



**UNIVERSIDAD DE BURGOS**

**Departamento de Biotecnología y Ciencia de los Alimentos**

**Methodology for the disaggregation of polytaxic cartographic units in  
intensive conventional soil maps.**

**Application towards the improvement of the Integrated Terroir Zoning of  
the Designation of Origin Campo de Borja**

Tesis doctoral

Alberto D. Lázaro López

Burgos, 2021





**UNIVERSIDAD  
DE BURGOS**

**Departamento de Biotecnología y Ciencia de los Alimentos**

**Methodology for the disaggregation of polytaxic cartographic units in  
intensive conventional soil maps.**

**Application towards the improvement of the Integrated Terroir Zoning of  
the Designation of Origin Campo de Borja**

Metodología para la disgregación de unidades cartográficas politáxicas en mapas  
de suelos convencionales intensivos.

Aplicación en la mejora de la Zonificación Integrada del Terroir de la  
Denominación de Origen Campo de Borja

Memoria de Tesis Doctoral presentada por D. Alberto D. Lázaro López  
para optar al grado de doctor por la Universidad de Burgos

Directores

Dra. M<sup>a</sup> Luisa González San José

Dr. Vicente D. Gómez Miguel

Burgos, 2021





## UNIVERSIDAD DE BURGOS

La Dra. M<sup>a</sup> Luisa González San José, Catedrática del área de Tecnología de los Alimentos de la Universidad de Burgos, y el Dr. Vicente Daniel Gómez Miguel, Profesor Titular del área de Edafología de la Universidad Politécnica de Madrid, en su calidad de directores de la tesis doctoral de título *“Methodology for the disaggregation of polytaxic cartographic units in intensive conventional soil maps. Application towards the improvement of the Integrated Terroir Zoning of the Designation of Origin Campo de Borja”* que presenta D. Alberto D. Lázaro López,

Informan:

Favorablemente la presentación de dicha tesis, ya que consideran que reúne las condiciones necesarias para su defensa, en cuanto a fase experimental (uso de información y manejo de datos) y elaboración de la memoria.

Y para que así conste y surta los efectos oportunos firman en Burgos a 28 de diciembre de 2021.

Fdo.: Dña. M<sup>a</sup> Luisa González San José      Fdo.: D. Vicente Daniel Gómez Miguel



Este trabajo de Tesis Doctoral ha sido  
realizado con:

La ayuda del programa para Formación  
de Profesores Universitarios (FPU)  
[número FPU17/05855] financiada por  
el Ministerio de Ciencia, Innovación y  
Universidades de España.

Y la ayuda para contratos predoctorales  
de la Universidad de Burgos 2017.





*I met a traveller from an antique land  
Who said:—Two vast and trunkless legs of stone  
Stand in the desert. Near them on the sand,  
Half sunk, a shatter'd visage lies, whose frown*

*And wrinkled lip and sneer of cold command  
Tell that its sculptor well those passions read  
Which yet survive, stamp'd on these lifeless things,  
The hand that mock'd them and the heart that fed.*

*And on the pedestal these words appear:  
«My name is Ozymandias, king of kings:  
Look on my works, ye mighty, and despair!»*

*Nothing beside remains: round the decay  
Of that colossal wreck, boundless and bare,  
The lone and level sands stretch far away.*

Percy Bysshe Shelley, «Ozymandias» (1817)

*A un viajero vi, de tierras remotas.  
Me dijo: hay dos piernas en el desierto,  
De piedra y sin tronco. A su lado cierto  
Rostro en la arena yace: la faz rota,*

*Sus labios, su frío gesto tirano,  
Nos dicen que el escultor ha podido  
Salvar la pasión, que ha sobrevivido  
Al que pudo tallarlo con su mano.*

*Algo ha sido escrito en el pedestal:  
«Soy Ozymandias, el gran rey. ¡Mirad  
Mi obra, poderosos! ¡Desesperad!»*

*La ruina es de un naufragio colosal.  
A su lado, infinita y legendaria  
Sólo queda la arena solitaria».*

Traducción de Fernando G. Toledo  
en ¡Vive este diario!  
(Roca Editorial, Barcelona, 2016)



## **Agradecimientos**

Durante la redacción del presente texto he sido titular de la ayuda FPU financiada por el Ministerio de Ciencia, Innovación y Universidades de España. Con anterioridad, había disfrutado de una ayuda para contratos predoctorales de la Universidad de Burgos. Ambos contratos ligados al Departamento de Biotecnología y Ciencia de los Alimentos de la Universidad de Burgos. Por ello, quiero hacer un reconocimiento expreso a estas instituciones por su labor.

A todos quienes habéis contribuido a hacer posible este trabajo, mi más sincera gratitud.



# TABLA DE CONTENIDOS

<i>Tabla de contenidos</i>	<i>xi</i>
<i>Abreviaturas y acrónimos</i>	<i>xvii</i>
<i>Índice de figuras</i>	<i>xxi</i>
<i>Índice de tablas</i>	<i>xxv</i>
<i>Summary</i>	<i>xxvii</i>
<i>Contextualización</i>	<i>1</i>
<i>Objetivos</i>	<i>5</i>
<i>Capítulo 1 De la Cartografía de Suelos Convencional a la Digital: antecedentes</i>	<i>7</i>
<b>1 Inventarios del Recurso Suelo</b>	<b>7</b>
<b>2 Cartografía de Suelos Convencional</b>	<b>7</b>
2.1 Objetivos, recursos y escalas	8
2.2 Delineaciones	9
2.3 Muestreo	11
2.4 Clasificación	13
2.5 Correlación	14
2.6 Presentación de los resultados	15
<b>3 Cartografía de Suelos Digital</b>	<b>17</b>
3.1 Definición de Cartografía de Suelos Digital	17
3.1.1 Bases teóricas	17
3.1.2 Objetivos y mejoras respecto a la Cartografía de Suelos Convencional	18
3.1.3 Organización de la Cartografía de Suelos Digital	19
3.2 Datos de suelos heredados	19
3.3 Fuentes de datos para coberturas continuas	21
3.3.1 Caracterización	21

3.3.2	Preprocesado	22
<b>3.4</b>	<b>Covariables</b>	<b>22</b>
3.4.1	Geomorfométricas	22
3.4.2	Reflectancia en el espectro electromagnético	24
3.4.3	Interpolación desde observaciones	24
3.4.4	Selección de covariables	25
<b>3.5</b>	<b>Muestreo</b>	<b>26</b>
<b>3.6</b>	<b>Tipología</b>	<b>28</b>
3.6.1	Propiedades o atributos del suelo	28
3.6.2	Clases de suelo y categorías de atributos	28
<b>3.7</b>	<b>Evaluación</b>	<b>30</b>
3.7.1	Precisión	31
3.7.2	Incertidumbre	33
<b>4</b>	<b>Bibliografía</b>	<b>33</b>
<b>Capítulo 2</b>		<b>39</b>
<b><i>Disaggregation of conventional soil maps: a review</i></b>		
<b>1</b>	<b>Introduction</b>	<b>39</b>
<b>2</b>	<b>Bibliographical Search</b>	<b>40</b>
<b>3</b>	<b>Disaggregation methodologies on selected papers</b>	<b>41</b>
<b>3.1</b>	<b>The conventional soil map</b>	<b>45</b>
3.1.1	Extent and location	45
3.1.2	Scale	46
3.1.3	Categorical detail	46
3.1.4	Cartographic detail	46
3.1.5	Scale gain during the disaggregation	46
<b>3.2</b>	<b>Environmental data sources and covariates</b>	<b>47</b>
3.2.1	Covariates and SCORPAN factors	47
3.2.2	Multi-scale covariates	49
3.2.3	Covariate selection	49
3.2.4	Covariate importance	51
<b>3.3</b>	<b>Classification of soil classes</b>	<b>52</b>
3.3.1	Leverage of the conventional map in the soil-landscape model	52

3.3.2	Assignment of soil categories: training sets	53
3.3.1	Sampling schemes over map units	54
3.3.2	Assignment of soil categories in two-stages	59
<b>3.4</b>	<b>Evaluation of soil-landscape models</b>	<b>61</b>
3.4.1	Accuracy evaluation from observations	61
3.4.2	Analysis of the Overall Accuracy	63
3.4.3	Accuracy evaluation from conventional maps	65
<b>4</b>	<b>Conclusions</b>	<b>66</b>
<b>5</b>	<b>References</b>	<b>67</b>
<b>Capítulo 3 A new methodology to divide conventional soil map units</b>		<b>73</b>
<b>1</b>	<b>Introduction</b>	<b>73</b>
<b>2</b>	<b>Material and methods</b>	<b>74</b>
2.1	Research site	74
2.2	Data sources	74
2.2.1	Legacy soil data: the conventional soil map	75
2.2.1	Environmental data: DEM and remote sensing	75
2.2.2	Ancillary data: structure and constructions	76
2.3	Data management, GIS, and statistical software	77
2.4	Procedure	77
2.4.1	Processing and scaling of data sources	77
2.4.2	Calculation of covariates	78
2.4.1	Analysis of covariates	78
2.4.1	Selection of covariates	78
2.4.2	Division of the initial Soil Map Units into new homogenous areas	80
2.4.3	Evaluation of the new homogeneous areas	81
<b>3</b>	<b>Results and discussion</b>	<b>82</b>
3.1	Processing of data sources	82
3.2	Analysis of covariates	82
3.2.1	Curvature variables	82
3.2.1	Erosion vs accumulation variables	83

3.2.2	Remote sensing variables	83
3.2.1	Topo-Climatic variables	83
<b>3.3</b>	<b>Selection of covariates</b>	<b>85</b>
<b>3.4</b>	<b>Evaluation of the new homogeneous areas</b>	<b>86</b>
3.4.1	Comparative of delineations	86
3.4.2	Comparative of divisions within each Soil Map Unit	86
<b>4</b>	<b>Conclusions</b>	<b>89</b>
<b>5</b>	<b>References</b>	<b>90</b>

**Capítulo 4 Disgregación del mapa de suelos convencional semi-detallado de la DO Campo de Borja** 93

<b>1</b>	<b>Introducción</b>	<b>93</b>
<b>2</b>	<b>Materiales y métodos</b>	<b>94</b>
2.1	Área de trabajo	94
2.2	Fuentes de datos	95
2.2.1	Mapa convencional de suelos	95
2.2.2	Observaciones de suelos heredadas	96
2.2.1	Covariables	98
2.2.2	Estructuras y construcciones misceláneas	99
2.3	Procedimiento	100
2.3.1	División de las SMU	100
2.3.2	Correlación entre divisiones y STU	102
2.3.3	Refinado de las delineaciones	102
2.3.4	Validación	103
<b>3</b>	<b>Resultados y discusión</b>	<b>104</b>
3.1	STU y nuevas SMU	104
3.2	Delineaciones	106
3.3	Precisión	108
<b>4</b>	<b>Conclusiones</b>	<b>110</b>
<b>5</b>	<b>Bibliografía</b>	<b>111</b>



<b>Capítulo 5</b>	<b>Aplicación del mapa de suelos disgregado en la Zonificación Integrada del Terroir vitícola de la DO Campo de Borja</b>	<b>115</b>
<b>1</b>	<b>Introducción</b>	<b>115</b>
<b>2</b>	<b>Materiales y métodos</b>	<b>117</b>
2.1	Área de trabajo	117
2.2	Fuentes de datos	117
2.2.1	Zonificación del suelo original: Mapa de suelos convencional con fases	117
2.2.2	Zonificación del suelo actualizada: Mapa de suelos disgregado	117
2.2.3	Observaciones de suelos heredadas	118
2.2.4	Mapa de zonificación del clima	118
2.2.5	Mapa de zonificación del medio	119
2.2.6	Registro Vitícola y Sistema de Información Geográfica de Parcelas Agrícolas	122
2.3	Procedimiento	122
2.3.1	Nuevas Unidades Homogéneas del Medio	124
2.3.2	Índice de Calidad	125
2.3.3	Índice de Ocupación	127
2.3.4	Categorización: clases de UHM por índices	127
2.3.5	Modelización: IO vs IC	128
<b>3</b>	<b>Resultados y discusión</b>	<b>128</b>
3.1	UHM con el mapa de suelos disgregado	128
3.2	IC	129
3.3	IO	132
<b>4</b>	<b>Conclusiones</b>	<b>134</b>
<b>5</b>	<b>Bibliografía</b>	<b>135</b>
	<b>General conclusions</b>	<b>137</b>
	<b>Anexo I Mapa del Inventario del Recurso Suelo disgregado de la Denominación de Origen Campo de Borja</b>	<b>139</b>
	<b>Anexo II Mapa de la Nueva Zonificación Integrada del Terroir de la Denominación de Origen Campo de Borja</b>	<b>147</b>



## ABREVIATURAS Y ACRÓNIMOS

En general, se usarán abreviaturas y acrónimos relacionados con los nombres en inglés en las partes escritas en castellano por ser lo más habitual en los textos y trabajos de este campo/área (ej. DSM: *Digital Soil Mapping*, Cartografía de Suelos Digital), con excepciones minoritarias (ej. IRS ~ SRI).

<b>AIC</b>	Criterio de Información de Akaike	Akaike Information Criterion
<b>ANN</b>	Redes Neuronales Artificiales	Artificial Neural Networks
<b>ASA</b>	Área media de las delineaciones	Average-size area
<b>ASD</b>	Tamaño medio de las delineaciones	Average-size delineation
<b>ASPECT</b>	Orientación	Aspect
<b>CCROS</b>	Curvatura transversal	Cross-sectional curvature
<b>CFS</b>	Selección de variables basada en la correlación	Correlation-based feature selection
<b>CI</b>	Índice de Coloración	Coloration Index
<b>CLARA</b>	Agrupación de grandes paquetes	Clustering Large Applications
<b>cLHS</b>	Muestreo de Hipercubo Latino Condicionado	Conditioned Latin Hypercube Sampling
<b>CLONG</b>	Curvatura longitudinal	Longitudinal curvature
<b>CMAXI</b>	Curvatura máxima	Maximum curvature
<b>CMINI</b>	Curvatura mínima	Minimum curvature
<b>CNIG</b>	Centro Nacional de Información Geográfica	National Geographic Information Center
<b>CPLAN</b>	Curvatura horizontal	Plan curvature
<b>CPROF</b>	Curvatura de perfil	Profile curvature
<b>CSM</b>	Cartografía de Suelos Convencional	Conventional Soil Mapping
<b>CTANG</b>	Curvatura tangencial	Tangential curvature
<b>DAH</b>	Calor anisotrópico diurno	Diurnal Anisotropic Heat
<b>DDIFF</b>	Diferencia entre DWSLOPE y SLOPE	Difference between DWSLOPE and SLOPE
<b>DISTV</b>	Distancia vertical a la red de cauces	Vertical distance to cannel network
<b>DO</b>	Denominación de Origen	Designation of Origin
<b>DSM</b>	Cartografía de Suelos Digital	Digital Soil Mapping

<b>DT</b>	Árboles de decisión		Decision Trees
<b>DWSLOPE</b>	Índice de pendiente cuesta abajo		Downslope index
<b>EAST</b>	Orientalización		Easternness
<b>EASTSP</b>	Interacción entre EAST y SLOPE		Interaction between EAST and SLOPE
<b>ELEVATION</b>	Altitud		Elevation
<b>ESA</b>	Agencia Espacial Europea		European Space Agency
<b>ETM+</b>	Visor temático mejorado, Landsat		Enhanced Thematic Mapper Plus, Landsat
<b>FAO</b>	Organización de las Naciones Unidas para la Alimentación y la Agricultura		Food and Agriculture Organization
<b>FEOUS</b>	Ratio simple entre las regiones del rojo y del SWIR		Simple ratio Red ~ SWIR
<b>FIA</b>	Fotointerpretación aérea	<b>API</b>	Aerial Photo Interpretation
<b>GNSS</b>	Sistema Global de Navegación por Satélite		Global Navigation Satellite System
<b>GPS</b>	Sistema de Posicionamiento Global		Global Positioning System
<b>GRASS</b>			Geographic Resources Analysis Support System
<b>IC</b>	Índice de Calidad		Quality Index
<b>IMR</b>	Índice de Máxima Reducción		Index of Maximum Reduction
<b>IO</b>	Índice de Ocupación		Occupation Index
<b>IR</b>	Infrarrojo		Infrared
<b>IRS</b>	Inventarios del Recurso Suelo	<b>SRI</b>	Soil Resources Inventories
<b>IUSS</b>	Unión Internacional de Ciencia del Suelo		International Union of Soil Sciences
<b>kNN</b>	k Vecinos más cercanos		k-Nearest Neighbour
<b>LIDAR</b>			Light Detection And Ranging
<b>LST</b>	Temperatura de la superficie terrestre		Land Surface Temperature
<b>MDA</b>	Disminución media de la precisión		Mean Decrease Accuracy
<b>MDT</b>	Modelo Digital del Terreno	<b>DEM</b>	Digital Elevation Model
<b>ML</b>	Aprendizaje automático		Machine Learning
<b>MLA</b>	Área Mínima Legible		Minimum Legible Area
<b>MLD</b>	Delineación Mínima Legible		Minimum Legible Delineation
<b>MLR</b>	Regresión logística multinomial		Multinomial Logistic Regression
<b>MODIS</b>	Espectrorradiómetro de imágenes de resolución moderada		Moderate Resolution Imaging Spectroradiometer

<b>MPI</b>	Índice de Protección Morfométrica	Morphometric Protection Index
<b>MRRTF</b>	Índice de multirresolución del aplanamiento de la cima de la cresta	Multiresolution Index of Ridge Top Flatness
<b>MRVBF</b>	Índice de multirresolución del aplanamiento del fondo del valle	Multiresolution Index of Valley Bottom Flatness
<b>MSE</b>	Error cuadrático medio	Mean Square Error
<b>MSS</b>	Sistema de escáner multiespectral, Landsat	Multispectral Scanner System, Landsat
<b>NASA</b>	Administración Nacional de Aeronáutica y el Espacio (EUA)	National Aeronautics and Space Administration
<b>NASIS</b>	Sistema de Información Nacional de Suelos (EUA)	National Soil Information System
<b>NCSS</b>	Cooperativa Nacional de Reconocimiento de Suelos (EUA)	National Cooperative Soil Survey
<b>NDVI</b>	Índice de Vegetación de Diferencia Normalizada	Normalized Difference Vegetation Index
<b>NIR</b>	Infrarrojo cercano	Near infrared
<b>NORTH</b>	Septentrionalidad	Northernness
<b>NORTHSP</b>	Interacción entre NORTH y SLOPE	Interaction between NORTH and SLOPE
<b>NRCS</b>	Servicio de Conservación de Recursos Naturales (EUA)	Natural Resources Conservation Service
<b>OA</b>	Precisión Global	Overall Accuracy
<b>OIV</b>	Organización Internacional de la Viña y el Vino	International Organisation of Vine and Wine
<b>OLA</b>	Área Óptima Legible	Optimum Legible Area
<b>OLD</b>	Delineación Óptima Legible	Optimum Legible Delineation
<b>OLI</b>	Operational Land Imager, Landsat	Operational Land Imager, Landsat
<b>OOB</b>	(Error) fuera de la bolsa	Out of Bag (error)
<b>PCA</b>	Análisis de Componentes Principales	Principal Component Analysis
<b>R<sup>2</sup></b>	Coefficiente de determinación	Coefficient of determination
<b>RDIFUS</b>	Radiación solar entrante difusa potencial	Diffuse Potential Incoming Solar Radiation
<b>RDIRECT</b>	Radiación solar entrante directa potencial	Direct Potential Incoming Solar Radiation
<b>REVI</b>	Registro Vitícola	
<b>RF</b>	Bosque aleatorio	Random Forest
<b>RI</b>	Índice de intensidad del rojo	Redness Index
<b>RMSE</b>	Raíz del error cuadrado medio	Root Mean Square Error
<b>SAGA</b>		System for Automated Geoscientific Analyzes
<b>SAVI</b>	Índice de vegetación ajustado al suelo	Soil Adjusted Vegetation Index

<b>SCA</b>	Área de captación estándar		Standard Catchment Area
<b>SI</b>	Índice de Salinidad		Salinity Index
<b>SIG</b>	Sistemas de Información Geográfica	<b>GIS</b>	Geographic Information System
<b>SIGPAC</b>	Sistema de Información Geográfica de Parcelas Agrícolas		
<b>SLOPE</b>	Pendiente		Slope
<b>SMU</b>	Unidad Cartográfica de Suelos		Soil Map Unit
<b>SOTER</b>	Programa para base de datos de suelos y del terreno		Soil and Terrain database programme
<b>SPI</b>	Índice de potencia de la corriente		Stream Power Index
<b>SSCA</b>	Área de captación modificada de SAGA		Modified Catchment Area by SAGA
<b>ST</b>			Soil Taxonomy
<b>STU</b>	Unidad Taxonómica de Suelo		Soil Taxonomy Unit
<b>STWI</b>	Índice Topográfico de Humedad de SAGA		Topographic Wetness Index by SAGA
<b>SVM</b>	Máquinas de vectores de soporte		Support Vector Machines
<b>SWIR</b>	Infrarrojo de onda corta		Short-wave infrared
<b>TCILOW</b>	Índice de clasificación del terreno para las tierras bajas		Terrain Classification Index for Lowlands
<b>TM</b>	Monitor temático, Landsat		Thematic Mapper, Landsat
<b>TPI</b>	Índice de Posición Topográfica		Topographic Position Index
<b>TRASP</b>	Índice de radiación solar - orientación		Solar-radiation Aspect Index
<b>TRI</b>	Índice de rugosidad del terreno		Terrain Ruggedness Index
<b>TWI</b>	Índice de Humedad Topográfica		Topographic Wetness Index
<b>UHM</b>	Unidad Homogénea del Medio	<b>HTU</b>	Homogeneous Terroir Unit
<b>UPSLOPE</b>	Pendiente en el área de captación modificada de SAGA		Modified Catchment Slope from SAGA TWI
<b>USDA</b>	Departamento de Agricultura de EUA		United States Department of Agriculture
<b>USGS</b>	Servicio Geológico de EUA		United States Geological Survey
<b>VIS</b>	Espectro visible		Visible Spectrum
<b>WOS</b>			Web of Science
<b>WRB</b>	Base referencial mundial del recurso suelo		World Reference Base for Soil Resources
<b>ZIT</b>	Zonificación Integrada del Terroir	<b>ITZ</b>	Integrated Terroir Zoning

# ÍNDICE DE FIGURAS

<i>Figura 1-1. Diagrama de las relaciones entre los elementos del detalle cartográfico en un mapa.</i>	16
<i>Figura 1-2. Abstracción de métodos de muestreo aleatorios: simple, estratificado simple y estratificado por fases.</i>	27
<i>Figura 1-3. Diagrama de líneas de acción en la Cartografía de Suelos Digital para clases de suelos en función de la procedencia de la información de los suelos y de su cobertura de la región del proyecto.</i>	30
<i>Figura 1-4. Ejemplo teórico de matriz de errores para una clasificación con clases de suelos.</i>	31
<i>Figure 2-1. Distribution of published articles with disaggregation projects by year (bars) and cumulative articles over years (line).</i>	41
<i>Figure 2-2. Flowchart of the disaggregation of a conventional soil map within the Digital Soil Mapping framework. The actions are organised into four sections: soil information, environmental data and covariates, classification, and evaluation. The primary route leverages the CSM map as the main data source for the classification and it is plotted in solid line. The secondary one takes the soil information from point observations while uses the map as a covariate and it is drawn in dotted line.</i>	44
<i>Figure 2-3. Location of the reviewed disaggregation applications by countries.</i>	45
<i>Figure 2-4. 10 most important covariates among reviewed disaggregation applications.</i>	51
<i>Figure 2-5. Distribution of the Overall Accuracy achieved in the reviewed disaggregation applications.</i>	64
<i>Figure 3-1. Location of the Designation of Origin Campo de Borja in NE Spain (left) and distribution of the selected ten soil map units within the conventional soil resource inventory map (right).</i>	74
<i>Figure 3-2. Spearman's correlation matrix of the covariates studied within the project. The four main associations.</i>	84
<i>Figure 3-3. Soil map units from the conventional soil resource inventory map (left), new divisions from the collection C1 (middle) and C2 (right). From top to bottom, the maps correspond to the soil map units 17C, 23C and 29C.</i>	88
<i>Figura 4-1. Mapa del inventario del recurso suelo convencional de la Denominación de Origen Campo de Borja y sus 46 unidades cartográficas de suelos sin fases.</i>	96

- Figura 4-2. Localización de las observaciones (perfiles de suelos y excluyendo sondeos) del inventario del recurso suelo convencional fijadas mediante muestreo dirigido en la Denominación de Origen Campo de Borja. \_\_\_\_\_ 97
- Figura 4-3. Diagrama de flujo de la metodología de disgregación mediante la división de las unidades cartográficas de suelos convencionales, la correlación de los grupos con las unidades taxonómicas de suelos de origen y el refinamiento de las delineaciones. \_\_\_\_\_ 101
- Figura 4-4. Unidades cartográficas de suelos con fases (un color graduado) del mapa de suelos convencional (izquierda); nuevas unidades cartográficas del mapa de suelos disgregado sin refinamiento, base (centro); y disgregado con refinamiento, final (derecha). De arriba a abajo, los mapas corresponden a las regiones de las unidades cartográficas de suelos originales 17C, 23C y 29C. \_\_\_\_\_ 105
- Figura 4-5. Gráfico de dispersión entre el número de delineaciones frente al Área Mínima Legible (área mínima de delineación en el mapa) que se aproxima a un límite asintótico cuando se alcanzan los valores de superficies mayores. \_\_\_\_\_ 107
- Figura 4-6. Gráfico de dispersión entre el Índice Máximo de Reducción frente al Área Mínima Legible (área mínima de delineación en el mapa) que se aproxima a un límite asintótico para los niveles más grandes de superficies. \_\_\_\_\_ 107
- Figura 4-7. Mapa de las 143 unidades cartográficas de suelos disgregadas y refinadas, 74 monotáxicas (el identificador comienza por M) y 69 politáxicas (por P). \_\_\_\_\_ 108
- Figura 5-1. Mapa de zonificación climática de la Denominación de Origen Campo de Borja (ver características de las zonas en la tabla Tabla 5-1). \_\_\_\_\_ 118
- Figura 5-2. Mapa de unidades litoestratigráficas para el inventario del recurso suelo convencional de la Denominación de Origen Campo de Borja. (ver descripción de las unidades litoestratigráficas en la tabla Tabla 5-2). \_\_\_\_\_ 120
- Figura 5-3. Distribución de los recintos de viñedos con datos tomados del Registro Vitícola y del Sistema de Información Geográfica de Parcelas Agrícolas de la DO Campo de Borja. \_\_\_\_\_ 123
- Figura 5-4. Diagrama de flujo de la Zonificación Integrada del Terroir vitícola y su evaluación en línea continua. Vías suplementarias de información y de comparación entre zonificaciones en línea discontinua. \_\_\_\_\_ 124
- Figura 5-5. Distribución del Índice de Calidad en la región del proyecto según las 5 clases definidas (eje principal) y la superficie abarcada por cada clase (eje secundario). \_\_\_\_\_ 129
- Figura 5-6. Representación cartográfica de los flujos desde la matriz de confusión entre las Unidades Homogéneas del Medio categorizadas según el Índice de Calidad. Por cada categoría de la



*nueva zonificación (tonalidades), se simboliza la categoría origen en el mapa original (intensidades). Se representa en intensidades más fuertes aquellas áreas que mantienen la categoría y en intensidades más ligeras las que han variado.*\_\_\_\_\_ 131

*Figura 5-7. Distribución del Índice de Ocupación en la región del proyecto según las 5 clases definidas (eje principal) y la superficie abarcada por cada clase (eje secundario).*\_\_\_\_\_ 133



## ÍNDICE DE TABLAS

<i>Tabla 1-1. Niveles de reconocimiento de suelos, adaptados desde varias fuentes.</i>	10
<i>Table 2-1. Basic cartographic, data sources and covariates information of the selected disaggregation projects.</i>	42
<i>Table 2-2. Classification and evaluation methods for the selected disaggregation projects.</i>	55
<i>Table 3-1. Main characteristics of the selected Soil Map Units from the conventional Soil Resource Inventory map of the Designation of Origin Campo de Borja.</i>	76
<i>Table 3-2. Soil Taxonomy Units content of the selected Soil Map Units from the conventional Soil Resource Inventory map of the Designation of Origin Campo de Borja.</i>	76
<i>Table 3-3. Symbol, name, and a brief description of the morphometry covariates of the project associated with the SCORPAN R factor.</i>	79
<i>Table 3-4. Symbol, name, and a brief description of the hydrology covariates of the project associated with the SCORPAN R factor.</i>	80
<i>Table 3-5. Symbol, name, and a brief description of the topo-climate covariates of the project associated with the SCORPAN C factor.</i>	80
<i>Table 3-6. Symbol, name, and a brief description of the remote sensing covariates of the project associated with the SCORPAN S factor.</i>	80
<i>Table 3-7. Collections of covariates selected in the analysis.</i>	86
<i>Table 3-8. Number of new divisions of each Soil Map Unit using the two collections of covariates.</i>	87
<i>Tabla 4-1. Distribución de las observaciones (perfiles de suelos) por Subgrupos de «Soil Taxonomy» y la superficie estimada para cada uno calculada según el porcentaje de las Unidades Taxonómicas de Suelo correspondientes en cada Unidad Cartográfica de Suelos.</i>	98
<i>Tabla 4-2. Colección de covariables usadas en la disgregación con su descripción y el factor SCORPAN al que se asocian.</i>	99
<i>Tabla 4-3. Características del detalle cartográfico de los mapas que intervienen en la disgregación.</i>	106
<i>Tabla 4-4. Valores de los índices de precisión categórica Precisión Global y Coeficiente de Kappa bajo todos los supuestos y para los tres mapas que intervienen en la disgregación: Final: mapa disgregado y refinado; Base: mapa disgregado sin refinamiento; y CSM: mapa de suelos convencional.</i>	109

<i>Tabla 5-1. Parámetros climáticos e índice bioclimáticos característicos de las zonas climáticas de la zonificación del clima.</i>	<i>119</i>
<i>Tabla 5-2. Unidades litoestratigráficas del inventario del recurso suelo convencional de la Denominación de Origen Campo de Borja.</i>	<i>120</i>
<i>Tabla 5-3. Propiedades y parámetros considerados en el Índice de Calidad, la zonificación parcial origen de los datos y su nivel de evaluación.</i>	<i>126</i>
<i>Tabla 5-4. Características cartográficas de la zonificación de referencia y la nueva zonificación.</i>	<i>128</i>
<i>Tabla 5-5. Matriz de confusión de la intersección entre las Unidades Homogéneas del Medio categorizadas según el Índice de Calidad del mapa de zonificación original (columnas) y nueva zonificación (filas), y sus parámetros globales.</i>	<i>130</i>
<i>Tabla 5-6. Modelo de Índice de Ocupación a partir de Índices de Calidad parciales de los parámetros considerados en la zonificación del terroir por regresión lineal con selección de variables en etapas por adición según el Criterio de Información de Akaike y con validación cruzada.</i>	<i>134</i>

## SUMMARY

In the current scenario of over-exploitation of natural resources and the growing need for their sustainable management, detailed soil maps are becoming increasingly necessary. Conventional soil mapping methods have shortcomings to meet this demand, yet they have been surpassed by new methodologies that take advantage of digitised environmental data and quantitative statistical techniques, which have been encompassed under a common framework referred to as "Digital Soil Mapping".

Among its lines of action is the "disaggregation" of polytaxic soil mapping units from conventional maps by disclosing the non-explicit soil distribution pattern, which entails a new cartography of higher detail. To this end, several strategies have been followed concerning the available data and classification techniques. The analysis of their applications revealed some limitations were detected in terms of precision and requirements.

The present study develops a new disaggregation methodology geared towards regions where the potential location of soil taxonomic units within map units is unclear. It relies on the detection of soil homogeneous areas with no prior definition using the unsupervised CLARA classification technique with Mahalanobis distance within a space of selected covariates and an expert knowledge-driven correlation of the areas with the taxonomic units.

The implementation on the 1:25,000 scale conventional soil map of the Designation of Origin Campo de Borja yielded a new 1:10,000 scale map with a larger number of delineations. The new map was subsequently used to elaborate a new integrated terroir zoning of the region, resulting in an increase of the number of homogeneous environmental units detected and a more detailed definition of them. These results provide a new ground for updating soil resource inventories with higher detail, and, in the specific application to the viticultural terroir, for developing further research to relate its features with the properties of products from them.



## CONTEXTUALIZACIÓN

El interés por conocer las características y la distribución de los suelos ha estado presente a lo largo del desarrollo contemporáneo. Los avances más relevantes en este campo han sido alcanzados mediante pautas científico-técnicas y se promovieron en momentos históricos marcados por las consecuencias de condiciones climáticas adversas (Rusia, sequía de finales del S.XIX) y desastres ambientales (Estados Unidos de América, *Dust Bowl* en el S. XX). Estos hechos tuvieron un fuerte impacto en las sociedades de la época y dieron lugar a desabastecimientos de alimentos, grandes movimientos migratorios o conflictos sociopolíticos.

En la actualidad, se está produciendo una explotación intensiva de los recursos edáficos, con exigencias de mayor productividad, unida a una situación de degradación prolongada y un cambio climático acelerado por la acción antrópica. En este contexto, existe una creciente demanda de información detallada de los suelos que permita alcanzar una gestión, manejo y monitorización optimizados y sostenibles desde ámbitos tan dispares como la agricultura o la ingeniería.

En el campo de la viticultura el objetivo de manejo óptimo y de precisión está vinculado íntimamente a la noción de «terroir» y su zonificación, esto es, la definición de zonas relativamente homogéneas del conjunto de factores del medio con influencia en el desarrollo de la vid, entre los que se incluye el suelo.

Dado el carácter del suelo como integrador de elementos concernientes a múltiples disciplinas, su concepción y estudio puede producirse desde diferentes aproximaciones.

Desde una perspectiva descriptiva los suelos se definen como cuerpos naturales compuestos por sólidos minerales y orgánicos, líquidos, y gases situados en la superficie terrestre, entre la litología y la atmósfera, y que, además, cuentan con la capacidad de soportar el desarrollo vegetal (biosfera) en un ambiente natural.

Desde una perspectiva genética es posible considerar los suelos por sus causas, como el producto de la interacción de unos factores activos (clima y organismos) sobre ciertos factores pasivos (relieve y roca) durante un tiempo determinado; y por los efectos producidos por estas, como el resultado de procesos de adiciones, transformaciones, transferencias y pérdidas de energía o materia.

En este trabajo se relacionan las consideraciones anteriores y usamos el suelo a través del resultado de las causas (factores del medio) y sus efectos (procesos de formación) que dan lugar a su conformación en capas de disposición sensiblemente horizontal o subparalela a la superficie del terreno (horizontes). Estos horizontes son distinguibles del material original y se diferencian entre sí por sus propiedades y características. El conjunto de horizontes de un suelo constituye su perfil y su importancia radica en que se establece frecuentemente como la unidad de estudio.

El resultado inmediato de esta diferenciación de horizontes es la variabilidad de suelos, tanto vertical o en profundidad (tipología de suelos), como horizontal o en superficie (geografía de suelos). La representación de la variabilidad en ambos sentidos en mapas constituye el objeto central del presente documento.

Esta representación se enmarca habitualmente dentro de proyectos de «Inventarios del Recurso Suelo» como colofón del «Reconocimiento de Suelos». Tradicionalmente, esta representación se ha llevado a cabo mediante medios eminentemente analógicos, calificados como «Cartografía de Suelos Convencional». Sin embargo, han sido superados recientemente con la llegada de nuevos métodos que aprovechan datos digitalizados y métodos estadísticos cuantitativos y que han sido englobados bajo un marco común denominado «Cartografía de Suelos Digital». Un estudio de la evolución desde la Cartografía de Suelos Convencional a la Digital y su contextualización se presenta en el Capítulo 1 de Antecedentes.

La Cartografía de Suelos Digital reúne distintas líneas de acción según el tipo de información de suelos empleada y el alcance de sus objetivos. Una de ellas es la «disgregación», que tiene como objetivo la generación de una nueva cartografía al revelar el patrón de distribución de los suelos no explícito dentro de las unidades cartográficas «politáxicas» o multi-componentes de los mapas de suelos convencionales. En última instancia, se persigue aprovechar al máximo los recursos y la información disponible ambiental y de suelos para alcanzar una representación de mayor detalle y más fidedigna de forma más eficiente y con ello dar respuesta a las demandas actuales. En el Capítulo 2 se aborda una revisión exhaustiva de las metodologías de disgregación publicadas en bases de datos científicas de referencia.

A partir del conocimiento extraído desde la revisión, en el Capítulo 3 se introduce una metodología novedosa de división de unidades cartográficas politáxicas mediante una técnica de clasificación no supervisada aplicada sobre datos digitales de parámetros ambientales, favorable en contextos con escaso conocimiento de la región del proyecto. Dicha metodología acomete también la



definición de un subconjunto de parámetros ambientales o covariables relevantes durante el proceso.

Sobre este sistema de división de unidades cartográficas politáxicas en el Capítulo 1 se desarrolla una metodología completa de disgregación, en la que las divisiones obtenidas son correlacionadas con clases de suelos. Esta metodología se aplica sobre el mapa de suelos convencional (1:25.000) de la Denominación de Origen Campo de Borja y da lugar a un nuevo mapa disgregado de mayor escala (1:10.000), que es finalmente validado.

El Capítulo 5 aborda el uso potencial del mapa de suelo disgregado con unidades en la «Zonificación Integrada del Terroir» para generar «Unidades Homogéneas del Medio» de mayor detalle frente a la zonificación proveniente del mapa de suelos convencional, beneficiándose de la ganancia de escala y la definición mayoritaria de unidades monotáxicas alcanzadas en la disgregación. La calidad de las nuevas Unidades Homogéneas del Medio es evaluada y contrastada con la señalada en el mapa de zonificación de referencia para entender el efecto del cambio introducido. Las nuevas Unidades Homogéneas del Medio abren la posibilidad de desarrollar una mejor gestión del recurso suelo en viñedos, en línea con la viticultura de precisión.



## OBJETIVOS

La presente tesis doctoral tiene como objetivos:

1. El desarrollo de una metodología novedosa de disgregación de mapas de suelos convencionales con el que aumentar su escala y detalle, que aproveche los recursos de información de suelos disponibles.
2. La implementación de la metodología de disgregación al mapa de suelos convencional de la Denominación de Origen Campo de Borja y su validación.
3. La aplicación del mapa disgregado obtenido como Inventario del Recurso Suelo en la Zonificación Integrada del Terroir vitícola de la Denominación de Origen Campo de Borja.
4. El contraste entre la nueva zonificación integrada del terroir desde el mapa disgregado de suelos frente a la zonificación de referencia y, con ello, la verificación del proceso de zonificación y su mejora.



# **CAPÍTULO 1 DE LA CARTOGRAFÍA DE SUELOS**

## **CONVENCIONAL A LA DIGITAL: ANTECEDENTES**

### **1 Inventarios del Recurso Suelo**

Los «Inventarios del Recurso Suelo» (IRS; en inglés *Soil Resources Inventories* [SRI]) son estudios que describen y cuantifican las características de los suelos, los clasifican según un sistema establecido, trazan sus límites geográficos, los organizan en unidades diferenciables y homogéneas según el sistema elegido, y además establecen predicciones sobre su comportamiento para usos específicos. De esta forma, los IRS abordan el análisis de la tipología de los suelos (variabilidad vertical) y de su distribución (variabilidad horizontal), así como de la correlación entre ambos aspectos, de la que resulta la cartografía de suelos (mapa de suelos). El producto final de los IRS incluye el mapa y una memoria donde se recoge la descripción e interpretación referidas a las diferentes unidades que constituyen la leyenda del mapa de suelos y sus conclusiones (Gómez Miguel, 2006).

Este tipo de estudio tiene dos objetivos que se complementan: uno científico y otro técnico. El primero está orientado a sintetizar el conocimiento sobre las relaciones entre los suelos y el resto de los componentes del medio ambiente. Por otro lado, su finalidad práctica busca incrementar la riqueza particular o el beneficio general de la sociedad a través de un mejor manejo y gestión de los suelos, condicionado por las demandas y limitaciones de las partes interesadas (Rossiter, 2000).

### **2 Cartografía de Suelos Convencional**

Tradicionalmente, la información recabada en los IRS se ha expresado a través de mapas categóricos, los cuales dividen el territorio en delineaciones con límites precisos y adscritas a clases recogidas en la leyenda del mapa, y que, en última instancia, estratifican el medio (Rossiter, 2000).

Su ejecución puede ser descrita en seis fases sucesivas: i) la definición de la escala en relación con los objetivos previstos y los recursos disponibles; ii) la delimitación de áreas por la semejanza de

sus factores y su subsiguiente definición; iii) la recopilación de datos sobre el terreno siguiendo un muestreo diseñado a partir de las delineaciones definidas en la segunda fase, acorde con la escala, los objetivos y los recursos fijados en la fase primera, y la posterior realización, descripción y el análisis de las observaciones definidas; iv) la clasificación de estas observaciones bajo un sistema establecido y la creación de «unidades taxonómicas de suelos» (en inglés *Soil Taxonomy Units* [STU]) adecuadas a la escala; v) la correlación entre delineaciones y observaciones y la consiguiente generación de «unidades cartográficas de suelos» (en inglés *Soil Map Units* [SMU]); y, finalmente, vi) la presentación de resultados, a través de los mapas y la memoria asociada.

Esta línea metodológica para la elaboración de mapas de suelos es referida como «Cartografía de Suelos Convencional» (en inglés *Conventional Soil Mapping* [CSM]).

## 2.1 Objetivos, recursos y escalas

En un mapa, a la relación existente entre lo representado y la realidad se le denomina escala y lleva asociada la definición de la representación. Este parámetro determina así tanto el detalle cartográfico, la precisión del ajuste entre las delineaciones, como el detalle categórico, el grado de la información que se muestra.

La finalidad con la que se ejecuta el IRS y la disponibilidad de recursos para realizarlo condicionan el alcance del estudio y el nivel de detalle de la representación necesarios. En concreto, las escalas más pequeñas, inferiores a 1:250.000, dan capacidad para realizar planificaciones estatales y, a su vez, permiten identificar áreas de interés sobre las que intensificar otros estudios. Las escalas superiores a 1:250.000 y hasta el entorno de 1:50.000 aportan la información adecuada para el uso y gestión de tierras de forma generalizada por parte de organismos regionales; y un incremento hasta 1:20.000 abre la posibilidad a la planificación de comunidades y asentamientos. Escalas mayores de 1:20.000 son las destinadas al desarrollo y al manejo de áreas de interés agrícola; y, cuando superan 1:12.000, específicamente para proyectos ejecutivos en explotaciones, fincas y parcelas. Los límites de estos intervalos pueden variar sensiblemente según las fuentes.

Considerando las propiedades de los rangos descritos, se han definido niveles de reconocimiento de suelos que buscan facilitar la transferencia de la interpretación y precisión de la información de estos. Por ejemplo, la «Cooperativa Nacional de Reconocimiento de Suelos» (en inglés *National Cooperative Soil Survey* [NCSS]) liderada por el «Servicio de Conservación de Recursos Naturales» (en inglés *Natural Resources Conservation Service* [NRCS]) del «Departamento de Agricultura de

EUA» (en inglés *United States Department of Agriculture* [USDA]) los denomina órdenes y establece 5 niveles (Soil Survey Division Staff, 2017); descritos también en términos de intensidad o tipos de reconocimientos en 6 niveles (Gómez Miguel, 2006; Rossiter, 2000). A cada uno de estos niveles le corresponden diferentes ajustes de los elementos que definen el IRS (Tabla 1-1).

Estos elementos, que serán introducidos en las secciones sucesivas, son interdependientes y están sujetos a la escala específica fijada dentro de los rangos señalados. En este sentido, la variación de la escala tiene una influencia directa en las características, con un impacto directo en el presupuesto final del IRS, como el número total de observaciones, por ejemplo, siendo indispensable alcanzar un equilibrio entre el detalle alcanzado, la inversión y el tiempo necesarios para el desarrollo del proyecto.

## 2.2 Delineaciones

El estudio de los suelos cuenta con la limitación de que estos no pueden ser observados la mayoría de las veces directamente, como un conjunto continuo, sino a través observaciones discretas y aisladas, en forma de calicatas y/o sondeos, y que además resultan destructivas.

Es por ello que convencionalmente se acude al modelo «CLORPT», que define en términos cualitativos el suelo en función de los denominados factores formadores: el clima (CL), los organismos (O), el relieve (R), el material parental (P), y el tiempo (T). Fue descrito por Jenny (1941) para desarrollar un modelo conceptual que pretende establecer una relación entre los suelos y su medio; de forma que en aquellas zonas cuyos factores y sus interacciones sean semejantes se espera que se desarrolle la misma tipología de suelos.

El cartógrafo atiende a la evaluación de estos factores formadores para completar el trazado de sus límites (delineaciones). En última instancia, esto da lugar a la agrupación de aquellas delineaciones con un contenido similar y la separación de aquellas más diferentes. En definitiva, da lugar a la creación del precedentes de las unidades cartográficas del mapa (Gómez Miguel, 2006).

Tabla 1-1. Niveles de reconocimiento de suelos, adaptados desde varias fuentes.

