digital soil mapping: A tutorial supported by R scripts. *Geoderma*, **338**.
- Burrough, P.A. y McDonnell, R.A. (1998) *Principles of Geographical Information Systems*. Oxford University Press.
- Burrough, P.A., McDonnell, R., McDonnell, R.A. y Lloyd, C.D. (2015) *Principles of geographical information systems*. Oxford university press.
- Cabido, D., Cabido, M., Garré, S. y Gorgas, J. (2003) Regiones naturales de la provincia de Córdoba.
- Cabido, M.R. y Zak, M.R. (1999) *Vegetación del norte de Córdoba*.
- Calvet, R. (2005) *Les pesticides dans le sol: conséquences agronomiques et environnementales*. France Agricole Editions.
- Cameletti, M., Lindgren, F., Simpson, D. y Rue, H. (2013) Spatio-temporal modeling of particulate matter concentration through the SPDE approach. *ASta Advances in Statistical Analysis*, **97**, 109–131.
- Carignano, C.A. (1999) Late Pleistocene to recent climate change in Córdoba Province, Argentina: Geomorphological evidence. *Quaternary International*, **57**, 117–134.
- Carignano, C., Kröhling, D., Degiovanni, S. y Cioccale, M. (2014) Geomorfología de la provincia de Córdoba (Argentina). *Relatorio del XIX Congreso Geológico Argentino Geología y Recursos Naturales de la Provincia de Córdoba*. Ed.: Ira., Chapter: *Geomorfología*, Publisher: *Asociación Geológica Argentina (AGA)*, Eds: *Martino R., A. Guereschi* pp. 747–822.
- Carsel, R.F. (1984) *Users manual for the pesticide root zone model (PRZM), release 1*. Environmental Research Laboratory, Office of Research and Development, US ....
- Celis, R., Cornejo, J., Hermosin, M.C. y Koskinen, W.C. (1998) Sorption of atrazine and simazine by model associations of soil colloids. *Soil Science Society of America Journal*, **62**, 165–171.
- Chaplain, V., Mamy, L., Vieublé-Gonod, L., Mougin, C., Benoit, P., Barriuso, E., et al. (2011) Fate of pesticides in soils: Toward an integrated approach of influential factors. *Pesticides in the modern world-Risks and benefits* p. IntechOpen.
- Charrad, M., Ghazzali, N., Boiteau, V., Niknafs, A. y Charrad, M.M. (2014) Package ‘nbclust’ *J. Stat. Softw*, **61**, 1–36.
- Chen, Y. (2013) New approaches for calculating Moran’s index of spatial autocorrelation. *PloS one*, **8**, e68336.
- Cheng, H.-H. (1990) Pesticides in the soil environment: processes, impacts, and modeling. p. Soil Science Society of America.
- Chessel, D., Dufour, A.B. y Thioulouse, J. (2004) The ade4 package-I-One-table methods. *R news*, **4**, 5–10.
- Córdoba, M.A., Bruno, C.I., Costa, J.L., Peralta, N.R. y Balzarini, M.G. (2016) Protocol for multivariate homogeneous zone delineation in precision agriculture. *biosystems*

- engineering*, **143**, 95–107.
- Correa Morales, J.C., Causil, B. y Javier, C. (2018) *Introducción a la estadística bayesiana: notas de clase*. Instituto Tecnológico Metropolitano.
- Cosby, B.J., Hornberger, G.M., Clapp, R.B. y Ginn, T. (1984) A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils. *Water resources research*, **20**, 682–690.
- Cressie, N. (1990) The origins of kriging. *Mathematical geology*, **22**, 239–252.
- Cressie, N. (1993) *Spatial statistics*. New York.
- Cressie, N. y Chan, N.H. (1989) Spatial modeling of regional variables. *Journal of the American Statistical Association*, **84**, 393–401.
- Cressie, N. y Wikle, C.K. (2015) *Statistics for spatio-temporal data*. John Wiley & Sons.
- Davidian, M. (2017) *Nonlinear models for repeated measurement data*. Routledge.
- Diggle, P.J. (2013) *Statistical analysis of spatial and spatio-temporal point patterns*. Chapman and Hall/CRC.
- Díaz-Zorita, M., Duarte, G.A. y Grove, J.H. (2002) A review of no-till systems and soil management for sustainable crop production in the subhumid and semiarid Pampas of Argentina. *Soil and Tillage Research*, **65**, 1–18.
- Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., et al. (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, **36**, 27–46.
- Dray, S., Saïd, S. y Débias, F. (2008) Spatial ordination of vegetation data using a generalization of Wartenberg's multivariate spatial correlation. *Journal of vegetation science*, **19**, 45–56.
- Durbin, J. (1970) Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables. *Econometrica (pre-1986)*, **38**, 410.
- Eberbach, P. (1998) Applying non-steady-state compartmental analysis to investigate the simultaneous degradation of soluble and sorbed glyphosate (N-(phosphonomethyl) glycine) in four soils. *Pesticide Science*, **52**, 229–240.
- Efron, B. y Hastie, T. (2016) *Computer age statistical inference: Algorithms, evidence, and data science*.
- Efron, B. y Tibshirani, R. (1997) Improvements on cross-validation: the 632+ bootstrap method. *Journal of the American Statistical Association*, **92**, 548–560.
- Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., et al. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.
- Elith, J., Leathwick, J.R. y Hastie, T. (2008) A working guide to boosted regression trees. *Journal of Animal Ecology*, **77**, 802–813.
- Fabrizzi, K.P., García, F.O., Costa, J.L. y Picone, L.I. (2005) Soil water dynamics, physical properties and corn and wheat responses to minimum and no-tillage systems in the southern Pampas of Argentina. *Soil and Tillage Research*, **81**, 57–69.

- Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., et al. (2007) The shuttle radar topography mission. *Reviews of geophysics*, **45**.
- Fenner, K., Canonica, S., Wackett, L.P. y Elsner, M. (2013) Evaluating pesticide degradation in the environment: blind spots and emerging opportunities. *science*, **341**, 752–758.
- Fernandes, M.C., Cox, L., Hermosín, M.C. y Cornejo, J. (2006) Organic amendments affecting sorption, leaching and dissipation of fungicides in soils. *Pest Management Science: formerly Pesticide Science*, **62**, 1207–1215.
- G. Vazquez-Amabile, B. A. Engel y D. C. Flanagan. (2006) Modeling and risk analysis of nonpoint-source pollution caused by atrazine using SWAT. *Transactions of the ASABE*, **49**, 667–678.
- Gaitan, J.J., Navarro, M.F., Tenti Vuegen, L.M., Pizarro, M.J. y Carfagno, P. (2017) *Estimación de la pérdida de suelo por erosión hídrica en la República Argentina*. Ediciones INTA.
- Geladi, P. y Kowalski, B.R. (1986) Partial least-squares regression: a tutorial. *Analytica chimica acta*, **185**, 1–17.
- Gelman, A. (2004) Parameterization and Bayesian modeling. *Journal of the American Statistical Association*, **99**, 537–545.
- De Gerónimo, E., Aparicio, V.C., Bárbaro, S., Portocarrero, R., Jaime, S. y Costa, J.L. (2014) Presence of pesticides in surface water from four sub-basins in Argentina. *Chemosphere*, **107**, 423–431.
- Gevao, B., Semple, K.T. y Jones, K.C. (2000) Bound pesticide residues in soils: a review. *Environmental pollution*, **108**, 3–14.
- Ghida Daza, C., Juárez-INTA, E.E.A.M., Sánchez, C. y Manfredi-INTA, E.E.A. (2009) *Zonas agroeconómicas homogéneas: Córdoba*. Instituto Nacional de Tecnología Agropecuaria, Buenos Aires (Argentina) ....
- Giannini Kurina, F., Balzarini, M., Rampoldi, A. y Hang, S. (2019a) Site-specific data on herbicide soil retention and ancillary environmental variables. *Data in Brief*, 104754.
- Giannini Kurina, F., Hang, S., Cordoba, M.A., Negro, G.J. y Balzarini, M.G. (2018) Enhancing edaphoclimatic zoning by adding multivariate spatial statistics to regional data. *Geoderma*, **310**, 170–177.
- Giannini Kurina, F., Hang, S., Macchiavelli, R. y Balzarini, M. (2019b) Spatial predictive modelling essential to assess the environmental impacts of herbicides. *Geoderma*, **354**, 113874.
- Gili, A.A. (2013) Modelación de la variación espacial de variables edáficas y su aplicación en el diseño de planes de muestreo de suelos.
- Gimsing, A.L. y dos Santos, A.M. (2005) Glyphosate. p. ACS Publications.
- Golub, G.H., Heath, M. y Wahba, G. (1979) Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, **21**, 215–223.
- Goovaerts, P. (1999) Geostatistics in soil science: state-of-the-art and perspectives.