Órdenes de reconocimiento de suelos NCSS <sup>1</sup>	Nivel de reconocimiento <sup>2</sup>	Objetivos	Densidad de observaciones, ½ obs/cm <sup>2</sup> de mapa (nº obs/km <sup>2</sup> )	MLA, ¼ cm <sup>2</sup> mapa (ha)	Tipo de STU ST, NRCS <sup>3</sup>	Tipo de SMU NRCS <sup>1</sup>	Rango de escalas óptimas
5º	Muy extensivo. Exploratorio	Planificación estatal y regional	Información indirecta complementada con un mínimo de un perfil por STU	250 - 4.000	Fases de grupos o taxa inferiores a la serie	Asociaciones; pocos complejos y grupos indiferenciados	1:250.000 - 1.000.000
5º	Muy extensivo. Generalizado	Planificación estatal y regional	0,05 - 0,006	250 - 2.250	Fases de grupos o taxa inferiores a la serie	Asociaciones; pocos complejos y grupos indiferenciados	1:250-000 - 750.000
4º	Extensivo. Reconocimiento	Uso y gestión de tierras de forma generalizada	1,25 - 0,05	10 - 250	Fases de subgrupos o taxa inferiores a la serie	Mayoría de asociaciones; algunos complejos, consociaciones y grupos indiferenciados	1:50.000 - 1:250.000
3º	Moderadamente intensivo. Semi-detallado	Planificación comunitaria y asentamientos	8 - 1,25	1,6 - 10	Fases de familias o taxa inferiores a la serie	Mayoría de asociaciones y complejos; algunas consociaciones y grupos indiferenciados	1:20.000 - 1:50.000
2º	Intensivo. Detallado	Uso y gestión de tierras; desarrollo agrario; planificación urbana	22 - 5	0,58 - 2,5	Fases de series de suelos; pocos componentes con taxa inferiores a la serie.	Consociaciones, complejos; pocas asociaciones y grupos indiferenciados	1:12.000 - 1:25.000
1º	Muy intensivo. Detallado	Parcelas, edificaciones individuales, revisiones y permisos requeridos por agencias reguladoras	125 - 14	0,1 - 0,9	Fases de series de suelos	Mayoría de consociaciones; algunos complejos	1:5.000 - 1:15.000
1º	Muy intensivo. Muy detallado	Parcelas experimentales; trabajos detallados de ingeniería	125	0,1	Fases de series de suelos	Mayoría de consociaciones; algunos complejos	> 1:5.000

<sup>1</sup> (Soil Survey Division Staff, 2017) ; <sup>2</sup> (Gómez Miguel, 2006; Rossiter, 2000);

<sup>3</sup> (Soil Survey Division Staff, 2014).

MLA: Área Mínima Legible; NCSS: Cooperativa Nacional de Reconocimiento de Suelos (EUA); SMU: Unidad cartográfica de suelos; ST: Soil Taxonomy; STU: Unidad taxonómica de suelos.



Para este fin se han utilizado habitualmente ortofotos, la «fotointerpretación aérea» (FIA; en inglés *Aerial Photo Interpretation* [API]) y el método directo. Aunque existe cierto solapamiento entre ellas, generalmente se suele acudir a la valoración de ortofotos para los reconocimientos a escalas pequeñas y muy pequeñas (Orden 5º y 4º), la FIA para aquellos a escalas pequeñas, medias y grandes (Orden 4º, 3º y 2º) y, por último, la división por observaciones en el método directo para las escalas muy grandes (Orden 1) (Tabla 1-1).

El proceso de delimitación de las delineaciones es laborioso y requiere de la dedicación de personal especializado durante periodos de tiempo prolongados. Esto supone un factor limitante y un consumo elevado de recursos, muy superior en la aplicación de método directo que en la FIA y de esta respecto a la valoración de ortofotos. Además, el modelo conceptual y el resultado dependen del especialista, esto es, de su conocimiento experto del entorno así como de su experiencia previa y su habilidad (Rossiter, 2000).

## 2.3 Muestreo

Durante el muestreo se recopilan de datos del terreno para la identificación, descripción y cuantificación de las diferentes tipologías de suelos. Los muestreos se caracterizan por dos componentes: la definición del número o densidad y el tipo de observaciones (relacionado con la variabilidad vertical), y el diseño de la distribución y localización de estas observaciones (relacionado con la variabilidad horizontal).

La «densidad de observaciones» (en inglés *inspection density*) en un IRS es un parámetro sobre el que no se ha establecido un convenio entre técnicos y científicos y los diferentes valores utilizados han sido justificados por la heterogeneidad edáfica del área de estudio, por la disponibilidad de recursos e incluso por la experiencia del cartógrafo (Gómez Miguel, 2006). Este parámetro regula el número de observaciones en relación con la escala del IRS, tomando habitualmente valores que oscilan entre 4 (Vink, 1963), 2,5 y  $\frac{1}{4}$  (Soil Survey Division Staff, 2017) y 1 y  $\frac{1}{2}$  (OIV, 2012) observaciones por cm<sup>2</sup> de mapa.

El número de observaciones obtenido se distribuye entre dos tipos: calicatas y sondeos. Las calicatas son aperturas del terreno lo suficientemente amplias como para que el edafólogo pueda acceder a ellas con facilidad. De esta forma es posible llevar a cabo la descripción completa del perfil y la toma de muestras, idealmente de cada horizonte, si bien en situaciones con limitación de recursos se reduce a horizontes específicos. Su principal fin es la identificación y descripción

de las distintas tipologías de suelo presentes. Por su parte, los sondeos son perforaciones efectuadas con barrenas, donde el acceso al perfil se restringe al contenido de las extracciones y de los que con frecuencia se muestrean los materiales superficiales. Por ello, su descripción también es limitada, pero en base a la información disponible de las calicatas es posible realizar su contraste. Se utilizan para cuantificar la ocupación superficial en las unidades cartográficas provisionales formadas por las delineaciones (Schoeneberger et al., 2012; Soil Survey Division Staff, 2017).

El reparto de observaciones entre calitas y sondeos responde a una relación también condicionada a la escala y que oscila entre un mínimo 28:100 respectivamente en escalas muy detalladas y hasta un máximo de 2:1 a escala 1:250.000 en IRS de reconocimiento. Esto se debe a que en este tipo de IRS la heterogeneidad de los suelos a organizar puede ser acotada en gran parte a través de la FIA, lo que permite reducir significativamente las observaciones destinadas a la cuantificación frente a las escalas más detalladas (Nieves Bernabé et al., 1985). En el reparto además influye la eficiencia edafológica, en donde se considera la experiencia del equipo, el tipo de paisaje o la época del año.

Las observaciones son completamente descritas y muestreadas. Las muestras recogidas son analizadas en el laboratorio, donde se les realizan las determinaciones físicas, químicas y fisicoquímicas que se consideren necesarias en función de los objetivos del estudio. Estas determinaciones deben realizarse de acuerdo con metodologías homologables (Burt y Soil Survey Division Staff, 2014; MAPA, 1986)

Por su parte, la distribución y localización de las observaciones son dirigidas por el criterio del cartógrafo. Para ello, considera la cualidad de las unidades cartográficas provisionales del mapa y busca la representación de todas mediante una serie de repeticiones limitada, aunque superior a tres.

Existen modelos para la organización de las series de repeticiones ubicándolas por líneas o transectos, bandas, sitios o áreas de muestreo; con preferencia de unos dependiendo de la escala. Alternativamente, se acude al método directo o una rejilla fija (en inglés *fixed grid*) en la que la distancia entre los vértices varía con la escala y que se utiliza sólo sobre determinadas unidades complejas sobre todo el área de muestreo.

## 2.4 Clasificación

La definición del contenido de las unidades cartográficas provisionales en términos de tipología de suelos se basa en la clasificación de todas las observaciones, a partir de los datos procedentes de las descripciones y de las analíticas. Esta clasificación se realiza acorde con un sistema previamente establecido.

Con este fin se han desarrollado múltiples sistemas de clasificación que pueden caracterizarse según la importancia que otorgan a la génesis del suelo en su estructuración, al ámbito geográfico al que han sido adaptados, a la objetividad y morfometría de sus criterios de división o a la especificidad de su propósito (Ségalen y Van Diepen, 1984). En términos generales, estos sistemas son estructurados en niveles o categorías, con grados de abstracción semejante, y organizados en clases sintéticas, que agrupan los suelos en características similares.

Aquellos sistemas que cuentan con mayor cobertura temática, son adecuadas para fines agronómicos y tienen con una difusión global son *Soil Taxonomy* (ST) (Soil Survey Division Staff, 1999), impulsada por NCSS y la agencia NRCS-USDA; y «Base referencial mundial del recurso suelo» (en inglés *World Reference Base for Soil Resources* [WRB]) (IUSS Working Group WRB, 2014), promovida por la «Organización de las Naciones Unidas para la Alimentación y la Agricultura» (en inglés *Food and Agriculture Organization* [FAO]) y la «Unión Internacional de Ciencia del Suelo» (en inglés *International Union of Soil Sciences* [IUSS]). Además, muchos de los sistemas regionales desarrollados con posterioridad se han basado sobre ellos.

Explícitamente, ambos sistemas han sido desarrollados con el fin de apoyar IRS, por lo que la separación entre sus clases generalmente tiene un sentido interpretativo y estas resultan idóneas para calificar los componentes de la leyenda del mapa, las denominadas STU. En última término, la nomenclatura no explicita las características, sino que estas se encuentran implícitas en la clase (Rossiter, 2000).

Así y todo, en ocasiones resulta interesante incorporar características no consideradas en el nivel taxonómico utilizado en las STU, pero que son importantes para describir el comportamiento o el manejo del suelo y que pueden ser cartografiadas de forma consistente. Esto se consigue a través de las fases, que son grupos funcionales fuera de la clasificación taxonómica que se adjuntan a cualquier nivel taxonómico para mejorar la descripción de la SMU y, por lo tanto, añaden información e incrementan la utilidad del IRS.

La elección entre un sistema u otro está condicionada por su estructura y las características asociadas a las clases, así como a la escala objetivo del IRS. En este sentido, ST es el único sistema válido a escalas grandes y muy grandes por acoger la mayor diversidad de suelos en la categoría más baja, la serie, que resulta necesaria para la agricultura ejecutiva.

## 2.5 Correlación

La correlación es un proceso iterativo en el que se relacionan las SMU provisionales (obtenidas por agrupación de delineaciones con factores similares) con su contenido en términos de las STU creadas (procedente de las observaciones realizadas) para generar las SMU definitivas y con ello la leyenda del mapa. En última instancia, resulta en la calificación y cuantificación del contenido de las SMU en términos de STU.

Las SMU definitivas pueden estar formadas por uno o más componentes o términos, es decir, pueden conformarse por una o más STU al nivel categórico fijado en función de la escala del IRS y ligado también a la complejidad e intrincamiento del medio, la densidad de observaciones realizada y a la propia experiencia del edafólogo. En este sentido, se distinguen SMU monotáxicas, cuando tienen un solo término, y politáxicas, con varios.

Las guías para los IRS de las diferentes escuelas han acuñado términos más específicos para describir las relaciones posibles entre las STU de una SMU. El *Soil Survey Manual* (Soil Survey Division Staff, 2017), referido principalmente a ST, distingue entre suelos similares y disímiles, en función de que permitan un mismo manejo o diferenciado respectivamente; y, a su vez, entre mayoritarias, dominantes e inclusiones. Las inclusiones son categorías fuera de las STU mayoritarias y de su gama de propiedades que están identificadas de forma reducida o donde su delimitación no es consistente (Van Wambeke y Forbes, 1986).

Con todos estos elementos se construyen los diferentes tipos de SMU: consociaciones, asociaciones, complejos, y grupos indiferenciados, ordenados de mayor a menor homogeneidad en sus componentes. Las primeras se definen cuando existe una categoría dominante complementada con otros suelos similares minoritarias organizadas uniformemente en todas las delineaciones. En presencia de categorías disímiles, se describen asociaciones si es posible identificar su patrón de distribución, pero no puede resolverse en el mapa debido al uso de una escala reducida; mientras que, en el caso de patrones intrincados difícilmente separables a mayor escala, se denominan complejos. En última instancia, si estos suelos disímiles no están asociados

geográficamente de manera consistente y, por lo tanto, no siempre aparecen juntos en las delineaciones de una misma unidad cartográfica, pero su interpretación y manejo sí son próximos para usos comunes, se describen como grupos indiferenciados.

La configuración de las SMU resulta de un equilibrio entre la composición en STU y el nivel taxonómico que alcanzan dentro del sistema de clasificación. Así ocurre que una disminución del rango del nivel taxonómico de las STU (categoría inferior) tiende a provocar una disgregación y un aumento del número de constituyentes, mientras que el aumento del rango (categoría superior) suele llevar aparejado su reducción. Ambos aspectos conforman el detalle categórico que, en última instancia, es dependiente de la escala y queda recogido en los órdenes o niveles de reconocimiento con niveles de STU y tipos de SMU óptimos (Tabla 1-1).

## 2.6 Presentación de los resultados

El producto final del proceso de cartografía dentro del IRS es un mapa que representa la distribución espacial de las categorías de suelos identificadas. Se estructura a partir de SMU de agrupaciones de delineaciones y compuestas por una o más STU, que, como se ha señalado, constituyen la leyenda del mapa.

El mapa de suelos que completa el IRS es un documento gráfico que, como tal, requiere unas cualidades asociadas principalmente con la legibilidad; y, a su vez, se trata de un documento orientado a la toma de decisiones, lo que exige unos determinados criterios de calidad.

En relación con el aspecto gráfico, se necesita asegurar la legibilidad de la representación de todos los elementos incluidos. Empíricamente, el grupo de «Adecuación de IRS» en *the Cornell University* ha situado el tamaño de la «Delineación Mínima Legible» (en inglés *Minimum Legible Delineation* [MLD]) de un mapa alrededor de  $0,4 \text{ cm}^2$  (Rossiter, 2000), ya que con un tamaño menor i) no se dispone de espacio interior suficiente para incluir una simbología legible; y ii) la proporción de delineación que es cubierta por la representación de su perímetro es significativa (Forbes et al., 1987). Para una escala dada, dicha delineación se corresponde con un «Área Mínima Legible» (en inglés *Minimum Legible Area* [MLA]) sobre el terreno. De igual forma, el tamaño de una «Delineación Óptima Legible» (en inglés *Optimum Legible Delineation* [OLD]) se ha fijado en cuatro veces el valor de la MLD, que se corresponde respectivamente con un «Área Óptima Legible» (en inglés *Optimum Legible Area* [OLA]).

El concepto de textura de un mapa está orientado a determinar la legibilidad de los mapas y está intrínsecamente relacionado con el detalle cartográfico. Ambos se refieren al tamaño de las delineaciones presentes en el mapa y su ajuste. Es posible establecer la evaluación de la textura o el detalle cartográfico a través de la densidad o distribución del tamaño de las delineaciones (Forbes et al., 1987). En ese sentido, se recomienda que el «tamaño medio de las delineaciones» (en inglés *Average-size delineation* [ASD]) supere el límite del MLD y se aproxime al OLD. Este balance se refleja en el «Índice de Máxima Reducción» (en inglés *Index of Maximum Reduction* [IMR]), que es el factor por el cual la escala del mapa podría reducirse antes de que el ASD fuera equivalente al MLD (Figura 1-1). Por consiguiente, cuando el ASD equivale al OLD, el IMR toma un valor de 2. Finalmente, en base a esta relación y conociendo el área de las delineaciones sobre el terreno y su área media (en inglés *Average-size area* [ASA]), sería posible resolver la escala óptima de publicación del mapa (Rossiter, 2000).

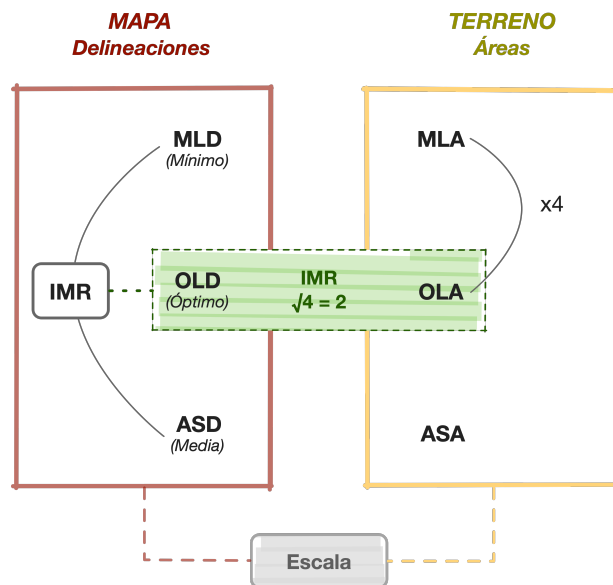


Figura 1-1. Diagrama de las relaciones entre los elementos del detalle cartográfico en un mapa. ASA: Área media de las delineaciones; ASD: Tamaño medio de las delineaciones; MLA: Área Mínima Legible; MLD: Área Mínima Legible; OLA: Área Óptima Legible; OLD: Delineación Óptima Legible; IMR: Índice de Máxima Reducción.

En cuanto a la calidad de la información proporcionada por el mapa, es un factor esencial de cara a su aplicabilidad. En base a dicha información se toman decisiones en áreas críticas como la planificación territorial, inversión para desarrollo agrícola, o la protección de recursos ambientales y su gestión, que está recogida en la legislación vigente.

Es posible llevar a cabo una evaluación de la calidad del mapa en términos de su predictibilidad. Esta se define como el grado de rigor en que el mapa y la leyenda representan la distribución y la

tipología de suelos en el paisaje, y puede entenderse desde dos perspectivas complementarias. Una, enfocada a la distribución, que atiende a la exactitud de la ubicación de los límites de las delineaciones en comparación con el terreno; frente a otra, orientada a la tipología, que considera la verosimilitud de la información recogida en la leyenda. El primer enfoque evalúa los errores de los límites entre delineaciones, mientras el segundo lo hace respecto a los errores de la clasificación dentro de las SMU (Forbes et al., 1987). En la medida en que discordancias en los límites llevan implícitas también desajustes en la composición, ambos tipos de errores guardan cierta interrelación.

El interés sobre la predictibilidad y la calidad de los mapas se centra especialmente en la tipología. Esto se debe a que su cuantificación resulta más directa, a través del contraste con observaciones, y esto permite una evaluación más objetiva.

## **3 Cartografía de Suelos Digital**

### **3.1 Definición de Cartografía de Suelos Digital**

Los avances en la creación y recopilación de datos ambientales a partir de nuevas tecnologías como la teledetección y en herramientas para su manejo como los «Sistemas de Información Geográfica» (SIG; en inglés *Geographic Information System* [GIS]) sumados al incremento en la potencia computacional (Rossiter, 2018) han impulsado el desarrollo y la evolución de nuevos métodos de cartografía de suelos basados en medios digitales (Minasny y McBratney, 2016) que han sido englobados bajo un marco común denominado «Cartografía de Suelos Digital» (en inglés *Digital Soil Mapping* [DSM]) (McBratney et al., 2003) o «Cartografía de Suelos Predictiva» (en inglés *Predictive Soil Mapping*) (Hengl y MacMillan, 2019).

DSM supone una nueva aproximación al paradigma suelo-ambiente, con la que se busca cuantificar las relaciones entre sus componentes. Este enfoque combina aspectos de múltiples disciplinas científicas, entre ellas la edafología, la edafometría, la geoestadística y la teledetección y los SIG.

#### *3.1.1 Bases teóricas*

En términos generales, la distribución de los suelos en el paisaje ha sido contextualizada como un caso particular del modelo universal de variación espacial (Hengl y MacMillan, 2019), que responde a una componente determinística (la tendencia), una componente espacialmente

correlacionada, y el ruido (variación a micro-escala y errores de medición). La parte determinística recoge la variación de los suelos que puede ser descrita mediante un modelo, relacionando las medidas observadas con variaciones en los factores con influencia. En CSM esa componente se desarrolla mediante relaciones descriptivas y subjetivas definidas por el cartógrafo; mientras que en DSM se establecen relaciones cuantitativas a través del uso de la geoestadística y «aprendizaje automático» (en inglés *Machine Learning* [ML]), donde es posible ponderar el error de las predicciones y donde también se incorpora la componente de la correlación espacial.

Para ello, DSM se basa en un nuevo esquema de factores, conceptualizado como «SCORPAN» (McBratney et al., 2003), que extiende el modelo CLORPT ya descrito y además considera la información y el conocimiento experto proveniente de cartografías previas (S) y la posición espacial (N) como factores.

En este nuevo modelo, los factores son representados por superficies de información continua de parámetros ambientales, los cuales se relacionan con las categorías y propiedades de suelos utilizando técnicas estadísticas robustas para generar el modelo cuantitativo. En última instancia, se obtiene la expresión óptima de una compresión validada de la relación suelo-ambiente (Hengl y MacMillan, 2019).

### *3.1.2 Objetivos y mejoras respecto a la Cartografía de Suelos Convencional*

Las principales mejoras de DSM respecto a CSM son: la cuantificación de las relaciones y su mayor objetividad; su reproducibilidad y fácil actualización; la cuantificación de los errores e incertidumbre sobre las predicciones; y una reducción de los costes asociados.

Se trata de un modelo cuantitativo y validado, y en la medida en la que no está basado exclusivamente en el conocimiento experto del cartógrafo existe una mayor objetividad. Por este motivo, es más fácilmente reproducible y actualizable, lo que permite mantener la consistencia entre miembros de los equipos de cartografía y optimizar los flujos de trabajo y análisis. Además, la modelización cuantitativa permite estimar los errores cometidos en las predicciones e incorporarlos en los métodos estadísticos.

Los métodos CSM aplicados durante el S. XX han requerido una importante inversión en recursos financieros y humanos altamente cualificados, que se incrementan exponencialmente con el aumento de su precisión y detalle (Rossiter, 2000). En DSM es posible aprovechar información



previa generada mediante CSM, a la vez que incorpora datos auxiliares desde nuevas fuentes, como la teledetección y la geomorfometría. Estos datos tienen una extensa cobertura y una amplia resolución espacial y espectral, su obtención es regular y cuentan con una amplia difusión y disponibilidad, lo que hace que DSM sea económicamente viable para grandes escalas.

### *3.1.3 Organización de la Cartografía de Suelos Digital*

Un proyecto de IRS mediante DSM puede estructurarse en las siguientes fases: i) la definición de la escala en relación con los objetivos previstos y los recursos disponibles; ii) recopilación de datos de suelos disponibles, o datos heredados; iii) recopilación de fuentes de datos para coberturas continuas; iv) derivación de covariables a partir de los datos recabados y su selección; v) recopilación de datos sobre el terreno, que incluye el diseño del muestreo y posteriormente la descripción y análisis de las observaciones realizadas; vi) la clasificación de estas observaciones bajo un sistema establecido y la creación de STU adaptadas a la escala; vii) generación de un modelo suelo-ambiente cuantitativo mediante métodos estadísticos adaptados que relaciona las observaciones con los datos ambientales del terreno, y la extrapolación de este modelo a toda la región de interés; y viii) presentación de los resultados.

En comparación, supone una actualización del procedimiento de cartografía convencional para adaptarlo a las nuevas necesidades en prácticamente todos sus aspectos, salvo lo relativo a la definición inicial de la escala en función de los objetivos y los recursos disponibles, la descripción de las observaciones en campo, el análisis de sus muestras en laboratorio, y su clasificación en un sistema establecido. Las novedades son descritas con mayor detenimiento a continuación en las siguientes secciones.

## **3.2 Datos de suelos heredados**

Los datos de suelos heredados están formados por el conjunto de la información, los datos y los conocimientos acumulados durante los procesos de IRS históricos y, en ese sentido, constituyen un legado con un interés intrínseco, no solo para ser preservados, sino del que generar un aprovechamiento de forma constructiva. En el contexto de DSM aportan información sobre suelos complementaria al proceso cartográfico con la que construir los modelos suelo-ambiente. Específicamente, se pueden distinguir cuatro tipos centrales: i) las observaciones y mediciones obtenidas en campo; ii) los mapas de clases de suelos de los IRS; iii) el conocimiento experto tácito; y iv) registros fotográficos y audiovisuales (Hengl y MacMillan, 2019).

La información referente a observaciones y mediciones tomadas en campo, principalmente en forma de sondeos y calicatas, incluye las descripciones de los perfiles y de su entorno, así como los datos analíticos derivados de las muestras recogidas. Suponen un tipo de dato primario, ya que proporcionan información relativamente directa sobre atributos y propiedades individuales de los suelos que varían con el paisaje. La calidad de todos estos datos es muy variable en términos de precisión, exhaustividad, relevancia y antigüedad, por lo que es necesario que sean utilizados considerando las implicaciones y limitaciones que les afectan. En este sentido, idealmente las observaciones están localizadas con exactitud mediante un «Sistema Global de Navegación por Satélite» (en inglés *Global Navigation Satellite System* [GNSS]) como el «Sistema de Posicionamiento Global» (en inglés *Global Positioning System* [GPS]), pero puede ocurrir que sus coordenadas hayan sido redondeadas en la publicación, que su ubicación sea descrita vagamente o que, directamente, carezcan de esa referencia. El nivel de minuciosidad y exhaustividad en la descripción y el muestreo varía en función del tipo de observación, siendo superior en las calicatas en comparación con los sondeos. Respecto a la relevancia de los atributos recogidos sobre las muestras, estos pueden provenir de análisis en laboratorio, siguiendo procedimientos controlados, o que fueran estimados por otros métodos durante el proceso de descripción. La fecha de captación también influye porque los criterios de análisis, rigor y clasificación se han ido actualizando con el transcurso del tiempo.

Los mapas de categorías de suelos de los IRS han supuesto tradicionalmente la principal fuente de representación y comunicación de la variación espacial de los suelos y sus tipologías. Estos mapas pueden ser utilizados por sí mismos como fuentes auxiliares de información para la mejora y la creación de nueva cartografía, si bien deben ser consideradas las limitaciones que presenten derivadas de su antigüedad y de los parámetros descritos para su elaboración (Capítulo 1, 2).

Complementariamente, existe un conocimiento experto tácito generado durante los reconocimientos que puede haber sido recogido en las memorias de IRS o de manera informal en diagramas y reglas de decisión, e incluso que no haya sido comunicado explícitamente. Este conocimiento representa la denominada «información blanda» (en inglés *soft data*), en contraste con los datos de las observaciones puntuales y los mapas.

Además, existe todo un registro histórico fotográfico y audiovisual con capacidad de aportar información auxiliar de gran valor y de completar carencias en las descripciones. Es el caso de todas las fotografías producidas durante la recopilación de observaciones en el campo y que

muestran los perfiles y su entorno. También de todo tipo de fotografías aéreas, como las series de ortofotos, con las que es posible observar la evolución de las regiones en el tiempo; o las series de imágenes estereoscópicas.

### 3.3 Fuentes de datos para coberturas continuas

#### 3.3.1 Caracterización

Para expresar los factores del paisaje, en las metodologías de DSM se hace uso de superficies de información ambiental continua. Con frecuencia son derivadas desde mediciones de teledetección de la superficie terrestre condicionadas por las características y localización de los sensores empleados.

Existen sensores captando información desde satélites, aerotransportados (aviones, drones) o proximales (cerca de la superficie, robotizados), con propiedades y aplicaciones particulares. Se distinguen por la i) resolución espectral, según la información o región del espectro que recojan, como gamma, espectro visible (en inglés *Visible Spectrum* [VIS]) o el infrarrojo (en inglés *Infrared* [IR]); ii) la resolución espacial, según el tamaño de malla de captación, que pueden ir desde pocos centímetros en los proximales y aerotransportados hasta kilómetros en los satélites; iii) la resolución temporal, según la periodicidad de medición, ya sea puntual o recurrente; y iv) la resolución radiométrica, según la discriminación de las variaciones en la radiación captada.

En el caso de la información satelital, han sido organizados programas de observación terrestre continuos como «Landsat» coordinado entre la «Administración Nacional de Aeronáutica y el Espacio» (en inglés *National Aeronautics and Space Administration* [NASA]) y «Servicio Geológico de EUA» (en inglés *United States Geological Survey* [USGS]) en EUA y que es el más longevo del mundo. Por su parte, la Agencia Espacial Europea (en inglés *European Space Agency* [ESA]) lleva a cabo el programa «Copernicus», dentro del cual destaca la misión «Sentinel», con información en el rango del VIS e IR a muy alta resolución. Ambos se encuentran entre los programas más relevantes de monitorización del globo en abierto (Radočaj et al., 2020), ya que distribuyen la mayor parte de su información libremente. Desde estas mediciones se generan principalmente imágenes de reflectancia en las regiones del espectro establecidas.

Otro producto obtenido a partir de teledetección son los «Modelos Digitales del Terreno» (MDT; en inglés *Digital Elevation Model* [DEM]), que representan la elevación de la cota del terreno a

través de una malla de valores. En la actualidad, existe una importante aplicación de aquellos MDT derivados desde vuelos con sensores *Light Detection And Ranging* (LIDAR) debido a su alto nivel de precisión y detalle, que alcanzan precisiones milimétricas.

### *3.3.2 Preprocesado*

Generalmente, estas fuentes de datos necesitan un procesamiento previo a su uso y que es complementario a las correcciones efectuadas antes de su distribución. Por ejemplo, en el momento de trabajar sobre áreas por imágenes solapadas con coberturas parciales, se ha señalado el uso de la igualación *Pseudo Invariant Features* para evitar límites pronunciados (Young et al., 2017).

Por otro lado, en los MDT se ha aplicado favorablemente el relleno de depresiones espurias (Wang y Liu, 2006), que fuerzan la obtención de un modelo hidrológico plenamente conectado y sin artefactos; así como filtros que buscan preservar las formas marcadas en las pendientes a la vez que eliminan las pequeñas variaciones que introducen ruido (Florinsky, 2017)(Stevenson et al., 2010).

## **3.4 Covariables**

Las superficies digitales continuas que cuantifican parámetros ambientales relacionados con los factores SCORPAN se denominan covariables. De manera común, las covariables son derivadas desde las fuentes de datos señaladas y recogidas en formato ráster. Se trata de imágenes con estructura de malla rectangular o matriz de píxeles, donde cada celda o píxel está georreferenciada y recoge el valor asociado a la variable en el espacio que cubre.

Las covariables son capaces de representar un rango muy amplio aspectos ambientales, y es posible categorizarlas en función de ellos.

### *3.4.1 Geomorfométricas*

Las covariables geomorfológicas agrupan los parámetros relacionados con la morfología de la superficie terrestres. En primer lugar, es posible distinguir aquellas derivadas directamente desde MDT, que son denominadas como «básicas» (en inglés *basic land-surface parameters*) por Olaya (Hengl y Reuter, 2009) o «primarias» según Wilson (2018).

Dentro de estas, Florinsky (2017) clasifica como «local» las basadas en el análisis de la vecindad de un punto o píxel dado (en inglés *neighbourhood*) mediante una «ventana de análisis» (en inglés *analysis window*) hasta una cierta distancia y en la dirección marcada por uno de los dos pares de secciones normales perpendiculares entre sí. Uno es el par de secciones normales de extremos mínimos y máximos, independientes de la gravedad, que define los parámetros de forma como las curvaturas máxima, media y mínima o como el «Índice de Forma» (en inglés *Shape Index*). El otro es el par de secciones definido por el vector de aceleración de la gravedad (coincidente con la pendiente) y su normal tangente a las curvas de nivel y que define los parámetros de flujo como la pendiente, la orientación, las curvaturas tangencial o de perfil, etc.

En contraste con las locales, Florinsky (2017) define las «no locales o regionales» (en inglés *non-local o regional*), que describen la posición relativa respecto al entorno de un punto o píxel dado en la superficie.

Generalmente, estas covariables se construyen estableciendo relaciones topológicas entre píxeles a través de los flujos de corrientes superficiales, en los que se considera la acción de la gravedad a favor de la pendiente, y se corresponden con propiedades hidrológicas.

Para lograr simular estas rutas pueden ser empleados múltiples algoritmos, que se dividen entre aquellos de dirección simple o múltiple. Los del primer grupo, como «D8» o «Rho8», calculan para cada píxel una única dirección en una malla rectangular, que resulta en valores múltiplos de 45°. Sin embargo, esta simplificación trae consigo limitaciones, en la medida en que la máxima pendiente real no tiene por qué ajustarse a esos valores. Como alternativa, se dirige la propagación del flujo a la vez entre varias celdas a través de los «algoritmos de dirección múltiple» (en inglés *Multiple Flow Direction*). En ocasiones esto da lugar a un exceso de dispersión respecto a la realidad. Del análisis de los flujos de corrientes se derivan covariables como la superficie de captación de la cuenca o divisoria de aguas (en inglés *catchment area*) o la distancia vertical a la red de drenaje desde cada punto.

Entre las regionales también cabría incluir las covariables de parámetros estadísticos básicos, como la media o la desviación estándar; así como otras más sofisticadas como el «Índice de Rugosidad» (en inglés *Surface Roughness Index*) o el «Índice de Posición Topográfica» (en inglés *Topographic Position Index [TPI]*), que cuantifica la diferencia entre el punto central y la media de los píxeles en la ventana de análisis.

En contraposición a las covariables primarias se distinguen las «compuestas» (Florinsky, 2017) o secundarias (Wilson, 2018), bien como una combinación entre covariables primarias o al incorporar al cálculo variables de dominios ajenos al MDT, por ejemplo, la radiación solar o la velocidad del viento. Aquí se consideran el «Índice de Humedad Topográfica» (en inglés *Topographic Wetness Index* [TWI]), los índices multiresolución de planicidad de los valles y de las crestas (en inglés *Multiresolution Index of Valley Bottom Flatness* y *Multiresolution Index of Ridge Top Flatness* respectivamente) (Gallant y Dowling, 2003), o el «calor anisotrópico diurno» (en inglés *Diurnal Anisotropic Heat*) (Boehner & Antonic in: Hengl y Reuter, 2009)

### *3.4.2 Reflectancia en el espectro electromagnético*

Imágenes con valores de reflectancia en ciertos rangos del espectro electromagnético han sido utilizadas directamente como covariables, habitualmente con la intención de capturar la variabilidad presente en la vegetación y en el suelo.

A partir de estas también se han construido índices asociados con propiedades y factores ambientales, o variables sintéticas mediante «Análisis de Componentes Principales» (en inglés *Principal Component Analysis* [PCA]) o *Tasseled Cap Transformation* (Soil Survey Division Staff, 2017). Cuentan con un uso extendido los índices relacionados con el vigor de la vegetación y su desarrollo (Bannari et al., 1995; Zarco Tejada et al., 2005), entre los que destaca el «Índice de Vegetación de Diferencia Normalizada» (en inglés *Normalized Difference Vegetation Index* [NDVI]). También, se han descrito índices ligados a características del suelo que potencialmente identifican la presencia en superficie de sales como carbonatos o sulfatos, óxidos de hierro y arcillas (Mulder et al., 2011; Viscarra Rossel et al., 2006). Ambos tipos de propiedades se basan fundamentalmente en las regiones del VIS e IR.

Por último, se han asociado mediciones de la concentración de radiación gamma en la superficie del suelo emitidas por los isótopos K, U y Th, responsables de la mayor parte de la radiación gamma natural, con propiedades mineralógicas, químicas y de textura de los materiales presentes (Dierke y Werban, 2013; Wilford y Minty, 2006) y parental (Gray et al., 2016)

### *3.4.3 Interpolación desde observaciones*

Las covariables también pueden ser resultado de la interpolación desde mediciones puntuales a todo un área de interés como, por ejemplo, ciertas propiedades del suelo desde observaciones

(Hengl et al., 2004) o variables climatológicas desde estaciones (Samanta et al., 2012; Webb et al., 2016).

#### 3.4.4 Selección de covariables

La construcción de covariables a partir de las fuentes de datos es relativamente asequible, en términos de disponibilidad y coste de las propias fuentes; y accesible, ya que su cálculo se realiza a través de software GIS específicos que en buena parte son de código abierto, por ejemplo *System for Automated Geoscientific Analyzes* (SAGA) (Conrad et al., 2015), *Geographic Resources Analysis Support System* (GRASS) (GRASS Development Team, 2019) o QGIS (QGIS Development Team, 2019).

Así es posible generar varias decenas de covariables con relativa facilidad. Sin embargo, aunque sea posible introducir todas aquellas covariables disponibles en la modelización, suele ser preferible seleccionar un subconjunto adaptado y óptimo. Esto se debe a que la inclusión de un alto número puede favorecer el sobreajuste y, si son poco informativas, también el aumento de la incertidumbre del modelo. Además, los modelos más simples resultan más fáciles de interpretar. La reducción del número de covariables también está dirigida a la optimización de los tiempos de procesamiento y cálculos (Soil Survey Division Staff, 2017).

En última instancia, se busca que las covariables seleccionadas sean capaces de revelar la varianza ambiental con la mínima redundancia, y donde su número se mantenga lo más bajo posible para lograr la máxima eficiencia.

Existen métodos semi-automáticos que permiten identificar subconjuntos de estas características y que pueden ser categorizados entre no supervisados y supervisados (Soil Survey Division Staff, 2017). Los no supervisados realizan la evaluación de forma independiente al modelado atendiendo a diversos criterios, y entre ellos se reconocen los análisis de correlación (Thompson et al., 2016), el «Factor de Índice Óptimo» (en inglés *Optimum Index Factor*) y PCA (Brungard et al., 2015; Soil Survey Division Staff, 2017). También, han sido descrito métodos de mayor sofisticación que se apoyan en el análisis de la varianza (Behrens et al., 2010).

Por su parte, la selección en los métodos supervisados está integrada e identifica las covariables óptimas que maximizan su capacidad de predicción. Estos incluyen la «selección recursiva» aditiva y sustractiva (en inglés *forward* y *backward recursive feature selection*) o la aplicación del algoritmo «Boruta», el cual extrae aquellas covariables cuya acción es significativamente mayor

que la de covariables aleatorias. Los métodos supervisados basados en «árboles de decisión» (en inglés *Decision Trees* [DT]) incluyen técnicas de selección específicas basadas en parámetros internos (Soil Survey Division Staff, 2017). Ambos métodos de reducción de covariables, supervisados y no supervisados, pueden ser utilizados conjuntamente.

Dado que los métodos tienen aproximaciones diferentes, los subconjuntos identificados varían entre unos y otros. En este sentido, el conocimiento edafológico resulta esencial para juzgar el sentido material ceñido al área de interés de los resultados obtenidos.

### 3.5 Muestreo

Las observaciones de suelos cumplen dos finalidades autónomas y de igual importancia en DSM. Pueden aportar información sobre el factor suelo con el que entrenar el modelo, así como participar en la evaluación del modelo obtenido.

Ambas acciones se enmarcan en la aplicación de métodos estadísticos y, por ese motivo, la distribución de las observaciones en los muestreos responde a parámetros similares, si bien se tratan de conjuntos separados e, idealmente, independientes entre sí. En este sentido, el diseño de un muestreo específico para DSM es preferiblemente aleatorio y orientado por datos, ya que de esta forma resulta más robusto y menos propenso al sesgo del cartógrafo o a la creación de correlaciones no deseadas. En contraposición, la construcción del muestreo en CSM sigue mayoritariamente un enfoque dirigido, donde las observaciones se han situado en lugares claves identificados por el cartógrafo (Capítulo 1, 2.3).

La organización del diseño del muestreo dependerá de diversos factores: el tamaño y la accesibilidad del área de interés, los objetivos de la modelización, la disponibilidad de recursos y tiempo, la variabilidad esperada de las características del suelo y el nivel de confianza y de precisión marcado (Soil Survey Division Staff, 2017). Para satisfacer los requisitos de las condiciones de aleatoriedad se aplican diferentes tipos de muestreo caracterizados por los límites, la forma de reparto y el tamaño fijado.

En cuanto a los límites, existen muestreos simples o estratificados, ya sean simples o por fases, y variaciones de mayor sofisticación (Figura 1-2). El muestreo simple es el más sencillo y resulta de establecer las localizaciones de las observaciones de forma independiente, en el que todo el espacio tiene la misma probabilidad de ser muestreado. En el estratificado simple se subdivide la



región y se aplican muestreos aleatorios individuales dentro de cada parcelación o estrato. Estos estratos suelen estar basados en geformas o propiedades particulares del paisaje, aunque también se ha acudido a métodos no supervisados, como *k-means* (Brus et al., 2006). Con ellos se busca delimitar zonas que guarden una relación próxima con las características del suelo que son objeto del modelado. Posteriormente, el reparto del muestreo individual puede ser fijo o en proporción al área, y con diferentes tamaños de muestras. Este sistema ofrece una alta precisión a un costo reducido, si bien depende de la idoneidad de los estratos definidos. Una versión reducida es la estratificación por fases, en el que se selecciona un subconjunto de estratos para los muestreos individuales. Dicha selección puede responder a criterios establecidos o también ser aleatoria, y persigue disminuir los costes, a expensas de una menor exactitud.

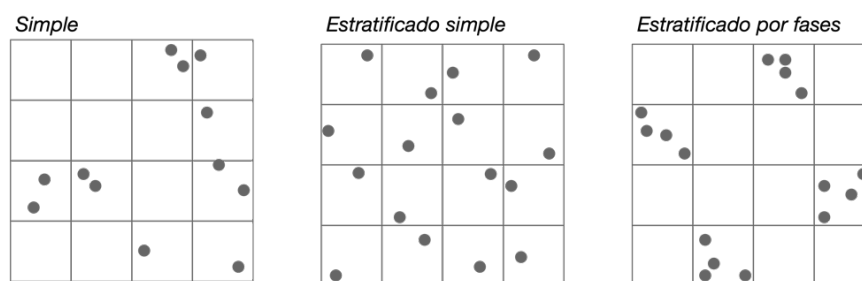


Figura 1-2. Abstracción de métodos de muestreo aleatorios: simple, estratificado simple y estratificado por fases.

Otra posibilidad de mayor sofisticación es el «Muestreo de Hipercubo Latino condicionado» (en inglés *Conditioned Latin Hypercube Sampling* [cLHS]). Se trata de un tipo especial de muestreo aleatorio estratificado donde, en lugar de prospectar el espacio geográfico, lo hace en el espacio definido por las covariables (n-dimensional), buscando maximizar la variabilidad capturada. Este tipo de muestreo es eficiente en la medida en que es capaz de representar la distribución multivariable de las covariables mediante tamaños de muestra relativamente pequeños (Minasny y McBratney, 2006).

Alternativamente, es frecuente aprovechar los datos heredados de observaciones obtenidos por muestreos dirigidos diseñados para cartografía convencional, o hacer uso de métodos sistemáticos como el de muestreo en malla fija. En ese sentido, es posible combinar observaciones capturadas bajo diseños diferentes, de forma que un muestreo preexistente sea complementado para mejorarlo.

## 3.6 Tipología

La modelización cuantitativa de las relaciones suelo-ambiente sobre la que gira DSM se establece mediante técnicas estadísticas. En ellas, los factores que configuran el medio son representados a través de covariables, las cuales son apiladas en una estructura en formato ráster, de modo que existe información de todas ellas para cada localización o píxel. Por su parte, la información del suelo capturada en las observaciones puntuales y sus datos son incluidos en los píxeles correspondientes para cada localización.

Siguiendo esta lógica, cualquier información referente al suelo que sea traducida a esta estructura puede ser incorporada en la función del modelo, ya sea a modo de variable dependiente o independiente, a saber, expresando relaciones internas entre atributos del suelo. Esto permite aprovechar la información heredada de los mapas de suelos convencionales existentes, e incluso los resultados de proyectos de reconocimiento de suelos por DSM anteriores.

El alcance de la modelización de las relaciones está sujeta a la capacidad de los métodos estadísticos para trabajar sobre múltiples tipos de valores y para generar diferentes salidas, ya sean continuos o categóricos. De esta forma se abre la posibilidad de generar mapas de categorías de suelos, pero también de propiedades continuas.

### 3.6.1 *Propiedades o atributos del suelo*

Esta aproximación supone la predicción de una propiedad del suelo como una capa de valores continuos. Para ello, se han utilizado con éxito diversos métodos de regresión e interpolación desde observaciones puntuales, por ejemplo, diferentes variantes de *kriging* (*regression, universal*) (Hengl et al., 2004).

Así, se han desarrollado experiencias favorables donde se ha inferido la Profundidad Efectiva del Suelo, la Capacidad de Retención de Agua Disponible o el contenido en Carbono Orgánico en el Suelo (Malone et al., 2009). En relación a la Materia Orgánica en los suelos, han sido completados con éxito proyectos que han cubierto amplias extensiones en ambientes específicos (Poggio et al., 2013) o estableciendo predicciones en 3D (Kempen et al., 2011; Poggio y Gimona, 2014).

### 3.6.2 *Clases de suelo y categorías de atributos*

Las clases de suelos pertenecientes a sistemas de clasificación establecidos y las categorías de atributos edáficos, unificadas ambas en adelante bajo el término clases de suelos, facilitan la

comprensión de los comportamientos de los suelos, y son asociadas de forma unívoca a los perfiles y, tradicionalmente, también a las unidades de los mapas convencionales.

La estructura de las imágenes ráster permite el tratamiento individualizado de las clases en cada píxel, y también permite asignarles múltiples probabilidades de categorías bajo la lógica «fuzzy» (Odeh et al., 1992). Con ello, además, es posible calcular la incertidumbre asociada a cada localización y derivar mapas desde la perspectiva de las clases mayoritarias o desde niveles de probabilidad menores.

Con el objetivo de asignar un número finito de clases a los píxeles se han empleado diversos métodos de clasificación. En aquellas ocasiones en las que no se ha tenido información precisa acerca de las clases se han aplicado métodos de clasificación no supervisados, como *k-means* o *c-means*, también en sus variantes fuzzy (Odeh et al., 1992; Zeraatpisheh et al., 2019). En cambio, cuando sí ha sido posible su definición, se han empleado métodos de clasificación supervisados como «k vecinos más cercanos» (en inglés *k-Nearest Neighbour* [kNN]), «análisis discriminatorio de máxima verosimilitud» (en inglés *Discrimatory Analysis de Maximum likelihood*), «regresión logística multinomial» (en inglés *Multinomial Logistic Regression* [MLR]), y algoritmos más sofisticados como «máquinas de vectores de soporte» (en inglés *Support Vector Machines* [SVM]) o «redes neuronales artificiales» (en inglés *Artificial Neural Networks* [ANN]), que se enmarcan en ML. Entre estos también se incluyen los DT, que cuentan con variantes como CART, C5.0 o C4.5; así como un método basado en su combinación, «bosque aleatorio» (en inglés *Random Forest* [RF]) (Brungard et al., 2015; Flynn et al., 2019; Heung et al., 2016).

Sobre esta base se han desarrollado varias líneas de acción (Figura 1-3). Una ha sido la construcción de modelos con información precisa desde las observaciones puntuales y que luego ha sido extendido a localizaciones intermedias no visitadas (Carré y Girard, 2002; Hengl et al., 2007), que cabría denominar como «generalización». Con esos principios, se han generado favorablemente nuevos mapas (Ramcharan et al., 2018) y actualizado inventarios del recurso suelo (Kempen et al., 2015) de gran cobertura.

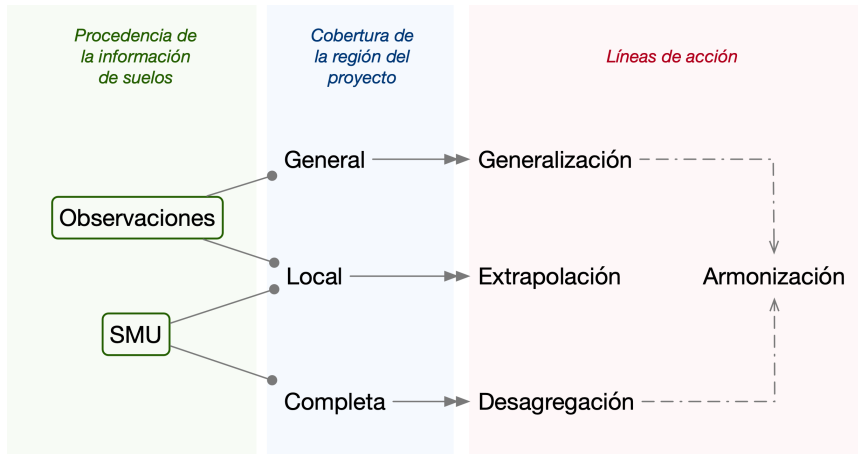


Figura 1-3. Diagrama de líneas de acción en la Cartografía de Suelos Digital para clases de suelos en función de la procedencia de la información de los suelos y de su cobertura de la región del proyecto.

SMU: Unidad Cartográfica del Suelo.

En otras ocasiones, dichos modelos han sido generados sobre áreas reducidas a una escala grande para capturar con gran detalle el patrón de distribución de los suelos y, después, han sido «extrapolados» a áreas desconocidas colindantes (Behrens et al., 2005; Grinand et al., 2008) o con condiciones ambientales semejantes (Angelini et al., 2020; Silva et al., 2016), donde se han logrado altas tasas de validación.

Por otro lado, para ciertas regiones hay disponible cartografía de suelos elaborada por CSM en las que se han definido SMU politáxicas, lo que limita su interpretación y aplicabilidad. En ellas es posible describir la variación espacial no explícita y aumentar así la escala, en lo que se conoce como «disgregación» (Sarmiento et al., 2017). Tanto la generalización (Wei et al., 2010) como la disgregación (Odgers et al., 2014), permiten además realizar la «armonización» de las zonas adyacentes. Esto es útil cuando esas áreas han sido cartografiadas en diferentes proyectos y se detectan desajustes.

### 3.7 Evaluación

La evaluación del modelo es una fase indispensable para asegurar la calidad de los resultados producidos. En última instancia, se trata de estimaciones de la realidad sujetas a un grado de error. Al ser construidos en base a técnicas estadísticas, sus predicciones pueden ser evaluadas cuantitativamente en términos de precisión e incertidumbre (Soil Survey Division Staff, 2017).

### 3.7.1 Precisión

Desde la perspectiva de la precisión, se evalúa la desviación de una predicción respecto a un valor real para una misma localización; y los métodos para su cuantificación difieren según se estudien clases o propiedades de suelos.

Para las clases de suelo se emplean métricas específicas que comparan el número de observaciones que son predichas correcta e incorrectamente, y que habitualmente se basan en el análisis de una matriz de errores (Figura 1-4).

La más sencilla es la «precisión global» (en inglés *Overall accuracy* [OA]) que resulta de calcular la proporción de observaciones correctamente clasificadas respecto al total y que se expresa en forma de porcentaje. También se define la «precisión del productor» (en inglés *Producer's Accuracy*) como la frecuencia con la que una clase sobre el terreno es clasificada como tal y se mide atendiendo al número de observaciones clasificadas correctamente respecto al número total de observaciones por cada clase (por columnas de la matriz de errores). Complementariamente, la «precisión del usuario» (en inglés *User's accuracy*) evalúa con qué frecuencia las clases en el mapa están realmente presentes sobre el terreno y se calcula como el número de predicciones correctas para una clase frente al número total de veces que se dan predicciones para esa clase (por filas de la matriz de errores). También es denominada como fiabilidad. El análisis de todas estas métricas de forma conjunta aporta una visión integral. Por ejemplo, puede ocurrir que ciertas clases alcancen ratios muy altos de una precisión, pero que éstos sean muy pobres para otras.

		Observaciones			
		Alfisol	Andisol	Inceptisol	Total
Predicciones	Alfisol	16	1	2	19
	Andisol	3	7	1	11
	Inceptisol	4	0	21	25
	Total	23	8	24	55

Figura 1-4. Ejemplo teórico de matriz de errores para una clasificación con clases de suelos.

Alternativamente, también se ha acudido al «Coeficiente de Kappa», otra métrica obtenida a partir de una prueba estadística en la que se evalúa el desempeño de la clasificación frente a valores asignados al azar. Este coeficiente toma valores en un rango entre -1 y 1, donde valores inferiores

a 0 indican que la clasificación es peor que valores aleatorios y próximos a 1, que son significativamente mejores (Brus et al., 2011).

Un rasgo distintivo de los sistemas de clasificación de suelos es que la división de las clases responde a criterios edafológicos que no son equivalentes entre sí; por ejemplo, en ST la semejanza en características edafológicas y funcionalidad entre entisoles e inceptisoles es mayor que entre cualquiera de ellos con los oxisoles. Esto ha llevado a intentar establecer una distancia taxonómica que permita modular los aciertos y errores fijados por las métricas (Carré y Girard, 2002; Minasny y McBratney, 2007).

En el caso de las propiedades del suelo, la precisión se evalúa típicamente mediante parámetros clásicos como el «error cuadrático medio» (en inglés *Mean Square Error* [MSE]), la «raíz del error cuadrado medio» (en inglés *Root Mean Square Error* [RMSE]), o el «coeficiente de determinación» ( $R^2$ , en inglés *Coefficient of determination*) (Soil Survey Division Staff, 2017).

Para ambos tipos de resultados también es posible acudir a medidas internas de rendimiento de los propios métodos estadísticos con los que se construyen los modelos, por ejemplo, el «error fuera de la bolsa» (en inglés *Out of Bag error* [OOB]) en RF y los modelos basados en DT (Håring et al., 2012).

El conjunto de observaciones empleado para la evaluación de la precisión de los modelos ha de ser exclusivo para este fin, ya que de lo contrario la evaluación quedaría invalidada. Idealmente, este conjunto es muestreado para la ocasión independientemente. Sin embargo, esto no es posible habitualmente y, en su lugar, se aprovechan datos heredados o se utilizan métodos alternativos de aprovechamiento de las observaciones disponibles.

Una opción es separar una porción de las observaciones para la validación, comúnmente entre el 10 y el 30%. Si este procedimiento de división de los datos entre conjuntos de entrenamiento y validación se repite sucesivamente durante n-iteraciones, de forma que cada observación sea utilizada una única vez para la validación y n-1 como entrenamiento, es posible calcular el error medio de todos los intentos, en lo que se conoce como validación cruzada (*cross-validation* en inglés). Dicha técnica permite hacer un uso más eficiente de las observaciones (Soil Survey Division Staff, 2017).

### 3.7.2 Incertidumbre

La incertidumbre es una estimación del intervalo de valores dentro de los cuales se afirma que se encuentra el valor real. Específicamente en las predicciones sobre clases de suelo, la incertidumbre puede ser cuantificada atendiendo a las probabilidades de pertenencias calculadas para cada píxel y mediante el «Índice de Confusión» (en inglés *Confusion Index*). Esta toma valores entre 0 y 1, siendo 0 el valor mínimo de confusión o incertidumbre (Ecuación 1-1).

$$CI = 1 - (\mu_n - \mu_{(n-1)})$$

Ecuación 1-1. Índice de Confusión, donde n es el rango de posibilidad de aparición.

## 4 Bibliografía

- Angelini, M.E., Kempen, B., Heuvelink, G.B.M., Temme, A.J.A.M., Ransom, M.D., 2020. Extrapolation of a structural equation model for digital soil mapping. *Geoderma* 367, 114226. <https://doi.org/10.1016/j.geoderma.2020.114226>
- Bannari, A., Morin, D., Bonn, F., Huete, A.R., 1995. A review of vegetation indices. *Remote Sens. Rev.* 13, 95–120. <https://doi.org/10.1080/02757259509532298>
- Behrens, T., Förster, H., Scholten, T., Steinrücken, U., Spies, E.-D., Goldschmitt, M., 2005. Digital soil mapping using artificial neural networks. *J. Plant Nutr. Soil Sci.* 168, 21–33. <https://doi.org/10.1002/jpln.200421414>
- Behrens, T., Zhu, A.-X., Schmidt, K., Scholten, T., 2010. Multi-scale digital terrain analysis and feature selection for digital soil mapping. *Geoderma* 155, 175–185. <https://doi.org/10.1016/j.geoderma.2009.07.010>
- Brungard, C.W., Boettinger, J.L., Duniway, M.C., Wills, S.A., Edwards, T.C., 2015. Machine learning for predicting soil classes in three semi-arid landscapes. *Geoderma* 239–240, 68–83. <https://doi.org/10.1016/j.geoderma.2014.09.019>
- Brus, D.J., de Gruijter, J.J., van Groenigen, J.W., 2006. Chapter 14 Designing Spatial Coverage Samples Using the k-means Clustering Algorithm, in: *Digital Soil Mapping: An Introductory Perspective*, Developments in Soil Science. Elsevier, pp. 183–192. [https://doi.org/10.1016/S0166-2481\(06\)31014-8](https://doi.org/10.1016/S0166-2481(06)31014-8)
- Brus, D.J., Kempen, B., Heuvelink, G.B.M., 2011. Sampling for validation of digital soil maps. *Eur. J. Soil Sci.* 62, 394–407. <https://doi.org/10.1111/j.1365-2389.2011.01364.x>
- Burt, R., Soil Survey Division Staff (Eds.), 2014. *Kellogg Soil Survey Laboratory Methods Manual*, 5.0. ed, Soil Survey Investigations Report. USDA-Natural Resources Conservation Service.
- Carré, F., Girard, M.C., 2002. Quantitative mapping of soil types based on regression kriging of taxonomic distances with landform and land cover attributes. *Geoderma* 110, 241–263. [https://doi.org/10.1016/S0016-7061\(02\)00233-1](https://doi.org/10.1016/S0016-7061(02)00233-1)

- Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J., 2015. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geosci. Model Dev.* 8, 1991–2007. <https://doi.org/10.5194/gmd-8-1991-2015>
- Dierke, C., Werban, U., 2013. Relationships between gamma-ray data and soil properties at an agricultural test site. *Geoderma* 199, 90–98. <https://doi.org/10.1016/j.geoderma.2012.10.017>
- Florinsky, I.V., 2017. An illustrated introduction to general geomorphometry. *Prog. Phys. Geogr. Earth Environ.* 41, 723–752. <https://doi.org/10.1177/0309133317733667>
- Flynn, T., van Zijl, G., van Tol, J., Botha, C., Rozanov, A., Warr, B., Clarke, C., 2019. Comparing algorithms to disaggregate complex soil polygons in contrasting environments. *Geoderma* 352, 171–180. <https://doi.org/10.1016/j.geoderma.2019.06.013>
- Forbes, T., Rossiter, D.G., Van Wambeke, A., 1987. Guidelines for evaluating the adequacy of Soil Resource Inventories, SMSS Technical Monograph. Soil Conservation Service. US Department of Agriculture (USDA), Washington D.C.
- Gallant, J.C., Dowling, T.I., 2003. A multiresolution index of valley bottom flatness for mapping depositional areas. *Water Resour. Res.* 39. <https://doi.org/10.1029/2002WR001426>
- GRASS Development Team, 2019. GRASS Geographic Resources Analysis Support System. Open Source Geospatial Foundation Project.
- Gray, J.M., Bishop, T.F.A., Wilford, J.R., 2016. Lithology and soil relationships for soil modelling and mapping. *Catena* 147, 429–440. <https://doi.org/10.1016/j.catena.2016.07.045>
- Grinand, C., Arrouays, D., Laroche, B., Martin, M.P., 2008. Extrapolating regional soil landscapes from an existing soil map: Sampling intensity, validation procedures, and integration of spatial context. *Geoderma* 143, 180–190. <https://doi.org/10.1016/j.geoderma.2007.11.004>
- Häring, T., Dietz, E., Osenstetter, S., Koschitzki, T., Schröder, B., 2012. Spatial disaggregation of complex soil map units: A decision-tree based approach in Bavarian forest soils. *Geoderma* 185–186, 37–47. <https://doi.org/10.1016/j.geoderma.2012.04.001>
- Hengl, T., Heuvelink, G.B.M., Stein, A., 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma* 120, 75–93. <https://doi.org/10.1016/j.geoderma.2003.08.018>
- Hengl, T., MacMillan, R.A., 2019. Predictive Soil Mapping with R. OpenGeoHub foundation, Wageningen, the Netherlands.
- Hengl, T., Reuter, H.I. (Eds.), 2009. *Geomorphometry: concepts, software, applications*, 1st ed. ed, Developments in soil science. Elsevier, Amsterdam, Netherlands; Oxford, UK; Boston [Mass.].
- Hengl, T., Toomanian, N., Reuter, H.I., Malakouti, M.J., 2007. Methods to interpolate soil categorical variables from profile observations: Lessons from Iran. *Geoderma* 140, 417–427. <https://doi.org/10.1016/j.geoderma.2007.04.022>
- Heung, B., Ho, H.C., Zhang, J., Knudby, A., Bulmer, C.E., Schmidt, M.G., 2016. An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. *Geoderma* 265, 62–77. <https://doi.org/10.1016/j.geoderma.2015.11.014>



- Gómez Miguel, V., 2006. Geología, Geomorfología y Edafología, Monografías. Atlas Nacional de España. Centro Nacional de Información Geográfica, Madrid.
- IUSS Working Group WRB, 2014. World reference base for soil resources 2014: international soil classification system for naming soils and creating legends for soil maps. FAO, Rome.
- Jenny, H., 1941. Factors of soil formation: a system of quantitative pedology. Dover, New York.
- Kempen, B., Brus, D.J., de Vries, F., 2015. Operationalizing digital soil mapping for nationwide updating of the 1:50,000 soil map of the Netherlands. *Geoderma* 241–242, 313–329. <https://doi.org/10.1016/j.geoderma.2014.11.030>
- Kempen, B., Brus, D.J., Stoorvogel, J.J., 2011. Three-dimensional mapping of soil organic matter content using soil type-specific depth functions. *Geoderma* 162, 107–123. <https://doi.org/10.1016/j.geoderma.2011.01.010>
- Malone, B.P., McBratney, A.B., Minasny, B., Laslett, G.M., 2009. Mapping continuous depth functions of soil carbon storage and available water capacity. *Geoderma* 154, 138–152. <https://doi.org/10.1016/j.geoderma.2009.10.007>
- MAPA, 1986. Métodos oficiales de análisis. Dirección General de Política Alimentaria. Ministerio de Agricultura, Pesca y Alimentación (MAPA).
- McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. *Geoderma* 117, 3–52. [https://doi.org/10.1016/S0016-7061\(03\)00223-4](https://doi.org/10.1016/S0016-7061(03)00223-4)
- Minasny, B., McBratney, A.B., 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Comput. Geosci.* 32, 1378–1388. <https://doi.org/10.1016/j.cageo.2005.12.009>
- Minasny, B., McBratney, A.B., 2007. Incorporating taxonomic distance into spatial prediction and digital mapping of soil classes. *Geoderma* 142, 285–293. <https://doi.org/10.1016/j.geoderma.2007.08.022>
- Minasny, B., McBratney, A.B., 2016. Digital soil mapping: A brief history and some lessons. *Geoderma* 264, 301–311. <https://doi.org/10.1016/j.geoderma.2015.07.017>
- Mulder, V.L., de Bruin, S., Schaepman, M.E., Mayr, T.R., 2011. The use of remote sensing in soil and terrain mapping — A review. *Geoderma* 162, 1–19. <https://doi.org/10.1016/j.geoderma.2010.12.018>
- Nieves Bernabé, M., Forcada Pérez, R., Gómez-Miguel, V.D., 1985. Precisión, escala y densidad de observaciones en los estudios de suelos. *Bol. Estac. Cent. Ecol.* 14, 47:56.
- Odeh, I.O.A., McBratney, A.B., Chittleborough, D.J., 1992. Soil Pattern Recognition with Fuzzy-c-means: Application to Classification and Soil-Landform Interrelationships. *Soil Sci. Soc. Am. J.* 56, 505–516. <https://doi.org/doi:10.2136/sssaj1992.03615995005600020027x>
- Odgers, N.P., Sun, W., McBratney, A.B., Minasny, B., Clifford, D., 2014. Disaggregating and harmonising soil map units through resampled classification trees. *Geoderma* 214–215, 91–100. <https://doi.org/10.1016/j.geoderma.2013.09.024>
- OIV, 2012. Resolution 423/2012: OIV Guidelines for vitiviniculture zoning methodologies on a soil and climate level.
- Poggio, L., Gimona, A., 2014. National scale 3D modelling of soil organic carbon stocks with uncertainty propagation — An example from Scotland. *Geoderma* 232–234, 284–299. <https://doi.org/10.1016/j.geoderma.2014.05.004>

- Poggio, L., Gimona, A., Brewer, M.J., 2013. Regional scale mapping of soil properties and their uncertainty with a large number of satellite-derived covariates. *Geoderma* 209–210, 1–14. <https://doi.org/10.1016/j.geoderma.2013.05.029>
- QGIS Development Team, 2019. QGIS Geographic Information System. Open Source Geospatial Foundation Project.
- Radočaj, D., Obhodaš, J., Jurišić, M., Gašparović, M., 2020. Global Open Data Remote Sensing Satellite Missions for Land Monitoring and Conservation: A Review. *Land* 9, 402. <https://doi.org/10.3390/land9110402>
- Ramcharan, A., Hengl, T., Nauman, T., Brungard, C., Waltman, S., Wills, S., Thompson, J., 2018. Soil Property and Class Maps of the Conterminous United States at 100-Meter Spatial Resolution. *Soil Sci. Soc. Am. J.* 82, 186. <https://doi.org/10.2136/sssaj2017.04.0122>
- Rossiter, D.G., 2000. *Methodology for Soil Resource Inventories*, 2nd Revised Version. ed, Lecture Notes & Reference. International Institute for Aerospace Survey & Earth Sciences (ITC).
- Rossiter, D.G., 2018. Past, present & future of information technology in pedometrics. *Geoderma* 324, 131–137. <https://doi.org/10.1016/j.geoderma.2018.03.009>
- Samanta, S., Pal, D.K., Lohar, D., Pal, B., 2012. Interpolation of climate variables and temperature modeling. *Theor. Appl. Climatol.* 107, 35–45. <https://doi.org/10.1007/s00704-011-0455-3>
- Schoeneberger, P.J., Wysocki, D.A., Benham, E.C., Soil Survey Division Staff, 2012. *Field book for describing and sampling soils*, Version 3.0. ed. Natural Resources Conservation Service (NRCS-USDA), Lincoln, NE.
- Ségalen, P., Van Diepen, C.A. (Eds.), 1984. *Projet de classification des sols*, Technical paper. International Soil Reference and Information Centre, Wageningen, The Netherlands.
- Silva, S.H.G., Menezes, M.D. de, Owens, P.R., Curi, N., 2016. Retrieving pedologist's mental model from existing soil map and comparing data mining tools for refining a larger area map under similar environmental conditions in Southeastern Brazil. *Geoderma* 267, 65–77. <https://doi.org/10.1016/j.geoderma.2015.12.025>
- Soil Survey Division Staff, 2017. *Soil Survey Manual*, USDA Handbook. U.S. Government Printing Office, Washington, D.C.
- Soil Survey Division Staff, 1999. *Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys*, 2nd Edition. ed, USDA Handbook. USDA-Natural Resources Conservation Service.
- Soil Survey Division Staff, 2014. *Keys to Soil Taxonomy*, 12th Edition. ed. USDA-Natural Resources Conservation Service, Washington, DC.
- Stevenson, J.A., Sun, X., Mitchell, N.C., 2010. Despeckling SRTM and other topographic data with a denoising algorithm. *Geomorphology* 114, 238–252. <https://doi.org/10.1016/j.geomorph.2009.07.006>
- Thompson, S.D., Nelson, T.A., Giesbrecht, I., Frazer, G., Saunders, S.C., 2016. Data-driven regionalization of forested and non-forested ecosystems in coastal British Columbia with LiDAR and RapidEye imagery. *Appl. Geogr.* 69, 35–50. <https://doi.org/10.1016/j.apgeog.2016.02.002>

- Van Wambeke, A., Forbes, T., 1986. Guidelines for using Soil Taxonomy in the names of soil map units, SMSS Technical Monograph. Soil Conservation Service. US Department of Agriculture (USDA).
- Vink, A.P.A., 1963. Planning of soil surveys in land development. International Institute for Land Reclamation and Improvement.
- Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J., Skjemstad, J.O., 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* 131, 59–75. <https://doi.org/10.1016/j.geoderma.2005.03.007>
- Wang, L., Liu, H., 2006. An efficient method for identifying and filling surface depressions in digital elevation models for hydrologic analysis and modelling. *Int. J. Geogr. Inf. Sci.* 20, 193–213. <https://doi.org/10.1080/13658810500433453>
- Webb, M.A., Hall, A., Kidd, D., Minansy, B., 2016. Local-scale spatial modelling for interpolating climatic temperature variables to predict agricultural plant suitability. *Theor. Appl. Climatol.* 124, 1145–1165. <https://doi.org/10.1007/s00704-015-1461-7>
- Wei, S., McBratney, A., Hempel, J., Minasny, B., Malone, B., D’Avello, T., Burras, L., Thompson, J., 2010. Digital Harmonisation of Adjacent Analogue Soil Survey areas - 4 Iowa Counties, in: *Proceedings of the 4th Global Workshop on Digital Soil Mapping. Presented at the 4th Global Workshop on Digital Soil Mapping*, p. 8.
- Wilford, J., Minty, B., 2006. Chapter 16 The Use of Airborne Gamma-ray Imagery for Mapping Soils and Understanding Landscape Processes, in: *Digital Soil Mapping: An Introductory Perspective, Developments in Soil Science*. Elsevier, pp. 207–610. [https://doi.org/10.1016/S0166-2481\(06\)31016-1](https://doi.org/10.1016/S0166-2481(06)31016-1)
- Wilson, J.P., 2018. *Environmental Applications of Digital Terrain Modeling, New Analytical Methods in Earth and Environmental Science*. John Wiley & Sons, Ltd, Chichester, UK.
- Young, N.E., Anderson, R.S., Chignell, S.M., Vorster, A.G., Lawrence, R., Evangelista, P.H., 2017. A survival guide to Landsat preprocessing. *Ecology* 98, 920–932. <https://doi.org/10.1002/ecy.1730>
- Zarco Tejada, P., Berjon, A., Lopez-Lozano, R., Miller, J., Martin, P., Cachorro, V., Gonzalez, M., Defrutos, A., 2005. Assessing vineyard condition with hyperspectral indices: Leaf and canopy reflectance simulation in a row-structured discontinuous canopy. *Remote Sens. Environ.* 99, 271–287. <https://doi.org/10.1016/j.rse.2005.09.002>
- Zeraatpisheh, M., Ayoubi, S., Brungard, C.W., Finke, P., 2019. Disaggregating and updating a legacy soil map using DSMART, fuzzy c-means and k-means clustering algorithms in Central Iran. *Geoderma* 340, 249–258. <https://doi.org/10.1016/j.geoderma.2019.01.005>



# **CAPÍTULO 2    DISAGGREGATION OF CONVENTIONAL SOIL MAPS: A REVIEW**

Este capítulo se corresponde con el artículo publicado (28/10/2021) en la revista *Soil Research* (10.1071/SR20288).

## **1 Introduction**

Soils fulfil essential environmental functions to sustain human life and, for this reason, the access to information on their vertical variability (soil typology) and horizontal distribution (soil geography) across the land is crucial to face the contemporary challenges (McBratney *et al.* 2014). Specifically, in agriculture and food security, there is a high demand for detailed information to support optimal land management.

The primary tools used to express this variability are the SRI and their corresponding maps. CSM methods applied during the 20th century have undertaken the division of the land into crisp delineations of the soil units recorded in the map legend based on qualitative soil-landscapes models. A large part of the cartography produced with CSM methods was developed on low-detail scales smaller than 1:25.000 and defined polytaxic SMU, which aggregate various components, a fact that imposes some limits on their specific application. Furthermore, CSM methods require a substantial investment in financial and highly qualified human resources, which increases exponentially on higher detail and accuracy levels (Rossiter 2000).

Recent advances in environmental data collection through technologies such as remote sensing and management tools such as GIS, together with the increase in computing power (Rossiter 2018) have fostered the development of new soil mapping methods based on digital data (Minasny and McBratney 2016). These have been brought together under a common framework called DSM (McBratney *et al.* 2003) or Predictive Soil Mapping (Scull *et al.* 2003).

DSM entails a new approach to the soil-landscape paradigm in which quantitative modelling of the relations between the elements is produced by means of statistical techniques (McBratney *et al.* 2003). The environmental domain is embodied by the SCORPAN factors, which extend the CLORPT model (Climate, Organisms, Relief, Parent material, and Age-time) and additionally include other

soil related information (S) and the spatial location (N) (McBratney *et al.* 2003). Data of these environmental factors stems mainly from remote sensing and geomorphometry in the form of continuous surfaces with a grid structure, hence there is information from all of them for every location or cell. Any soil information embedded in this structure can be incorporated then into the model function, allowing predictions for every cell. In this way, results of category predictions resemble monotaxic or single-component SMUs and also the mapping procedure can profit from legacy data of previous CSM cartography. Besides, the high availability and extensive coverage of continuous data sources used in DSM make this framework economically feasible for large scales (Hengl and MacMillan 2019).

A line of action of DSM is specifically aimed at harnessing the information available in legacy soil maps to generate a new cartography that enlarges the cartographic detail, in accordance with the categorical level that was captured in the original SRI (Gagkas and Lilly 2019). It purposes to depict the internal and non-explicit spatial distribution pattern of soils within polytaxic SMUs, and it is referred to as disaggregation (Nauman and Thompson 2014; Sarmiento *et al.* 2017). Its ultimate result is a coverage based on the most likely soil class per pixel (Nauman and Thompson 2014).

This broad review aims to collate information about soil disaggregation methodologies published to date, to compare them and to highlight those approaches reported to be advantageous as well as to identify future improvements for the development of more efficient methodologies. To our knowledge, this is the first review of this research line.

## **2 Bibliographical Search**

The bibliographical research was performed in the databases Web of Science (WOS) and Scopus with the following criteria: a) search in title, abstract and keywords, (TITLE-ABS-KEY); b) terms and booleans: “disaggregation” AND (“soil map” OR “soil class” OR “digital soil mapping”); and c) all years and all document types.

Results from each database were cross-checked, resulting in a final set of 80 publications. A comprehensive review of this set revealed that only 25 papers (32% of the total) cover soil disaggregation methodologies, modifications, or new implementations (27 applications in total). Whereas the rest for the most part perform other DSM lines of action on soil classes, such as

generalisation or updating, and, to a lesser extent, a few address soil properties but mention some of the keywords. The indicated 27 applications (Table 2-1) are the subject of this review.

The analysis by date of publication suggests that the interest in developing soil disaggregation methods is relatively recent (Figure 2-1). Although the first paper dates from 2001 (Bui and Moran 2001), the largest number of papers have been released from 2014 onwards. That year 3 seminal articles were published (Heung *et al.* 2014; Nauman and Thompson 2014; Odgers *et al.* 2014) and since then a major growth occurred in 2019.

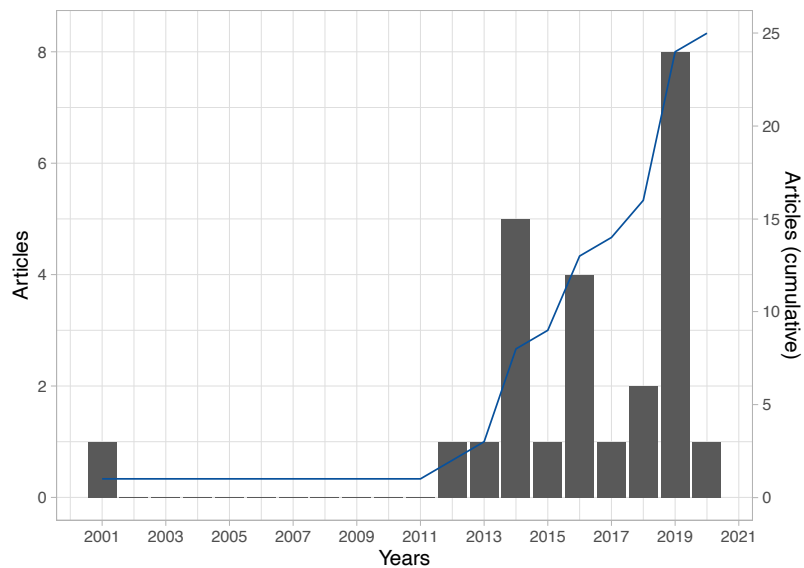


Figure 2-1. Distribution of published articles with disaggregation projects by year (bars) and cumulative articles over years (line).

### 3 Disaggregation methodologies on selected papers

The disaggregation methodologies analysed were tailored to specific SRIs. Nevertheless, all of them share common elements of DSM and disaggregation, which makes a comparative study feasible.

In broad terms DSM projects adhere to a general schema (Figure 2-2) that begins with a phase of data management of both soil and environmental datasets from the project region. On the environmental side, gathering and pre-processing of continuous data sources is undertaken in order to derive the environmental covariates and, where appropriate, the selection thereof. Afterwards, a quantitative soil-landscape model is generated using statistical classification techniques. In the final phase, the result is evaluated.

Table 2-1. Basic cartographic, data sources and covariates information of the selected disaggregation projects.

Reference	Location	Extent (km <sup>2</sup> )	Scale number	Soil classification system	Taxa	Grid size		Gain over sugg. Scale (%)	Data sources	Data sources pre-processing	S (n°)	C (n°)	O (n°)	R (n°)	P (n°)	A (n°)	N (n°)	Covariate selection method	
						Sugg- from scale <sup>1</sup>	Actual												
Bui and Moran 2001	New South Wales, Australia	339,000	500,000	Australian Soil Class.		250	250	0	DEM (AUSLIG)		0	0	0	1	0	0	0		
	Manilla, New South Wales, Australia	18,700	500,000	Australian Soil Class.		250	250	0	DEM (AUSLIG) Landsat MSS		0	0	4	1	0	0	0		
	Goondwindi, Queensland, Australia	245,000	250,000	Fluvial facies		125	250	-100	DEM (AUSLIG) Landsat MSS Digital lithology	DEM filter by a varying analysis window	0	0	4	3	1	0	0	0	
Häring <i>et al.</i> 2012	Bavaria, Germany	64,000	25,000	WRB	1*	12.5	10	20	DEM (LIDAR)	Removed settlements and roads	0	0	0	23	0	0	0	0	Relieff
Subburayalu and Slater 2013	Monroe County, Ohio, USA	1,174	24,000	ST	Series	12	10	17	DEM Geologic map Vegetation map Precipitation		0	2	1	22	1	0	0	0	CFS
Heung <i>et al.</i> 2014	Langley, British Columbia, Canada	5,472	25,000	Parent material classes		12.5	100	-70	DEM		0	0	0	27	0	0	0	0	
Nauman <i>et al.</i> 2014	Organ Pipe Cactus National Monument, Arizona, USA	1,340	24,000	ST	Series	12	30	-150	DEM (USGS NED) ASTER		14	0	15	25	0	0	0	0	PCA
Nauman and Thompson 2014	Pocahontas County, West Virginia, USA	4,000	24,000	ST	Series	12	30	-150	DEM (USGS NED) Landsat Geocover Geologic survey	Landsat PCA	0	0	4	9	1	0	0	0	
Odgers <i>et al.</i> 2014	Charters Towers, Queensland, Australia	68,000	250,000	Australian Soil Class.	Series	125	30	76	DEM (STRM) Landsat Gamma radiometric		0	0	7	13	3	0	0	0	
Subburayalu <i>et al.</i> 2014	Monroe County, Ohio, USA	1,174	24,000	ST	Series	12	10	17	DEM Geologic map Vegetation map Precipitation		0	2	1	22	1	0	0	0	CFS
Holmes <i>et al.</i> 2015	Western Australia, Australia	2,500,000	24,000 <sup>2</sup>	Western Australia Class.	Groups	12	85	-608	DEM Landsat TM		0	0	2	11	2	0	1	0	
Chaney <i>et al.</i> 2016	CONUS	8,000,000	24,000 <sup>2</sup>	ST	Series	12	30	-150	NLCD-CONUS Gamma radiometric		0	0	1	12	3	0	0	0	
Heung <i>et al.</i> 2016	Langley, British Columbia, Canada	5,472	25,000 <sup>2</sup>	Canadian Soil Class.	Great groups	12.5	100	-700	DEM Landsat TM	3 successive mean filters (3x3, 3x3, 5x5) PCA of geomorphic indexes	0	3	4	13	0	0	0	0	PCA
Liu <i>et al.</i> 2016	Manitoba, Canada	7,000	1,000,000	Canadian Soil Class.	Subgroups	500	90	82	DEM (STRM) MODIS		0	3	2	3	0	0	0	0	



Reference	Location	Extent (km <sup>2</sup> )	Scale number	Soil classification system	Taxa	Grid size		Gain over sugg. Scale (%)	Data sources	Data sources pre-processing	S (n <sup>o</sup> )	C (n <sup>o</sup> )	O (n <sup>o</sup> )	R (n <sup>o</sup> )	P (n <sup>o</sup> )	A (n <sup>o</sup> )	N (n <sup>o</sup> )	Covariate selection method
						Sugg. from scale <sup>1</sup>	Actual											
Pahlavan-Rad <i>et al.</i> 2016	Golestan Province, Iran	850	50,000 <sup>2</sup> ST		Series	25	30	-20	DEM (Aster GDEM) Landsat 5 TM API soil map		1	0	2	11	0	0	0	
Sarmiento <i>et al.</i> 2017	Serra Gaúcha, Brazil	673	50,000 Brazilian Soil Class.		Suborders	25	30	-20	DEM Landsat TM		0	0	9	12	0	0	0	
Machado <i>et al.</i> 2018	Porto Alegre, Brazil	497	50,000 Brazilian Soil Class.		Orders Subgroups	25	15	40	DEM		0	0	0	32	0	0	0	CFS
Vincent <i>et al.</i> 2018	Brittany, France	27,000	250,000 WRB		1 <sup>st</sup> 2 <sup>nd</sup>	125	50	60	DEM Land use map Gamma radiometric Soil property interpolat.		2	0	1	4	3	0	0	
Ellili Bargaoui <i>et al.</i> 2019	Ille-et-Vilaine, Brittany, France	6,848	250,000 WRB		1 <sup>st</sup> 2 <sup>nd</sup>	125	50	60	Same as Vincent <i>et al.</i> , (2018)		2	0	1	4	3	0	0	
Costa <i>et al.</i> 2019	Águas Frias, Santa Catarina, Brazil	27.7	25,000 Brazilian Soil Class.		Suborders	12.5	1	92	DEM		0	0	0	3	0	0	0	
Flynn <i>et al.</i> 2019a	Swartland, South Africa	3.66	250,000 WRB South African Soil Class.		Soil depths classes	125	10	92	DEM (ALOS-2) Sentinel 2A Lithology map	Sinks filling of DEM	0	0	3	11	1	0	0	Manually by importance
Flynn <i>et al.</i> 2019b	KwaZulu-Natal, South Africa	16.9	250,000 South African Soil Class.		Series	125	30	76	DEM (ALOS-2) Sentinel 2A		0	0	5	18	0	0	0	
Gagkas and Lilly 2019	Scotland, United Kingdom	78,763	250,000 WRB		HOST (Hydrology)	125	100	20	DEM WorldClim 2 20 Bedrock and quaternary deposits maps		0	4	1	27	2	0	0	Manually by importance
Jamshidi <i>et al.</i> 2019	Shiraz, Iran	165	50,000 ST		Subgroups	25	10	60	DEM Sentinel 2	2 filters (3x3, 5x5) on geomorphic indexes Temporal series	0	0	3	11	0	0	0	
Møller <i>et al.</i> 2019	Denmark	43,000	1,000,000 WRB		2 <sup>nd</sup>	500	30.4	94	DEM Landsat 8 Geological map Soil property interpolat.		0	6	12	20	4	0	0	
Zeraatpisheh <i>et al.</i> 2019	Borujen, Iran	860	1,000,000 ST		Great Groups Subgroups	500	30	94	DEM (Aster GDEM) Landsat ETM+		0	0	5	10	0	0	0	
Ellili-Bargaoui <i>et al.</i> 2020	Brittany, France	67,756	250,000 WRB			125	50	60	DEM Land use map Gamma radiometric Soil property interpolat.		2	0	1	4	3	0	0	

<sup>1</sup> Hengl 2006; <sup>2</sup> Lowest scale number in the range of used conventional soil maps.

CFS: Correlation-based feature selection; CONUS: Conterminous United States; DEM: Digital Elevation Model; HOST: Hydrology of Soil Types; PCA: Principal Component Analysis; ST: Soil Taxonomy; WRB: World Reference Base for Soil Resources.

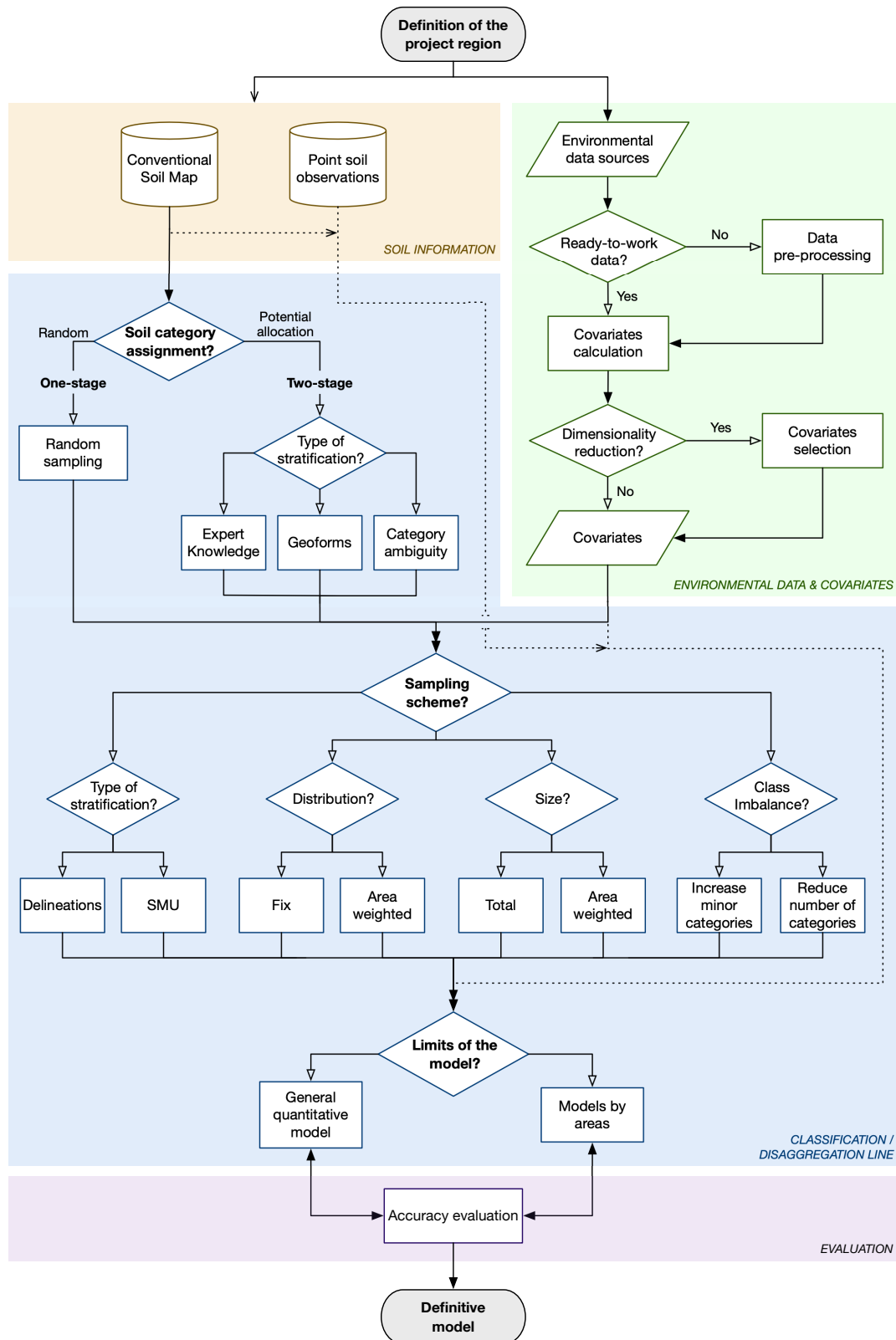


Figure 2-2. Flowchart of the disaggregation of a conventional soil map within the Digital Soil Mapping framework. The actions are organised into four sections: soil information, environmental data and covariates, classification, and evaluation. The primary route leverages the CSM map as the main data source for the classification and it is plotted in solid line. The secondary one takes the soil information from point observations while uses the map as a covariate and it is drawn in dotted line. SMU: Soil Map Unit

Specifically, the research line of the disaggregation (Figure 2-2) requires a conventional soil map concerning the soil dataset as a starting point. This map has intrinsic characteristics in terms of scale and cartographic and categorical detail, as well as other ones that are linked to the project region, such as the extent or the location. Such characteristics also extend to the disaggregated map. Ultimately, the distinctive feature of the disaggregation is how the conventional soil map is leveraged during the classification as well as the sampling scheme and the assignment of soil categories to support it.

All these key aspects are addressed throughout the review in dedicated sections that provide overviews on the variants explored and quantify their implementations.

## 3.1 The conventional soil map

### 3.1.1 Extent and location

With regard to the conventional soil maps subject to the disaggregation methodologies, they covered areas of a wide range of sizes (Table 2-1): relatively small (<10,000 km<sup>2</sup>), moderately large (10,000-100,000 km<sup>2</sup>), and large (>100,000 km<sup>2</sup>) areas, the former being the majority. These results agree with the trend reported by Holmes et al., (2015).

On the other hand, they spread across all the continents (Figure 2-3). They mainly took place in the United States of America and Australia (five each), with notable developments in Brazil, Canada, France, and Iran (three each).

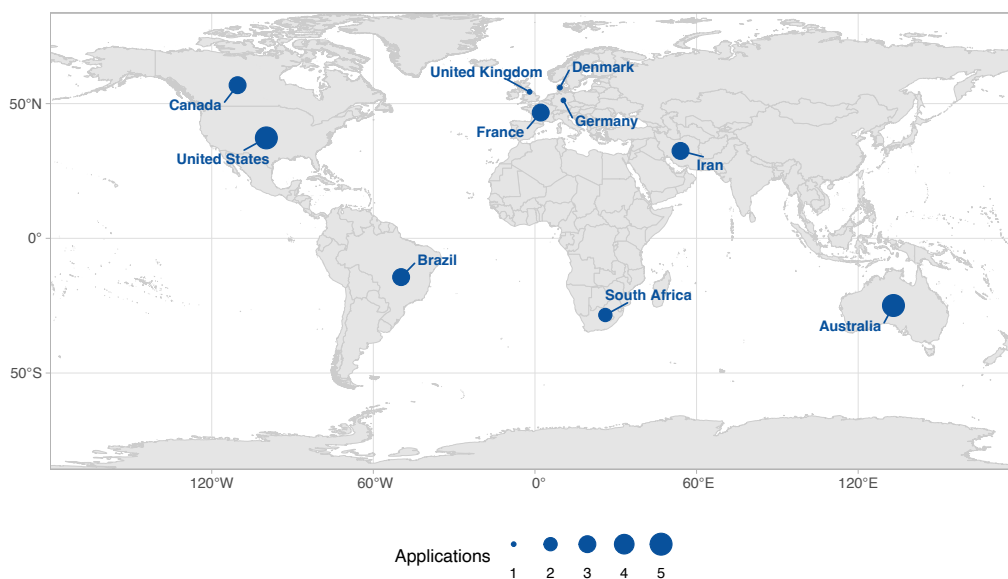


Figure 2-3. Location of the reviewed disaggregation applications by countries.

### *3.1.2 Scale*

The scale of the maps determines the scope in which its information can be used. In this regard, the National Resources Conservation Service of the United States Department of Agriculture (USDA NRCS) organised 5 orders, i.e., ranges of scales, with the aim of facilitating the transference of information accuracy and interpretation provided (Soil Survey Division Staff 2017). According to these orders, 10 applications featured very intensive and intensive conventional soil maps (orders 1 and 2), with scales ranging up to 1:25,000. Another 12 showed extensive ones (orders 3 and 4), between 1:50,000 and 1:250,000 scales. Finally, a few (five) were focused on very extensive conventional soil maps (order 5), from 1:500,000 to 1,000,000 scales (Table 2-1).