- Geoderma*, **89**, 1–45.
- Goovaerts, P. (2001) Geostatistical modelling of uncertainty in soil science. *Geoderma*, **103**, 3–26.
- Gorgas, J. y Tassile, J. (2003) Recursos naturales de la provincia de Córdoba. *INTA, Córdoba*.
- Goring, C.A.I., Laskowski, D.A., Hamaker, J.W. y Meikle, R.W. (1975) Principles of pesticide degradation in soil. *Environmental dynamics of pesticides* pp. 135–172. Springer.
- Gregorich, E.G. y Carter, M.R. (2007) *Soil sampling and methods of analysis*. CRC press.
- Grimm, R., Behrens, T., Märker, M. y Elsenbeer, H. (2008) Soil organic carbon concentrations and stocks on Barro Colorado Island—Digital soil mapping using Random Forests analysis. *Geoderma*, **146**, 102–113.
- Grünewald, S. y Walraven, J.C. (2001) Parameter-study on the influence of steel fibers and coarse aggregate content on the fresh properties of self-compacting concrete. *Cement and Concrete Research*, **31**, 1793–1798.
- Gustafson, D.I. (1989) Groundwater ubiquity score: a simple method for assessing pesticide leachability. *Environmental Toxicology and Chemistry: An International Journal*, **8**, 339–357.
- Hamby, D.M. (1994) A review of techniques for parameter sensitivity analysis of environmental models. *Environmental monitoring and assessment*, **32**, 135–154.
- Hance, R.J. (1976) Adsorption of glyphosate by soils. *Pesticide Science*, **7**, 363–366.
- Hang, S., Andriulo, A., Sasal, C., Nassetta, M., Portela, S. y Cañas, A.I. (2010) Integral study of atrazine behavior in field lysimeters in Argentinean humid pampas soils. *Chilean J. Agric. Res*, **70**, 104–112.
- Hang, S., Barriuso, E. y Houot, S. (2003) Behavior of 14 C-atrazine in Argentinean topsoils under different cropping managements. *Journal of environmental quality*, **32**, 2216–2222.
- Hang, S., Barriuso, E. y Houot, S. (2005) Atrazine behaviour in the different pedological horizons of two Argentinean non-till soil profiles. *Weed research*, **45**, 130–139.
- Hang, S., Bocco, M. y Sereno, R. (2000) [Adsorption of atrazine in profiles of Argentina soils under no-tillage systems]. [Spanish]. *Agrochimica*.
- Hang, S., Houot, S. y Barriuso, E. (2007) Mineralization of 14C-atrazine in an entic haplustoll as affected by selected winter weed control strategies. *Soil and Tillage Research*, **96**, 234–242.
- Hang, S., Mercuri, P., Díaz-Zorita, M., Havrylenko, S. y Barriuso, E. (2011) Satellite images as a tool to identify accelerated atrazine mineralization in soils. *Crop protection*, **30**, 663–670.
- Hang, S., Negro, G., Becerra, A. y Rampoldi, A.E. (2015) *Suelos de Córdoba: Variabilidad de las propiedades del horizonte superficial*. Jorge Omar Editorial, Córdoba, Argentina.

- Havrylenko, S.B., Bodoque, J.M., Srinivasan, R., Zucarelli, G. V y Mercuri, P. (2016) Assessment of the soil water content in the Pampas region using SWAT. *Catena*, **137**, 298–309.
- Hengl, T., Heuvelink, G.B.M. y Rossiter, D.G. (2007) About regression-kriging: From equations to case studies. *Computers and Geosciences*, **33**, 1301–1315.
- Hengl, T., Heuvelink, G.B.M. y Stein, A. (2004) A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma*, **120**, 75–93.
- Hengl, T., de Jesus, J.M., Heuvelink, G.B.M., Gonzalez, M.R., Kilibarda, M., Blagotić, A., et al. (2017) SoilGrids250m: Global gridded soil information based on machine learning. *PLoS one*, **12**, e0169748.
- Hengl, T., Nussbaum, M., Wright, M.N., Heuvelink, G.B.M. y Gräler, B. (2018) Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*, **6**, e5518.
- Hengl, T., Rossiter, D.G. y Stein, A. (2003) Soil sampling strategies for spatial prediction by correlation with auxiliary maps. *Soil Research*, **41**, 1403–1422.
- Hennig, C. (2007) Cluster-wise assessment of cluster stability. *Computational Statistics & Data Analysis*, **52**, 258–271.
- Hodson, J. y Williams, N.A. (1988) The estimation of the adsorption coefficient (Koc) for soils by high performance liquid chromatography. *Chemosphere*, **17**, 67–77.
- Holland, J.M. (2004) The environmental consequences of adopting conservation tillage in Europe: reviewing the evidence. *Agriculture, ecosystems & environment*, **103**, 1–25.
- Houot, S., Topp, E., Yassir, A. y Soulas, G. (2000) Dependence of accelerated degradation of atrazine on soil pH in French and Canadian soils. *Soil Biology and Biochemistry*, **32**, 615–625.
- Huang, J., Malone, B.P., Minasny, B., McBratney, A.B. y Triantafyllis, J. (2017) Evaluating a Bayesian modelling approach (INLA-SPDE) for environmental mapping. *Science of the Total Environment*, **609**, 621–632.
- Imbellone, P.A. y Teruggi, M.E. (1993) Paleosols in loess deposits of the Argentine Pampas. *Quaternary International*, **17**, 49–55.
- Iriondo, M.H. y García, N.O. (1993) Climatic variations in the Argentine plains during the last 18,000 years. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **101**, 209–220.
- Isaaks, E.H. y Srivastava, R.M. (1989) *An introduction to applied geostatistics*. Oxford university press.
- Jarsún, B., Gorgas, J., Zamora, E., Bosnero, H., Lovera, E., Ravelo, A., et al. (2006) Los suelos de Córdoba. *Agencia Córdoba Ambiente e Instituto Nacional de Tecnología Agropecuaria, EEA Manfredi. Córdoba, Argentina*.
- Jenks, B.M., Roeth, F.W., Martin, A.R. y McCallister, D.L. (1998) Influence of surface and subsurface soil properties on atrazine sorption and degradation. *Weed Science*, **46**, 132–138.