### *3.1.3 Categorical detail*

Most of the reviewed studies worked on conventional maps of soil categories from ordinary soil classification systems (Table 2-1), such as the ST (Soil Survey Division Staff 1999) and WRB (IUSS Working Group WRB 2014). However, sometimes SMUs were defined by categories of soil attributes such as parent material (Heung *et al.* 2014), soil depth (Flynn *et al.* 2019a) or hydrological properties (Gagkas and Lilly 2019).

### *3.1.4 Cartographic detail*

An important consideration is the quality of the original maps from the perspective of scale and cartographic detail agreement. This was evidenced in the disaggregation by the same approach of two maps from the same region at different scales, but within the same survey order. The larger-scale map did not provide adequate cartographic detail and a better model resulted from the smaller-scale one (Møller *et al.* 2019).

### *3.1.5 Scale gain during the disaggregation*

The unrefined outcome of the disaggregation process is an image with a grid structure, where each cell or pixel is associated with a probability of occurrence of one or several soil units. In the new disaggregation context, conventional delineations of monotaxic SMUs might be built from the aggregated representation of those pixels most probably associated with every soil unit. The reconfiguration of the existing delineations implies a decrease in the minimum legible delineation and the average-size delineation and, accordingly, an increase in the optimal scale of publication. All of these variations are interesting effects, but no explicit reference was usually made to them or to a new target scale in the applications studied.

Hengl (2006) noted a general relation between conventional vector maps and those in raster format, so that an optimal scale of the former would correspond to a grid size of the latter, given the level of visual perception. Based on this association, it is possible to derive an optimal scale of the disaggregated map from its actual grid size and, conversely, a suitable grid size for the representation of the CSM map in raster format from its scale. The comparison either in terms of scale or grid size between the two maps, the conventional and its disaggregation, would allow computing the improvement achieved during the disaggregation process (Flynn *et al.* 2019a).

The revision of the methodologies pointed out a positive gain in scale for the majority of them, with an average increase of around 60%. That means, for example, that a CSM map originally developed at 1:250,000 scale could be transformed to 1:100,000 following a disaggregation based on a grid size of 50 m. In a few cases the optimal scale of the disaggregated map was lower than the original one, given a grid size that would not be fine enough to capture the environmental variations (Table 2-1).

Furthermore, in most applications the optimal scale of the disaggregated map was listed in the same survey order as the one from the conventional map and, therefore, the categorical level would continue to be appropriate in the former. There are a few exceptions where the scale gap was too wide (Møller *et al.* 2019; Zeraatpisheh *et al.* 2019) and the categorical level was uneven. This imbalance might lead to the aggregation level of information not being able to reflect existing variations, i.e., small sets of distinct pixels may be overshadowed and not discernible during a global view of the new map.

## 3.2 Environmental data sources and covariates

In order to build the quantitative soil-landscape model, a wide array of covariates is created for every project to depict the environment in terms of the SCORPAN factors. For this purpose, most of the considered methodologies employed the open-source GIS applications SAGA (Conrad *et al.* 2015) and GRASS GIS (GRASS Development Team 2019).

### 3.2.1 Covariates and SCORPAN factors

The R factor (relief) was mainly characterised by geomorphometric covariates that operate at different levels: either local, such as slope, orientation, and different curvature; regional, such as catchment area, downslope flow path length and TPI; or as a combination of both, e.g.,

Topographic Wetness Index (TWI), Multiresolution Index of Valley Bottom Flatness (MRVBF) and Floodplain Index (Florinsky 2016). They were derived out of DEMs from sources with regional coverage, such as AUSLIG (Bui and Moran 2001), USGS-NED (Nauman and Thompson 2014) and those generated from LIDAR; or global, such as ALOS-2 (Flynn *et al.* 2019a), ASTER (Pahlavan-Rad *et al.* 2016) and STRM (Liu *et al.* 2016). In this regard, predictions with SRTM-derived covariates were noted to be more suitable than those derived from ASTER (Massawe *et al.* 2018). DEMs were often processed prior to their usage by filtering (Bui and Moran 2001) as well as by filling in spurious sinks (Silva *et al.* 2016; Flynn *et al.* 2019a) to reduce the artefacts that created noise in the derived covariates and, in turn, in the classification. Following the same goal, pixels representing roads and settlements were also excluded (Häring *et al.* 2012). Overall, this type of covariate is present in all methodologies and predominates within them (Table 2-1).

The covariates associated with O factor (organisms) mainly consisted of cover maps and vegetation indices derived out of reflectance images from remote sensing (Nauman and Thompson 2014). The main sources for reflectance images were the various sensors present on the Landsat satellites, such as the Multispectral Scanner System (MSS) (Bui and Moran 2001), Thematic Mapper (TM) (Sarmiento *et al.* 2017), Enhanced Thematic Mapper Plus (ETM+) (Zeraatpisheh *et al.* 2019) or Operational Land Imager (OLI) (Levi 2017). Images from the Moderate Resolution Imaging Spectroradiometer (MODIS) have been also incorporated (Liu *et al.* 2016) and, more recently, images from Sentinel 2 too (Flynn *et al.* 2019b). They all essentially cover VIS, near IR (NIR) and short-wave IR (SWIR) spectral regions. Operation of these images depends on the pre-processing level at which they were released and controls on them focus on cloud cover (Young *et al.* 2017). The O factor, for its part, appeared in most of the methodologies and represented a significant share of the covariates in each.

Those covariates associated with the remaining SCORPAN factors were used to a lesser extent and accounted for a small portion in each project. The P factor (parent material) was often depicted by conventional maps of lithology or parent material. Gamma radiation measures have been also adopted to this end, taking advantage of their indirect connection to the mineralogy of the soil surface materials (Holmes *et al.* 2015). Covariates related to the C factor (climate) were available from datasets such as WorldClim (Gagkas and Lilly 2019) or interpolations from weather stations (Scull *et al.* 2005). Likewise, for the S factor (soil), surfaces of soil attributes interpolated from soil sites were employed (Vincent *et al.* 2018). Besides this, indices derived from remote sensing images were used (Nauman *et al.* 2014), as well as CSM maps directly (Pahlavan-Rad *et al.* 2016;

Leenaars *et al.* 2020). Lastly, the N factor (space) had a marginal usage, while no covariate was explicitly associated with the A factor (age).

Data from each source were retrieved with varying grid sizes and then rescaled to a common reference, usually based on the DEM. The grid size of the projects ranges from 1 to 250 m, with a median of 30 m (Table 2-1).

### *3.2.2 Multi-scale covariates*

The adjustment of the pixel size together with the analysis window during the calculation of the covariates enables to capture the terrain characteristics that are revealed at different scales (Smith *et al.* 2006). This way, it is possible to incorporate multi-scale spatial information into the soil-landscape model.

Such procedures are embedded directly in regional geomorphometric covariates, such as the MRVBF (Gallant and Dowling 2003), or through specific algorithms with varying analysis windows (Wood 1996). Besides, they can be introduced a posteriori in a generalised way through the application of local filters (Heung *et al.* 2017; Levi 2017).

Complementary, multi-scale information in the time domain was integrated through multi-annual time series of indices based on reflectance values, which allow detecting constant patterns over the variability linked to the environmental noise (Levi 2017; Jamshidi *et al.* 2019). Although, its adoption is not yet widespread.

### *3.2.3 Covariate selection*

The collection of covariates specific to each project changes in response to the characteristics of each location.

In this regard, it was reasoned by expert-knowledge the preferential use of geomorphometric covariates associated with the R factor in a glacial landscape, where soil variability is mainly tied to the geomorphology and the deposit of materials (Heung *et al.* 2014). However, it was pointed out its limitations in contrasting environments mainly conditioned by the parent material (Flynn *et al.* 2019b). In general, the higher detail of DEMs allows for greater granularity of the spatial variability captured by the geomorphometric covariates compared to, for example, a conventional parent material map (Collard *et al.* 2014). All these considerations might be coupled with

cartographic factors, i.e., the overall availability of DEMs and the software used to derive the full set of covariates.

Furthermore, some studies relied on data-driven techniques to identify the most relevant covariates in the entire set for revealing the variability within the soil map. One approach was the unsupervised methods, which don't depend on the classification procedure itself, such as PCA (Nauman *et al.* 2014; Heung *et al.* 2016, 2017), or, more sophisticated, on the analysis of the variance (Behrens *et al.* 2010). A serious disadvantage of PCA lies in the fact that it gives more weights to covariates with high variability, regardless of whether they adequately represent the environment (Behrens *et al.* 2010) and, thus, the discriminating power of the chosen set depends to some extent on the initial selection made by the surveyor.

Another approach for dimensionality reduction was supervised methods such as the correlation-based feature selection (CFS) (Subburayalu *et al.* 2014; Machado *et al.* 2018). Unlike PCA, CFS heuristically assesses the possible subsets of covariates by evaluating their predictive performance against the degree of internal correlation. It then selects preferentially the one that maximises the former while lowering the latter. It's implemented, for example, in the Weka software (Frank *et al.* 2016).

In this line of model evaluation, embedded feature selection techniques were also employed such as the importance given to covariates in DT-based methods. This is calculated according to the proportion of leaves in a tree that require a given covariate at some point across the if-then decision rules (Odgers *et al.* 2014; Vincent *et al.* 2018). For the RF algorithm, an ensemble method of DT, another way of measuring the covariate importance was also favourably reported in terms of the Mean Decrease Accuracy (MDA), which gives the average accuracy for the predictor minus the decrease in accuracy after its permutation (Gagkas and Lilly 2019).

Based on this metric, a manual stepwise selection of covariates was carried out in recursive attempts until the accuracy level started to decrease significantly (Flynn *et al.* 2019a; Gagkas and Lilly 2019). The final set was the one with the lowest classification error in terms of the mean OOB error rate (Gagkas and Lilly 2019). This was a time-consuming action (Flynn *et al.* 2019a) that was pointed out may have a minimal impact on the RF performance, as this classifier may cope with potentially non-informative predictor covariates (Heung *et al.* 2014; Gagkas and Lilly 2019). This type of feature selection was performed after a dimensionality reduction through PCA too, albeit with poor effects on the accuracy level (Flynn *et al.* 2019a).



On average, 19 covariates on every project were created before any type of selection up to a maximum of 54 (Table 2-1). The ultimate set of covariates used to build the models depended on the approach taken and the criteria established. On the basis of the aforementioned statements, a thorough data-driven selection supported by expert knowledge may be well suited for methodologies other than those using DT-based classification methods. Given the qualities of the latter to deal with non-informative covariates, the use of expert knowledge might focus on improving computational performance for these cases. Likewise, the importance metric might be used as a proxy for examining the environmental relations.

### 3.2.4 Covariate importance

Some of the covariates were identified as important within the various studies (Figure 2-4). The elevation is at the top of the list in terms of the number of instances and, also, it recurrently ranks among the most important. Other highlighted features in this domain are TPI and the relative elevation to the nearest stream or channel network.

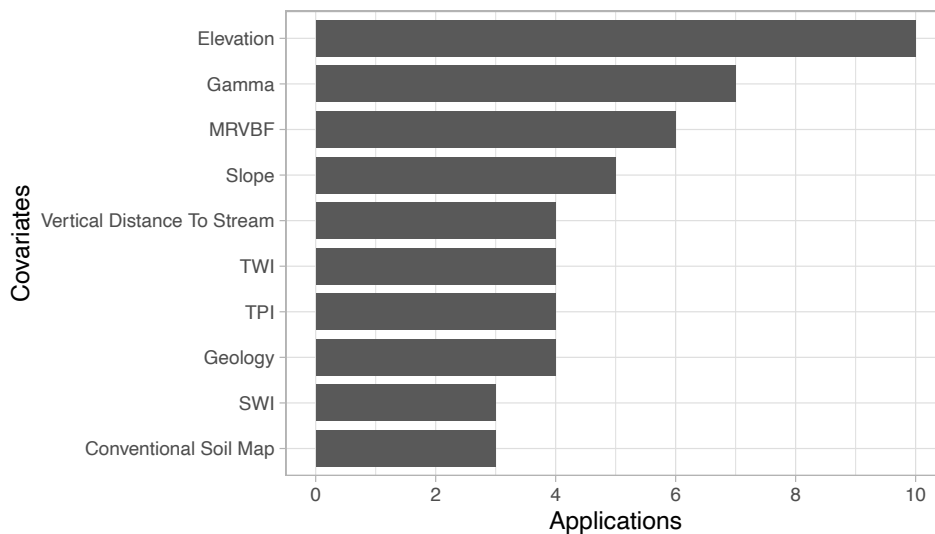


Figure 2-4. 10 most important covariates among reviewed disaggregation applications. MRVBF: Multiresolution Index of Valley Bottom Flatness; STWI: Índice Topográfico de Humedad de SAGA; TPI: Índice de Posición Topográfica; TWI: Índice Topográfico de Humedad.

In second place come the series of gamma radiometric covariates which include K, Th, U isotopes or the K/Th ratio, despite their relatively small grid size which would be potentially less effective in capturing spatial variations. Next, another group could comprise MRVBF, TWI, SAGA modified TWI (STWI) and the slope, i.e., morphometric and hydrologic covariates associated with material erosion and deposit accumulation areas. When categorical covariates such as the conventional soil maps or the bedrock geology or parent material maps have been used, they have consistently

taken a prominent role in comparison to the rest of covariates. This would suggest their value as well as a way to implicitly exploit the expert knowledge captured within this type of maps.

Even though categorical maps of land use associated with the O factor have had a positive response, continuous covariates of this factor such as NDVI or Soil Adjusted Vegetation Index (SAVI) have usually performed poorly. Møller et al. (2019) rationalised it on the grounds that the scale of the input map couldn't potentially accommodate the detailed pattern of the vegetation. In this context, a noteworthy drawback concerning vegetation-related continuous covariates is the fact that they are heavily affected by anthropogenic activities, which can lead to distortions in their patterns, e.g., indexes based on reflectance values from agricultural areas (Liu *et al.* 2016). Hence, the potential benefits of integrating time-series into the studies.

The preponderance of the R factor within the set of the most important, but also among the project collections, could be ascribed to pedological criteria, due to its capacity to interrelate with other factors, such as S and C (Hengl and Reuter 2009). In this regard, it was reported that there was no clear trend towards the designation of local vs regional geomorphometric parameters as important covariates (Häring *et al.* 2012). However, the majority of those listed here are of type regional. Conversely, the group of curvatures and the ASPECT were of low interest as happened in Møller et al. (2019).

### 3.3 Classification of soil classes

#### 3.3.1 *Leverage of the conventional map in the soil-landscape model*

The proposed disaggregation methodologies took distinct approaches to capitalise on the information in CSM maps.

Often, CSM maps and their SMUs have been used as the source for defining soil category information. When references for these categories came only from the map legend or they were deemed uncertain, unsupervised classification methods, such as k-means (Bui and Moran 2001) or its fuzzy logic variant (Zeraatpisheh *et al.* 2019), were applied to outline the distribution. Whereas when more precise knowledge on topo-sequences was available, expert knowledge-based rules were used (Bui and Moran 2001). A more advanced approach to both strategies is the supervised classification and Machine Learning methods, which were implemented in most of the analysed methodologies (89%) (Table 2-2). Among these, the various DT algorithms and RF, an

ensemble of DT-based models, stand out due to their wide adoption, supported by the combination of complex environmental models with accessible interpretation (covariate importance) and high validation rates. These advantages come at the cost of intense computational effort and expertise to manage it compared to the less demanding unsupervised methods. They are featured instead by simpler models but subject to expert knowledge correlations between the output and soil classes.

One complete disaggregation process was developed on DT called "Disaggregation and Harmonisation of Soil Map Units Through Resampled Classification Trees" (DSMART) (Odgers *et al.* 2014). Through it, multiple occurrences of DT are created by performing independent samples that are then combined, so that the aggregation of category probabilities reveals the final pattern.

Otherwise, some methodologies considered the CSM map as a categorical covariate that acts as an independent variable within the model (Pahlavan-Rad *et al.* 2016). It enhances a classification based on the soil information derived directly from point observations in the field. To this end, supervised methods such as the aforementioned RF or MLR were applied (Table 2-2). A variant of this system took direct advantage of the map's delineations to restrict the scope of the supervised models, so that a single model was generated for each SMU from the observations contained therein (Häring *et al.* 2012). In addition, a single methodology integrated aspects of the two aforesaid strategies (Vincent, 2018). It primarily defines soil category information from the CSM and then uses the SMU distribution as an independent covariate.

Most of the considered methodologies relied on the implementations of the above-mentioned statistical techniques in the open-source programming languages R (R Core Team 2020) or Python (Van Rossum and Drake 2009) and which are available within their many packages and modules.

### *3.3.2 Assignment of soil categories: training sets*

When CSM maps are used as ancillary data, soil categories are defined from field observations and their information is assigned to those individually intercepted pixels in order to create the training set to build the supervised models (Hengl *et al.* 2007). Generally, these observations come from legacy data of pre-existing SRIs (Häring *et al.* 2012; Liu *et al.* 2016). They were collected through guided samplings, and, for this reason, their locations mostly respond to key places identified by the cartographer under their bias and the conditions established. For one methodology (Pahlavan-Rad *et al.* 2016) new observations were acquired using the cLHS system (Minasny and McBratney

2006). This is a special type of stratified random sampling where locations are determined on the space defined by the covariates ( $n$ -dimensional) looking to maximise the captured variability. It is efficient by being able to represent the multivariate distribution of the covariates using relatively small sample sizes (Minasny and McBratney 2006).

Alternatively, the soil information used to compose the training sets is extracted directly from the CSM maps, by proportionally assigning the categories listed in the map legend to randomly sampled pixels within the SMU delineations. The main advantages of the map mining approach are associated with its capacity to arbitrarily generate large samples (Heung *et al.* 2017) multiple times (Gagkas and Lilly 2019), which captures a higher degree of variability in the landscape as well as in the multivariate space considered (Heung *et al.* 2017). A way of translating these benefits into observation-based methodologies involves taking pixels within circular buffers over points (Silva *et al.* 2016). An inherent issue in mining approach is connected to the variability and purity within the SMU itself at certain scales. In this regard, high detail studies reported substantially higher levels of accuracy (Heung *et al.* 2014, 2017).

A comparison between predictions established with soil category information from field point observations versus those from SMU delineations revealed that the accuracies rates were consistently higher in the latter regardless of the model used, including the aforementioned DT, MLR and RF with variations (Heung *et al.* 2017). Indeed, only monotaxic SMUs were considered in this case to build the models for the entire area. Further research could potentially compare these results against the evaluation of a disaggregation based only on polytaxic SMUs versus one based on soil observations either legacy or newly sampled. This would provide another assessment of the conditions for a better performance regarding the two types of disaggregation procedures.

Unsupervised methods or expert knowledge-based rules do not require the definition of training sets. Instead, they rely respectively on delineations and proportions described in the map legend for each SMU as well as on all the supplementary information available to search over the entire area and thus establishing the final allocation (Zeraatpisheh *et al.* 2019).

### *1.1.1 Sampling schemes over map units*

The methodologies that leverage the inner boundaries of CSM maps followed different random sampling schemes in order to create the training sets for supervised classifications.

Table 2-2. Classification and evaluation methods for the selected disaggregation projects.

Reference	Classification method	Soil data source	Stages	Soil class assignment	Evaluation	Soil obs. for evaluation	Ratio obs. to project extent (obs/km <sup>2</sup> )		Overall Accuracy	
							Regular	Nearest neighbour	Up to 3 <sup>rd</sup> most probable	
Bui and Moran 2001	EK Rules	CSM map	1	Manually by report and expert knowledge						
	K-means	CSM map	1	Manually by report and expert knowledge						
Håring <i>et al.</i> 2012	DT C5.0 (x10)	CSM map	2	Stratified area-weighted random sampling by facies; 25% of facies						
	RF. Conditional Inference DT (x500)	Profile	1	Balance of soil class proportions within SMUs	Probability > 0.7 Out-of-bag error	1,820	0.028	70		
Subburayalu and Slater 2013	DT C4.5 / J48	CSM map	2	6 predominant classes (87% of the map legend); kNN disambiguation; 5 pixels by polygon; equal size sampling	Soil observations	164	0.14	34		
	RF (750)	CSM map	1	3 methods: equal class sampling; 5 pixels by polygon, stratified, area weighted random sampling by SMU (only monotaxic SMUs)	Soil observations Consistency with monotaxic SMUs; proportions with polittaxicas SMUs Out-of-bag error	307	0.056	78	85	
Nauman <i>et al.</i> 2014	DT CART (x500) Ensemble	CSM map	2	EK rules over 1~3 covariates; stratified, area-weighted random sampling by STU; resample with replacement for each tree; equal class balancing	Soil observations	53	0.04	69.8		
Nauman and Thompson 2014	DT CART (x100) Ensemble	CSM map	2	EK rules over 6 covariates; stratified, area weighted random sampling by STU with replacement; class balance by modified area	Soil observations Training set agreement Visual SSURGO comparison	87	0.022	24	44	64 (3x3)
Odgers <i>et al.</i> 2014	DSMART: DT See5 (x100)	CSM map	1	m = 15	Soil observations Confusion Index	285	0.004	22.5		50
Subburayalu <i>et al.</i> 2014	DT C4.5/J48 NS-PDT Clust-PDT	CSM map	1	5 pixels by polygon; 6 predominant classes (87%); equal size sampling	Soil observations	106	0.09	39		
Holmes <i>et al.</i> 2015	DSMART: DT C4.5	CSM map	1		Soil observations Confusion Index	43,351	0.017	20		41
Chaney <i>et al.</i> 2016	RF	CSM map	1	Non-overlapping target boxes (30x30km) + buffer; 50,000 pixels for each target; stratified, area weighted random sampling by SMU	Soil observations	29,4746	0.037	17	48	

Reference	Classification method	Soil data source	Stages	Soil class assignment	Evaluation	Soil obs. for evaluation	Ratio obs. to project extent (obs./km <sup>2</sup> )	Overall Accuracy		
								Regular	Nearest neighbour	Up to 3 <sup>rd</sup> most probable
Heung <i>et al.</i> 2016	DT CART kNN LMT MLR ANN NSC RF SVM-Linear SVM-RBF	CSM map	1	Monotaxic SMUs: 1) equal-class sampling, 2) by-polygon sampling, 3) area-weighted sampling, 4) area-weighted sampling ROS (duplicating number of samples for minority classes)	Soil observations Consistency with monotaxic SMUs, proportions with politáxicas SMUs	262	0.048	52	66	
Liu <i>et al.</i> 2016	Similarity-based linear regression	Profile	1	560 profiles + SMU delineations	Soil observations	186	0.027	67		
Pahlavan-Rad <i>et al.</i> 2016	RF (100-200-700) MLR (x100)	Profile	1	cLHS	Soil observations (by cLHS) Leave-group out cross validation	99	0.116	44		
Sarmiento <i>et al.</i> 2017	DT	CSM map	2	EK rules (LandMapr + slope + SMU); stratified, area-weighted random sampling by zones (10% of pixels)	Soil observations Proportions Visual analysis	233	0.346	34	48	
Machado <i>et al.</i> 2018	DT C4.5/48	CSM map	2	2 set of EK rules: well drainage areas (LandMapr + slope) and poor drainage areas (TWI); stratified random sampling by STU with area weighted (2.5 pixels/ha)	Soil observations	72	0.145		80	
Vincent <i>et al.</i> 2018	DSMART: DT C4.5 (x50)	CSM map	2	EK rules by code over: soil parent material, waterlogging index, TPI, SMU; stratified random sampling by EK polygons with area weighted (at least 1 obs./5 ha) + by legacy polygons (m = 30)	Soil observations Visual analysis 1:25,000 map	3,761	0.139	19.5	30.8	67 (3x3)
Elili Bargaoui <i>et al.</i> 2019	Same as Vincent <i>et al.</i> , 2018	CSM map	2	Same as Vincent <i>et al.</i> , 2018	Independent soil observations: stratified sampling split in overall, strata, hillslope section Evaluation of 4 categories: parent material, soil type, drainage, depth Partial, mean, and strict purity	405	0.059	60		
Costa <i>et al.</i> 2019	DT /48	CSM map	2	EK rules: Stratified random sampling by map units with area weighted	Soil observations Proportions	20	0.722	44.4	52.8	
Flynn <i>et al.</i> 2019a	DSMART: DT C5.0 (x100) DSMART RF DSMART MLR	CSM map	2	Geomorphons and expert delineation by API lagged by slope position; m = 15	Soil observations: by cLHS (46) and expert location (47)	93	25.41	68		90

Reference	Classification method	Soil data source	Stages	Soil class assignment	Evaluation	Soil obs. for evaluation	Ratio obs. to project extent (obs./km <sup>2</sup> )	Overall Accuracy		
								Regular	Nearest neighbour	Up to 3 <sup>rd</sup> most probable
Flynn <i>et al.</i> 2019b	DSMART + kNN, NSC, LDA, MRR, C5.0, RF, SGB, SVM-Linear, SVM-RBF, MLP (NN)	CSM map	2	Topographic Index, Landform Classification (TPIC); m = 15	Soil observations	145	8.58	70		
Gagkas and Lilly 2019	RF (x500)	CSM map	1	Different TS per iteration; area weighted random sampling by HOST 1 class (*1, 5, 10 pixel/km <sup>2</sup> ); Random assignment by proportions within polytaxic SMUs	Soil observations Confusion Index	718	0.009	34		52
Jamshidi <i>et al.</i> 2019	DSMART: DT C5.0 (x50)	CSM map	1	m = 15	Soil observations Confusion Index	78	0.473	68		
Møller <i>et al.</i> 2019	RF (100) DT C5.0 (x50)	CSM map	2	EK rules (wetlands + soil texture)	Soil observations	777	0.018	18		
Zeraatpisheh <i>et al.</i> 2019	FCM (7-8, x100) K-Means (7-8, x100) DSMART (x10)	CSM map	1	m = 15	Soil observations (by cLHS) Shannon Entropy Index	125	0.145	33		36
Elliil-Bargaoui <i>et al.</i> 2020	DSMART: DT C5.0 (x100) DSMART + EK rules DSMART + soil observations	CSM map	2	EK rules as Vincent <i>et al.</i> , 2018; m = 30; soil observations	Soil observations (independent) V-measure Shannon Entropy Index	755	0.011	18	23.4	50.2

ANN: Artificial neural networks; cLHS: Conditioned Latin Hypercube sampling; CSM map: Conventional Soil Mapping map; Clust-PDT: Clustering possibilistic DT; DT: Decision tree; EK rules: Expert-knowledge rules; FCM: Fuzzy c-means; kNN: k-nearest neighbour; LDA: Linear discriminatory analysis; LMT: Logistic model tree ; MLP: Multilayer perceptron; MLR: Multinomial logistic regression; MRR: Multinomial ridge regression; NSC: Nearest shrunken centroid; NS-PDT: Non-specificity possibilistic DT; RF: Random forest; SGB: Stochastic gradient boosting; SVM-Linear: Linear support vector machines; SVM-RBF: support vector machine with radial basis function.

Some set the limits of the SMUs as a reference, and based on them, area-weighted stratified samples were taken. The sample size was set either in pixel ratios per area unit (Gagkas and Lilly 2019) or in proportions over a maximum value in accordance with the map area (Sarmiento *et al.* 2017). The latter was motivated as a balance between the computational feasibility and appropriate characterisation of the relationships among soils and the environment. In addition, a fixed size sampling for each category was tested (Heung *et al.* 2016).

Other methodologies performed random sampling narrowed down to the delineation limits regardless the SMUs, e.g., the DSMART system (Odgers *et al.* 2014), in which a fixed number of pixels per delineation is sampled ( $m$  parameter). This factor has been given two main values in later applications, i.e., 15 (Holmes *et al.* 2015; Flynn *et al.* 2019b; Jamshidi *et al.* 2019; Zeraatpisheh *et al.* 2019) and 30 (Vincent *et al.* 2018; Ellili Bargaoui *et al.* 2019). The DSMART methodology has been enhanced by combining its sampling method with a random selection weighted by delineation areas (Vincent *et al.* 2018).

Some disaggregation studies compared the aforementioned methods of stratification and allocation of samples. Altogether, the area-weighted allocation scheme increased accuracy and consistency both when categories (Heung *et al.* 2016) or polygons (Møller *et al.* 2019) were involved. On sampling boundaries, Gao *et al.* (2020) recently suggested that delineations should be preferred over categories, because the latter could introduce higher variability into the models. Although, this has not yet been explicitly tested. Finally, it was noted the influence of the sample size on the classification errors in RF, achieving fewer misclassifications at the highest density (Gagkas and Lilly 2019).

A common concern during random sampling over categories is class imbalance, meaning the under-representation of minor categories in the training sets. This fact limited the predictive ability of the models (Subburayalu *et al.* 2014; Levi 2017) and was the stated reason for low evaluation or validation ratios (Levi 2017). Two possible strategies to overcome the cited inconvenience are the relative increase of observations in the least represented categories, and the reduction of the number of categories to be modelled (Brungard *et al.* 2015; Levi 2017).

By repeating samples from smaller categories, it is possible to increase their amount (Heung *et al.* 2016). Alternatively, the rebalancing of larger categories can be achieved by the modification of the weighting criterion (Nauman and Thompson 2014) and by levelling their size using a fixed value (Subburayalu and Slater 2013).



To reduce the number of categories, Subburayalu and Slater (2013) selected those most representative; Heung et al. (2014) used higher categorical levels, and Häring et al. (2012) built individual models per group of SMUs.

### *3.3.3 Assignment of soil categories in two-stages*

The DSMART methodology considers the CSM map with its legend for the disaggregation process without any further information. On that premise, all training sets are built by stratified random sampling and every selected pixel within a SMU is randomly assigned to one of the soil categories listed therein (Odgers *et al.* 2014). Additionally, the authors noted the potential for making use of knowledge on the studied region to circumscribe the category distribution and thus improve the training sets.

The first implementations of DSMART were conducted over large areas under distinct environmental conditions and different scale and categorical levels (Holmes *et al.* 2015; Chaney *et al.* 2016). However, these studies pointed out similar limitations in terms of accuracy and consistency and noted again, as a future improvement, the assignment of soil categories to pixels supported by environmental references.

In this way, subsequent disaggregation approaches will try to solve the cited limitations by developing new methodologies in two stages: firstly, areas potentially associated with soil categories are defined in order to label the matching pixels, and from these, secondly, supervised models are created. These strategies profit information from the CSM legend and expert knowledge and seek to reduce the incidence of misleading allocations.

A methodology contemporary to DSMART (Nauman and Thompson 2014) took in the environmental information available about the STU (USDA-NRCS Soil Survey Geographic Database, SSURGO) to establish rules, based on some geomorphometric covariates, in order to identify the potential areas within the SMU in which each STU is present. These areas are the basis for drawing the training sets up with stratified random sampling. The authors pointed out that this methodology gave accurate results.

Recent DSMART applications described expert knowledge-based rules using soil databases, such as DoneSol (Vincent *et al.* 2018; Ellili Bargaoui *et al.* 2019), for this purpose. Analogous DT-based methodologies also implemented this same approach (Sarmiento *et al.* 2017; Costa *et al.* 2019).

Occasionally, the rules have been tailored to a relevant factor with distinct areas in the study, e.g., the soil drainage capacity (Machado *et al.* 2018).

Møller *et al.* (2019) did not demarcate potential areas for the classes when defining soil-landscape rules, but instead some factors were taken into account to strip part of the variability within the SMUs. Delineations of a small number of SMUs were manually reshaped with regard to soil textural classes and the extent of wetlands supplied via additional maps. By doing so, the implementation of rules was later compared to the genuine DSMART procedure, and it was shown that they enhanced the model accuracy.

These expert knowledge-based rules are intimately related to the concept of geofoms, whose boundaries demark the action of pedological and hydrological processes of transport, accumulation and leaching of materials (Evans 2012). Accordingly, some of the analysed methodologies have outlined their divisions by means of a TPI classification (Flynn *et al.* 2019b) or using landscape analysis tools such as LandMapr (Machado *et al.* 2018). Afterwards, they were manually associated with the soil categories based on the existing knowledge on the area (Lelyk *et al.*, in: Arrouays *et al.* 2020). Thus, the demarcated geographical area enables the subsequent identification of a region in the covariate space, in contrast to the soil-landscape rules.

Besides these, other systems have been developed to stratify the landscape into geofoms, although they have not been incorporated to this end in disaggregation methodologies. These also comprise clustering algorithms such as Self-Organizing Data Analysis Technique (Ehsani and Quiel 2009), fuzzy k-means clustering (Irvin *et al.* 1997; Schmidt and Hewitt 2004), and nested-means (Iwahashi and Pike 2007). The application of these systems and its comparison against the approaches already adopted could be a matter for future research. In broad terms, these methods are subject to the scale and the analysis window provided, which may or may not be appropriate for the soil variability present, and require a considerable amount of computational power.

There is also a novel system built atop computational vision, a new paradigm beyond differential geometry and the analysis window, called geomorphons (Jasiewicz and Stepinski 2013) that have been recently introduced into a disaggregation methodology (Flynn *et al.* 2019a). Based on the generalization of simple ternary patterns that conform to particular terrain morphologies, geomorphons allow classifying and mapping landform elements from a DEM with high computational efficiency and robustness (Jasiewicz and Stepinski 2013). In a first attempt, these geomorphons were grouped and manually related to soil depth categories (Flynn *et al.* 2019a),

although a newer implementation has been carried out through statistical procedures (Flynn *et al.* 2020). Both studies pointed out how stratification by geomorphons improves the accuracy over the genuine DSMART implementation. Ultimately, it has been shown that proposed two-stage methods boost the classification performance.

In addition to the two strategies discussed above in terms of soil-landscape rules and geoforms, some variants for handling the ambiguity in soil class allocation were also carried out using *k*-Nearest Neighbours (Subburayalu and Slater 2013). They rely on knowledge of the potential location of dominant STUs of each SMU in the covariate space to re-assign the closest instances. The minor classes were not correctly predicted though, mostly due to a class imbalance problem that caused suboptimal predictive information.

Rather than the two-stage sequence, another approach to overcome the shortcomings of soil category assignment adopted variants of a supervised classification algorithm capable of working with fuzzy distributions rather than individual categories, such as the Possibilistic DT group (Subburayalu *et al.* 2014). Hence, the entire content of the legend for each SMU is directly involved and it is the algorithm itself that accounts for the partitions.

### 3.4 Evaluation of soil-landscape models

The development of a new method commonly concludes with the validation or evaluation phase, which targets the degree of error and uncertainty in the model predictions. In the reviewed disaggregation methodologies, this evaluation process focused on the categorical accuracy, that is, the similarity between the predicted and real value at the same location. Comparisons with field point observations and other categorical maps were made to this end.

#### 3.4.1 Accuracy evaluation from observations

All the reviewed applications utilised soil point observations to evaluate their results (Table 2-2). These soil observations came mostly from legacy data of previous SRIs collected through guided sampling, similarly to the training sets for the classification. Some papers showed alternatives on the basis of new sets of observations randomly located, which benefit robust applications of the statistical methods. For this purpose, the cLHS system was adopted (Zeraatpisheh *et al.* 2019) and also combined with locations selected by expert knowledge (Flynn *et al.* 2019a). Besides, stratified random sampling over physiographic units (Håring *et al.* 2012) or multiple levels of landscape

units, from general to specific: global, by strata, and by transects along the slope (Ellili Bargaoui *et al.* 2019) were employed. The high cost of these alternatives limits their implementation and explains the low number of studies that introduce them.

The set of observations employed in assessing the model accuracy must be exclusive to this purpose, otherwise it would invalidate the evaluation. In the case of methodologies that extract the soil information from CSM maps, the full set of observations can be used. On the contrary, when the disaggregation method relies on them for the classification, two different sets are required. Thus, if a single set either legacy or new is available, it must be split into training and validation subsets. Several subset divisions can be repeated over n-iterations to maximise the benefit from the available data by applying some of the variants of cross-validation, e.g., allocating 70% and 30% to the subsets in each iteration (Pahlavan-Rad *et al.* 2016; Levi 2017).

The evaluation of soil categories involves specific metrics that compare the number of observations that are predicted successfully and incorrectly. The most common metric of this type in the reviewed methodologies is the OA, which comes from the error matrix as the proportion of correct predictions out of the full set of observations and, thus, ranges from 0 to 100%. Another measure used is the Kappa coefficient, a statistical test that assesses the performance of the classification against a random assignment (Heung *et al.* 2014; Liu *et al.* 2016; Pahlavan-Rad *et al.* 2016; Flynn *et al.* 2019a, 2019b; Gagkas and Lilly 2019; Ellili-Bargaoui *et al.* 2020). This coefficient ranges from -1 to 1, where values below 0 indicate that the classification is worse than a random assignment and those close to 1 indicate that it is significantly better (Brus *et al.* 2011). Further proposals are evaluation parameters of the statistical technique used to build the disaggregation models, e.g., OOB error in DT and RF-based models (Häring *et al.* 2012) based on bootstrap aggregation or bagging.

A more comprehensive analysis proposal considered the agreement on multiple categories besides soil class, including the parent material, drainage, or presence of redoximorphic conditions and the depth class (Ellili Bargaoui *et al.* 2019). On this basis, various purity values were determined: partial, in respect of one of the criteria; average, by weighing the whole set; and strict, if all criteria were met simultaneously.

The level of reliability of legacy data is not even and it impacts directly on the evaluation results (Nauman and Thompson 2014). When such sets of soil observations were involved (Table 2-2), some studies considered the nearest neighbourhood of the points as well, usually within analysis

windows of 3x3 pixels, to account for potential location imprecisions (Heung *et al.* 2014, 2016; Machado *et al.* 2018). On average, this resulted in an improvement of the calculated OA (Table 2-2). In addition, categories up to the third most likely per-pixel on the disaggregated map were accepted in order to yield an accuracy value somewhat analogous to the one obtained at the same point from a conventional map, where all the categories registered in a polytaxic SMU would be taken into account (Nauman and Thompson 2014; Vincent *et al.* 2018; Ellili-Bargaoui *et al.* 2020). This would allow mitigating errors among adjacent categories in the classification system, while also opening up comparisons between both maps. In the reviewed studies, it implied a substantial increase in the accuracy (Table 2-2). For these reasons, the usage of these variations together with the original OA should be justified and, thus, they should be regarded as supplementary and informative tools to better understand where miss-classifications may originate rather than a full replacement.

Another spatial element with influence on the evaluation process is the presence of speckled pixels (Gagkas and Lilly 2019), which were corrected by filtering at a certain level of generalization (Leenaars *et al.* 2020).

Occasionally, because of the lack of enough observations in the registered categories, the evaluation of the disaggregation methodology was not fully carried out. Instead, the number of categories was reduced by aggregating them on a lower level, although it led to the degradation of the categorical detail in the final map (Pahlavan-Rad *et al.* 2016). Furthermore, the introduction of the taxonomic distance between STUs (Minasny and McBratney 2007) has also been discussed to assess misclassifications more accurately in future developments (Ellili-Bargaoui *et al.* 2020), rather than simply taking categories as stand-alone.

### *3.4.2 Analysis of the Overall Accuracy*

With regard to the OA of the reviewed applications, the values varied over a wide range of 17 to 78% (Figure 2-5).

These values were presented as positive for each case. Despite this, no common threshold of acceptance in terms of this parameter was identified during the review. From a statistical viewpoint, disaggregations with an OA above the one obtained from a fully random assignment might be regarded as advantageous and, thereby, the threshold would be conditioned by the

average number of components within the SMUs. However, no further analysis was possible on the extracted data.

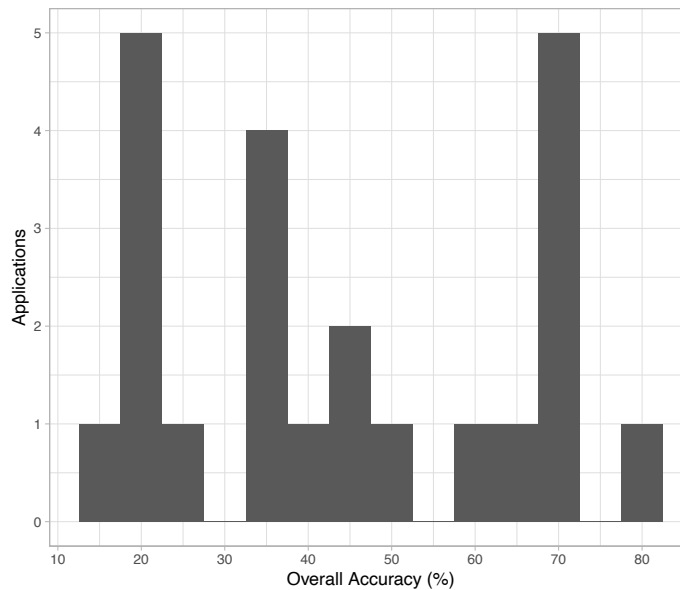


Figure 2-5. Distribution of the Overall Accuracy achieved in the reviewed disaggregation applications.

In turn, OA values stem from the interaction between the many factors analysed above. Due to the number of applications together with non-normal distributions of the OA and the features used to quantify those factors influencing the disaggregations (Table 2-1; Table 2-2), an optimal model of the OA was not feasible. However, certain trends were discerned when taking into account each feature individually.

Heung et al. (2017) noted that higher accuracies were achieved when disaggregating large-scale maps following a common methodology within the same region, and a similar insight was drawn from the set of reviewed applications. Unfortunately, no quality metrics of the conventional maps, such as the number of components within SMUs, the Index of Maximum Reduction, and the Average-Size Delineation, were available in order to explore a connection with the cartographic detail as reported by Møller et al. (2019). These outcomes might be associated with smaller map unit areas that bring along a homogenisation of the variability to be organized. Ultimately, this would facilitate the disaggregation process, that is, the regionalisation of that variability. Furthermore, pixel size or scale gain were not valid predictors of the OA.

Overall, no association was found neither between the number of covariates in use and the OA. It was significant for DSMART-based applications though, where a higher number of covariates was

correlated to a larger OA. This distinction may be related to the fact that DSMART is built on top of DT models, which were reported to effectively deal with non-informative covariates.

Likewise, the methodologies showing the highest OA applied DT-based classification techniques (Häring et al. 2012; Heung et al. 2014). For their part, retrieving soil information from point observations yielded higher OA values on average than from the CSM maps. It is worth noting that only three applications were identified from the former compared to 24 from the latter, which also exhibited a much larger deviation and the highest OA value of all. Hence, this assertion would not be conclusive given the current data.

Regarding the methodologies that define soil categories from the maps, no difference was found between single-stage versus two-stage methods. Were this analysis narrowed down to only DSMART-based methodologies, two-stage achieved on average a higher OA ratio. That is consistent with the focus on development of new two-stage methodologies that has taken place.

All in all, these results are a preliminary assessment insofar as they may be biased by the conditions of the initial data and do not allow determining a preferential way forward in every analysed aspect. This again raises the importance of comparative studies on the same location based on polytaxic SMUs.

### *3.4.3 Accuracy evaluation from conventional maps*

The comparison with categorical CSM maps was also carried out to evaluate the disaggregated maps. A qualitative approach results from the visual agreement between the distribution pattern on the disaggregation versus the baseline CSM map (Nauman and Thompson 2014; Sarmiento *et al.* 2017), detailed CSM maps available in part of the project region (Vincent *et al.* 2018) or expert knowledge (Sarmiento *et al.* 2017).

In quantitative terms, Ellili Bargaoui et al. (2020) proposed a reinterpretation of the V-measure concept created by Nowosad and Stepinski (2018) as a metric to evaluate regionalisation analysis, considering delineations by themselves. This metric is associated with two intermediate measures: homogeneity, which assesses the well-fitting of the regions from one map into the zones from another, and completeness, which assesses the well-fitting of the zones from the latter into the regions from the former.

An alternative is the Shannon Entropy Index (Ellili-Bargaoui *et al.* 2020), which was used to evaluate the spatial correspondence between the divisions of the two maps involved. An analogous index was derived by means of an analysis window on pixels in the disaggregated map as a measure of heterogeneity and ultimately used as an adjustment term for OA (Zeraatpisheh *et al.* 2019).

Moreover, other authors compared the proportions of the new divisions and their categories against those contained in the CSM map legend (Heung *et al.* 2016). The drawback lies in the fact that the precision of these values is subject to the expert knowledge of the surveyors, and that sometimes they are provided in ranges which are relatively wide (Møller *et al.* 2019).

## **4 Conclusions**

A set of published papers on Digital Soil Mapping shares a common aim of leveraging the information available in conventional soil maps to generate a new cartography that enlarges the cartographic detail. They all comprise an active research line referred to as disaggregation, which arises an increasing interest.

Most of these methodologies have built their quantitative soil-landscape models by using intensive and extensive conventional maps as the main source for defining and assigning soil categories. Many of these have achieved promising performance on the grounds of supervised classification techniques, with emphasis on decision trees and random forest algorithms and those methods based upon them as DSMART.

Category assignment and sampling for building the training sets are critical tasks that have a strong influence on the classification output. Variations in most recent applications of disaggregation and DSMART have focused on improving these two key aspects following a two-stage strategy. The proposed improvements in terms of expert-knowledge rules and geofoms patterns to potentially allocate categories as well as sampling limits, distribution, and size, have given rise to improved accuracy values on their implementations. Further research on the comparison between them is deemed necessary to identify the best performers and in which conditions. Moreover, the commented results open the possibility to develop new disaggregation methodologies based on combinations of those highlighted parameters and strategies.



The disaggregation offers within the Digital Soil Mapping framework a complementary approach to the generalization from-scratch and from point soil observations; one that makes the most of available soil information and environmental sources and that may provide detailed soil maps. And even when accuracy rates were relatively low, they might be seen as an initial attempt to better understand the project region and to identify areas of interest to be mapped further with another strategy.

## 5 References

- Arrouays D, Poggio L, Salazar Guerrero OA, Mulder VL. 2020. Digital soil mapping and GlobalSoilMap. Main advances and ways forward. *Geoderma Regional* **21**: e00265. DOI: 10.1016/j.geodrs.2020.e00265
- Behrens T, Zhu A-X, Schmidt K, Scholten T. 2010. Multi-scale digital terrain analysis and feature selection for digital soil mapping. *Geoderma* **155**: 175–185. DOI: 10.1016/j.geoderma.2009.07.010
- Brungard CW, Boettinger JL, Duniway MC, Wills SA, Edwards TC. 2015. Machine learning for predicting soil classes in three semi-arid landscapes. *Geoderma* **239–240**: 68–83. DOI: 10.1016/j.geoderma.2014.09.019
- Brus DJ, Kempen B, Heuvelink GBM. 2011. Sampling for validation of digital soil maps. *European Journal of Soil Science* **62**: 394–407. DOI: 10.1111/j.1365-2389.2011.01364.x
- Bui EN, Moran CJ. 2001. Disaggregation of polygons of surficial geology and soil maps using spatial modelling and legacy data. *Geoderma* **103**: 79–94. DOI: 10.1016/S0016-7061(01)00070-2
- Chaney NW, Wood EF, McBratney AB, Hempel JW, Nauman TW, Brungard CW, Odgers NP. 2016. POLARIS: A 30-meter probabilistic soil series map of the contiguous United States. *Geoderma* **274**: 54–67. DOI: 10.1016/j.geoderma.2016.03.025
- Collard F, Kempen B, Heuvelink GBM, Saby NPA, Richer de Forges AC, Lehmann S, Nehlig P, Arrouays D. 2014. Refining a reconnaissance soil map by calibrating regression models with data from the same map (Normandy, France). *Geoderma Regional* **1**: 21–30. DOI: 10.1016/j.geodrs.2014.07.001
- Conrad O, Bechtel B, Bock M, Dietrich H, Fischer E, Gerlitz L, Wehberg J, Wichmann V, Böhner J. 2015. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geoscientific Model Development* **8**: 1991–2007. DOI: 10.5194/gmd-8-1991-2015
- Costa JJF, Giasson E, Silva EB da, Campos AR, Machado IR, Bonfatti BR, Bacic ILZ. 2019. Individualization of soil classes by disaggregation of physiographic map polygons. *Pesquisa Agropecuária Brasileira* **54**: e00290. DOI: 10.1590/s1678-3921.pab2019.v54.00290
- Ehsani AH, Quiel F. 2009. A semi-automatic method for analysis of landscape elements using Shuttle Radar Topography Mission and Landsat ETM+ data. *Computers & Geosciences* **35**: 373–389. DOI: 10.1016/j.cageo.2007.09.019

- Ellili Bargaoui Y, Walter C, Michot D, Saby NPA, Vincent S, Lemerrier B. 2019. Validation of digital maps derived from spatial disaggregation of legacy soil maps. *Geoderma* **356**: 113907. DOI: 10.1016/j.geoderma.2019.113907
- Ellili-Bargaoui Y, Malone BP, Michot D, Minasny B, Vincent S, Walter C, Lemerrier B. 2020. Comparing three approaches of spatial disaggregation of legacy soil maps based on the Disaggregation and Harmonisation of Soil Map Units Through Resampled Classification Trees (DSMART) algorithm. *SOIL* **6**: 371–388. DOI: 10.5194/soil-6-371-2020
- Evans IS. 2012. Geomorphometry and landform mapping: What is a landform? *Geomorphology* **137**: 94–106. DOI: 10.1016/j.geomorph.2010.09.029
- Florinsky IV. 2016. *Digital Terrain Analysis in Soil Science and Geology*. Elsevier. DOI: 10.1016/B978-0-12-804632-6.09988-0
- Flynn T, Rozanov A, de Clercq W, Warr B, Clarke C. 2019a. Semi-automatic disaggregation of a national resource inventory into a farm-scale soil depth class map. *Geoderma* **337**: 1136–1145. DOI: 10.1016/j.geoderma.2018.11.003
- Flynn T, Rozanov A, Ellis F, de Clercq W, Clarke C. 2020. Farm-scale soil patterns derived from automated terrain classification. *CATENA* **185**: 104311. DOI: 10.1016/j.catena.2019.104311
- Flynn T, van Zijl G, van Tol J, Botha C, Rozanov A, Warr B, Clarke C. 2019b. Comparing algorithms to disaggregate complex soil polygons in contrasting environments. *Geoderma* **352**: 171–180. DOI: 10.1016/j.geoderma.2019.06.013
- Frank E, Hall MA, Witten IH. 2016. The WEKA Workbench. Online Appendix. *Data Mining: Practical Machine Learning Tools and Techniques*
- Gagkas Z, Lilly A. 2019. Downscaling soil hydrological mapping used to predict catchment hydrological response with random forests. *Geoderma* **341**: 216–235. DOI: 10.1016/j.geoderma.2019.01.048
- Gallant JC, Dowling TI. 2003. A multiresolution index of valley bottom flatness for mapping depositional areas. *Water Resources Research* **39**. DOI: 10.1029/2002WR001426
- Gao H, Zhang X, Wang L, He X, Shen F, Yang L. 2020. Selection of training samples for updating conventional soil map based on spatial neighborhood analysis of environmental covariates. *Geoderma* **366**: 114244. DOI: 10.1016/j.geoderma.2020.114244
- GRASS Development Team. 2019. *GRASS Geographic Resources Analysis Support System*. Open Source Geospatial Foundation Project
- Häring T, Dietz E, Osenstetter S, Koschitzki T, Schröder B. 2012. Spatial disaggregation of complex soil map units: A decision-tree based approach in Bavarian forest soils. *Geoderma* **185–186**: 37–47. DOI: 10.1016/j.geoderma.2012.04.001
- Hengl T. 2006. Finding the right pixel size. *Computers & Geosciences* **32**: 1283–1298. DOI: 10.1016/j.cageo.2005.11.008
- Hengl T, MacMillan RA. 2019. *Predictive Soil Mapping with R*. OpenGeoHub foundation: Wageningen, the Netherlands
- Hengl T, Reuter HI (eds). 2009. *Geomorphometry: concepts, software, applications*. Elsevier: Amsterdam, Netherlands; Oxford, UK; Boston [Mass.]

- Hengl T, Toomanian N, Reuter HI, Malakouti MJ. 2007. Methods to interpolate soil categorical variables from profile observations: Lessons from Iran. *Geoderma* **140**: 417–427. DOI: 10.1016/j.geoderma.2007.04.022
- Heung B, Bulmer CE, Schmidt MG. 2014. Predictive soil parent material mapping at a regional-scale: A Random Forest approach. *Geoderma* **214–215**: 141–154. DOI: 10.1016/j.geoderma.2013.09.016
- Heung B, Ho HC, Zhang J, Knudby A, Bulmer CE, Schmidt MG. 2016. An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. *Geoderma* **265**: 62–77. DOI: 10.1016/j.geoderma.2015.11.014
- Heung B, Hodúl M, Schmidt MG. 2017. Comparing the use of training data derived from legacy soil pits and soil survey polygons for mapping soil classes. *Geoderma* **290**: 51–68. DOI: 10.1016/j.geoderma.2016.12.001
- Holmes KW, Griffin EA, Odgers NP. 2015. Large-area spatial disaggregation of a mosaic of conventional soil maps: evaluation over Western Australia. *Soil Research* **53**: 865. DOI: 10.1071/SR14270
- Irvin BJ, Ventura SJ, Slater BK. 1997. Fuzzy and isodata classification of landform elements from digital terrain data in Pleasant Valley, Wisconsin. *Geoderma* **77**: 137–154. DOI: 10.1016/S0016-7061(97)00019-0
- IUSS Working Group WRB. 2014. *World reference base for soil resources 2014: international soil classification system for naming soils and creating legends for soil maps*. FAO: Rome
- Iwahashi J, Pike RJ. 2007. Automated classifications of topography from DEMs by an unsupervised nested-means algorithm and a three-part geometric signature. *Geomorphology* **86**: 409–440. DOI: 10.1016/j.geomorph.2006.09.012
- Jamshidi M, Delavar MA, Taghizadehe-Mehrjerdi R, Brungard C. 2019. Disaggregation of conventional soil map by generating multi realizations of soil class distribution (case study: Saadat Shahr plain, Iran). *Environmental Monitoring and Assessment* **191**: 769. DOI: 10.1007/s10661-019-7942-x
- Jasiewicz J, Stepinski TF. 2013. Geomorphons — a pattern recognition approach to classification and mapping of landforms. *Geomorphology* **182**: 147–156. DOI: 10.1016/j.geomorph.2012.11.005
- Leenaars JGB, Elias E, Wösten JHM, Ruiperez-González M, Kempen B. 2020. Mapping the major soil-landscape resources of the Ethiopian Highlands using random forest. *Geoderma* **361**: 114067. DOI: 10.1016/j.geoderma.2019.114067
- Levi MR. 2017. Neighborhood Size of Training Data Influences Soil Map Disaggregation. *Soil Science Society of America Journal* **81**: 354–368. DOI: 10.2136/sssaj2016.08.0258
- Liu F, Geng X, Zhu A -xing, Fraser W, Song X, Zhang G. 2016. Soil polygon disaggregation through similarity-based prediction with legacy pedons. *Journal of Arid Land* **8**: 760–772. DOI: 10.1007/s40333-016-0087-7
- Machado IR, Giasson E, Campos AR, Costa JFF, Silva EB da, Bonfatti BR. 2018. Spatial Disaggregation of Multi-Component Soil Map Units Using Legacy Data and a Tree-Based Algorithm in Southern Brazil. *Revista Brasileira de Ciência do Solo* **42**. DOI: 10.1590/18069657rbcS20170193

- Massawe BHJ, Subburayalu SK, Kaaya AK, Winowiecki L, Slater BK. 2018. Mapping numerically classified soil taxa in Kilombero Valley, Tanzania using machine learning. *Geoderma* **311**: 143–148. DOI: 10.1016/j.geoderma.2016.11.020
- McBratney AB, Field DJ, Koch A. 2014. The dimensions of soil security. *Geoderma* **213**: 203–213. DOI: 10.1016/j.geoderma.2013.08.013
- McBratney AB, Mendonça Santos ML, Minasny B. 2003. On digital soil mapping. *Geoderma* **117**: 3–52. DOI: 10.1016/S0016-7061(03)00223-4
- Minasny B, McBratney AB. 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Computers & Geosciences* **32**: 1378–1388. DOI: 10.1016/j.cageo.2005.12.009
- Minasny B, McBratney AB. 2007. Incorporating taxonomic distance into spatial prediction and digital mapping of soil classes. *Geoderma* **142**: 285–293. DOI: 10.1016/j.geoderma.2007.08.022
- Minasny B, McBratney AB. 2016. Digital soil mapping: A brief history and some lessons. *Geoderma* **264**: 301–311. DOI: 10.1016/j.geoderma.2015.07.017
- Møller AB, Malone B, Odgers NP, Beucher A, Iversen BV, Greve MH, Minasny B. 2019. Improved disaggregation of conventional soil maps. *Geoderma* **341**: 148–160. DOI: 10.1016/j.geoderma.2019.01.038
- Nauman TW, Thompson JA. 2014. Semi-automated disaggregation of conventional soil maps using knowledge driven data mining and classification trees. *Geoderma* **213**: 385–399. DOI: 10.1016/j.geoderma.2013.08.024
- Nauman TW, Thompson JA, Rasmussen C. 2014. Semi-Automated Disaggregation of a Conventional Soil Map Using Knowledge Driven Data Mining and Random Forests in the Sonoran Desert, USA. *Photogrammetric Engineering & Remote Sensing* **80**: 353–366. DOI: 10.14358/PERS.80.4.353
- Nowosad J, Stepinski TF. 2018. Spatial association between regionalizations using the information-theoretical V-measure. *International Journal of Geographical Information Science* **32**: 2386–2401. DOI: 10.1080/13658816.2018.1511794
- Odgers NP, Sun W, McBratney AB, Minasny B, Clifford D. 2014. Disaggregating and harmonising soil map units through resampled classification trees. *Geoderma* **214–215**: 91–100. DOI: 10.1016/j.geoderma.2013.09.024
- Pahlavan-Rad MR, Khormali F, Toomanian N, Brungard CW, Kiani F, Komaki CB, Bogaert P. 2016. Legacy soil maps as a covariate in digital soil mapping: A case study from Northern Iran. *Geoderma* **279**: 141–148. DOI: 10.1016/j.geoderma.2016.05.014
- R Core Team. 2020. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing: Vienna, Austria
- Rossiter DG. 2000. *Methodology for Soil Resource Inventories*. International Institute for Aerospace Survey & Earth Sciences (ITC)
- Rossiter DG. 2018. Past, present & future of information technology in pedometrics. *Geoderma* **324**: 131–137. DOI: 10.1016/j.geoderma.2018.03.009

- Sarmiento EC, Giasson E, Weber EJ, Flores CA, Hasenack H. 2017. Disaggregating conventional soil maps with limited descriptive data: A knowledge-based approach in Serra Gaúcha, Brazil. *Geoderma Regional* **8**: 12–23. DOI: 10.1016/j.geodrs.2016.12.004
- Schmidt J, Hewitt A. 2004. Fuzzy land element classification from DTMs based on geometry and terrain position. *Geoderma* **121**: 243–256. DOI: 10.1016/j.geoderma.2003.10.008
- Scull P, Franklin J, Chadwick OA. 2005. The application of classification tree analysis to soil type prediction in a desert landscape. *Ecological Modelling* **181**: 1–15. DOI: 10.1016/j.ecolmodel.2004.06.036
- Scull P, Franklin J, Chadwick OA, McArthur D. 2003. Predictive soil mapping: a review. *Progress in Physical Geography* **27**: 171–197. DOI: 10.1191/0309133303pp366ra
- Silva SHG, Menezes MD de, Owens PR, Curi N. 2016. Retrieving pedologist's mental model from existing soil map and comparing data mining tools for refining a larger area map under similar environmental conditions in Southeastern Brazil. *Geoderma* **267**: 65–77. DOI: 10.1016/j.geoderma.2015.12.025
- Smith MP, Zhu A-X, Burt JE, Stiles C. 2006. The effects of DEM resolution and neighborhood size on digital soil survey. *Geoderma* **137**: 58–69. DOI: 10.1016/j.geoderma.2006.07.002
- Soil Survey Division Staff. 1999. *Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys*. USDA-Natural Resources Conservation Service
- Soil Survey Division Staff. 2017. *Soil Survey Manual*. U.S. Government Printing Office: Washington, D.C.
- Subburayalu SK, Jenhani I, Slater BK. 2014. Disaggregation of component soil series on an Ohio County soil survey map using possibilistic decision trees. *Geoderma* **213**: 334–345. DOI: 10.1016/j.geoderma.2013.08.018
- Subburayalu SK, Slater BK. 2013. Soil Series Mapping By Knowledge Discovery from an Ohio County Soil Map. *Soil Science Society of America Journal* **77**: 1254. DOI: 10.2136/sssaj2012.0321
- Van Rossum G, Drake FL. 2009. *Python 3 Reference Manual*. CreateSpace: Scotts Valley, CA
- Vincent S, Lemercier B, Berthier L, Walter C. 2018. Spatial disaggregation of complex Soil Map Units at the regional scale based on soil-landscape relationships. *Geoderma* **311**: 130–142. DOI: 10.1016/j.geoderma.2016.06.006
- Wood J. 1996. The Geomorphological Characterisation of Digital Elevation Models. PhD thesis, University of Leicester
- Young NE, Anderson RS, Chignell SM, Vorster AG, Lawrence R, Evangelista PH. 2017. A survival guide to Landsat preprocessing. *Ecology* **98**: 920–932. DOI: 10.1002/ecy.1730
- Zeraatpisheh M, Ayoubi S, Brungard CW, Finke P. 2019. Disaggregating and updating a legacy soil map using DSMART, fuzzy c-means and k-means clustering algorithms in Central Iran. *Geoderma* **340**: 249–258. DOI: 10.1016/j.geoderma.2019.01.005



# **CAPÍTULO 3 A NEW METHODOLOGY TO DIVIDE CONVENTIONAL SOIL MAP UNITS**

Este capítulo se presenta en el formato del artículo de título equivalente que está en proceso de publicación.

## **1 Introduction**

Soils fulfil essential functions in the biosphere, and they are currently facing multiple challenges to their conservation (Koch et al., 2013). In this scenario of a growing need for sustainable management and monitoring of natural resources, the demand for highly detailed information on soil resources and its spatial distribution has increased remarkably within the precision agriculture framework.

A new common framework named DSM (McBratney et al., 2003) use available information generated by CSM methods, while incorporating data from new sources, such as remote sensing and geomorphometry. In this regard, a dataset built from legacy and new sources can be combined with DSM methods to describe the non-explicit spatial variation in conventional SMU as potentially homogeneous areas (Odeh et al., 1992) or new SMU, in what is known as disaggregation (Sarmiento et al., 2017). It would allow to break down politaxic or multi-component SMUs (Nauman and Thompson, 2014) and to transform conventional SRI from medium scales into larger scales.

This work is focused on the development of a new methodology that allows capturing soil variations within an area, and so facilitating the detection of smaller potential homogeneous areas within politaxic SMU. It does so by using a special unsupervised classification technique for large datasets and a data-driven workflow, whereby the data i) assists in selecting which environmental parameters are the most relevant and ii) also guides the number of divisions and their distributions within the covariate and spatial universe.

More than five million ha of SRI have been carried out across Spain, primarily within the Viticultural Terroir Zoning framework, at 1:50,000 and 1:25,000 scales and with politaxic SMU. Main actors in the wine sector are currently demanding more detailed cartography with

monotaxic SMU at larger scales (generally greater than 1:10,000), seeking to tackle the characterization and management of terroir units and precision viticulture.

The study was implemented in a wine region of north-eastern Spain, where a group of ten SMU, representing a wide surface, were considered.

## 2 Material and methods

### 2.1 Research site

The study was carried out into the Spanish wine Designation of Origin (DO, in Spanish *Denominación de Origen* [DO]) Campo de Borja, a relevant wine region with 6,416.14 ha of vineyard (data from Regulatory Council of the DO Campo de Borja, 2019), located at the Northeast of Spain in the Autonomous Community of Aragon (Figure 3-1).

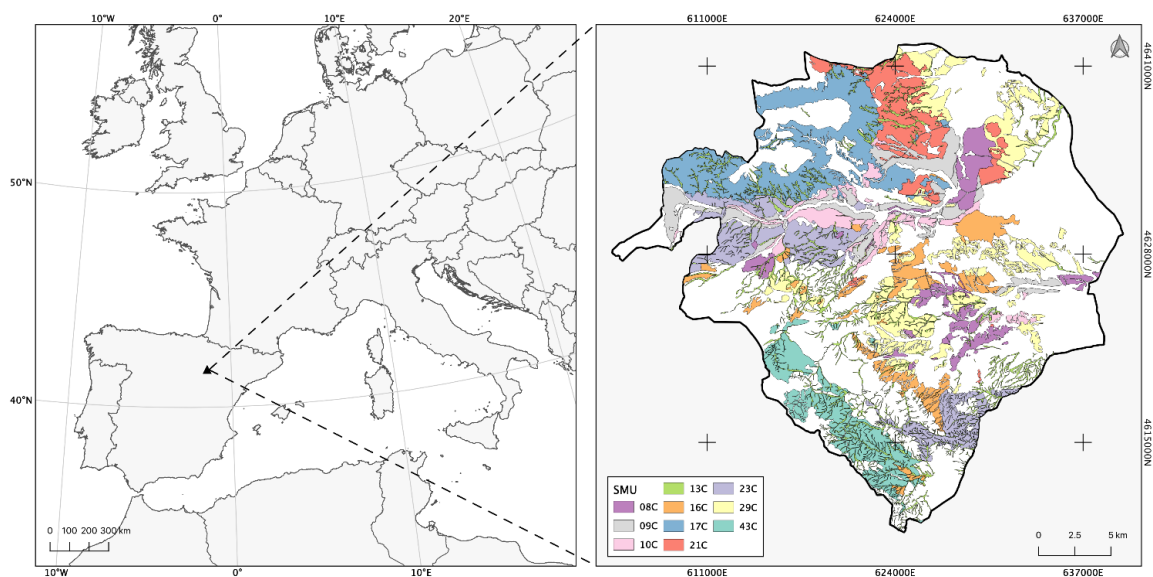


Figure 3-1. Location of the Designation of Origin Campo de Borja in NE Spain (left) and distribution of the selected ten soil map units within the conventional soil resource inventory map (right). SMU: Soil Map Units.

### 2.2 Data sources

Three data sources were used: the legacy soil data of the SRI of DO Campo de Borja (Gómez-Miguel and Sotés, 2015); the environmental data of a DEM and High-Resolution imagery from remote sensing of the project region; along with ancillary data of human-made structures and construction.



### *2.2.1 Legacy soil data: the conventional soil map*

The conventional SRI was produced at 1:25,000 scale following the Soil Survey Manual (Soil Survey Division Staff, 1993) ) and the ST (Soil Survey Division Staff, 2014). Delineations were based on Aerial Photo Interpretation (API) and fieldworks with an observation density of 3.8 observations/km<sup>2</sup>. The conventional SRI map gathers SMUs of associations and complexes (46) with phases (343) and whose content in STU is formed by soil series (20). When the series could not be defined, families together with the specific lithology were assigned (122). The STUs belong to four Orders (Alfisol, Inceptisol, Entisol and Mollisol), to eight Suborders (Xeralf, Aqualf, Aquept, Xerept, Aquent, Fluvent and Orthent) and to twelve Great Groups (Haploxeroll, Palexeralf, Haploxeralf, Rhodoxeralf, Epiaqualf, Epiaquept, Haploxerept, Calcixerept, Fluvaquent, Epiaquent, Xerofluvent, Xerorthent).

This study was conducted on ten SMUs of the conventional SRI without phases. They covered more than 50% of the total surface of the DO (Figure 3-1) and they are representative of the existing environments. Their main characteristics are summarized in Table 3-1 and Table 3-2.

#### *1.1.1 Environmental data: DEM and remote sensing*

A DEM made by the National Geographic Information Center (CNIG) from its first-coverage LIDAR flight data of 5 m grid size was used. This grid size was selected to ensure a target scale of the new delineations suitable for a very intensive SRI with consociations (around 1:10,000), following the general relation between conventional vector maps and those in raster format noted by Hengl (2006).

High-Resolution imagery from Sentinel-2 satellites were considered, with a spectral range extends from VIS to SWIR and a variable grid size of 10, 20 and 60 m depending on band domain. It was employed specifically an image of the first half of February 2019, which was distributed by ESA with a processing level 2A.

This mission was taken based on the similarity of its spatial resolution to the one already set by the DEM, spectral requirements from edaphic and bioclimatic indices (Mulder et al., 2011), its temporal resolution and determining factors of viticulture and land use within the area. Thus, it was restricted to a time frame that excluded the growing cycle of the vine (Flynn et al., 2019; Vaudour et al., 2015).

Table 3-1. Main characteristics of the selected Soil Map Units from the conventional Soil Resource Inventory map of the Designation of Origin Campo de Borja.

SMU	Area		Field observations (FO)			FO Density (FO/Km <sup>2</sup> )	Surface/FO (ha/FO)
	ha	% Total DO	Soil profiles Backhoe pits	Auger hole Pits	Total FO		
08C	2.260	4	25	75	100	4,4	22,6
09C	2.933	5	61	65	126	4,3	23,3
10C	1.570	2	22	51	73	4,6	21,5
13C	3.394	5	33	115	148	4,4	22,9
16C	2.598	4	27	82	109	4,2	23,8
17C	5.543	9	44	180	224	4,0	24,7
21C	2.852	4	36	82	118	4,1	24,2
23C	3.753	6	27	128	155	4,1	24,2
29C	5.114	8	36	170	206	4,0	24,8
43C	2.915	5	20	110	130	4,5	22,4
<b>Total</b>	<b>32.932</b>	<b>52</b>	<b>331</b>	<b>1058</b>	<b>1389</b>	<b>Mean: 4,3</b>	<b>Mean: 23,5</b>

DO: Designation of Origin; SMU: Soil Map Unit.

Table 3-2. Soil Taxonomy Units content of the selected Soil Map Units from the conventional Soil Resource Inventory map of the Designation of Origin Campo de Borja.

SMU	STU 1		STU 2		STU 3		STU 4	
	Label*	%	Label*	%	Label*	%	Label*	%
08C	J05	30	J08	25	J04	20	J07	15
09C	J04	35	J08	30	J02	15	J07	10
10C	J03	45	J06	25	J09	20	HxaT1F	10
13C	J20	30	J03	25	HxaC2FV	20	XfnT1FV	15
16C	CxpT3T	45	PxaC1T	25	CxpP1T	15	XonT3FY	10
17C	J17	30	J16	25	J15	20	J14	15
21C	J11	35	J12	25	XfnT2Y2	20	XonT2Y2	15
23C	J13	35	J19	30	HxaC1Y7	15	XonT3Y7	10
29C	J10	40	J11	25	XonT4Y2	15	HxaC2Y2	10
43C	XonT1M2	35	HxaT1M2	25	CxpT1M2	20	PxaT1M2	10

\*Labels starting with J indicated soil series. Labels formed by acronyms indicate soil families regarding their lithology (last letters). SMU: Soil Map Unit; STU: Soil Taxonomy Unit.

### 2.2.2 Ancillary data: structure and constructions

Layers with information about state roads, conventional railways, and populations from the National Topographic Base (CNIG) and the National Cadastre were employed.

## 2.3 Data management, GIS, and statistical software

The project data has been managed from a database developed for this purpose and that implements a specific modelling for SRI in Spain and the viticultural terroir (Lázaro-López et al., 2018), based on the Soil and Terrain database programme (SOTER) and the National Soil Information System (NASIS) schemas. The database uses PostgreSQL (13) and the extension PostGIS (3.1) as its database management system engine.

The methodology proposed in this work was developed with R (R Core Team, 2019). In this context, specialized packages were used to connect to the database: 'RPostgres' (Wickham et al., 2019); and to subsequently manage data: 'sf' (Pebesma, 2018), 'raster' (Hijmans, 2019), 'bigmemory' (Kane et al., 2013) (Kane et al., 2013) and 'Tidyverse' (Wickham et al., 2019). Data sources were processed using the GIS software SAGA (Conrad et al., 2015) through the 'RSAGA' package (Brenning et al., 2018). Finally, statistical analyses have relied principally on 'ClusterR' (Mouselimis, 2019), 'Nbclust' (Charrad et al., 2014) and 'foreach' (Microsoft Corporation and Weston, 2019) packages and the 'QGIS' (QGIS Development Team, 2019) software.

## 2.4 Procedure

### *2.4.1 Processing and scaling of data sources*

Data collected from the different sources used in this study were processed to erase artefacts and avoid conflicts coming from scale differences.

On the one hand, it was executed a sequence of sink filling and then filtering on the DEM, following previous experiences of DEM processing (Flynn et al., 2019; Hengl and Reuter, 2009; Taghizadeh-Mehrjardi et al., 2015; Wilson and Gallant, 2000). First, the sink filling algorithm developed by Wang and Liu (2006) was selected because it enforces the drainage network to be fully connected, without any depressions or spurious artefacts (Gallant and Austin, 2015; Hengl and Reuter, 2009; Wilson and Gallant, 2000). Then, a filter with the Denoise algorithm (Sun et al., 2007) was implemented to preserve sharpened edges on slopes while smoothing of the small variations of the flattest areas, and so ultimately, to reduce noise on the DEM (Stevenson et al., 2010).

Concerning remote sensing, although images from Sentinel-2 were obtained with an optimal processing level, they were re-scaled considering the finer grid size of DEM to get them aligned.

Furthermore, features of miscellaneous areas detectable to larger scales were not considered during the initial demarcation of the conventional SRI. For this reason, in some cases, original delineations needed to be tinkered to integrate these new areas (Håring et al., 2012). Hence, available information of state roads, conventional railways, and populations from the National Topographic Base (CNIG) and the National Cadastre, were used.

#### *2.4.2 Calculation of covariates*

Thirty-seven different covariates were derived from environmental data sources to quantify geomorphological, hydrological, topo-climatic, and edaphic parameters associated with the SCORPAN factors. They were calculated using several SAGA modules with default parameters and considering previously published recommendations (see references in Table 3-3 to Table 3-6).

It should be noted that covariates related to the vegetation (O factor) were omitted to avoid the distortions arising from the different anthropic land uses, typical of agricultural regions. In addition, covariates directly related to parent material (P factor) were also not studied since the information on this attribute was deemed to be integrated into the SMUs through the API, which took into account the lithology.

#### *1.1.1 Analysis of covariates*

The entire set of calculated covariates was analyzed quantitatively and qualitatively. On the one hand, the quantitative analysis was based on Pearson and Spearman correlations as they allow assessing linear correlations and the strength of monotonic relationships respectively (Thompson et al., 2016). On the other hand, the qualitative analysis was conducted by visual exploration seeking to check the physical meaning of each variable and to detect noise in the data. These two complementary analyses allowed understanding the behaviour of the covariates in the specific area under study and to distinguish the most useful set.

#### *1.1.2 Selection of covariates*

At this step, the most useful covariables to depict new divisions within original SMUs are potentially selected based on the results of the quantitative and qualitative analysis previously accomplished. This collection of covariates needs to better disclose the environmental variance while keeping redundancy to a minimum and their number low in order to obtain a simple but efficient methodology. All the distinct groups should be feature too. For this reason, this is crucial to the proposed methodology.

Table 3-3. Symbol, name, and a brief description of the morphometry covariates of the project associated with the SCORPAN R factor.

Symbol	Covariate name and description
ASPECT	<b>Aspect.</b> Polar variable: 0° point to N and, clockwise, 90° to E, 180° to S and 270° to W.
CCROS	<b>Cross-sectional curvature.</b> Tangential curvature intersecting with the plane defined by the surface normal and a tangent to the contour, perpendicular to maximum gradient direction. It's calculated with a multi-scale approach (Florinsky, 2017; Wood, 1996).
CLONG	<b>Longitudinal curvature.</b> Curvature intersecting with the plane defined by the surface normal and maximum gradient direction calculated with a multi-scale approach (Florinsky, 2017; Wood, 1996). A counterpart of CPROF.
CMAXI	<b>Maximum curvature,</b> in any direction calculated with a multi-scale approach (Florinsky, 2017; Wood, 1996).
CMINI	<b>Minimum curvature,</b> in the perpendicular direction to the direction of maximum curvature calculated with a multi-scale approach (Florinsky, 2017; Wood, 1996).
CPLAN	<b>Plan curvature.</b> Horizontal curvature, intersecting with the XY plane with a multi-scale approach (Florinsky, 2017; Wood, 1996).
CPROF	<b>Profile curvature.</b> Curvature intersecting with the plane defined by Z-axis and the maximum gradient direction calculated with a multi-scale approach (Florinsky, 2017; Wood, 1996).
CTANG	<b>Tangential curvature.</b> A counterpart of CCROS calculated on a single scale.
DWSLOPE	<b>Downslope index.</b> Downslope controls on local drainage (Hjerdt et al., 2004).
DDIFF	<b>Difference between DWSLOPE and SLOPE</b> (Hjerdt et al., 2004).
EAST	<b>Easternness.</b> East-West linear variable derived from ASPECT. Range from 1 (East) to -1 (West)
EASTSP	<b>Interaction between EAST and SLOPE.</b> Equals to $\sin(\text{aspect}) * \text{slope}$ .
ELEVATION	<b>Elevation</b>
MPI	<b>Morphometric Protection index</b> (Yokoyama et al., 2002).
MRRTF	<b>Multiresolution Index of Ridge Top Flatness</b> (Gallant and Dowling, 2003).
MRVBF	<b>Multiresolution Index of Valley Bottom Flatness</b> (Gallant and Dowling, 2003).
NORTH	<b>Northernness.</b> North-South linear variable derived from ASPECT. Range from 1 (North) to -1 (South).
NORTHSP	<b>Interaction between NORTH and SLOPE.</b> Equals to $\cos(\text{aspect}) * \text{slope}$ .
SLOPE	<b>Slope.</b>
TPI	<b>Topographic Position Index.</b> Difference between central and mean window elevations calculated in a multi-scale approach.
TRI	<b>Terrain Ruggedness Index.</b> The mean difference between a central elevation and points within the analysis window.
UPSLOPE	<b>Modified Catchment Slope from SAGA TWI.</b>

Table 3-4. Symbol, name, and a brief description of the hydrology covariates of the project associated with the SCORPAN R factor.

Symbol	Covariate name and description
DISTV	<b>Vertical distance to channel network</b> (m).
SCA	<b>Standard catchment area</b> (m <sup>2</sup> ). The unit measure depends on DEM). It's also described as an indirect estimator of night temperature and air humidity (C factor) (Hengl and Reuter, 2009).
SSCA	<b>Standard catchment area for SAGA TWI</b> . Same attributes as SCA.
SPI	<b>Stream Power Index</b> . It's based on SCA and SLOPE.
STWI	<b>Topographic Wetness Index by SAGA</b> . It uses SSCA.
TCILOW	<b>Terrain Classification Index for Lowlands</b> . Range from 0 to 1, where high values correspond to low landforms.
TWI	<b>Topographic Wetness Index</b> . It uses SCA.

Table 3-5. Symbol, name, and a brief description of the topo-climate covariates of the project associated with the SCORPAN C factor.

Symbol	Covariate name and description
DAH	<b>Diurnal Anisotropic Heat</b> regarding the aspect with the maximum total heat surplus (also, R factor) (Hengl and Reuter, 2009).
LST	<b>Land Surface Temperature</b>
RDIFBUS	<b>Diffuse Potential Incoming Solar Radiation</b>
RDIRECT	<b>Direct Potential Incoming Solar Radiation</b> .
TRASP	<b>Solar-radiation Aspect Index</b> (Roberts and Cooper, 1989).

Table 3-6. Symbol, name, and a brief description of the remote sensing covariates of the project associated with the SCORPAN S factor.

Symbol	Covariate name and description
CI	<b>Coloration Index</b> (Escadafal, 1993).
FEOUS	<b>Simple ratio Red ~ SWIR</b> (Roecker et al., 2010).
RI	<b>Redness index</b> (Escadafal and Huete, 1991)
SI	<b>Salinity Index</b> (Al-Khaier, 2003).

### 2.4.3 Division of the initial Soil Map Units into new homogenous areas

Previously published papers have applied non-supervised classification techniques to reveal homogenous areas of soil-forming factors, on which soil-landscapes relations can be established (Bui and Moran, 2001; Burrough, 2000; Odeh et al., 1992; Zare et al., 2016).

On these notions, the non-supervised classification method “Clustering Large Applications” (CLARA) with Mahalanobis distance was applied to divide each of the studied SMUs into new and smaller divisions made up of several delineations.

CLARA is geared towards large datasets, where it calculates the optimal group centroids for fixed-size samples and, subsequently, re-evaluates the quality of the clusters by considering the dissimilarity between centroids and all observations. These centroids are medoids, actual data points, and their usage reduced the influence of outliers and noise from the covariates (Mouselimis, 2019). Moreover, Mahalanobis distance has been highlighted for application to soil data (Odeh et al., 1992), due to its invariability in terms of scale and its effect of decreasing data redundancy by accounting for the correlation between covariates.

Similarly to that noted by Häring et al. (2012), the division within the SMU boundaries (without phases) allowed both integrating the lithology information provided by the API process, at least indirectly, and reducing the global variability. This action aimed to overcome the potential drawback from a recent study (Flynn et al., 2019), where covariates derived from DEM and images of remote sensing, such as those used in this study, could extract relations between soil and landscapes, but not those between soil and parent material in contrasting areas.

The target number of new divisions for each SMU was established by a data-driven approach. Several quantitative indexes (Charrad et al., 2014) suggested the optimal group number within a range based on the map legend, where the minimum value matched the number of STU and the maximum the sum of STU and inclusions. The criterion followed in the study sought the maximum number of divisions within the indicated range.

#### *2.4.4 Evaluation of the new homogeneous areas*

Finally, the evaluation of the divisions using the unsupervised CLARA method on the basis of the selected collection of covariates is conducted.

The outcome is analysed from the perspective of the delineations, that is, the total number, the distribution of their areas as well as the average value and then compared with the source map. Besides, their spatial distribution pattern is studied. In this way, it is possible to identify the gain during the division process as well as the incidence on specific areas.

Furthermore, the environmental meaning of the divisions within the context of each SMU is examined. For this purpose, the standardised medoids of each division are taken, as they may be considered characteristic and representative of them, together with their visualisation on high-resolution orthophotos. Then, they are associated to geomorphs and landscape features potentially associated with pedological processes.

## **3 Results and discussion**

### **3.1 Processing of data sources**

The indicated transformations yielded to important changes in the data sources. A remarkable effect was observed after the of drainage sink and the filtering for the reduction of noise and geomorphometric artefacts, while the structure of the DEM remained stable. This was readily apparent in its histogram and when calculating covariates from each of the sources. Moreover, delineations number increased more than 300%, (from 3,713 to 12,596) due to divisions generated by new unsorted areas.

### **3.2 Analysis of covariates**

The most interesting results come from the quantitative analysis. Four main covariate associations were detected involving 27 of the 37 covariates studied (Figure 3-2), as described below. The other ten variables were occasionally and slightly correlated with the four mentioned groups or among them.

#### *3.2.1 Curvature variables*

Most of the curvature covariates except for CTANG were positively correlated among them, while were inversely correlated with CPLAN. This fact agreed with the theoretical relation between the pairs of curvature covariates noted by Florinsky (2017): CCROS and CPLAN, CLONG and CPROF, and CMAXI and CMINI; and also with the visual differences between those multi-scale curvature covariates against CTANG, the one derived from a single scale method.

On the other hand, the main group of curvature covariates was negatively correlated with DDIFF, TCILOW and TWI. The last two topographic covariates showed similar behaviour between them,



i.e., distortion in the steeper or higher areas, but adequate distribution in the areas of lower slope and altitude.

### *1.1.3 Erosion vs accumulation variables*

The slope covariates (DWSLOPE, SLOPE and UPSLOPE) were positively correlated among them together with DISTV and MPI, hydrologic and morphometric covariates respectively. Whereas, they all were inversely correlated with the topo-climatic covariates LST, RDIFFUS, and TRASP; the morphometric MRRTF and MRVBF; and the topographic STWI; all of them positively correlated among them. That is, covariates pointing out erosion areas and transport rates were clustered, like did those related to deposit accumulation.

This correlation between morphometric and topo-climatic covariates is closely associated with ELEVATION and ASPECT, and it might be explained considering the strong altitudinal gradient typical of the project region. Thus, the lower the altitude, the lesser the slope and the erosion.

DWSLOPE and UPSLOPE assess catchment-area measurements and thereby they remain unaffected by sharp edges and punctual anomalies, which are largely an expression of anthropogenic actions on terraces and plots. In contrast, other covariates cannot avoid but portray them. Meanwhile, LST and, to a lesser extent, RDIFFUS are minimal modifications of ELEVATION, without a meaningful contribution.

### *3.2.2 Remote sensing variables*

The four remote sensing covariates were positively correlated. It is interesting to note that, although FEOUS was described to detect ferrous iron (Roecker et al., 2010), in this study it seems to mainly evidence lighter areas, which could be associated with the presence of carbonated materials on the soil surface after the erosion process and the truncation of the soil profile (endopedons Bk and Ck exposed by erosive or other anthropic processes). On the other hand, CI and RI, both suggested indicators of reddish areas by the presence of clay materials (Flynn et al., 2019), did not clearly detect specific areas.

### *1.1.4 Topo-Climatic variables*

Two of the five topo-climatic variables, DAH and RDIRECT, were highly correlated with each other, as well as with ASPECT. This is consistent with the consideration of DAH a covariate within R

factor (Hengl and Reuter, 2009, pp. 195–226), because it is calculated using ASPECT, but aiming to avoid its polarity.

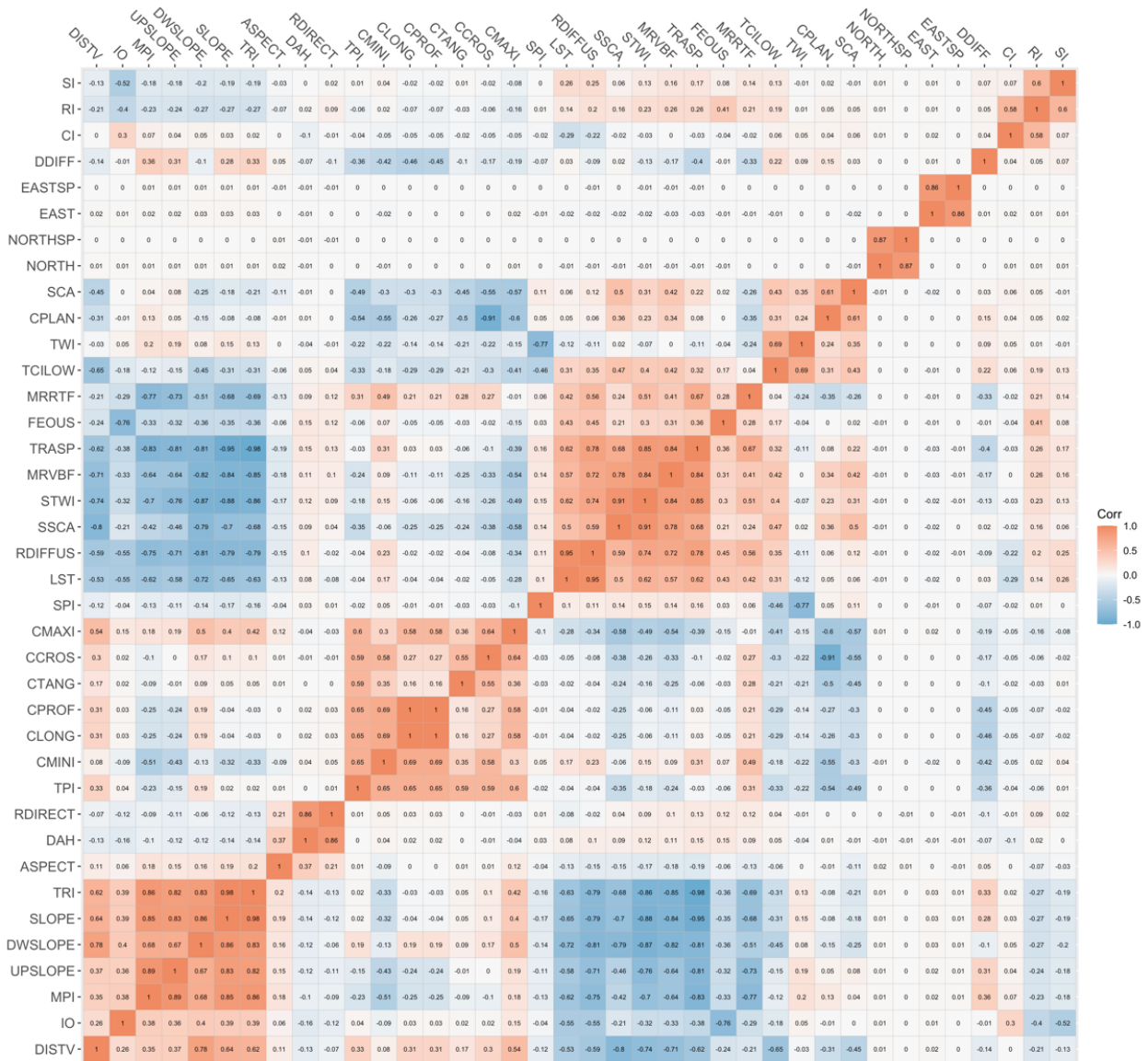


Figure 3-2. Spearman's correlation matrix of the covariates studied within the project. The four main associations.

ASPECT: Aspect; CCROS: Cross-sectional curvature; CI: Coloration Index; CLONG: Longitudinal curvature; CMAXI: Maximum curvature; CMINI: Minimum curvature; CPLAN: Plan curvature; CPROF: Profile curvature; CTANG: Tangential curvature; DAH: Diurnal Anisotropic Heat; DDIFF: Difference between DWSLOPE and SLOPE; DISTV: Vertical distance to cannel network; DWSLOPE: Downslope index; EASTSP: Interaction between EAST and SLOPE; FEOUS: Simple ratio Red ~ SWIR; LST: Land Surface Temperature; MPI: Morphometric Protection Index; MRRTF: Multiresolution Index of Ridge Top Flatness; MRVBF: Multiresolution Index of Valley Bottom Flatness; NORTH: Northerness; NORTHSP: Interaction between NORTH and SLOPE; RDIFFUS: Diffuse Potential Incoming Solar Radiation; RDIRECT: Direct Potential Incoming Solar Radiation; RI: Redness Index; SCA: Standard Catchment Area; SI: Salinity Index; SLOPE: Slope; SPI: Stream Power Index; SSCA: Modified Catchment Area by SAGA; STWI: Topographic Wetness Index by SAGA; SWIR: Short-wave infrared; TCLOW: Terrain Classification Index for Lowlands; TPI: Topographic Position Index; TRASP: Solar-radiation Aspect Index; TRI: Terrain Ruggedness Index; TWI: Topographic Wetness Index; UPSLOPE: Modified Catchment Slope from SAGA TWI.

### 3.3 Selection of covariates

Following the criteria established, the most relevant covariates within each association were selected. Particularly, in the curvatures group two different sets with comparable properties could be defined. Based on them, two similar collections were built, but using a set of curvatures of their own (Table 3-7). Ultimately, these collections are intended to define the space where the internal areas of the SMUs are demarcated and it is in the future analysis of these divisions that the whole process, including the collection, will be validated. Having two collections provides contrast of the process and offers insight into the effect of the cartographer's expert

In terms of the common covariates, first the association of erosion and deposit accumulation is considered. Although DWSLOPE and UPSLOPE were correlated among them, the qualitative analysis pointed out that their combination allowed contrasting different drainage network from several analysis windows. In addition, STWI was also included because of the altitudinal gradient characteristic of the project region its action in low areas, complementary to the other covariates.