- Jenny, H. (1994) *Factors of soil formation: a system of quantitative pedology*. Courier Corporation.
- Johnson, D.E. (1998) *Applied multivariate methods for data analysts*. Duxbury press Pacific Grove, CA.
- De Jonge, H., De Jonge, L.W., Jacobsen, O.H., Yamaguchi, T. y Moldrup, P. (2001) Glyphosate sorption in soils of different pH and phosphorus content. *Soil Science*, **166**, 230–238.
- Jury, W.A., Focht, D.D. y Farmer, W.J. (1987) Evaluation of pesticide groundwater pollution potential from standard indices of soil-chemical adsorption and biodegradation 1. *Journal of environmental quality*, **16**, 422–428.
- Kah, M., Beulke, S. y Brown, C.D. (2007) Factors influencing degradation of pesticides in soil. *Journal of agricultural and food chemistry*, **55**, 4487–4492.
- Kanevski, M., Timonin, V., Pozdnukhov, A. y Ritter, G. (2009) *Machine learning for spatial environmental data: theory, applications, and software*. EPFL press.
- Khan, S.U. (1980) *Pesticides in the soil environment*. Elsevier.
- Klute, A. (1986) Water retention: laboratory methods. *Methods of soil analysis: part 1—physical and mineralogical methods*, 635–662.
- Kohonen, T. (1982) Self-organized formation of topologically correct feature maps. *Biological cybernetics*, **43**, 59–69.
- Koskinen, W.C. y Clay, S.A. (1997) Factors affecting atrazine fate in north central US soils. *Reviews of environmental contamination and toxicology* pp. 117–165. Springer.
- 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.
- Krainski, E.T. y Lindgren, F. (2013) The R-INLA tutorial: SPDE models Warning: work in progress... Suggestions are welcome to elias@ r-inla. org.
- Krutz, L.J., Shaner, D.L., Accinelli, C., Zablutowicz, R.M. y Henry, W.B. (2008) Atrazine dissipation in s-triazine-adapted and nonadapted soil from Colorado and Mississippi: implications of enhanced degradation on atrazine fate and transport parameters. *Journal of environmental quality*, **37**, 848–857.
- Kuhn, M. y Johnson, K. (2013) *Applied predictive modeling*. Springer.
- Lammoglia, S.-K., Brun, F., Quemar, T., Moeys, J., Barriuso, E., Gabrielle, B., et al. (2018) Modelling pesticides leaching in cropping systems: Effect of uncertainties in climate, agricultural practices, soil and pesticide properties. *Environmental Modelling & Software*.
- Lark, R.M. (2000) A comparison of some robust estimators of the variogram for use in soil survey. *European Journal of soil science*, **51**, 137–157.
- Leistra, M. y Green, R.E. (1990) Efficacy of Soil-Applied Pesticides. *Pesticides in the Soil Environment: Processes, Impacts and Modeling*, **2**, 401–428.

- Li, Y. (2010) Can the spatial prediction of soil organic matter contents at various sampling scales be improved by using regression kriging with auxiliary information? *Geoderma*, **159**, 63–75.
- Li, J., Heap, A.D., Potter, A. y Daniell, J.J. (2011) Application of machine learning methods to spatial interpolation of environmental variables. *Environmental Modelling & Software*, **26**, 1647–1659.
- Lindgren, F. y Rue, H. (2015a) Bayesian Spatial Modelling with *R* - **INLA**. *Journal of Statistical Software*, **63**.
- Lindgren, F. y Rue, H. (2015b) Bayesian spatial modelling with R-INLA. *Journal of Statistical Software*, **63**, 1–25.
- Lindgren, F., Rue, H. y Lindström, J. (2011) An explicit link between Gaussian fields 670 and Gaussian Markov random fields: the SPDE approach (with discussion). *JR 671 Stat Soc, Series B*, **73**, 423–498.
- Linn, D.M., Carski, T.H., Brusseau, M.L. y Chang, F.-H. (1993) Sorption and degradation of pesticides and organic chemicals in soil.
- Loeppert, R.H., Inskeep, W.P. y Sparks, D.L. (1996) Methods of Soil Analysis. Part 3—Chemical Methods. *Soil Science Society of America, Inc. and American Society of Agronomy, Inc*, **677**.
- Long, D.S. (1998) Spatial autoregression modeling of site-specific wheat yield. *Geoderma*, **85**, 181–197.
- Longley, P.A., Goodchild, M.F., Maguire, D.J. y Rhind, D.W. (2005) *Geographic information systems and science*. John Wiley & Sons.
- Lovelace, R., Nowosad, J. y Muenchow, J. (2019) *Geocomputation with R*. CRC Press.
- Mamy, L., Barriuso, E. y Gabrielle, B. (2005) Environmental fate of herbicides trifluralin, metazachlor, metamitron and sulcotrione compared with that of glyphosate, a substitute broad spectrum herbicide for different glyphosate-resistant crops. *Pest Management Science: formerly Pesticide Science*, **61**, 905–916.
- 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.
- Manzur, A. (1997) Dinámicas evolutivas de suelos en Atum Pampa, sierras pampeanas, Córdoba, Argentina. *Multequina*, 67–83.
- Marín-Benito, J.M., Pot, V., Alletto, L., Mamy, L., Bedos, C., Barriuso, E., et al. (2014) Comparison of three pesticide fate models with respect to the leaching of two herbicides under field conditions in an irrigated maize cropping system. *Science of the Total Environment*, **499**, 533–545.
- Matérn, B. (1986) Spatial variation, vol. 36. *Lecture Notes in Statistics*, **2**.
- McBratney, A.B., Mendonça Santos, M.L. y Minasny, B. (2003) *On digital soil mapping*.
- McBride, M. y Kung, K.-H. (1989) Complexation of glyphosate and related ligands with iron (III). *Soil Science Society of America Journal*, **53**, 1668–1673.

- McGlamery, M.D. y Slife, F.W. (1966) The adsorption and desorption of atrazine as affected by pH, temperature, and concentration. *Weeds*, 237–239.
- McGrath, S.P., Sanders, J.R. y Shalaby, M.H. (1988) The effects of soil organic matter levels on soil solution concentrations and extractabilities of manganese, zinc and copper. *Geoderma*, **42**, 177–188.
- Mehlich, A. (1984) Mehlich 3 soil test extractant: A modification of Mehlich 2 extractant. *Communications in soil science and plant analysis*, **15**, 1409–1416.
- Meng, J., Zhang, H., Evans, G.D. y Duan, X. (2003) Studies on assembly and structure of supramolecular glyphosate intercalated magnesium-aluminum layered double hydroxide. *Chemical Journal of Chinese Universities*, **24**, 1315–1319.
- Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C.C., et al. (2014) e1071: Misc functions of the Department of Statistics (e1071), TU Wien. *R package version*, **1**.
- Minasny, B. y McBratney, A.B. (2006) A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Computers and Geosciences*, **32**, 1378–1388.
- Minasny, B. y McBratney, A.B. (2016) Digital soil mapping: A brief history and some lessons. *Geoderma*, **264**.
- Molenberghs, G., Verbeke, G. y Demétrio, C.G.B. (2007) An extended random-effects approach to modeling repeated, overdispersed count data. *Lifetime data analysis*, **13**, 513–531.
- Montgomery, D.R. (2007) Soil erosion and agricultural sustainability. *Proceedings of the National Academy of Sciences*, **104**, 13268–13272.
- Montoya, J.C., Bedmar, F., Daniel, P.E. y Costa, J.L. (2003) Sorption of atrazine and three of its degradation products in different soils and tillage systems. *International Workshop on Groundwater Risk Assessment at Contaminated Sites [GRACOS] and Integrated Soil and Water Protection [SOWA]. 2. 2003 03 20-21, 20-21 de marzo de 2003. Tubingen. DE.*
- Montoya, J.C., Costa, J.L., Liedl, R., Bedmar, F. y Daniel, P. (2006) Effects of soil type and tillage practice on atrazine transport through intact soil cores. *Geoderma*, **137**, 161–173.
- Moreale, A. y Van Bladel, R. (1979) SOIL INTERACTIONS OF HERBICIDE-DERIVED ANILINE RESIDUES: A THERMODYNAMIC APPROACH1. *Soil Science*, **127**, 1–9.
- Morgan, R.P.C. (2009) *Soil erosion and conservation*. John Wiley & Sons.
- Mortensen, J.L. (1963) Complexing of Metals by Soil Organic Matter 1. *Soil Science Society of America Journal*, **27**, 179–186.
- Nannipieri, P., Ascher, J., Ceccherini, M., Landi, L., Pietramellara, G. y Renella, G. (2003) Microbial diversity and soil functions. *European journal of soil science*, **54**, 655–670.
- Neitsch, S.L., Arnold, J.G. y Srinivasan, R. (2002) Pesticides fate and transport predicted by the soil and water assessment tool (SWAT). *Atrazine, Metolachlor and*