As for the remote sensing covariates, it was feasible to work with one of them without losing much information due to their correlation. FEOUS was selected based on the qualitative analysis, its environmental meaning and spatial pattern.

Although the climatic variables were not very correlated, only DAH was adopted. In this way, the daily heating associated with the orientation of maximum insolation was considered. The rest of the covariates had very coarse regionalisations.

From the results of the correlation between curvature covariates, two options seemed to be suitable. The first one was the selection of one covariate from the group of canonical planes (CPLAN, CPROF) and another from the group of normal planes (CCROS, CLONG). They might be CCROS and CPROF, as the pattern of CCROS better outlines differences between ridges and valleys than CPLAN; and CPROF would be the valid counterpart. They would be complemented by CMAXI, the least correlated covariate with them. The second option was the selection of the extreme curvatures (CMAXI and CMINI) together with the least correlated covariate with them too, CTANG. Eventually, each variant contributed to its corresponding final collection. The first option from curvatures, composed by CCROS, CPROF and CMAXI, together with the bundle of common covariates formed the Collection One (C1), while the Collection Two (C2) was composed by the second choice of curvatures, namely CMAXI, CMINI and CTANG (Table 3-7).

The assembled collections comprise mostly covariates associated with SCORPAN R factor (six out of eight), with a balance between surfaces of curvatures against erosion and deposit accumulation. The two others were one topo-climatic covariate (C factor) and one remote sensing (S factor).

Table 3-7. Collections of covariates selected in the analysis.

Type of variable	Covariable collections names	
	C1	C2
<b>Curvatures</b>	CCROS + CPROF + CMAXI	CMAXI + CMINI + CTANG
<b>Erosion and accumulation</b>	DWSLOPE + UPSLOPE + STWI	
<b>Climate</b>	DAH	
<b>Remote sensing</b>	FEOUS	

CCROS: Cross-sectional curvature; CMAXI: Maximum curvature; CMINI: Minimum curvature; CPROF: Profile curvature; CTANG: Tangential curvature; DAH: Diurnal Anisotropic Heat; DWSLOPE: Downslope index; FEOUS: Simple ratio Red ~ SWIR; STWI: Topographic Wetness Index by SAGA; UPSLOPE: Modified Catchment Slope from SAGA TWI.

### 3.4 Evaluation of the new homogeneous areas

#### 3.4.1 Comparative of delineations

A notable increase in the number of new delineations was observed from both sets of delineations in comparison with those obtained in the conventional SRI map. The initial 7,554 delineations corresponding to the ten studied SMUs increased to 220,671 in C1 and 244,889 in C2.

These results indicated that C2 rendered around 10% more delineations than C1. However, the difference was uneven among SMUs, while C2 produced more delineations in the SMU 09C, 10C, 16C and 43C, C1 rendered more delineations in the remainder. Regardless of the collection of covariates in use, many of the new delineations had areas ranging between 25 and 75 m<sup>2</sup>, equivalent to the size associated with one and three pixels. These delineations or groups of pixels could be considered speckled areas and, therefore, as noise across the map. This fact was particularly relevant in C2, which explains why the average-size area of its delineations was smaller (1,304 m<sup>2</sup>) than the one defined by C1 (1,447 m<sup>2</sup>).

#### 3.4.2 Comparative of divisions within each Soil Map Unit

The number of new divisions generated from each collection of covariates was similar, although slightly higher in C2 (46) than in C1 (44) (Table 3-8). Hence, the proposed methodology increased notably the internal divisions of the studied SMUs.

Table 3-8. Number of new divisions of each Soil Map Unit using the two collections of covariates.

		SMUs of the conventional SRI										
		08C	09C	10C	13C	16C	17C	21C	23C	29C	43C	Total
Collection of covariates	C1	4	4	4	4	4	5	4	5	6	4	44
	C2	4	7	6	4	4	4	4	4	4	6	46

SMU: Soil Map Unit; SRI: Soil Resource Inventory.

The combined analysis of the medoids of each new division paired with the visualization of the new delineations over high-resolution orthophotos (CNIG) suggests that their new divisions had an own spatial and physical meaning within the SMUs and so they could be considered new SMU on their own.

Overall, the boundaries of both sets of divisions overlapped, and only a few mid-sized discrepancies were consistently detected in those SMUs split into a different number of potential new SMUs. In this regard, there were 3 distinct scenarios.

As an example to illustrate the overlaps between the new divisions, the results from SMU 08C were taken. The SMU 08C extends over the level of a river terrace and it was divided into four new areas. One of them was characterised by high values of DWSLOPE, CCROS (in C2) and CMAXI (in C1) and low values of UPSLOPE and STWI, being an area with potential high energetic loss, convex and divergent, but with low internal slope and catchment area. Another division showed inverse values to the first division discussed above and highlighted the terrace escarpments. The other two new areas were located where the terrace is wider, but they were mainly opposed by their FEOUS values, having low and high values respectively. This outcome seemed to be associated with surficial horizons with different levels of calcium carbonate, a fact quite common in the studied area. However, it should be noted that some crops, specially irrigated crops, may distort it.

A higher number of new divisions in C2 than in C1 were generated in the SMUs 09C and 10C, which extend over two different older levels of a river terrace than the SMU 08C. This fact was related to some divisions with extreme values of CMAXI and CMINI, although without a clear physical meaning. A similar outcome was observed in the SMU 43C, where the higher number of C1 divisions was largely due to the interaction between CMAXI and FEOUS.

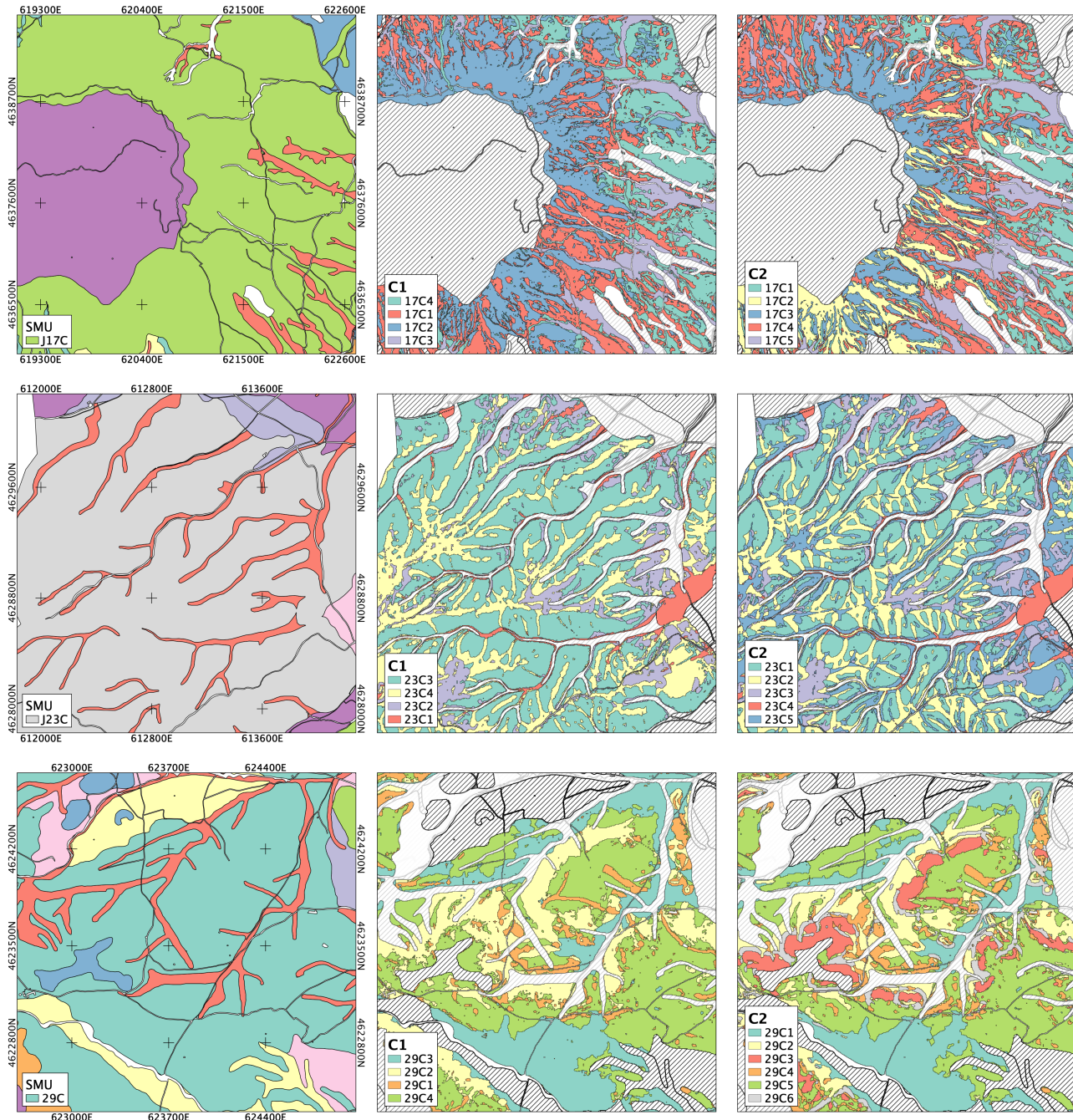


Figure 3-3. Soil map units from the conventional soil resource inventory map (left), new divisions from the collection C1 (middle) and C2 (right). From top to bottom, the maps correspond to the soil map units 17C, 23C and 29C. SMU: Soil Map Unit.

Conversely, C1 collection allowed detecting more divisions than C2 in the case of the SMU 17C, which is defined by hillsides coming from a horizontal surface located above its surroundings (Figure 3-3). The distinction might be caused by to the interaction associated with ASPECT factor. In this way, C1 allowed differential areas of greater insolation (DAH) and less steep slope (CPROF) against another with opposite values. Instead, this scenario was mostly driven by curvature

factors in the case of the SMU 23 C, which collects areas of pronounced relief and high elevation and differences by flows distribution are predominant (Figure 3-3). Here, C1 allowed distinguishing two divisions with high DWSLOPE and opposite CCROS values, one divergent and the other convergent. In addition, the coupling of CCROS and CPROF resulted in the differentiation of a new division characterised by ridges and deceleration zones. Besides, new concave zones of low altitude and higher deposits were detected by high STWI and low CMAXI values. The differentiation by flows and by the transition of higher erosion to deposits and accumulation areas also occurred in the new divisions of the SMU 29C when the covariable collection C2 was used (Figure 3-3).

Altogether, the new divisions created in the SMUs have given rise to new smaller and more homogeneous areas with a strong environmental meaning. The comparison of the disagreements between the set from each collection has also revealed that the specific curvature covariates have had a noteworthy influence on the division process. C1 assembled new divisions with clearer definition and the delineations showed limited noise and minor incidents. In contrast, C2 drew more delineations, but they were in the form of noise rather than finer detail, and when it resulted in a larger number of divisions, they showed less clear environmental meaning. From the previous comments, the C1 collection was selected to render the better divisions of the studied SMUs.

## **4 Conclusions**

A new methodology capable to render smaller, more homogeneous, and well delimited new divisions within the SMUs of conventional soil maps has been developed. Within this process, a quantitative analysis (correlation study) together with a qualitative analysis (visual exploration of environmental meaning) of the covariates are crucial tasks to tailor them to the environmental factors that characterise the project region and to benefit from the expert knowledge on these parameters. The unsupervised classification method CLARA with Mahalanobis distance is useful to obtain satisfactory results, and the comparative analysis of the geometries and environmental meaning of the new divisions is a valid criterion to choose a definitive covariates collection. Nonetheless, the proposed methodology ought to be tested in different regions, and it could be complemented with additional covariates related to soil attributes and with sources of ancillary data. Moreover, this study opens a new approach towards initiating the disaggregation of intensive conventional soil maps.

## 5 References

- Brenning, A., Bangs, D., Becker, M., 2018. RSAGA: SAGA geoprocessing and terrain analysis.
- Bui, E.N., Moran, C.J., 2001. Disaggregation of polygons of surficial geology and soil maps using spatial modelling and legacy data. *Geoderma* 103, 79–94. [https://doi.org/10.1016/S0016-7061\(01\)00070-2](https://doi.org/10.1016/S0016-7061(01)00070-2)
- Burrough, P.A., 2000. High-resolution landform classification using fuzzy k-means. *Fuzzy Sets Syst.* 16.
- Charrad, M., Ghazzali, N., Boiteau, V., Niknafs, A., 2014. NbClust: An R package for determining the relevant number of clusters in a data set. *J. Stat. Softw.* 61, 1–36. <https://doi.org/10.18637/jss.v061.i06>
- Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J., 2015. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geosci. Model Dev.* 8, 1991–2007. <https://doi.org/10.5194/gmd-8-1991-2015>
- Escadafal, R., 1993. Remote sensing of soil color: Principles and applications. *Remote Sens. Rev.* 7, 261–279. <https://doi.org/10.1080/02757259309532181>
- Florinsky, I.V., 2017. An illustrated introduction to general geomorphometry. *Prog. Phys. Geogr. Earth Environ.* 41, 723–752. <https://doi.org/10.1177/0309133317733667>
- Flynn, T., Rozanov, A., de Clercq, W., Warr, B., Clarke, C., 2019a. Semi-automatic disaggregation of a national resource inventory into a farm-scale soil depth class map. *Geoderma* 337, 1136–1145. <https://doi.org/10.1016/j.geoderma.2018.11.003>
- Flynn, T., van Zijl, G., van Tol, J., Botha, C., Rozanov, A., Warr, B., Clarke, C., 2019b. Comparing algorithms to disaggregate complex soil polygons in contrasting environments. *Geoderma* 352, 171–180. <https://doi.org/10.1016/j.geoderma.2019.06.013>
- Gallant, J.C., Austin, J.M., 2015. Derivation of terrain covariates for digital soil mapping in Australia. *Soil Res.* 53, 895. <https://doi.org/10.1071/SR14271>
- Gallant, J.C., Dowling, T.I., 2003. A multiresolution index of valley bottom flatness for mapping depositional areas. *Water Resour. Res.* 39. <https://doi.org/10.1029/2002WR001426>
- Gómez-Miguel, V.D., Sotés, V., 2015. Zonificación del Terroir: Estudio de Suelos y Ordenación del Cultivo de la Vid en la DO Campo de Borja (Zaragoza). Universidad Politécnica de Madrid (UPM), Madrid.
- Häring, T., Dietz, E., Osenstetter, S., Koschitzki, T., Schröder, B., 2012. Spatial disaggregation of complex soil map units: A decision-tree based approach in Bavarian forest soils. *Geoderma* 185–186, 37–47. <https://doi.org/10.1016/j.geoderma.2012.04.001>
- Hengl, T., 2006. Finding the right pixel size. *Comput. Geosci.* 32, 1283–1298. <https://doi.org/10.1016/j.cageo.2005.11.008>
- Hengl, T., Reuter, H.I. (Eds.), 2009. *Geomorphometry: concepts, software, applications*, 1st ed. ed, *Developments in soil science*. Elsevier, Amsterdam, Netherlands; Oxford, UK; Boston [Mass.].
- Hijmans, R.J., 2019. raster: Geographic data analysis and modeling.



- Hjerdt, K.N., McDonnell, J.J., Seibert, J., Rodhe, A., 2004. A new topographic index to quantify downslope controls on local drainage. *Water Resour. Res.* 40. <https://doi.org/10.1029/2004WR003130>
- Kane, M.J., Emerson, J., Weston, S., 2013. Scalable strategies for computing with massive data. *J. Stat. Softw.* 55, 1–19.
- Koch, A., McBratney, A., Adams, M., Field, D., Hill, R., Crawford, J., Minasny, B., Lal, R., Abbott, L., O'Donnell, A., Angers, D., Baldock, J., Barbier, E., Binkley, D., Parton, W., Wall, D.H., Bird, M., Bouma, J., Chenu, C., Flora, C.B., Goulding, K., Grunwald, S., Hempel, J., Jastrow, J., Lehmann, J., Lorenz, K., Morgan, C.L., Rice, C.W., Whitehead, D., Young, I., Zimmermann, M., 2013. Soil Security: Solving the Global Soil Crisis. *Global Policy* 4, 434–441. <https://doi.org/10.1111/1758-5899.12096>
- Microsoft Corporation, Weston, S., 2019. doParallel: Foreach parallel adaptor for the “parallel” package.
- Mouselimis, L., 2019a. ClusterR: Gaussian mixture models, k-means, mini-batch-kmeans, k-medoids and affinity propagation clustering.
- Mouselimis, L., 2019b. Functionality of the ClusterR package. *Funct. Clust. Package*. URL [https://cran.r-project.org/web/packages/ClusterR/vignettes/the\\_clusterR\\_package.html](https://cran.r-project.org/web/packages/ClusterR/vignettes/the_clusterR_package.html) (accessed 7.30.19).
- Mulder, V.L., de Bruin, S., Schaepman, M.E., Mayr, T.R., 2011. The use of remote sensing in soil and terrain mapping — A review. *Geoderma* 162, 1–19. <https://doi.org/10.1016/j.geoderma.2010.12.018>
- Odeh, I.O.A., McBratney, A.B., Chittleborough, D.J., 1992. Soil Pattern Recognition with Fuzzy-c-means: Application to Classification and Soil-Landform Interrelationships. *Soil Sci. Soc. Am. J.* 56, 505–516. <https://doi.org/doi:10.2136/sssaj1992.03615995005600020027x>
- Pebesma, E., 2018. Simple features for r: Standardized support for spatial vector data. *R J.* 10, 439–446. <https://doi.org/10.32614/RJ-2018-009>
- QGIS Development Team, 2019. QGIS Geographic Information System. Open Source Geospatial Foundation Project.
- Roberts, D.W., Cooper, S.V., 1989. Concepts and techniques of vegetation mapping, in: *Land Classifications Based on Vegetation: Applications for Resource Management*, INT-GTR. USDA Forest Service, Ogden, UT, pp. 90–96.
- Roecker, S.M., Howell, D.W., Haydu-Houdeshell, C.A., Blinn, C., 2010. Conventional Soil Survey and Digital Soil Mapping Approaches, in: Boettinger, J.L. (Ed.), *Digital Soil Mapping: Bridging Research, Environmental Application, and Operation*, Progress in Soil Science. Springer, Dordrecht [Netherlands]; London, pp. 369–384.
- Soil Survey Division Staff, 1993. *Soil Survey Manual*, USDA Handbook. U.S. Government Printing Office, Washington, D.C.
- Soil Survey Division Staff, 2014. *Keys to Soil Taxonomy*, 12th Edition. ed. USDA-Natural Resources Conservation Service, Washington, DC.
- Stevenson, J.A., Sun, X., Mitchell, N.C., 2010. Despeckling SRTM and other topographic data with a denoising algorithm. *Geomorphology* 114, 238–252. <https://doi.org/10.1016/j.geomorph.2009.07.006>

- Sun, X., Rosin, P.L., Martin, R., Langbein, F., 2007. Fast and Effective Feature-Preserving Mesh Denoising. *IEEE Trans. Vis. Comput. Graph.* 13, 925–938. <https://doi.org/10.1109/TVCG.2007.1065>
- Taghizadeh-Mehrjardi, R., Nabiollahi, K., Minasny, B., Triantafilis, J., 2015. Comparing data mining classifiers to predict spatial distribution of USDA-family soil groups in Baneh region, Iran. *Geoderma* 253–254, 67–77. <https://doi.org/10.1016/j.geoderma.2015.04.008>
- Thompson, S.D., Nelson, T.A., Giesbrecht, I., Frazer, G., Saunders, S.C., 2016. Data-driven regionalization of forested and non-forested ecosystems in coastal British Columbia with LiDAR and RapidEye imagery. *Appl. Geogr.* 69, 35–50. <https://doi.org/10.1016/j.apgeog.2016.02.002>
- Vaudour, E., Costantini, E., Jones, G.V., Mocali, S., 2015. An overview of the recent approaches to terroir functional modelling, footprinting and zoning. *SOIL* 1, 287–312. <https://doi.org/10.5194/soil-1-287-2015>
- Wang, L., Liu, H., 2006. An efficient method for identifying and filling surface depressions in digital elevation models for hydrologic analysis and modelling. *Int. J. Geogr. Inf. Sci.* 20, 193–213. <https://doi.org/10.1080/13658810500433453>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H., 2019a. Welcome to the Tidyverse. *J. Open Source Softw.* 4, 1686. <https://doi.org/10.21105/joss.01686>
- Wickham, H., Ooms, J., Müller, K., 2019b. RPostgres: “rcpp” interface to “PostgreSQL.”
- Wilson, J.P., Gallant, J.C., 2000. *Terrain Analysis: Principles and Applications*. Wiley & Sons.
- Wood, J., 1996. *The Geomorphological Characterisation of Digital Elevation Models* (PhD thesis). University of Leicester.
- Yokoyama, R., Shirasawa, M., Pike, R.J., 2002. Visualizing Topography by Openness: A New Application of Image Processing to Digital Elevation Models. *Photogramm. Eng. Remote Sens.* 257–265.
- Zare, E., Huang, J., Triantafilis, J., 2016. Identifying soil landscape units at the district scale by numerically clustering remote and proximal sensed data. *Comput. Electron. Agric.* 127, 510–520. <https://doi.org/10.1016/j.compag.2016.07.001>

# **CAPÍTULO 4 DISGREGACIÓN DEL MAPA DE SUELOS CONVENCIONAL DE LA DO CAMPO DE BORJA**

## **1 Introducción**

La disgregación es una línea de acción de DSM que engloba aquellas metodologías que buscan revelar el patrón de distribución espacial de los suelos no explícito dentro de las SMU politáxicas, y que lo hacen aprovechando la información disponible en mapas de suelos heredados. El objetivo es generar una nueva cartografía que amplíe el detalle cartográfico de acuerdo con el nivel categórico capturado en el IRS original (Lázaro-López et al., 2021). Para ello, se utilizan técnicas estadísticas de clasificación de forma que se generan modelos cuantitativos de las categorías de suelos en base a los parámetros ambientales continuos que representan a los factores SCORPAN, y se obtienen predicciones para cada localización o celda del ráster, así como la incertidumbre asociada.

En los últimos años se ha producido un aumento notable de la publicación de metodologías de disgregación. La mayoría se han basado en técnicas de clasificación supervisadas, especialmente DT y RF (Lázaro-López et al., 2021), y entre las que destaca DSMART (Odgers et al., 2014). En este tipo de técnicas, el muestreo y la asignación de categorías para construir los conjuntos de entrenamiento son tareas críticas que tienen una fuerte influencia en el resultado. En este sentido, las variaciones en las aplicaciones más recientes de disgregación y DSMART se han centrado en mejorar estos dos aspectos clave siguiendo estrategias de dos etapas. En una primera se definen las áreas potencialmente asociadas a las categorías de suelo para etiquetar los píxeles coincidentes, y a partir de ellas, en segundo lugar, se crean los modelos supervisados. De esta forma se aprovecha la información disponible en las leyendas de los mapas de CSM y el conocimiento experto de cartógrafos y edafólogos, a la vez que se reduce la incidencia de asignaciones erróneas (Lázaro-López et al., 2021).

Para la definición de las áreas potencialmente asociadas a categorías de suelos de la primera etapa se han seguido distintas aproximaciones en las diferentes metodologías. Una de ellas se refiere a las reglas de conocimiento experto sobre las covariables geomorfométricas, esto es, se asignan categorías de suelos a regiones del espacio de las covariables designadas por rangos y esto se

traduce en zonas geográficas delimitadas con las que se estratifica el paisaje (Costa et al., 2019; Nauman y Thompson, 2014; Sarmiento et al., 2017; Vincent et al., 2018). De forma alternativa, se ha recurrido a geoformas, cuyos límites espaciales están asociados a la acción de procesos edafológicos e hidrológicos, y partir de ellas se han definido regiones en el espacio de las covariables que han sido asociadas manualmente con categorías de suelos (Flynn et al., 2019; Machado et al., 2018).

Además de estos, existen otros sistemas para estratificar el paisaje en geoformas que aún no han sido probadas en metodologías de disgregación, como los algoritmos de agrupación no supervisada fuzzy k-means clustering (Irvin et al., 1997; Schmidt y Hewitt, 2004) o nested-means (Iwahashi y Pike, 2007).

El objetivo de esta publicación es definir un método alternativo de disgregación de mapas de suelos convencionales con SMU politáxicas que esté dirigido por datos, adaptado al entorno, y que no demande un conocimiento detallado previo, esto es, se base en técnicas de clasificación no supervisadas de agrupación, en línea con métodos de estratificación del paisaje en geoformas. Con este fin, se aprovecha el método de división de SMU para detectar áreas homogéneas mediante CLARA con distancia de Mahalanobis descrito en el Capítulo 2 y que después son asociadas a STU. El diseño de este sistema de disgregación además tiene la posibilidad de dar soporte a estrategias de dos pasos y servir así de forma dual.

La metodología es puesta en práctica en la DO Campo de Borja, una región vitivinícola de España, que dispone de un mapa de suelos convencional de un IRS intensivo. En el contexto del sector vitivinícola existe la demanda de una cartografía con SMU monotáxicas de mayor detalle (con escalas próximas a 1:10.000), buscando abordar la caracterización y gestión de los terroir y la Viticultura de Precisión. En este sentido, se persigue que el resultado sea equiparable al existente y que, por tanto, se genere una salida de datos vectorizada y adaptada a la escala del mapa.

## **2 Materiales y métodos**

### **2.1 Área de trabajo**

El área de trabajo es la DO Campo de Borja (Figure 3-1), situada al noreste de España en la Comunidad Autónoma de Aragón (41° 55' 28.9" a 41° 37' 15" N y 1° 44' 34.5" a 1° 19' 1.6" O). Se

trata de una región vitivinícola histórica con una superficie de 653,37 km<sup>2</sup> en la que se localizan 6.416,14 ha de viñedo (2019) y que se extiende por el somontano del Moncayo (2.316 msnm) hasta las terrazas del río Ebro en un rango de elevaciones desde 250 a 900 msnm. La geología está caracterizada por ambas unidades estructurales, con materiales de todas las edades en donde predominan formaciones del Mesozoico con alternancia de calizas, margas y areniscas; y su relieve está condicionado por el encajamiento de la red fluvial, especialmente en la parte más occidental de la región. La climatología en la zona es de tipo continental con influencia invernal atlántica, en donde destaca el viento regional del noroeste denominado Cierzo (frío y seco). Los contrastes térmicos diurnos y estacionales son muy acusados, con temperaturas extremas, y presenta precipitaciones escasas, con un promedio de 350 mm en las zonas bajas y de 450 mm en las más elevadas.

## 2.2 Fuentes de datos

### 2.2.1 *Mapa convencional de suelos*

El mapa de suelos convencional del IRS a disgregar procede del proyecto de zonificación integrada del terroir de la DO (Gómez-Miguel y Sotés, 2015). Fue realizado a escala 1:25.000 mediante metodología USDA (Soil Survey Division Staff, 1993) (Figura 4-1), donde las delineaciones se basaron en fotointerpretación aérea (API) y en trabajo de campo con una densidad de 3,8 observaciones/km<sup>2</sup> entre calicatas y sondeos. Está compuesto por SMU politáxicas de asociaciones y complejos (46) con fases (343) y con un contenido de STU clasificadas según ST (Soil Survey Division Staff, 2014). En su mayoría integran entre 4 y 6 componentes, entre STU principales e inclusiones, con los porcentajes de ocupación obtenidos a partir del conjunto de las observaciones (calicatas y sondeos). Las STU comprenden series de suelos (20) y, cuando estas no pudieron ser definidas, familias asignadas a su litología específica (122).

En ellas se identifican cuatro órdenes (Alfisol, Inceptisol, Entisol y Mollisol), ocho subórdenes (Xeralf, Aqualf, Aquept, Xerept, Aquent, Fluvent y Orthent) y doce grupos (Haploxeroll, Palexeralf, Haploxeralf, Rhodoxeralf, Epiaqualf, Epiaquept, Haploxerept, Calcixerept, Fluvaquent, Epiaquent, Xerofluvent, Xerorthent). Además, el mapa recoge zonas significativamente alteradas por la acción antrópica (estructuras y construcciones) como áreas misceláneas.

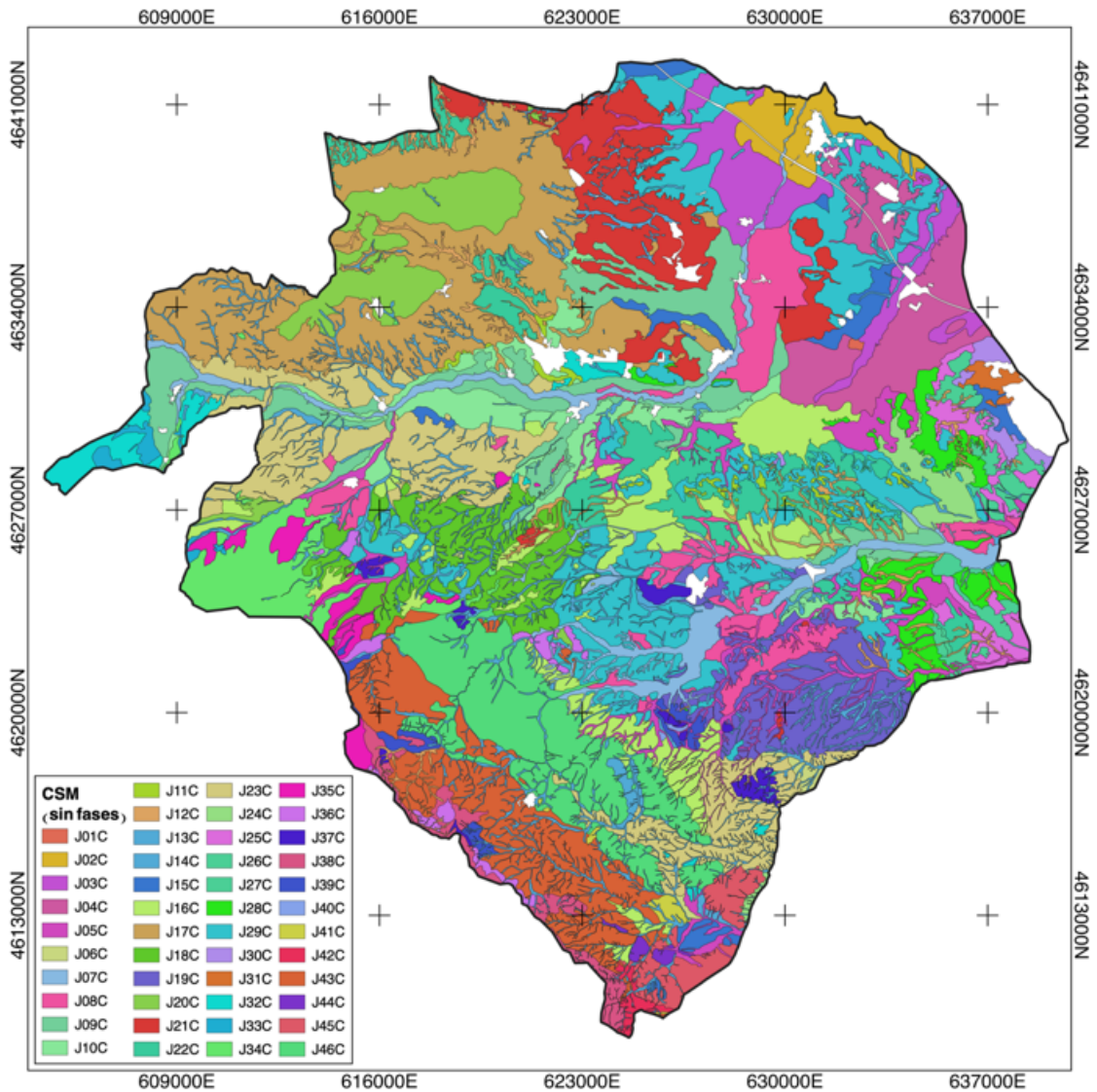


Figura 4-1. Mapa del inventario del recurso suelo convencional de la Denominación de Origen Campo de Borja y sus 46 unidades cartográficas de suelos sin fases. CSM: Cartografía de Suelos Convencional.

### 2.2.2 Observaciones de suelos heredadas

Se utilizan 519 observaciones puntuales de suelos procedentes del conjunto de observaciones del proceso de cartografía del IRS, identificadas con perfiles de suelos descritos, muestreados y analizados en todos sus horizontes (Figura 4-2). Las observaciones también están clasificadas según ST y están asociados a alguna de las STU recogidas en la leyenda del mapa.

La distribución de las observaciones por STU se hizo por muestreo dirigido y no es regular, si bien guarda una gran relación con la distribución de las superficies estimadas para las clases en base a la leyenda del mapa, calculadas según el porcentaje de las STU por cada SMU agregado (Tabla 4-1).

Por ello, cabría considerar que las observaciones son representativas del mapa para su utilización durante el proceso de cartografía digital a pesar de su origen en un muestreo dirigido.

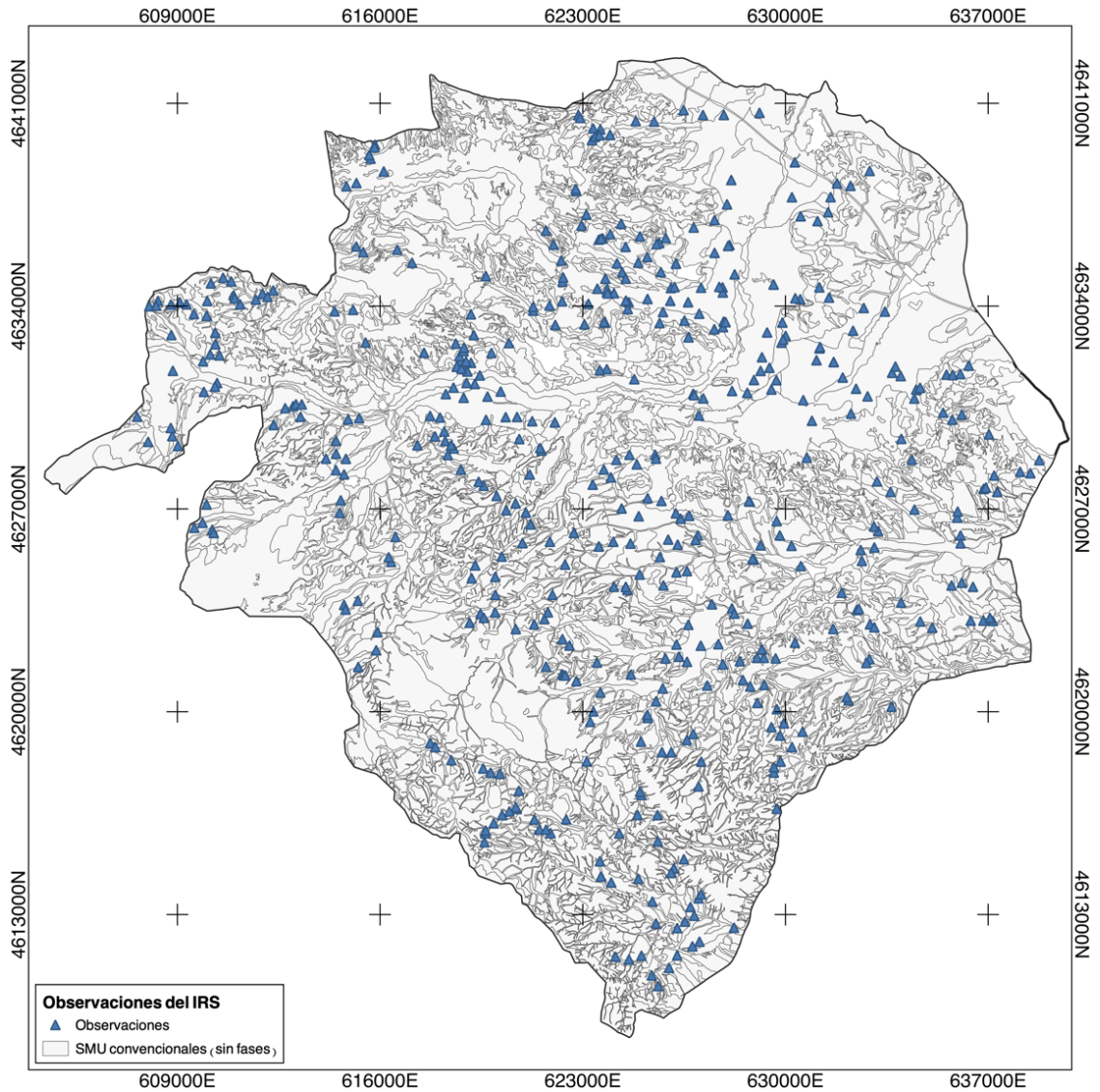


Figura 4-2. Localización de las observaciones (perfiles de suelos y excluyendo sondeos) del inventario del recurso suelo convencional fijadas mediante muestreo dirigido en la Denominación de Origen Campo de Borja.

IRS: Inventario del Recurso Suelo, SMU: Unidad Cartográfica de Suelos

Tabla 4-1. Distribución de las observaciones (perfiles de suelos) por Subgrupos de «Soil Taxonomy» y la superficie estimada para cada uno calculada según el porcentaje de las Unidades Taxonómicas de Suelo correspondientes en cada Unidad Cartográfica de Suelos.

Subgrupo de ST	Obs. (nº)	Perfiles por clase (%)	Área (ha)	Área por clase (%)	Diferencia entre perfiles y área (%)
<b>Typic Calcixerept</b>	229	49,9	31.858,85	49,2	0,7
<b>Calcic Haploxeralf</b>	71	15,5	12.129,34	18,7	-3,2
<b>Typic Xerofluvent</b>	54	11,8	4.246,96	6,6	5,2
<b>Typic Xerorthent</b>	51	11,1	7.251,19	11,2	-0,1
<b>Typic Haploxeralf</b>	24	5,2	4.100,79	6,3	-1,1
<b>Petrocalcic Calcixerept</b>	8	1,7	1.006,32	1,6	0,1
<b>Lithic Xerorthent</b>	5	1,1	252,02	0,4	0,7
<b>Calcic Palexeralf</b>	4	0,9	1.440,6	2,2	-1,3
<b>Typic Palexeralf</b>	4	0,9	479,93	0,7	0,2
<b>Petrocalcic Palexeralf</b>	2	0,4	286,83	0,4	0
<b>Calcic Rhodoxeralf</b>	2	0,4	56,3	0,1	0,3
<b>Typic Epiaquept</b>	1	0,2	160,89	0,2	0
<b>Aquic Haploxeralf</b>	1	0,2	68,15	0,1	0,1
<b>Gypsic Haploxerept</b>	1	0,2	68,15	0,1	0,1
<b>Cumulic Haploxeroll</b>	1	0,2	111,09	0,2	0
<b>Petrocalcic Rhodoxeralf</b>	1	0,2	75,11	0,1	0,1
<b>Total</b>	459		<b>63.592,52</b>	<b>Media</b>	<b>0,11</b>

ST: Soil Taxonomy; STU: Unidad Taxonómica de Suelo.

### 1.1.1 Covariables

Se utilizó una colección de 8 covariables como parámetros ambientales asociados a los factores SCORPAN (McBratney et al., 2003) (Tabla 4-2) a partir de un análisis y selección sobre un conjunto inicial de 38 covariables geomorfométricas, hidrológicas, topo-climáticas y edáficas (Capítulo 3).

Esta aplicación específica de la metodología de disgregación busca un nuevo mapa de suelos compatible con un IRS cuya escala óptima esté en torno a 1:10.000. Por ello, el tamaño de malla de todos los ráster se ha ajustado a 5 m, tomando como referencia la relación general entre los mapas vectoriales convencionales y los de formato ráster señalada por Hengl (2006).



Tabla 4-2. Colección de covariables usadas en la disgregación con su descripción y el factor SCORPAN al que se asocian.

Factor SCORPAN	Covariable	Descripción
<b>Morfométricas (R)</b>	<i>Curvatura transversal (CCROS)</i>	Curvatura definida por la intersección del plano normal a la superficie y la tangente al contorno. Cuantifica la convergencia de flujos de mayor cota por gravedad y lateralmente en los suelos cuando los valores son negativos, y la divergencia en los positivos. Su sentido geomorfológico se relaciona con la presencia de crestas y valles estrechos (Florinsky, 2017). Se calcula integrando múltiples escalas (Wood, 1996).
	<i>Curvatura de perfil (CPROF)</i>	Curvatura con la intersección del plano definido por el eje Z y la máxima pendiente. Valores positivos describen perfiles convexos y los negativos perfiles cóncavos (Florinsky, 2017). Se calcula integrando múltiples escalas (Wood, 1996).
	<i>Curvatura máxima (CMAXI)</i>	Máxima curvatura en cualquier dirección. Valores positivos corresponden con unidades alongadas convexas, como crestas, mientras que valores negativos se relacionan con unidades cóncavas (Florinsky, 2017). Se calcula integrando múltiples escalas (Wood, 1996).
<b>Hidrológicas (R)</b>	<i>Pendiente cuesta abajo (DWSLOPE)</i>	Índice topográfico que cuantifica el control de la pendiente en la red de drenaje. Los valores tienden a ser bajos para perfiles cóncavos y altos para perfiles convexos. Se ve menos afectado por la resolución del MDT que otros índices locales (Hjerdt et al., 2004).
	<i>Pendiente en el área de captación modificada de SAGA (UPSLOPE)</i>	Pendiente en el área de captación modificada del Índice Topográfico de Humedad de SAGA.
	<i>Índice Topográfico de Humedad de SAGA (STWI)</i>	Similar al Índice Topográfico de Humedad, pero basado en un área de captación modificada. Señala el potencial de humedad, mayor para las zonas bajas y de acumulación.
<b>Climática (C)</b>	<i>Calor anisotrópico diurno (DAH)</i>	Medida del calentamiento diurno relativo en función de la orientación de máxima insolación
<b>Suelo (S)</b>	<i>Ratio simple entre las regiones del rojo y del SWIR (FEOUS)</i>	Permite detectar áreas ricas en hierro ferroso (Roecker et al., 2010).

### 2.2.3 Estructuras y construcciones misceláneas

Durante la delineación inicial del mapa de suelos convencional no fueron considerados aquellos elementos asociados a las áreas que son detectables a escalas superiores y cuya integración favorece el proceso de disgregación (Häring et al., 2012). Como representación de esos elementos, se utilizó información disponible de carreteras, autopistas, ferrocarriles y poblamientos desde la Base Topográfica Nacional del CNIG y el Catastro Nacional.

## 2.3 Procedimiento

La metodología de disgregación que se propone comienza con la búsqueda de áreas potencialmente homogéneas de suelos mediante la división de las SMU del mapa convencional con la técnica de clasificación no supervisada de agrupación CLARA con distancia de Mahalanobis (Capítulo 3). Las divisiones o grupos resultado de este tipo de técnicas no están asociados a categorías de suelos, como sí ocurre en los métodos supervisados, y por ello es necesario acometer la correlación entre ambos elementos. Para ello, se recurre a la relación entre las ubicaciones de perfiles y medoides de los grupos en el espacio de las covariables (Figura 4-3).

Con el fin de ajustar el resultado a una salida vectorial y a criterios de publicación estándar, se vectorizan los grupos y se refinan las delineaciones de forma recursiva hasta alcanzar un equilibrio entre el MLA, ASA y la escala del mapa disgregado. Todo este proceso es finalmente validado utilizando observaciones de suelos heredadas del proceso cartográfico convencional.

### 2.3.1 División de las SMU

A diferencia de los métodos supervisados que dan lugar a modelos suelo ~ paisaje complejos directamente a costa de esfuerzos computacionales intensos y de conocimientos técnicos para gestionarlos, los modelos no supervisados como el obtenido mediante CLARA son más sencillos, aunque están sujetos a correlaciones entre grupos no definidos y categorías.

En primer lugar, se analizan las covariables cuantitativa y cualitativamente para seleccionar aquella colección reducida que mejor revele la variación ambiental y edáfica, a la vez que mantenga al mínimo la redundancia interna y el número total de covariables.

Durante el proceso se utilizan las SMU sin fases, ya que son estas las que definen el contenido en términos STU y las fases aportan características de interés particulares del cartógrafo, pero que no están recogidas en la clasificación taxonómica. De estas también se excluyen aquellas zonas recogidas en las capas de estructuras y construcciones, que son designadas como nuevas áreas misceláneas.

El número de nuevas divisiones para cada SMU es fijado en base a varios índices cuantitativos (Charrad et al., 2014) de forma que el valor óptimo esté dentro de un rango basado en la leyenda del mapa (Capítulo 3).