*Trifluralin in the Sugar Creek Watershed: BRC Report, 3.*

- Odeh, I.O.A., Chittleborough, D.J. y McBratney, A.B. (1991) Elucidation of soil-landform interrelationships by canonical ordination analysis. *Geoderma*, **49**, 1–32.
- Pang, N., Wang, T. y Hu, J. (2016) Method validation and dissipation kinetics of four herbicides in maize and soil using QuEChERS sample preparation and liquid chromatography tandem mass spectrometry. *Food chemistry*, **190**, 793–800.
- Panigatti, J.L. (2015) Aspectos de la erosión de los suelos en Argentina. *Ed. AACs. Buenos Aires*.
- Pebesma, E.J. (2004) Multivariable geostatistics in S: the gstat package. *Computers & Geosciences*, **30**, 683–691.
- Petach, M.C., Wagenet, R.J. y DeGloria, S.D. (1991) Regional water flow and pesticide leaching using simulations with spatially distributed data. *Geoderma*, **48**, 245–269.
- Pfeiffer, M. (2010) Groundwater Ubiquity Score (GUS). *Pesticide Training Resources* [www.ptrpest.com](http://www.ptrpest.com), 40–41.
- Piccolo, A., Celano, G. y Conte, P. (1996) Adsorption of glyphosate by humic substances. *Journal of agricultural and food chemistry*, **44**, 2442–2446.
- 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.
- Poggio, L., Gimona, A., Spezia, L. y Brewer, M.J. (2016) Bayesian spatial modelling of soil properties and their uncertainty: The example of soil organic matter in Scotland using R-INLA. *Geoderma*, **277**, 69–82.
- Prasad, A.M., Iverson, L.R. y Liaw, A. (2006) Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems*, **9**, 181–199.
- R Development Core Team, R. (2011) R: A Language and Environment for Statistical Computing (ed RDC Team). *R Foundation for Statistical Computing*, **1**, 409.
- Rampazzo, N., Todorovic, G.R., Mentler, A. y Blum, W.E.H. (2013) Adsorption of glyphosate and aminomethylphosphonic acid in soils. *International Agrophysics*, **27**, 203–209.
- Rampoldi, A., Hang, S. y Barriuso, E. (2008) Glyphosate mineralization: Effect of temperature and soybean and corn crop residues. *Chilean journal of agricultural research*, **68**, 13–20.
- Rampoldi, E.A., Hang, S. y Barriuso, E. (2011) The Fate of Glyphosate in Crop Residues. *Soil Science Society of America Journal*, **75**, 553.
- Raudenbush, S.W. y Bryk, A.S. (2002) *Hierarchical linear models: Applications and data analysis methods*. Sage.
- Reich, B.J. y Fuentes, M. (2007) A multivariate semiparametric Bayesian spatial modeling framework for hurricane surface wind fields. *The Annals of Applied Statistics*, **1**, 249–264.
- Di Rienzo, J.A., Casanoves, F., Balzarini, M.G., Gonzalez, L., Tablada, M. y Robledo,

C.W. (2019) InfoStat.

- Rocca, R.J., Redolfi, E.R. y Terzariol, R.E. (2006) Características geotécnicas de los loess de Argentina. *Revista Internacional de Desastres Naturales, Accidentes e Infraestructura Civil*, **6**, 149–166.
- Romagnoli, M., Portapila, M., Rigalli, A., Maydana, G., Burgués, M. y García, C.M. (2017) Assessment of the SWAT model to simulate a watershed with limited available data in the Pampas region, Argentina. *Science Of The Total Environment*, **596**, 437–450.
- Roudier, P. (2011) *clhs: a R package for conditioned Latin hypercube sampling*. R package version 0.5-1.
- Rue, H. y Held, L. (2005) *Gaussian Markov random fields: theory and applications*. CRC press.
- Rue, H., Martino, S. y 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.
- Sakamoto, Y., Ishiguro, M. y Kitagawa, G. (1986) Akaike information criterion statistics. *Dordrecht, The Netherlands: D. Reidel*, **81**.
- Sarmah, A.K., Close, M.E. y Mason, N.W.H. (2009) Dissipation and sorption of six commonly used pesticides in two contrasting soils of New Zealand. *Journal of Environmental Science and Health Part B*, **44**, 325–336.
- Sasal, M.C., Demonte, L., Cislighi, A., Gabioud, E.A., Oszust, J.D., Wilson, M.G., et al. (2015) Glyphosate loss by runoff and its relationship with phosphorus fertilization. *Journal of agricultural and food chemistry*, **63**, 4444–4448.
- Sasal, M.C., Wilson, M.G. y NA, G. (2008) Escurrimiento superficial y pérdidas de nutrientes y glifosato en secuencias de cultivo. *Agricultura sustentable Actualización técnica, Seria extensión*.
- Schabenberger, O. y Gotway, C.A. (2005) *Statistical methods for spatial data analysis*. CRC press.
- Schwarz, G. (1978) Estimating the dimension of a model. *The annals of statistics*, **6**, 461–464.
- Silva, V., Mol, H.G.J., Zomer, P., Tienstra, M., Ritsema, C.J. y Geissen, V. (2019) Pesticide residues in European agricultural soils – A hidden reality unfolded. *Science of the Total Environment*, **653**, 1532–1545.
- Sinkkonen, S. y Paasivirta, J. (2000) Degradation half-life times of PCDDs, PCDFs and PCBs for environmental fate modeling. *Chemosphere*, **40**, 943–949.
- Skjemstad, J.O., Reicosky, D.C., Wilts, A.R. y McGowan, J.A. (2002) Charcoal carbon in US agricultural soils. *Soil Science Society of America Journal*, **66**, 1249–1255.
- Soulas, G. y Lagacherie, B. (2001) Modelling of microbial degradation of pesticides in soils. *Biology and fertility of Soils*, **33**, 551–557.
- Sparks, D.L., Helmke, P.A. y Page, A.L. (1996) *Methods of soil analysis: Chemical methods*. SSSA.

- Spurlock, F., Clayton, M., Das, S.K. y Troiano, J. (2006) Modeling herbicide movement to ground water in irrigated sandy soils of the san Joaquin Valley, California. *Water, Air, and Soil Pollution*, **176**, 93–111.
- Srinivasan, R., Zhang, X. y Arnold, J. (2010) SWAT ungauged: hydrological budget and crop yield predictions in the Upper Mississippi River Basin. *Transactions of the ASABE*, **53**, 1533–1546.
- Stolpe, N.B. y Shea, P.J. (1995) Alachlor and atrazine degradation in a Nebraska soil and underlying sediments. *Soil Science*, **160**, 359–370.
- Stow, C.A., Reckhow, K.H. y Qian, S.S. (2006) A Bayesian approach to retransformation bias in transformed regression. *Ecology*, **87**, 1472–1477.
- Stroup, W.W. (2016) *Generalized linear mixed models: modern concepts, methods and applications*. CRC press.
- Sun, B., Zhou, S. y Zhao, Q. (2003) Evaluation of spatial and temporal changes of soil quality based on geostatistical analysis in the hill region of subtropical China. *Geoderma*, **115**, 85–99.
- Targulian, V.O. y Sokolova, T.A. (1996) Soil as a biotic/abiotic natural system: a reactor, memory, and regulator of biospheric interactions. *Eurasian Soil Science*, **29**, 30–41.
- Theodoridis, S. y Koutroumbas, K. (2009) Pattern recognition. 2003. *Elsevier Inc*.
- Topp, E., Gutzman, D.W., Millette, J., Gamble, D.S. y Bourgoin, B. (1995) Rapid mineralization of the herbicide atrazine in alluvial sediments and enrichment cultures. *Environmental Toxicology and Chemistry: An International Journal*, **14**, 743–747.
- Vaysse, K. y Lagacherie, P. (2015) Evaluating Digital Soil Mapping approaches for mapping GlobalSoilMap soil properties from legacy data in Languedoc-Roussillon (France). *Geoderma Regional*, **4**.
- Verbeke, G. y Molenberghs, G. (2009) *Linear mixed models for longitudinal data*. Springer Science & Business Media.
- Wackernagel, H. (2013) *Multivariate geostatistics: an introduction with applications*. Springer Science & Business Media.
- Wang, X., Ryan, Y.Y. y Faraway, J.J. (2018) *Bayesian Regression Modeling with INLA*. Chapman and Hall/CRC.
- Warner, L.L. y Diab, R.D. (2002) Use of geographic information systems in an environmental impact assessment of an overhead power line. *Impact Assessment and Project Appraisal*, **20**, 39–47.
- Wauchope, R.D. y Myers, R.S. (1985) Adsorption-Desorption Kinetics of Atrazine and Linuron in Freshwater-Sediment Aqueous Slurries 1. *Journal of environmental quality*, **14**, 132–136.
- Weber, J.B., Wilkerson, G.G. y Reinhardt, C.F. (2004) Calculating pesticide sorption coefficients (Kd) using selected soil properties. *Chemosphere*, **55**, 157–166.
- Webster, R. y Oliver, M.A. (2007) *Geostatistics for environmental scientists*. John Wiley & Sons.