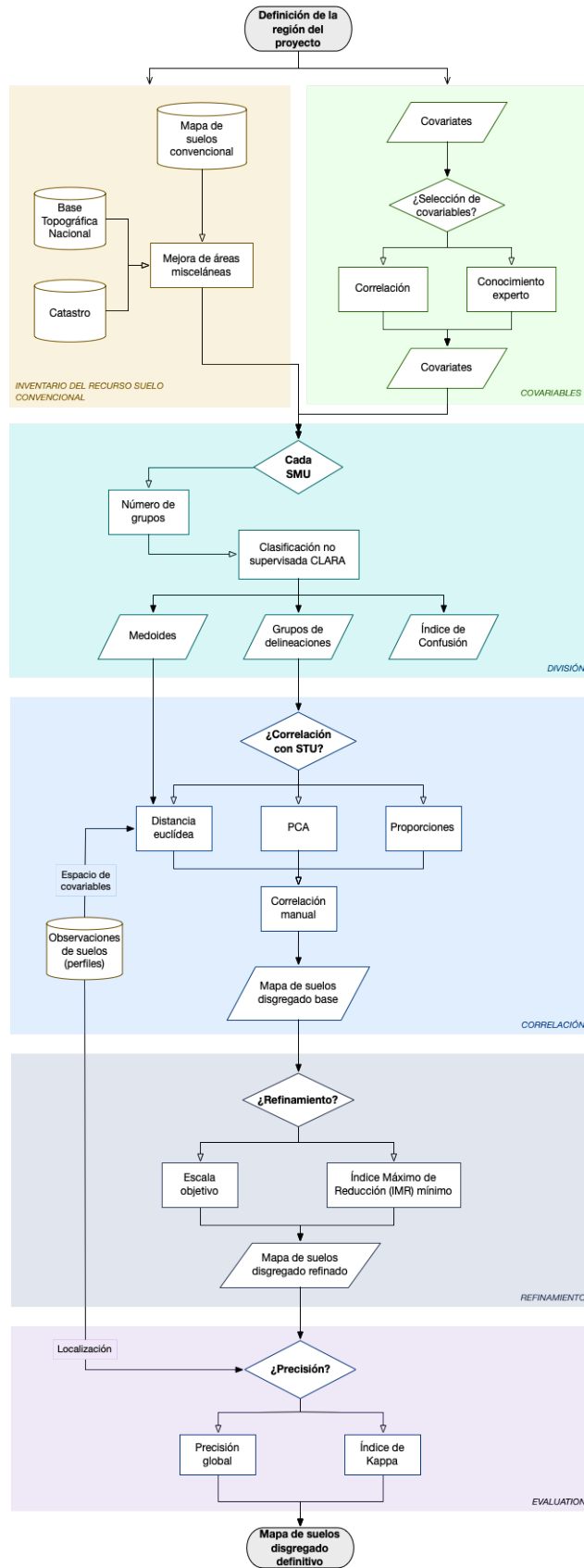


Figura 4-3. Diagrama de flujo de la metodología de disgregación mediante la división de las unidades cartográficas de suelos convencionales, la correlación de los grupos con las unidades taxonómicas de suelos de origen y el refinamiento de las delineaciones. SMU: Unidad Cartográfica de Suelos; STU: Unidad Taxonómica de Suelo.

En última instancia, este método es capaz de delimitar nuevas divisiones o grupos de delineaciones más pequeñas y homogéneas dentro de las SMU, pero que no están adscritas a categorías de suelos.

### *2.3.2 Correlación entre divisiones y STU*

El contenido categórico de los grupos delimitados dentro de las SMU se define en función de las STU recogidas en la leyenda del mapa para cada una a través de un proceso de correlación apoyado en datos y dirigido por conocimiento experto.

En él se tomaron como representantes de los grupos, a sus medoides, y de las STU, a aquellas observaciones de suelos asociadas a estas. Así, se consideraron en el espacio de las covariables: i) la distancia euclídea entre los medoides y las STU, como la media de las distancias a las observaciones de suelos asociadas a cada una; ii) la distribución de todos ellos caracterizada mediante PCA; y iii) los porcentajes de ocupación de la superficie de la SMU designados en la leyenda frente a los obtenidos por los grupos.

Con todo ello se pretendió que la distancia entre los grupos y las STU fuera la menor posible a la vez que los niveles de ocupación fuesen aproximados.

De esta forma queda establecida la composición en términos de STU de unas potenciales SMU que pueden ser tanto monotáxicas como politáxicas, cuando la división no ha dado lugar a suficientes grupos o la caracterización se juzga con un margen alto de incertidumbre, por ejemplo, la distancia a un medoide es similar para varias STU. Cuando esto ocurre, las nuevas ratios de ocupación asumen las proporciones dispuestas en la leyenda del mapa convencional. Con este procedimiento de correlación se asegura que todas las STU del mapa original quedan recogidas en el nuevo mapa y que cada grupo queda adscrito al menos a una STU.

### *2.3.3 Refinado de las delineaciones*

Durante el proceso cartográfico se debe asegurar la legibilidad de la representación de todos los elementos incluidos en el mapa, que convencionalmente en los mapas vectoriales se evalúa a través de la distribución del tamaño de las delineaciones presentes (Forbes et al., 1987). En este sentido, se ha tomado como referencia de esta distribución el ASA y el MLA de las delineaciones según las consideraciones de *The Cornell University* (Capítulo 1, 2.6), los cuáles guardan un

equilibrio que se refleja en el IMR. Dichas referencias condicionan, a su vez, la escala óptima de publicación del mapa resultante (Rossiter, 2000).

Con el objetivo de acondicionar el mapa vectorizado para una óptima visualización, se llevó a cabo un proceso de refinamiento de las delineaciones que consistió en el filtrado y la disolución de todas aquellas menores al MLA de la escala de publicación fijada y que, por lo tanto, no serían perceptibles. Al ser geometrías derivadas de la matriz regular de los ráster, se consideró como una única delineación aquellas geometrías independientes de un mismo grupo con contacto en su perímetro, esto es, en diagonal, puesto que visual y físicamente responderían a una misma entidad.

El refinamiento supone la adscripción del espacio correspondiente a una delineación colindante siguiendo unos criterios espaciales: aquella de menor superficie, pero superior a la considerada. Así se favorecería la aglomeración de grupos de aislados diferenciados de su entorno, manteniendo en lo posible las distinciones existentes. Redundaría, además, en la disminución de la dispersión respecto al ASA en la distribución de las superficies y, con ello, del IMR. En última instancia, llevaría aparejado una mejora de los atributos para la publicación del mapa.

Las acciones de filtrado, asignación y disolución se aplicaron en este orden de forma recursiva con un incremento del MLA correspondiente a la superficie de una celda de malla en cada ocasión hasta alcanzar la escala de publicación objetivo y reducir el IMR hasta un valor favorable, lo más próximo a 2 (Rossiter, 2000).

#### *2.3.4 Validación*

La fase final contempla la validación de las predicciones del mapa disgregado y refinado y, en última estancia, de la metodología. Esta evaluación se planteó en términos de precisión categórica, es decir, de concordancia entre el valor previsto y el real en el mismo lugar, y se realizó a través de las métricas específicas de categorías OA y Coeficiente de Kappa, en las que se contrastan las SMU y sus categorías frente a las observaciones de suelos.

Estas métricas se calcularon sobre el mapa final, disgregado y refinado; y también sobre el mapa base disgregado, pero no refinado; y el mapa convencional original; de forma que se analiza tanto el proceso completo como las acciones individuales de división y refinado.

Las observaciones de suelos utilizadas en la validación son heredadas del proceso convencional de cartografía, por lo que no se conoce con certeza la precisión con la que fueron capturadas sus

localizaciones. Este potencial error implícito ha sido incorporado en las métricas de precisión categóricas para metodologías de disgregación considerando los vecindarios próximos de las observaciones durante las intersecciones con las SMU. Específicamente, en esta metodología se aplica un vecindario de 3x3 píxeles.

Por otro lado, sólo aquellas intersecciones donde el valor de la observación es igual al de la predicción son tomadas como válidas. Sin embargo, las diferencias entre las clases de suelos en los sistemas de clasificación de suelos no son equivalentes y los errores detectados pueden proceder de observaciones de clases próximas o, por el contrario, muy diferentes. La evaluación del grado de diferencia y si ocurren especialmente en los niveles de más detalle, como subgrupos, familias o litologías, permitiría matizar el error de la clasificación, en línea con un sistema más avanzado como la distancia taxonómica (Capítulo 1, 3.7). Con ese fin, se calculan aquellas intersecciones que hubiesen sido válidas con una única modificación en dichos niveles.

Ambos supuestos se consideraron para apoyar la interpretación del valor estricto de las métricas.

## **3 Resultados y discusión**

### **3.1 STU y nuevas SMU**

El proceso de división de las 46 SMU sin fases del mapa convencional basado en la colección de covariables seleccionadas dio lugar a 207 nuevos grupos o áreas homogéneas con significado edáfico. En este sentido, estas áreas homogéneas guardan semejanza con las regiones delimitadas por las fases dentro de las SMU, pero con un detalle superior (Figura 4-4). Sólo en ciertos casos concretos los grupos respondieron a demarcaciones de cultivos o alteraciones antrópicas, principalmente asociados a las zonas de fondos de valles o a los cauces.

Durante la correlación posterior de esos grupos con las 142 STU de la leyenda del mapa, la mayor parte (63%) se asociaron a una única STU y constituyen potencialmente nuevas SMU monotáxicas. La consolidación final de las SMU, en las que se registraron aquellas combinaciones únicas de STU entre todos los grupos, permitió describir un total de 143 nuevas SMU. De ellas, el 52% son SMU monotáxicas, compuestas por un única STU, y el 36% politáxicas con 2 STU, por ejemplo, una STU principal con una secundaria. El 12% restante son SMU politáxicas con 3 o más componentes. De forma conjunta, las nuevas SMU del mapa disgregado reúnen una media de 1,6 STU en cada una.

En contraste, las 46 SMU del mapa convencional reúnen de media 5,6 STU, donde la mayor parte de ellas aúnan entre 4 y 6 STU (82%) contando con las inclusiones.

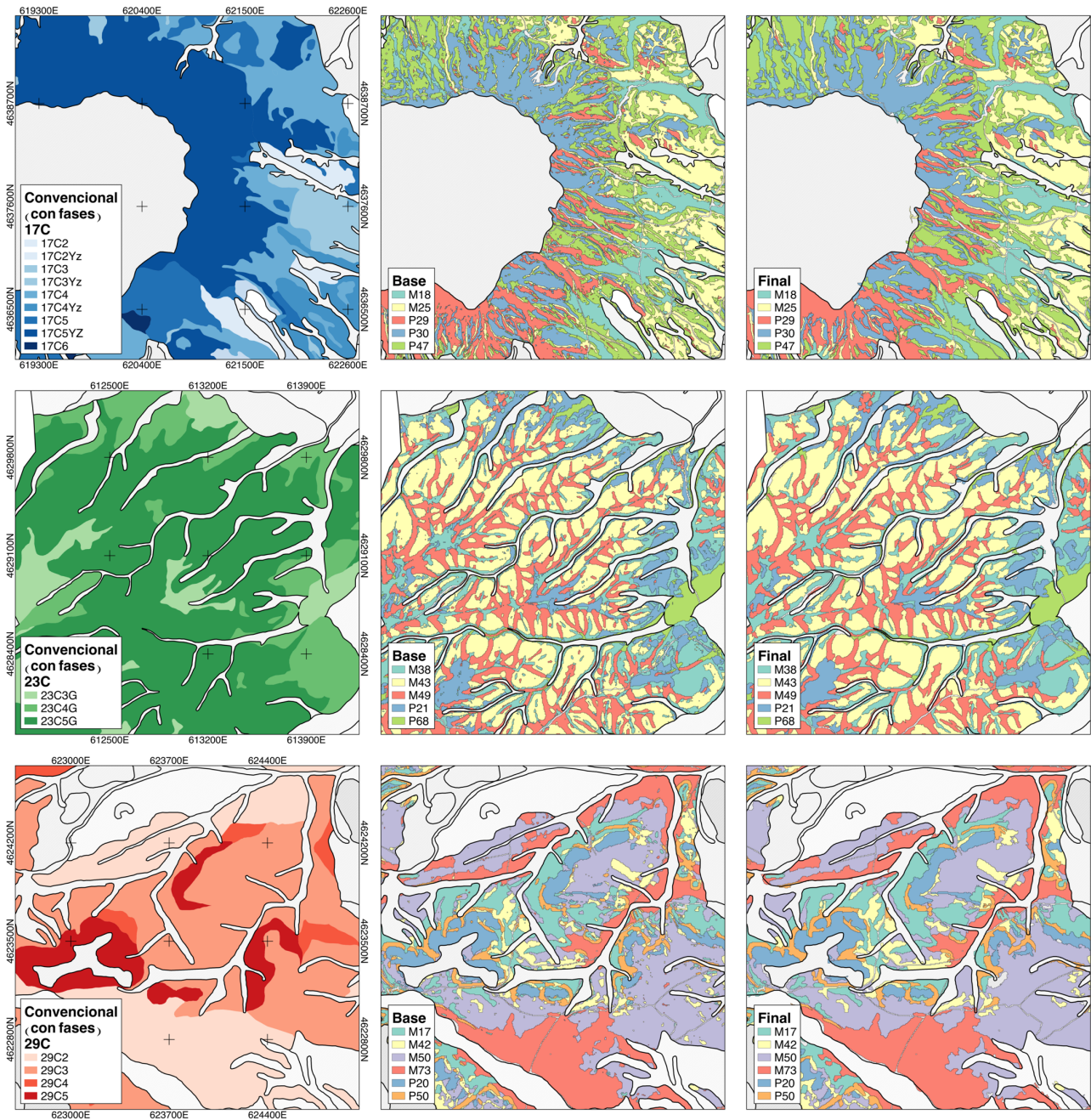


Figura 4-4. Unidades cartográficas de suelos con fases (un color graduado) del mapa de suelos convencional (izquierda); nuevas unidades cartográficas del mapa de suelos disgregado sin refinamiento, base (centro); y disgregado con refinamiento, final (derecha). De arriba a abajo, los mapas corresponden a las regiones de las unidades cartográficas de suelos originales 17C, 23C y 29C.

## 3.2 Delineaciones

El mapa convencional de suelos original con fases está formado por 3.184 delineaciones excluyendo las áreas misceláneas (Tabla 4-3). Tras las etapas de división y correlación, el mapa disgregado base pasó a contar con 288.612 delineaciones que componían las nuevas SMU disgregadas. En el análisis de la distribución de la superficie de estas delineaciones se detectó que estaba sesgada hacia la izquierda, lo que significaba que existía un gran número de delineaciones de pequeño tamaño de entre 25 y 75 m<sup>2</sup>, correspondientes a 1 y 3 píxeles (Figura 4-4). Esta circunstancia ya había sido observada previamente durante la definición de la metodología de división y su implementación sobre una selección de 10 SMU (0). Como resultado se produce un desequilibrio en la ratio entre el MLA y la ASA del nuevo mapa disgregado, que se reflejaba en un IMR muy elevado (Tabla 4-3). Cabe destacar que esta situación también ocurría en el mapa convencional de partida.

Tabla 4-3. Características del detalle cartográfico de los mapas que intervienen en la disgregación.

Mapa	Delineaciones (nº)	MLA (m <sup>2</sup> )	ASA (ha)	IMR	Escala óptima de publicación <sup>1</sup>
<b>Convencional con fases</b>	3.142	2.014	20.39	10.06	1:35.700
<b>Base. Disgregado sin refinamiento</b>	288.612	25	0.21	9.23	1:3.647
<b>Final. Disgregado y refinado</b>	38.445 (+1.223%)	1.825	1.59 (-93%)	2.96 (-71%)	1:9.993

<sup>1</sup>Cornell University (Capítulo 12.6)

ASA: Área media de las delineaciones; IMR: Índice de Máxima Reducción; MLA: Área Mínima Legible.

Con todos estos elementos se vio la necesidad de acometer el refinamiento de las delineaciones para la edición del nuevo mapa bajo unos parámetros de publicación óptimos. La ejecución recursiva del refinamiento partió del mapa disgregado y en cada ocasión se fue aumentando el MLA y disolviendo todas aquellas delineaciones menores que él. De esta forma, se generaron múltiples conjuntos de delineaciones definidos por los niveles de la MLA.

Con el aumento del MLA se consiguió una disminución paulatina del número de delineaciones (Figura 4-5), así como del IMR (Figura 4-6), que fue más pronunciada en los primeros niveles que en los superiores. Al mismo tiempo, la escala aumentó progresivamente. Todos ellos son efectos buscados (Figura 4-4) y que se encuentran dentro de los objetivos marcados, a saber, alcanzar la escala del entorno de 1:10.000 y un IMR próximo a 2.



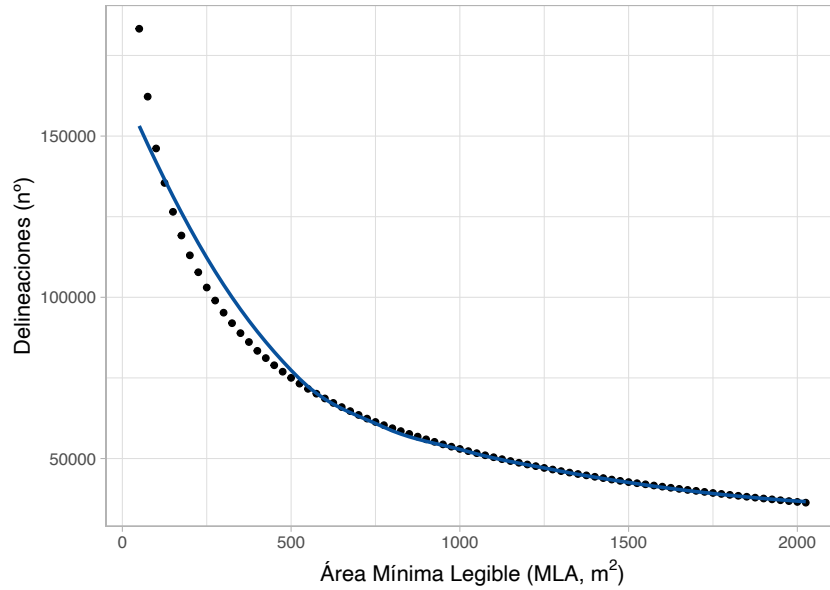


Figura 4-5. Gráfico de dispersión entre el número de delineaciones frente al Área Mínima Legible (área mínima de delineación en el mapa) que se aproxima a un límite asintótico cuando se alcanzan los valores de superficies mayores.

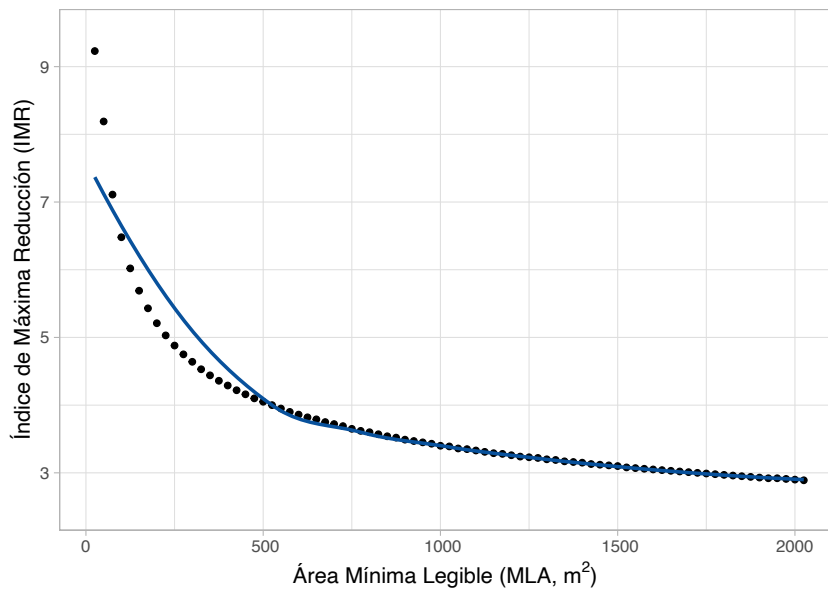


Figura 4-6. Gráfico de dispersión entre el Índice Máximo de Reducción frente al Área Mínima Legible (área mínima de delineación en el mapa) que se aproxima a un límite asintótico para los niveles más grandes de superficies.

Atendiendo a las consideraciones establecidas, el nivel entre todos los conjuntos refinados cuya escala asociada fue la más próxima a la escala objetivo fue aquel con un MLA de 1.825 m². Por su parte, su IMR correspondiente fue de 2,96, uno de los más bajos en la curva entre IMR y MLA, que presentaba un límite casi asintótico al aumentar el MLA (Figura 4-6). Por ese motivo, se determinó este MLA como el nivel óptimo de refinamiento y con él se obtuvo el mapa final disgregado y refinado (Figura 4-7). En última instancia, el ASA original de 20 ha se había visto reducido hasta

las 1,6 ha en el mapa disgregado y refinado (Tabla 4-3). En base a estos parámetros cabe destacar que la escala óptima de publicación del mapa convencional es 1:35.700, algo más pequeña que la de referencia (1:25.000).

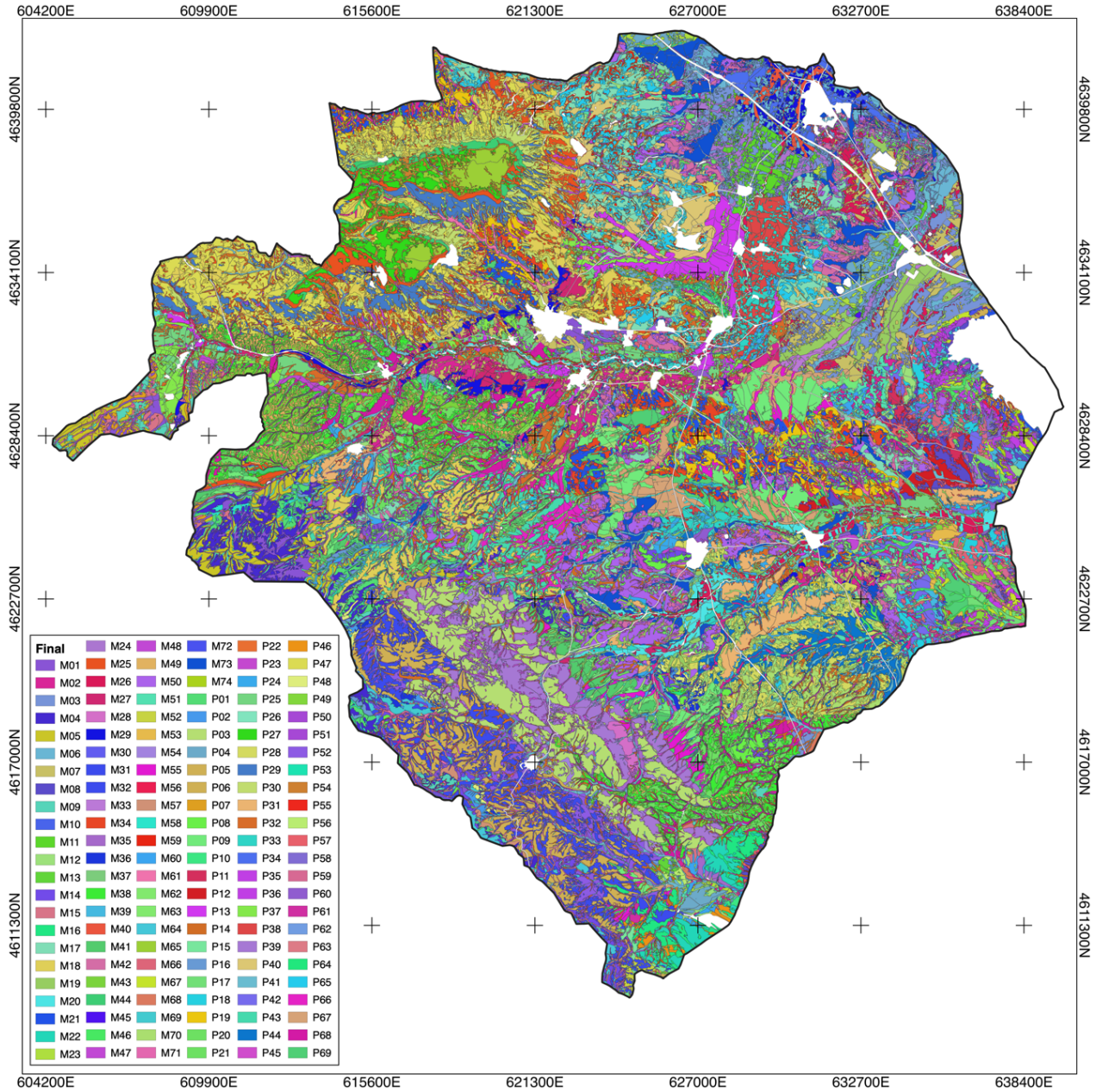


Figura 4-7. Mapa de las 143 unidades cartográficas de suelos disgregadas y refinadas, 74 monotáxicas (el identificador comienza por M) y 69 politáxicas (por P).

### 3.3 Precisión

Las métricas de precisión categórica para la validación fueron calculadas para todos los supuestos, a saber, i) el mapa final disgregado y refinado, la versión base disgregada sin refinamiento y el mapa convencional; ii) la intersección con la observación puntual o con un vecindario próximo

alrededor de esta; y iii) la coincidencia estricta en los niveles de mayor detalle de la STU o permitiendo una única variación. En total, se generaron 12 combinaciones diferentes de supuestos y para todas ellas se calcularon las dos métricas, OA y el Coeficiente de Kappa.

Los valores para el supuesto principal, esto es, estricto y para la localización puntual de las observaciones respecto al mapa disgregado y refinado, señalan un resultado favorable con limitaciones (Tabla 4-4). El Coeficiente de Kappa alcanza un valor de 0,22, que se encuentra en el rango superior a 0 e indicaría una concordancia moderada. Por su parte, la OA es del 24% y queda dentro del rango de los valores obtenidos en metodologías comparables en términos de condiciones de partida, escala y detalle categórico de una clasificación taxonómica (series de suelos), aunque es sensiblemente inferior a los resultados medios obtenidos (Capítulo 23.4.2).

Tabla 4-4. Valores de los índices de precisión categórica Precisión Global y Coeficiente de Kappa bajo todos los supuestos y para los tres mapas que intervienen en la disgregación: Final: mapa disgregado y refinado; Base: mapa disgregado sin refinamiento; y CSM: mapa de suelos convencional.

Supuestos	Precisión Global (%)			Coeficiente de Kappa		
	Final	Base	CSM	Final	Base	CSM
<b>Estricta + Punto</b>	24	26	0.87	0.23	0.24	0.87
<b>Estricta + Vecindario</b>	32	36	0.88	0.31	0.35	0.88
<b>Variación + Punto</b>	43	43	0.92	0.42	0.42	0.92
<b>Variación + Vecindario</b>	46	50	0.92	0.45	0.49	0.92

A diferencia de la metodología propuesta, estas se basaron en un sistema de clasificación supervisados, en su mayoría DT y RF, que dan lugar modelos suelo ~ paisaje complejos con ratios de validación positivos, pero que exigen esfuerzos computacionales intensos. Por el contrario, los modelos no supervisados son más sencillos, pero están sujetos a la correlación entre grupos no definidos y categorías.

Cuando se toma en consideración el vecindario próximo de las observaciones en un marco equivalente a 3x3 píxeles (7,5 metros de radio), la OA aumenta hasta el 32%. Esto supone una mejora notable que cabría asociarla en parte a los errores de precisión en la localización de las observaciones y al muestreo dirigido. Por otro lado, cuando se toman como válidas aquellas intersecciones con una única variación en los subgrupos, familias o litologías, la OA aumenta significativamente hasta el 43%. Ambos supuestos con condiciones algo menos restrictivas que

dan lugar a aumentos de la OA muestran un potencial de mejora de los valores de validación de la metodología. El efecto del primero podría corregirse usando nuevos conjuntos de observaciones aleatorias. Mientras el del segundo sería orientativo de carencias en las covariables para representar el paisaje, explícitamente de las relacionadas con la litología (factor P de SCORPAN) y con el suelo (factor S de SCORPAN). También de la incertidumbre incorporada en el proceso de correlación por conocimiento experto.

Los mismos parámetros respecto al mapa disgregado sin refinamiento muestran un ligero aumento (Tabla 4-4), posiblemente relacionado con la integración de áreas de pequeño tamaño, pero que contaban con sentido edafológico. Esto indicaría que el proceso de refinamiento tiene un coste pequeño en términos de precisión, que sería compensado por el beneficio en la visualización de la cartografía. Sin embargo, esta diferencia no aparece entre los dos mapas en el supuesto de variación única de la STU, lo que apoyaría que esas áreas con sentido edafológico disueltas son efectivamente variaciones en los niveles de más detalle de la STU. En este sentido, cabría hacer su delimitación de forma más consistente cubriendo las carencias en covariables litológicas y relacionadas con los suelos.

En comparación, el mapa convencional alcanza una OA del 86% en el supuesto principal y resulta muy superior a la obtenida en el mismo supuesto en mapa disgregado y refinado. Sin embargo, el número medio de STU por SMU es de 5,6 frente a 1,6 respectivamente, lo que facilita que las intersecciones con las observaciones sean correctas en el mapa original y con ello aumente su índice, esto es, cualquiera de las STU registradas en la SMU sería válida en cualquier punto interior. En términos de precisión unitaria por STU, no existiría diferencia entre ambos mapas.

## **4 Conclusiones**

La nueva metodología de disgregación de unidades cartográficas politáxicas en mapas convencionales de suelos es alternativa a la mayoritaria, que se basan en técnicas de clasificación supervisadas, y simplifica el proceso y las técnicas en uso.

Su implementación sobre el mapa de suelos convencional a escala 1:25.000 de la DO Campo de Borja ha dado lugar un nuevo mapa de suelos disgregado a escala aproximada a 1:10.000 con resultados favorables. En términos globales, se logra:

- i) El incremento sustancial en el número de delineaciones en el nuevo mapa (+ 1.223%), que ven reducido su tamaño medio notablemente (-93%).
- ii) El aumento de las SMU identificadas, que se triplican en número.
- iii) La reducción del número medio de STU por SMU, que pasan de las 5,6 en el mapa convencional a las 1,6 en el mapa disgregado. Especialmente, en el 52 % de los casos la valoración de las nuevas SMU es con un único término (monotáxicas).

Estos efectos son expresión de la capacidad del método para detectar áreas homogéneas con sentido edáfico. Por sus características, se considera que su aplicación puede orientarse hacia regiones donde no se conozca con detalle la situación potencial de las STU dentro de las SMU y que, además, tiene potencial para dar soporte a nuevas estrategias de dos pasos.

La estrategia de refinamiento del mapa una vez vectorizado para adaptarlo a criterios de publicación óptimos se considera beneficiosa en la medida de que tiene una escasa influencia en los valores de OA y, por el contrario, consigue un equilibrio en la textura del mapa, reduciendo el IMR un 71% y situándolo en 2.96.

No obstante, los valores de los índices de OA (24%) y Coeficiente de Kappa (0.23) obtenidos no son especialmente altos. El análisis de estos índices sobre los diferentes mapas según varios supuestos con menores restricciones, vecindario próximo y una aproximación a la distancia taxonómica, señala posibles vías de mejora como la incorporación de covariables litológicas o relacionadas con el factor S (suelos) de SCORPAN o la validación a través de nuevos conjuntos de observaciones aleatorios. Además, la metodología propuesta debería probarse en nuevas regiones con características ambientales diferentes.

## 5 Bibliografía

- Charrad, M., Ghazzali, N., Boiteau, V., Niknafs, A., 2014. NbClust: An R package for determining the relevant number of clusters in a data set. *J. Stat. Softw.* 61, 1–36. <https://doi.org/10.18637/jss.v061.i06>
- Costa, J.J.F., Giasson, E., Silva, E.B. da, Campos, A.R., Machado, I.R., Bonfatti, B.R., Bacic, I.L.Z., 2019. Individualization of soil classes by disaggregation of physiographic map polygons. *Pesqui. Agropecuária Bras.* 54, e00290. <https://doi.org/10.1590/s1678-3921.pab2019.v54.00290>
- Florinsky, I.V., 2017. An illustrated introduction to general geomorphometry. *Prog. Phys. Geogr. Earth Environ.* 41, 723–752. <https://doi.org/10.1177/0309133317733667>

- Flynn, T., van Zijl, G., van Tol, J., Botha, C., Rozanov, A., Warr, B., Clarke, C., 2019. Comparing algorithms to disaggregate complex soil polygons in contrasting environments. *Geoderma* 352, 171–180. <https://doi.org/10.1016/j.geoderma.2019.06.013>
- Forbes, T., Rossiter, D.G., Van Wambeke, A., 1987. Guidelines for evaluating the adequacy of Soil Resource Inventories, SMSS Technical Monograph. Soil Conservation Service. US Department of Agriculture (USDA), Washington D.C.
- Gómez-Miguel, V.D., Sotés, V., 2015. Zonificación del Terroir: Estudio de Suelos y Ordenación del Cultivo de la Vid en la DO Campo de Borja (Zaragoza). Universidad Politécnica de Madrid (UPM), Madrid.
- Häring, T., Dietz, E., Osenstetter, S., Koschitzki, T., Schröder, B., 2012. Spatial disaggregation of complex soil map units: A decision-tree based approach in Bavarian forest soils. *Geoderma* 185–186, 37–47. <https://doi.org/10.1016/j.geoderma.2012.04.001>
- Hengl, T., 2006. Finding the right pixel size. *Comput. Geosci.* 32, 1283–1298. <https://doi.org/10.1016/j.cageo.2005.11.008>
- Hjerdt, K.N., McDonnell, J.J., Seibert, J., Rodhe, A., 2004. A new topographic index to quantify downslope controls on local drainage. *Water Resour. Res.* 40. <https://doi.org/10.1029/2004WR003130>
- Irvin, B.J., Ventura, S.J., Slater, B.K., 1997. Fuzzy and isodata classification of landform elements from digital terrain data in Pleasant Valley, Wisconsin. *Geoderma* 77, 137–154. [https://doi.org/10.1016/S0016-7061\(97\)00019-0](https://doi.org/10.1016/S0016-7061(97)00019-0)
- Iwahashi, J., Pike, R.J., 2007. Automated classifications of topography from DEMs by an unsupervised nested-means algorithm and a three-part geometric signature. *Geomorphology* 86, 409–440. <https://doi.org/10.1016/j.geomorph.2006.09.012>
- Lázaro-López, A., González-SanJosé, M.L., Gómez-Miguel, V., 2021. Disaggregation of conventional soil maps: a review. *Soil Res.* <https://doi.org/10.1071/SR20288>
- Machado, I.R., Giasson, E., Campos, A.R., Costa, J.J.F., Silva, E.B. da, Bonfatti, B.R., 2018. Spatial Disaggregation of Multi-Component Soil Map Units Using Legacy Data and a Tree-Based Algorithm in Southern Brazil. *Rev. Bras. Ciênc. Solo* 42. <https://doi.org/10.1590/18069657rbc20170193>
- McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. *Geoderma* 117, 3–52. [https://doi.org/10.1016/S0016-7061\(03\)00223-4](https://doi.org/10.1016/S0016-7061(03)00223-4)
- Nauman, T.W., Thompson, J.A., 2014. Semi-automated disaggregation of conventional soil maps using knowledge driven data mining and classification trees. *Geoderma* 213, 385–399. <https://doi.org/10.1016/j.geoderma.2013.08.024>
- Odgers, N.P., Sun, W., McBratney, A.B., Minasny, B., Clifford, D., 2014. Disaggregating and harmonising soil map units through resampled classification trees. *Geoderma* 214–215, 91–100. <https://doi.org/10.1016/j.geoderma.2013.09.024>
- Roecker, S.M., Howell, D.W., Haydu-Houdeshell, C.A., Blinn, C., 2010. Conventional Soil Survey and Digital Soil Mapping Approaches, in: Boettinger, J.L. (Ed.), *Digital Soil Mapping: Bridging Research, Environmental Application, and Operation*, Progress in Soil Science. Springer, Dordrecht [Netherlands]; London, pp. 369–384.
- Rossiter, D.G., 2000. *Methodology for Soil Resource Inventories*, 2nd Revised Version. ed, Lecture Notes & Reference. International Institute for Aerospace Survey & Earth Sciences (ITC).

- Sarmiento, E.C., Giasson, E., Weber, E.J., Flores, C.A., Hasenack, H., 2017. Disaggregating conventional soil maps with limited descriptive data: A knowledge-based approach in Serra Gaúcha, Brazil. *Geoderma Reg.* 8, 12–23. <https://doi.org/10.1016/j.geodrs.2016.12.004>
- Schmidt, J., Hewitt, A., 2004. Fuzzy land element classification from DTMs based on geometry and terrain position. *Geoderma* 121, 243–256. <https://doi.org/10.1016/j.geoderma.2003.10.008>
- Soil Survey Division Staff, 1993. *Soil Survey Manual, USDA Handbook*. U.S. Government Printing Office, Washington, D.C.
- Soil Survey Division Staff, 2014. *Keys to Soil Taxonomy, 12th Edition*. ed. USDA-Natural Resources Conservation Service, Washington, DC.
- Subburayalu, S.K., Slater, B.K., 2013. Soil Series Mapping By Knowledge Discovery from an Ohio County Soil Map. *Soil Sci. Soc. Am. J.* 77, 1254. <https://doi.org/10.2136/sssaj2012.0321>
- Vincent, S., Lemercier, B., Berthier, L., Walter, C., 2018. Spatial disaggregation of complex Soil Map Units at the regional scale based on soil-landscape relationships. *Geoderma* 311, 130–142. <https://doi.org/10.1016/j.geoderma.2016.06.006>
- Wood, J., 1996. *The Geomorphological Characterisation of Digital Elevation Models (PhD thesis)*. University of Leicester.





# **CAPÍTULO 5 APLICACIÓN DEL MAPA DE SUELOS DISGREGADO EN LA ZONIFICACIÓN INTEGRADA DEL TERROIR VITÍCOLA DE LA DO CAMPO DE BORJA**

## **1 Introducción**

El «terroir vitícola» se refiere a la noción de «área geográfica delimitada» relacionada con su idoneidad para la producción de vinos de calidad, y cuenta con una tradición ancestral (Gómez-Miguel, 2011) que tiene su expresión en el marco legal actual en figuras como las DO (Vaudour et al., 2015).

A partir del concepto de terroir, y debido a la carencia de datos básicos enológicos y culturales georreferenciados en las distintas unidades del medio, se ha definido el concepto restrictivo de «terroir ambiental» (Gómez-Miguel, 2011). Se trata de un aspecto que incluye únicamente a los elementos del medio natural y del agroecosistema vitícola, esto es, las características específicas del suelo, la geología, el clima, la topografía, y las relativas a la planta y la viticultura. En adelante, las menciones genéricas al terroir referirán al concepto restrictivo de terroir ambiental.

Dentro del terroir, el suelo desempeña una figura central como factor intermediario del resto de factores ambientales y por su influencia sobre la vid y sus productos, especialmente a través de sus propiedades. Además, es el que presenta el mayor nivel de variabilidad espacial de los factores ambientales (Gómez-Miguel, 2011).

Para la caracterización y delimitación de los terroir ha sido recomendada la técnica de la zonificación (OIV, 2012). Esta consiste en la investigación del territorio con el fin de repartirlo en zonas relativamente homogéneas: con similares potencialidades y limitaciones para el desarrollo de la vid en base a los elementos del medio, atendiendo a sus propiedades y respecto a su distribución geográfica (Fregoni et al., 2003).

Los métodos para elaborar estudios de zonificación han estado ligados al desarrollo de técnicas para la caracterización del medio. Se iniciaron con análisis descriptivos y justificaciones históricas, y tradicionalmente han tomado la forma de mapas temáticos de aptitudes y usos, o de modelos del paisaje (Gómez-Miguel, 2011). Los avances en dichas técnicas han llevado a que hoy en día la

«Organización Internacional de la Viña y el Vino» (OIV) recomiende nuevos modelos complejos, fruto de la combinación de los anteriores (OIV, 2012). Entre estos se sitúa la metodología de Zonificación Integrada del Terroir (ZIT, en inglés *Integrated Terroir Zoning* [ITZ]) (Gómez-Miguel y Sotés, 2002, 1992, 2015).

La ZIT combina tres zonificaciones parciales de factores ambientales: suelo, clima y medio (geomorfología y litología), cuya integración da lugar a la delimitación y caracterización de «Unidades Homogéneas del Medio» (UHM, en inglés *Homogeneous Terroir Units* [HTU]). En ellas la vocación vitícola puede ser expresada de forma cuantitativa a través del Índice de Calidad (IC), calculado para cada una de las UHM (Gómez-Miguel y Sotés, 2015).

Cada una de las zonificaciones parciales atienden a parámetros específicos y a metodologías propias de las disciplinas en las que se encuadran. En el caso específico de la zonificación del suelo, responde al mapa de un IRS y tradicionalmente se ha generado por métodos convencionales. La adopción de metodologías de DSM para el IRS, que superan las convencionales, y el aprovechamiento de la cartografía generada a través de ellas, con capacidad de aumentar el detalle y que presentan SMU mayoritariamente monotáxicas, tendrían el potencial de mejorar la zonificación parcial del suelo y, con ello, el resultado de la zonificación del terroir.

En aquellas regiones que cuenten con una zonificación del terroir basada en un IRS por métodos convencionales, una vía de mejora aprovechando el marco de DSM es la línea de acción de la disgregación de suelos. De esta manera, es posible exprimir la información ya generada de suelos junto a nuevas fuentes de datos para aumentar el detalle cartográfico del IRS con una inversión reducida de recursos materiales. La DO Campo de Borja es una región que reúne estas condiciones, después del esfuerzo inicial de un estudio completo de zonificación de los terroir vitícolas (Gómez-Miguel y Sotés, 2015), y cuenta con un potencial importante de mejora.

El objetivo del trabajo que se presenta en este capítulo es generar una nueva zonificación del terroir siguiendo la metodología de la ZIT para la DO Campo de Borja que mejore la zonificación original disponible basándose en la mejora de su IRS como zonificación parcial del suelo. Adicionalmente, se evalúa dicha acción. La mejora del IRS se consigue con la disgregación de su mapa de suelos convencional con SMU politáxicas en un nuevo mapa de SMU mayoritariamente monotáxicas. El resultado último es un mapa con nuevas UHM de mayor detalle que el original y mejor valoradas cuantitativamente. Además, se consigue el objetivo de caracterizar potencialmente aquellas UHM de mayor ocupación y aquellas de mayor calidad dentro de la DO

aprovechando la definición aportada por el nuevo mapa de zonificación. En última instancia, este conjunto de acciones permite fijar un procedimiento para cubrir parte de la demanda de cartografía detallada en el marco que establece la Viticultura de Precisión.

## **2 Materiales y métodos**

### **2.1 Área de trabajo**

El área de trabajo es la DO Campo de Borja (Capítulo 3 y Capítulo 4).

### **2.2 Fuentes de datos**

#### *2.2.1 Zonificación del suelo original: Mapa de suelos convencional con fases*

El mapa de suelos convencional con fases del IRS a escala 1:25.000 que fue usado originalmente como zonificación del suelo procede del propio proyecto de Zonificación Integrada del Terroir de la DO (Gómez-Miguel y Sotés, 2015) (Capítulo 4).

Las SMU (politáxicas) de este mapa son fases (343) de asociaciones y complejos (46) y la mayoría de sus STU (142) han sido clasificadas a nivel de serie según ST (Soil Survey Division Staff, 2014). El contenido de la mayor parte de las SMU está formado por 4 y 6 términos (media 5,6) en los que también figuran las inclusiones (Figura 4-1).

#### *2.2.2 Zonificación del suelo actualizada: Mapa de suelos disgregado*

El mapa de suelos disgregado a partir del mapa de suelos convencional se utiliza para la nueva zonificación del suelo de mayor detalle. Fue generado con la metodología de división por la técnica CLARA y la correlación con clases de suelo, más un refinamiento posterior a una escala aproximada 1:10.000 (Capítulo 4).

Lo forman 143 SMU con un contenido global en el conjunto de sus STU idéntico al del mapa de suelos convencional según ST (Figura 4-7). La mayor parte son SMU monotáxicas (52%) y el resto politáxicas, que reúnen de media 1,6 STU incluyendo inclusiones.

### 2.2.3 Observaciones de suelos heredadas

Se utiliza el conjunto de 519 observaciones puntuales de suelos definido durante la desagregación del mapa convencional (Capítulo 42.2.2), que se formó a partir del total observaciones del proceso de cartografía del IRS y que fueron fijadas mediante un muestreo dirigido (Figura 4-2). Están clasificadas de acuerdo con ST y asociadas a alguna de las STU recogidas en la leyenda del mapa.

### 2.2.4 Mapa de zonificación del clima

El mapa de zonificación del clima procede del proyecto original del zonificación integrada del terroir de la DO en estudio (Gómez-Miguel y Sotés, 2015). En él se divide la región del proyecto en 9 unidades (Figura 5-1) con unas características climáticas propias definidas a partir de parámetros climáticos e índices bioclimáticos (Tabla 5-1).

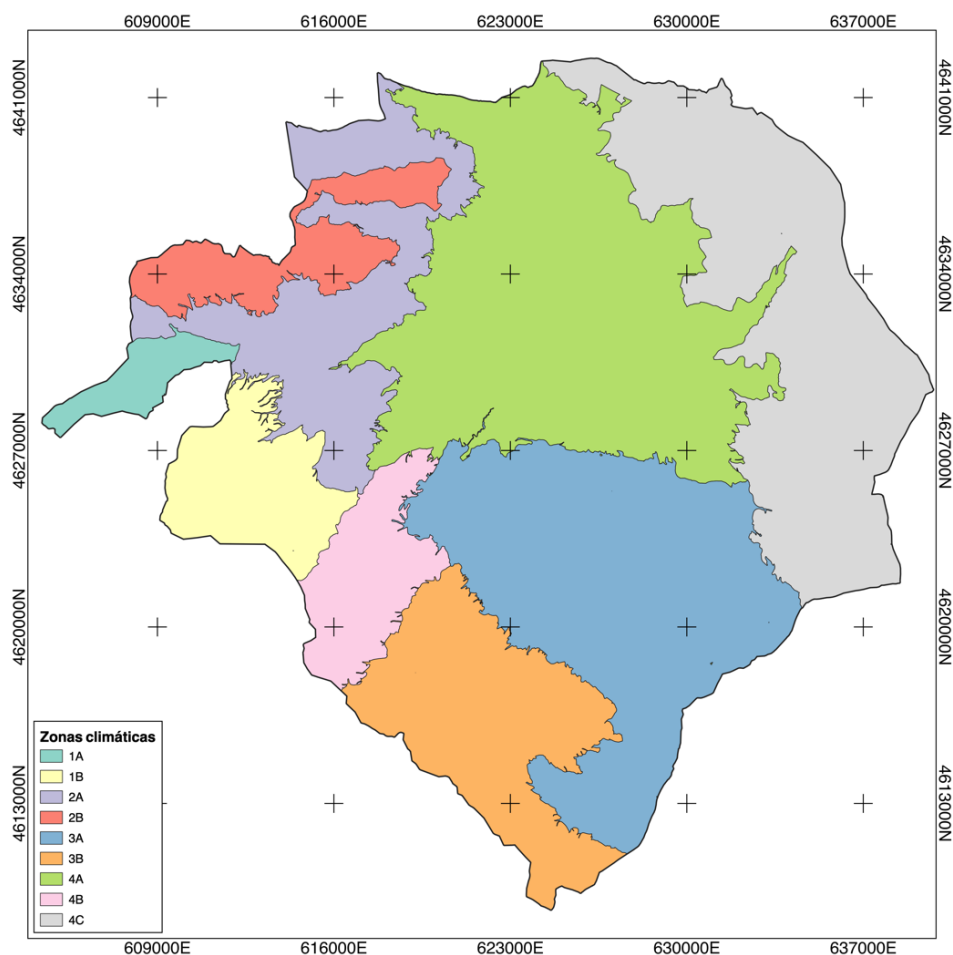


Figura 5-1. Mapa de zonificación climática de la Denominación de Origen Campo de Borja (ver características de las zonas en la tabla Tabla 5-1). (Fuente de la información: Gómez-Miguel y Sotés, 2015)

Tabla 5-1. Parámetros climáticos e índice bioclimáticos característicos de las zonas climáticas de la zonificación del clima.