- Weed, D.A.J., Kanwar, R.S., Stoltenberg, D.E. y Pfeiffer, R.L. (1995) Dissipation and distribution of herbicides in the soil profile. *Journal of Environmental Quality*, **24**, 68–79.
- West, T.O. y Post, W.M. (2002) Soil organic carbon sequestration rates by tillage and crop rotation. *Soil Science Society of America Journal*, **66**, 1930–1946.
- West, B.T., Welch, K.B., Galecki, A.T. y Edition, S. (2014) *Linear mixed models: a practical guide using statistical software*. Chapman and Hall/CRC.
- Willmott, C.J. (1981) On the validation of models. *Physical geography*, **2**, 184–194.
- Witten, I.H., Frank, E., Hall, M.A. y Pal, C.J. (2016) *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.
- Wu, J., Nofziger, D.L., Warren, J.G. y Hattey, J. a. (2003) Modeling Ammonia Volatilization from Surface-Applied Swine Effluent. *Soil Science Society of America Journal*, **67**, 1.
- Yakovchenko, V.P., Sikora, L.J. y Millner, P.D. (1998) Carbon and nitrogen mineralization of added particulate and macroorganic matter. *Soil Biology and Biochemistry*, **30**, 2139–2146.
- Zachar, D. (2011) *Soil erosion*. Elsevier.

## Anexo

### Anexo I: Descripción de variables edafoclimáticas

Cuadro 14: Variables de sitio edáficas climáticas y topográficas. Córdoba, Argentina.

Variable	Unidades	Descripción
ID_2	-	Código de identificación
X UTM20	m	Sistema de coordenadas Universal Transversal de Mercator zona 20.
Y UTM20	m	
pH	-	pH en agua 1:2,5 (suelo:agua)
CE	dS m <sup>-1</sup>	Conductividad Eléctrica en agua 1:2,5 (suelo:agua)
COT	g kg <sup>-1</sup>	Carbono Orgánico Total por combustión húmeda por 1N K <sub>2</sub> Cr <sub>2</sub> O <sub>7</sub> , método Walkley y Black (Sparks, Helmke y Page, 1996)
NT	% p:p	Nitrógeno Total, método Kjeldahl (Sparks et al., 1996)
Mn	mg kg <sup>-1</sup>	Magnesio extractable por Mehlic-3 (Mehlich, 1984)
Cu	mg kg <sup>-1</sup>	Cobre extractable por Mehlic-3 (Mehlich, 1984)
Zn	mg kg <sup>-1</sup>	Zinc extractable por Mehlic-3 (Mehlich, 1984)
CC	%	Capacidad de Campo, 300 kPa en olla de presión (Klute, 1986)
Arena	%	Contenido de arena, método de pipeta de Robinson (Sparks et al., 1996)
Limo	%	Contenido de limo, método de pipeta de Robinson (Sparks et al., 1996)
Arcilla	%	Contenido de arcilla, método de pipeta de Robinson (Sparks et al., 1996)
Al(Ox)	%	Óxidos de Aluminio (Loeppert, Inskeep y Sparks, 1996)
Fe(Ox)	%	Óxidos de Hierro (Loeppert et al., 1996)
P	ppm	Fósforo extractable por Bray y Kurtz, medición por colorimetría (Sparks et al., 1996)
K	ppm	Potasio intercambiable, medición por fotometría de llama (Sparks et al., 1996)
Ca	ppm	Calcio intercambiable, complexometría (Sparks et al., 1996)
Na	ppm	Sodio intercambiable, fotometría de llama (Sparks et al., 1996)
Mg	ppm	Magnesio intercambiable, complexometría (Sparks et al., 1996)
CIC	Cmol kg <sup>-1</sup>	Capacidad de Intercambio Catiónico (Sparks et al., 1996)
Elevación	m.s.n.m	Elevación, Modelo Digital de Elevación, STRM (Farr et al., 2007)
Pendiente	%	Pendiente derivada de DEM STRM (Farr et al., 2007)
pp	mm	Precipitaciones acumuladas anual, BIOCLIM (Booth et al., 2014)
Tmax	°C	Temperatura máxima anual, BIOCLIM (Booth et al., 2014)
Tmin	°C	Temperatura mínima anual, BIOCLIM (Booth et al., 2014)
Tm	°C	Temperatura media anual, BIOCLIM (Booth et al., 2014)
TvsPP	°C mm <sup>-1</sup>	Cociente entre Tm y pp, indicador de balance hídrico
Uso	-	Uso del suelo al momento del muestreo