Parámetros climáticos	Zonas Climáticas						
	1A	1B; 3B; 4B	2A	2B	3A	4A	4C
TMaxA	39,0	38,0	40,0	39,3	41,0	41,0	43,0
TMaxm	24,7	24,8	26,4	26,7	27,5	27,4	26,8
TMax	17,4	17,6	19,6	19,4	20,8	20,3	20,1
Tm	12,1	12,2	13,8	14,1	14,9	13,5	13,2
Tmin	6,7	6,8	8,0	8,6	8,9	9,5	9,8
Tmimm	1,0	1,3	2,2	3,4	3,6	4,3	4,7
TminA	-15,0	-12,0	-10,0	-5,7	-10,0	-10,0	-10,0
PRE	429,1	515,7	362,2	396,8	376,3	398,0	324,2
ETPTho	691,0	690,1	748,7	750,4	794,8	798,6	729,0
ETPHar	1036,0	1044,4	1129,6	1098,2	1187,9	1138,4	1115,9
ITE	1295,8	1368,7	1775,2	1787,3	2047,1	2075,6	1663,8
IH	1826,5	1908,7	2302,4	2262,6	2573,4	2549,4	2347,4
PB	2,38	3,8	5,6	5,8	6,8	6,9	5,0
IC	10,1	5,8	8,8	8,0	9,7	9,4	9,8
IB	14,9	7,4	15,6	14,6	18,2	17,4	15,5
PA	233	204	236	241	247	247	224
PL	202	169	193	214	203	217	220
SAL	24	27	32	21	33	25	6
RES	76	93	59	83	75	86	44
DAC	371,5	183,0	386,5	353,6	418,5	400,6	404,8

(Fuente: Gómez-Miguel y Sotés, 2015)

DAC: déficit hídrico acumulado hasta septiembre; TMaxA: temperatura máxima absoluta; TMaxm: media de las temperaturas máximas absolutas diarias; TMax: media de las temperaturas máximas medias diarias; Tm: media de las temperaturas medias diarias; Tmin: media de las temperaturas mínimas medias diarias; Tmimm: media de las temperaturas mínimas absolutas diarias; TminA: temperatura mínima absoluta; PRE: precipitación media anual; ETPTho: evapotranspiración de Thornthwaite; ETPHar: evapotranspiración de Hargreaves; ITE: Integral Térmica de Winkler o Integral Térmica Eficaz; IH: Índice de Huglin; PB: Producto Heliotérmico de Branjas; IC: Índice de Constantinescu; IB: Índice de Hidalgo; PA: periodo activo medio de la vid; PL: periodo libre de heladas de Emberger; SAL: diferencia entre el inicio del PA y el PL; RES: diferencia entre el inicio del PA y la fecha en la que la reserva de agua en el suelo se agota.

### 2.2.5 Mapa de zonificación del medio

La zonificación del medio original del proyecto de zonificación se compuso a partir de la FIA y el mapa geológico elaborado para el IRS de la DO (Gómez-Miguel y Sotés, 2015). Respecto al mapa geológico (Figura 5-2), cuenta con 50 unidades litoestratigráficas (Tabla 5-2). Mientras, a través de la FIA se categorizaron las áreas respecto a la pendiente; la altitud; la presencia de hidromorfía, de yesos y sales solubles, de afloramientos rocosos, o de terrazas y/o bancales.

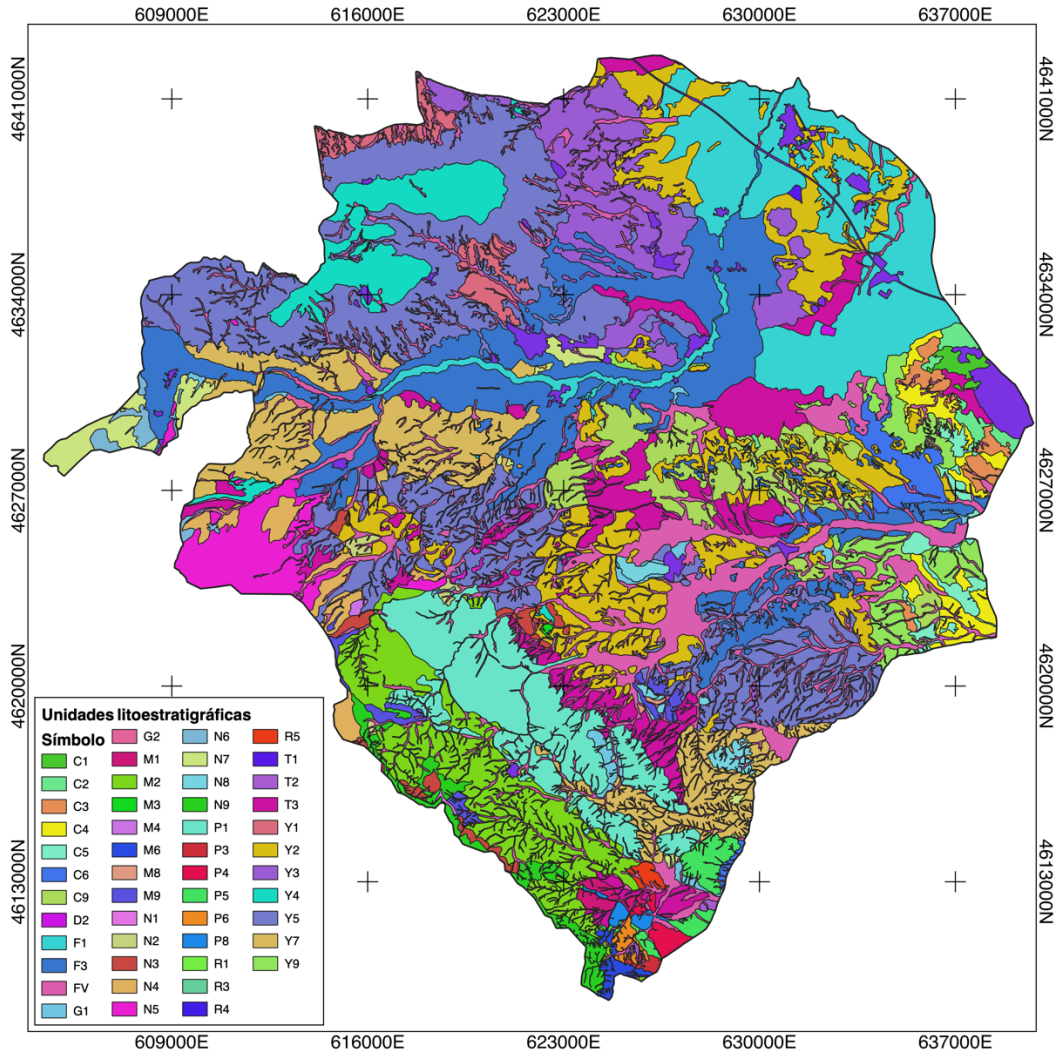


Figura 5-2. Mapa de unidades litoestratigráficas para el inventario del recurso suelo convencional de la Denominación de Origen Campo de Borja. (ver descripción de las unidades litoestratigráficas en la tabla Tabla 5-2Tabla 5-1).  
(Fuente de la información: Gómez-Miguel y Sotés, 2015)

Tabla 5-2. Unidades litoestratigráficas del inventario del recurso suelo convencional de la Denominación de Origen Campo de Borja.

Símbolo	Descripción	Unidad cronoeestratigráfica
C1	Margas y dolomicritas con intercalaciones de yesos y yesos	Aquitaniense-Chatteniense
C2	Lutitas rojas con intercalaciones de yesos tabulares y nodulares	Aquitaniense (Ageniense)
C3	Yesos tabulares y nodulares	Aragoniense inferior
C4	Lutitas rojas y lutitas carbonatadas grises y láminas de areniscas grises y micáceas	Aragoniense inferior
C5	Yesos tabulares con láminas de dolomicritas	Aragoniense inferior
C6	Conglomerados, areniscas y lutitas. Conglomerados heterolíticos	Aragoniense
C9	Yesos con silex, arcillas y limos yesíferos rojos (Fm. Monteagudo)	Pontiense/Vindoboniense
D2	Depósitos lagunares, áreas endorreicas y navas	Holoceno
F	Aluviales de los afluentes principales	Pleistoceno-Holoceno
F1	Modelado del río Ebro	Pleistoceno-Holoceno
F2	Modelado de los afluentes principales (3er orden)	Pleistoceno-Holoceno

Símbolo	Descripción	Unidad cronoestratigráfica
F3	Modelado de los afluentes principales (Huecha y Huechaseca)	Pleistoceno-Holoceno
FV	Aluviales de fondos de valle menores	Holoceno
G1	Coluviones	Pleistoceno-Holoceno
G2	Coluviones	Pleistoceno-Holoceno
M1	Conglomerados cuarcíticos	Bundsandstein
M2	Conglomerados y areniscas (paleocanales)	Bundsandstein
M3	Alternancia de areniscas y lutitas rojas	Bundsandstein
M4	Lutitas rojas con intercalaciones de areniscas	Bundsandstein
M6	Dolomías, calizas dolomíticas tableadas y margas (Ud. Carbonatada superior)	Muschelkalk
M8	Ofitas	Keuper
M9	Arcillas varioladas y yesos con intercalaciones de margas arenosas	Keuper
N1	Dolomías grises bien estratificadas (Fm. Imón)	Rethiense
N2	Brechas sedimentarias y de colapso, dolomías y carniolas (Fm. Cortes de Tajuña)	Hettangiense
N3	Calizas microcristalinas localmente oolíticas, bien estratificadas	Sinemuriense
N4	Margas con intercalaciones de calizas arcillosas	Toarciense/Pliensbachiense
N5	Calizas y margas	Calloviense/Bajociense
N6	Margas arenosas y calizas con esponjas en la base	Oxfordiense
N7	Areniscas microconglomerados y arcillas	Porlandiense/Kimmeridgiense
N8	Dolomicritas y micritas negras con laminaciones	Sinemuriense inferior y medio
N9	Brechas calcico dolomíticas y dolomías vacuolares y carniolas	Hettangiense/Rethiense
P1	Cuarcitas, areniscas y pizarras	Tremadociense
P3	Areniscas, lutitas (Fm. Bolloncillos)	Fameniense
P4	Areniscas gruesas, medios y microconglomerados (Miembro Filluelo)	Fameniense
P5	Areniscas medias, gruesas, limolitas gris-azuladas (Miembro Valdeinglés)	Fameniense
P6	Limolitas gris-azuladas y areniscas finas (Miembro Coscojar)	Fameniense
P8	Brechas y conglomerados muy cementados, areniscas gruesas, lutitas y areniscas	Bundsandstein superior
R1	Arcillas varioladas yesos y calizas arcillosas	Keuper
R3	Margas grises con laminaciones de areniscas	Muschelkalk
R4	Lutitas, areniscas, limolitas y yesos	Keuper
R5	Lutitas y areniscas de tonos rojos-vinosos	Bundsandstein
T1	Conos de deyección	Plioceno-Pleistoceno
T2	Abanicos Aluviales	Plioceno-Pleistoceno
T3	Glacis	Plioceno-Pleistoceno
WM	Áreas Misceláneas	SD
Y1	Yesos con sílex, arcillas y limos yesíferos rojos (Fm. Monteagudo)	Pontiense/Vindoboniense
Y2	Arcillas calcáreas y limos de tonos rojizos y grises con frecuentes intercalaciones de caliza	Vindoboniense (Burdigaliense-Aquitaniense)
Y3	Arcillas calcáreas y limos de tonalidades pardo-rojizas y grises (Fn. Tudela)	Pontiense/Vindoboniense
Y4	Calizas compactas, a veces arcillosas y silíceas (C. de la Muela)	Pontiense
Y5	Materiales detríticos con niveles carbonatados	Pontiense
Y7	Conglomerados, areniscas y limolitas rojizas, intercalaciones de calizas	Pontiense-Burdigaliense
Y9	Yesos tabulares con capas de micrita, dolomicritas con niveles de yesos y lutitas rojas	Aragoniense inferior (Burdigaliense)

(Fuente de la información: Gómez-Miguel y Sotés, 2015)

## *2.2.6 Registro Vitícola y Sistema de Información Geográfica de Parcelas Agrícolas*

El «Registro Vitícola» (REVI) es un inventario de datos de las explotaciones por parcela que reúne aspectos agronómicos, propiedades vitícolas, y los identificadores únicos de cada parcela entre otros.

Por su parte, el «Sistema de Información Geográfica de Parcelas Agrícolas» (SIGPAC) es otro registro en el que se delimita e identifica de forma unívoca los cultivos declarados por agricultores y ganaderos en cualquier régimen de ayudas públicas que esté relacionado con la superficie cultivada o aprovechada por el ganado.

La unión de ambas fuentes permite construir un conjunto de datos donde las características y propiedades del viñedo estén asociados a una componente geométrica. A partir de aquí es posible, por ejemplo, conocer la distribución de las parcelas de viñedos o específicamente aquellas con una variedad de vid concreta. En el presente estudio se obtiene este conjunto reducido a los viñedos de la DO Campo de Borja. En total, agrupa 10.046 geometrías de recintos con una superficie de 6.252,2 ha, que en su mayoría están dedicadas al cultivo de las variedades Garnacha y Tempranillo (Figura 5-3).

## **2.3 Procedimiento**

La metodología de ZIT aborda la delimitación y caracterización de los terroir a través de la integración de zonificaciones parciales asociadas a factores del medio que influyen en el desarrollo de la vid, específicamente las tratadas en los apartados anteriores: clima, del medio (litología y geomorfología) y del suelo. Esto supone la intersección de los mapas categóricos correspondientes a cada una de ellas para dar lugar a unidades cartográficas con combinaciones únicas de todos los atributos, las UHM (Figura 5-4). En el presente estudio se implementa esta metodología haciendo uso de las zonificaciones parciales de clima y medio originales y del mapa de suelos disgregado para generar la nueva zonificación y sus UHM.



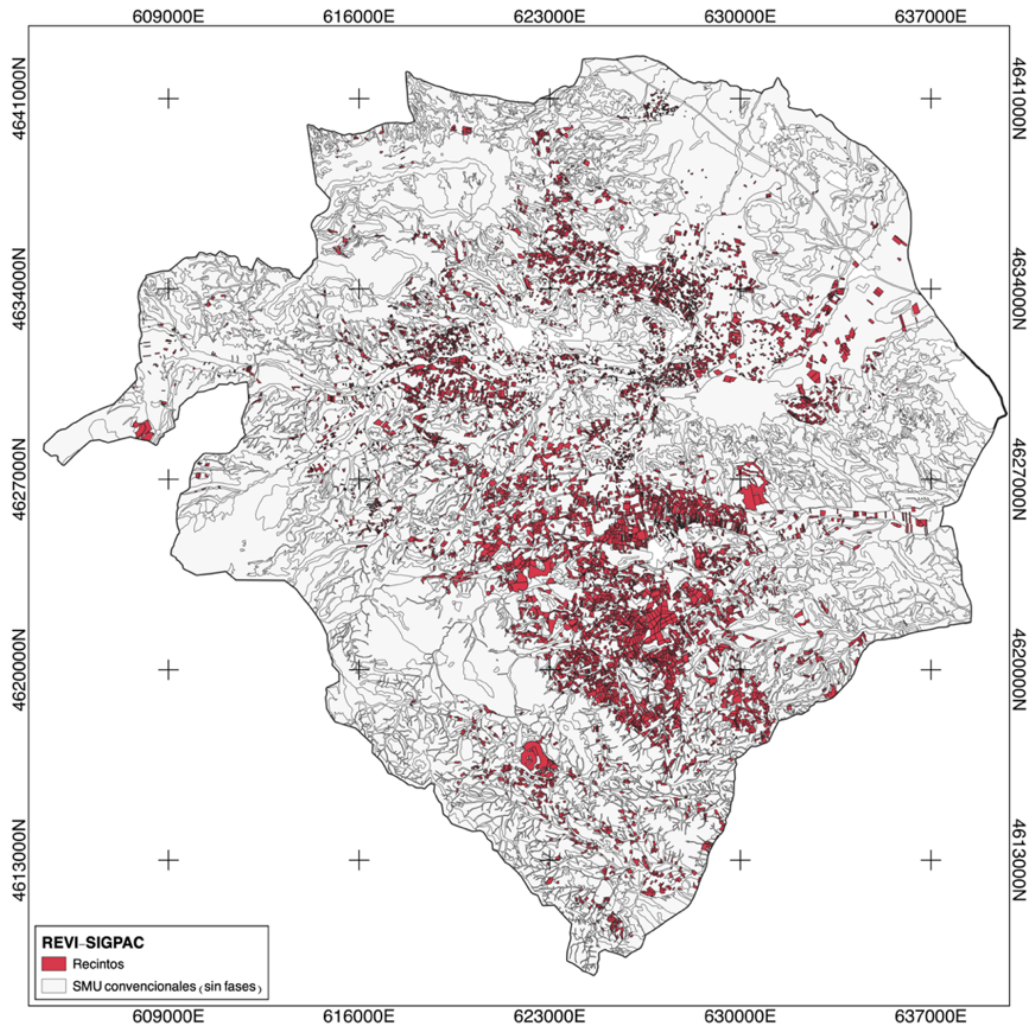


Figura 5-3. Distribución de los recintos de viñedos con datos tomados del Registro Vitícola y del Sistema de Información Geográfica de Parcelas Agrícolas de la DO Campo de Borja. REVI: Registro Vitícola; SIGPAC: Sistema de Información Geográfica de Parcelas Agrícolas.

Posteriormente, se valora cuantitativamente la calidad vitícola de las UHM definidas según el IC, cuyo resultado es un valor numérico que se proyecta sobre un rango de 0 a 100. De forma paralela, se evalúa la distribución de parcelas de viñedos por UHM en lo que se denomina IO, en el que se estima el porcentaje de ocupación para cada UHM.

Para cada índice y a partir de su distribución en la región del proyecto se establece una categorización no lineal en 5 clases, que tiene como objetivo facilitar el manejo y su análisis. Este sistema de evaluación se aplica individualmente tanto a la zonificación de los terroir original como a la nueva zonificación generada desde el mapa de suelos disgregado. Además, se genera un modelo del IO a partir del IC que busca entender la distribución de los viñedos a partir de la cuantificación de los parámetros ambientales facilitada por los IC parciales.

En base a estos resultados, se contrastan por un lado las adscripciones de las clases de IC para ambos mapas desde una matriz de confusión a partir de su intersección espacial. El análisis de los flujos de remodelación en la zonificación, esto es, de los cambios en las valoraciones de calidad vitícola según el IC en las mismas áreas entre los mapas, busca estimar el efecto provocado por la mejora de la zonificación del suelo cuando se mantienen constantes las otras dos zonificaciones complementarias. Por otro, se analizan las clases del IO en la nueva zonificación atendiendo al modelo elaborado con el fin de estudiar las características potenciales de aquellas UHM de mayor uso.

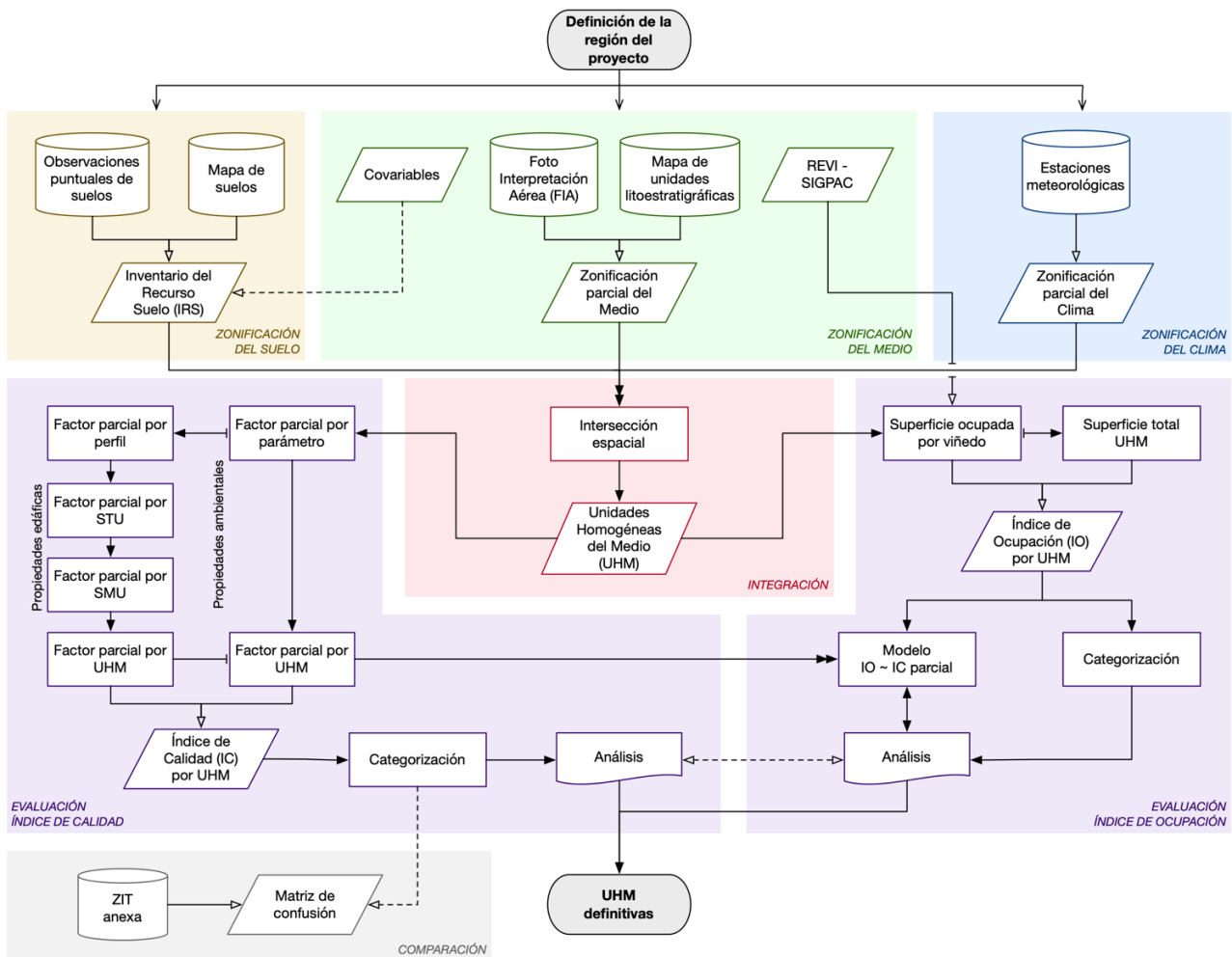


Figura 5-4. Diagrama de flujo de la Zonificación Integrada del Terroir vitícola y su evaluación en línea continua. Vías suplementarias de información y de comparación entre zonificaciones en línea discontinua.

### 2.3.1 Nuevas Unidades Homogéneas del Medio

Los terroir ambientales o UHM expresan, como se ha dicho, aquellas zonas homogéneas para los factores ambientales con influencia sobre el desarrollo del viñedo y se obtienen mediante la

intersección de los mapas categóricos correspondientes a las zonificaciones de cada uno de los factores considerados, específicamente, del clima, del medio y del suelo.

La caracterización de estas áreas responde así a parámetros asociados a dichos factores y que son considerados durante las zonificaciones parciales. De esta manera, dentro de una UHM se esperan valores constantes de los parámetros como, por ejemplo, el periodo activo medio de la vida para el clima, la pendiente en el medio, o las STU descritas para el suelo.

De los distintos factores del terroir ambiental, el suelo ocupa una figura central debido a la interacción que mantiene con el resto. Además, como ha sido mencionado, presenta el mayor nivel de variabilidad espacial de todos ellos, que se refleja en la mayor granularidad del mapa de suelos. En este sentido, la delimitación de las UHM suelen responder a los límites demarcados por delineaciones de las SMU, que ya toman como referencia las estratificaciones de la litología, geomorfología e, indirectamente a través de ellos, del clima. El resultado obtenido es un mapa de UHM.

### *2.3.2 Índice de Calidad*

El IC es un índice paramétrico para la evaluación de las UHM que fue desarrollado inicialmente desde la zonificación vitícola de la DO Ribera del Duero (Gómez-Miguel y Sotés, 1992) y que ha sido puesto a punto con las implementaciones de la metodología de referencia en sucesivos proyectos de zonificación (Gómez-Miguel y Sotés, 2002, 2011, 2010, 2006, 2004) hasta su adaptación actual a la DO Campo de Borja.

En él se considera un grupo de propiedades edáficas y otro grupo de parámetros ambientales que contribuyen de forma individual con un factor al valor del IC de la UHM (Tabla 5-3). Cada propiedad o parámetro está categorizada, bien en rangos cuando son variables continuas o por las propias clases en las categóricas, y a cada una de éstas se les ha asignado un grado sobre el valor máximo del factor en el supuesto más favorable. De esta manera, el valor de una propiedad está recogido en una categoría con un valor de factor asociado y que es, en último término, el factor de contribución de dicha propiedad al IC.

Las propiedades edáficas se evalúan a nivel del perfil según los datos de campo y laboratorios disponibles de los horizontes, y mediante un sistema paramétrico multiplicativo se calcula después el IC del perfil. A continuación, es posible estimar un IC parcial de cada STU mediante la agregación de los IC por perfil en base a la STU a la que están adscritos. En aquellas UHM definidas

sobre una SMU monotáxica, el IC parcial corresponde al de la STU que engloban. En el caso de las politáxicas, se pondera el IC parcial de cada STU por su porcentaje de participación en la SMU. Con todo, se obtiene un IC parcial de propiedades edáficas por UHM.

De forma complementaria, se evalúan los parámetros ambientales a nivel de UHM y se calcula otro IC parcial ambiental. La agregación de ambos índices parciales resulta en el IC definitivo por UHM. Por último, se considera la distribución del IC en la región del estudio y se proyecta sobre un rango de 0 a 100, correspondientes a los valores mínimos y máximos, para facilitar su contraste.

Tabla 5-3. Propiedades y parámetros considerados en el Índice de Calidad, la zonificación parcial origen de los datos y su nivel de evaluación.

Zonificación parcial	Parámetro	Evaluación
<b>Suelo</b>	Clases texturales en el perfil	<b>Propiedades edáficas (Perfil &gt; STU &gt; SMU &gt; UHM)</b>
	Caliza activa en el perfil	
	Materia orgánica del epipedon	
	Profundidad del solum	
	Capacidad de intercambio catiónico del endopedon	
	Acidez de cambio en aluminio en el epipedon	
	Porcentaje de aluminio intercambiable en el epipedon	
	Concentración y porcentaje de potasio intercambiable en el epipedon	
	Concentración y porcentaje de magnesio intercambiable en el epipedon	
	Relación potasio / magnesio en el epipedon	
	Relación calcio / magnesio en el epipedon	
	Conductividad eléctrica medida en el extracto de la pasta de saturación más alta en el perfil	
	Porcentaje de sodio intercambiable más alto en el perfil	
	Balace hídrico del perfil	
<b>Medio</b>	Litología	<b>Parámetros ambientales (UHM)</b>
	Pendiente	
	Yesos y sales más solubles	
	Afloramientos rocosos, gravas y pedregosidad	
	Hidromorfía	
	Altitud	
	Abancalamiento	
<b>Clima</b>	Zona climática	

STU: Unidad Taxonómica de Suelo; SMU: Unidad Cartográfica de Suelos; UHM: Unidad Homogénea del Medio.

Las reglas paramétricas y acciones descritas para el cálculo del IC se han implementado a través de un algoritmo de reglas de decisión en SQL como una extensión de la base de datos. A través de

este índice cuantitativo y objetivo de la zonificación se busca entender la evolución de zonificación original a microzonificación.

### *2.3.3 Índice de Ocupación*

El IO relaciona la superficie ocupada por el viñedo para un área determinada respecto a su superficie total. Las referencias de los viñedos son tomadas desde el conjunto de datos construidos por el REVI, con los datos agronómicos y las propiedades del viñedo, y de su geometría asociada desde el SIGPAC. El cálculo de la superficie ocupada se obtiene por la intersección de esas geometrías de las parcelas con el mapa de UHM de referencia y como resultado de ello un índice cuantitativo.

La distribución del viñedo, especialmente en el momento actual de reducción de superficie de cultivo en las regiones con tradición vitícola, responde a una selección progresiva basada en la experiencia de los viticultores de aquellas parcelas más favorables para producir uvas de más calidad (Gómez Sánchez, 1995; Gómez-Miguel, 2011; Vaudour et al., 2015). Por ello, el análisis de aquellas UHM donde se concentran los viñedos permite conocer las características potenciales que determinan la producción actual.

### *2.3.4 Categorización: clases de UHM por índices*

Ambos índices calculados son categorizados por separado en 5 clases mediante una técnica de agrupación no supervisada jerárquica según la distancia euclídea y por el método de centroides considerando sus distribuciones en la región del proyecto. Esta acción permite crear agrupaciones consistentes en distribuciones no lineales, esto es, que no responden a intervalos o tamaños regulares, y busca facilitar el contraste de los resultados.

Por un lado, es posible intersecar los mapas de zonificación con sus UHM categorizadas según su potencialidad señalado por el IC y establecer una comparación a través de una matriz de confusión. En ella las columnas y filas recogen las clases en el mapa original frente a la nueva zonificación respectivamente y las celdas contienen la superficie total bajo la condición de cruzamiento. De esta manera, una columna representa para las áreas adscritas a cierta clase del mapa de zonificación original cómo han sido clasificadas en las clases de la nueva zonificación. Inversamente, una fila representa la distribución de las áreas para una clase del mapa de la nueva zonificación en las clases del mapa original. El análisis de ambas perspectivas manifiesta el sentido de los cambios introducidos por la nueva zonificación a raíz del uso del mapa disgregado de suelos.

De forma complementaria, el estudio de aquellas UHM con mayor concentración de viñedos atendiendo al IO permite conocer las características mayoritarias que potencialmente determinan su valoración actual, en el sentido de producción y calidad.

### 2.3.5 Modelización: IO vs IC

En la modelización del IO se toman los factores de IC parciales correspondientes a los parámetros del terroir evaluados, de forma que se facilite la consideración de todos ellos con independencia de que en origen sean variables categóricas o continuas. A partir de aquí se construye un modelo de regresión lineal optimizado por selección de variables en etapas por adición según el Criterio de Información de Akaike (AIC) (Kuhn y Johnson, 2013) y con validación cruzada del 75% con 50 fases.

Esta modelización tiene como objeto revelar aquellos factores de mayor influencia en la distribución de los viñedos para la región del proyecto en un momento dado.

## 3 Resultados y discusión

### 3.1 UHM con el mapa de suelos disgregado

El proceso de integración de las tres zonificaciones parciales por intersección da lugar a las unidades cartográficas de la nueva zonificación que cuentan con mayor detalle y granularidad y son de menor tamaño (Tabla 5-4). En total, se identifican 4.460 combinaciones únicas de propiedades y parámetros de suelos, clima y medio que resultan en las nuevas UHM frente a las 957 del mapa de zonificación original. Por su parte, el número de delineaciones pertenecientes a estas unidades en el nuevo mapa aumentan desde 3.142 hasta 38.445.

Tabla 5-4. Características cartográficas de la zonificación de referencia y la nueva zonificación.

Mapa	UHM (nº)	Delineaciones (nº)	ASA (ha)	Escala óptima de publicación <sup>1</sup>
<b>Zonificación original (con mapa de suelos convencional)</b>	957	3.124	20.39	1:35.700
<b>Nueva zonificación (con mapa de suelos disgregado)</b>	4.460 (+466%)	38.445 (+1.207%)	1.59 (-93%)	1:9.993

<sup>1</sup>Cornell University (Capítulo 12.6)

ASA: Área media de las delineaciones; UHM: Unidad Homogénea del Medio.

### 3.2 IC

Para cada una de las nuevas UHM se obtiene un valor de IC. La distribución de este parámetro en la región del proyecto resulta sesgada hacia valores muy bajos y donde unas pocas UHM concentran los valores más altos (Figura 5-5).

La categorización de dicha distribución en 5 clases mediante la clasificación jerárquica da lugar a agrupaciones con separaciones muy marcadas, desde la clase 1 con la media más alta que disminuye notablemente en la clase 2 y en menor medida entre las clases sucesivas 3, 4 y 5 (Figura 5-5).

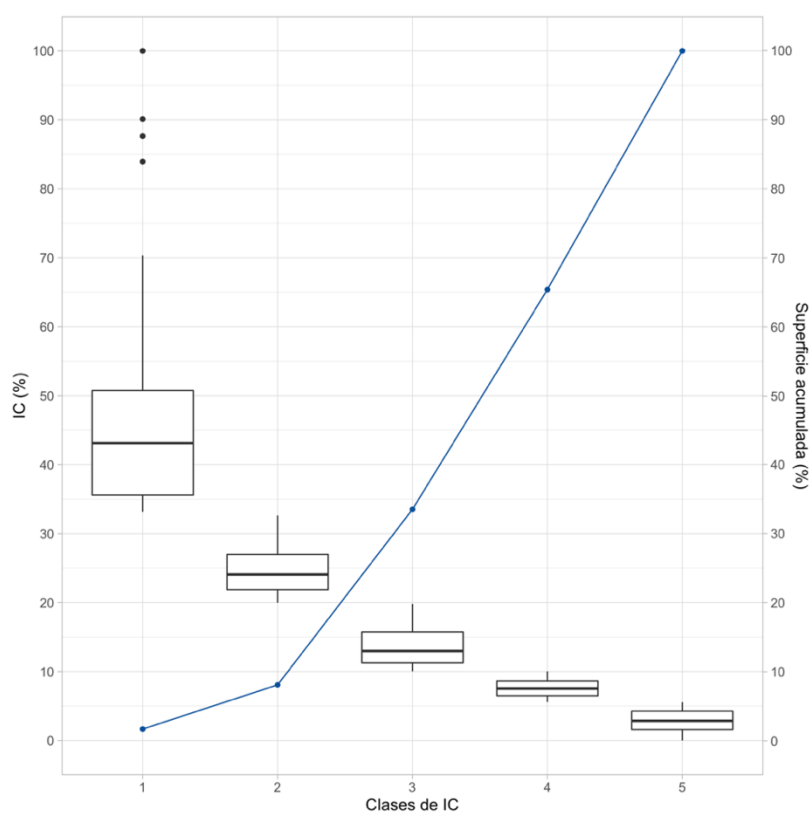


Figura 5-5. Distribución del Índice de Calidad en la región del proyecto según las 5 clases definidas (eje principal) y la superficie abarcada por cada clase (eje secundario). IC: Índice de Calidad.

Específicamente, la clase 1 reúne 59 nuevas UHM con los mayores valores de IC. Esas UHM están definidas sobre 14 SMU diferentes del mapa disgregado y las cuales contienen 17 STU distintas que se corresponden en su mayoría con subcategorías de los subgrupos Haploxeralf cálcico y típico, Calcixerepts típico y Palexeralf cálcico. Específicamente, la mayoría de las UHM se extienden en dos unidades monotáxicas, M50 y M01, que recogen Haploxeralf cálcicos sobre diferentes litologías.

Este conjunto de UHM se sitúa sobre 15 tipos de litoestratigrafías diferentes, si bien la mayoría se concentra sobre arcillas calcáreas y limos de tonos rojizos y grises con frecuentes intercalaciones de caliza (Y2), areniscas y microconglomerados (N7). En términos climáticos, predominan las zonas 3A y 4A, que se caracterizan por las temperaturas y los índices bioclimáticos más elevados unidos a un periodo libre de heladas de mayor duración, y a la vez por mayores restricciones hídricas ligadas a una menor precipitación y a un tiempo reducido para el agotamiento de la reserva de agua en el suelo durante el periodo vegetativo.

La superficie correspondiente a cada una de las 5 clases es desigual. Del total de 61.423 ha pertenecientes a la DO (se excluyen las áreas misceláneas), aproximadamente un 2% está recogida por la clase 1 y un 6% por la clase 2, mientras el 92% restante se divide de forma similar entre las otras clases (Figura 5-5). En contraste con el mapa de zonificación original, supone un aumento substancial de la clase 1 (600% aproximadamente) así como una disminución del 30% de la clase 2 y variaciones ligeras del resto de clases (Tabla 5-5).

Tabla 5-5. Matriz de confusión de la intersección entre las Unidades Homogéneas del Medio categorizadas según el Índice de Calidad del mapa de zonificación original (columnas) y nueva zonificación (filas), y sus parámetros globales.

		Superficie de las clases del IC en la zonificación original (ha)					Total (ha)	Precisión del usuario (%)
		1	2	3	4	5		
Superficie de las clases del IC en la nueva zonificación (ha)	1	75,49	879,14	57,74	5,00	2,60	<b>1.019,97</b>	0,074
	2	60,78	2.341,36	1.351,62	160,84	21,33	<b>3.935,93</b>	0,595
	3	2,27	1.700,35	8.462,34	5.052,71	416,49	<b>15.634,16</b>	0,541
	4	0,42	183,43	3.781,98	11.895,04	3.711,68	<b>19.572,55</b>	0,608
	5	20,52	28,69	1.181,90	4.141,52	15.877,08	<b>21.249,71</b>	0,747
Total (ha)		<b>159,48</b>	<b>5.132,97</b>	<b>14.835,58</b>	<b>21.255,11</b>	<b>20.029,18</b>	<b>61.412,32</b>	
Precisión del productor (%)		0,473	0,456	0,570	0,560	0,793	<b>Precisión Global (%)</b>	<b>0,629</b>

La matriz de confusión construida a partir de la intersección entre las UHM categorizadas de ambos mapas de zonificación muestra un grado de coincidencia alto, ya sea atendiendo al 63% de OA o al valor 0,48 del Coeficiente de Kappa (Tabla 5-5).



Desde la perspectiva parcial de cómo resulta la nueva zonificación respecto al punto de partida, existe un muy alto nivel de concordancia entre las clases 5, aquellas con los valores de IC más bajos. Este hecho unido a una variación mínima de la superficie total de categoría es un reflejo de la estabilidad de la delimitación y cuantificación de los terroir potencialmente menos favorables. La concordancia también es elevada, aunque es algo menor, en las clases 2, 3 y 4.

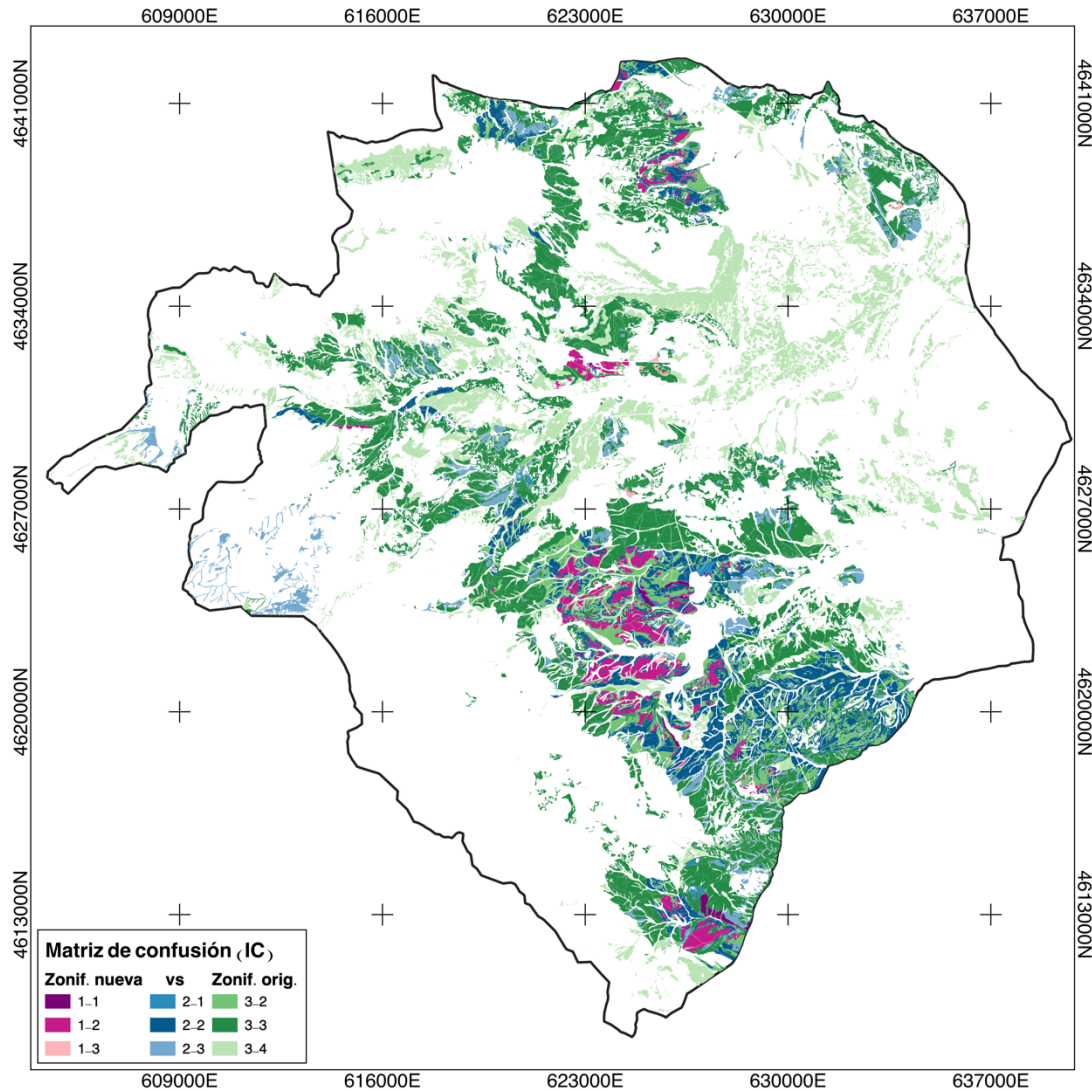


Figura 5-6. Representación cartográfica de los flujos desde la matriz de confusión entre las Unidades Homogéneas del Medio categorizadas según el Índice de Calidad. Por cada categoría de la nueva zonificación (tonalidades), se simboliza la categoría origen en el mapa original (intensidades). Se representa en intensidades más fuertes aquellas áreas que mantienen la categoría y en intensidades más ligeras las que han variado.

IC: Índice de Calidad

Por el contrario, tal concordancia resulta singularmente baja en la clase 1. Aquí casi la mitad de la delimitación de dicha categoría en el mapa de zonificación original se mantiene en la nueva, pero la superficie total aumenta muy notablemente. Específicamente, la matriz de confusión muestra un trasvase desde categorías inferiores, muy notablemente la 2. A efectos de la zonificación puede

ser expresión de un proceso por el que se perfilan las áreas e identifican zonas de mayor IC dentro de ellas. Un flujo semejante por el que una cuantía relativamente importante de superficie de una categoría inferior aumenta de valoración se repite en las clases 2 y 3 (Figura 5-6). En este sentido, es un resultado favorable de la nueva zonificación. Entre las clases 4 y 5 existe un intercambio con un valor neto bajo y que podría denotar del mismo modo la reestructuración de las delineaciones.

### 3.3 IO

De igual forma que con el IC, se obtiene un valor de IO para cada una de las nuevas UHM, a partir de los datos de las parcelas de viñedos dentro de la región del proyecto registrados en el REVI y el SIGPAC. Se observa que aproximadamente un 43% de las 4.460 nuevas UHM están ocupadas en algún grado por viñedos. Sin embargo, el reparto no es homogéneo y la distribución del IO se encuentra sesgada hacia valores bajos con unas pocas UHM con niveles altos de ocupación (Figura 5-7).

Una distribución con estas características refleja cómo la localización del cultivo no se ha producido de forma aleatoria, sino condicionada bajo ciertos criterios de los viticultores. Parte de esos criterios estarían relacionados con aquellos factores ambientales con influencia en el desarrollo de la vid y que son integrados en el concepto de terroir. En este sentido, no existe correlación directa y alta entre el IO y el IC (0,16). Sin embargo, la modelización del IO a partir de los factores de IC parciales de los parámetros resulta significativa (p valor:  $< 2.2e^{-16}$ ) con una explicación de la varianza leve ( $R^2: 0.142$ ) (Tabla 5-6). Esto es, los parámetros ambientales considerados en el IC dan una explicación limitada pero relevante de cómo se caracteriza la distribución del viñedo. Específicamente, se identifican como factores significativos (p valor  $< 0.05$ ) en positivo: la pendiente, el clima, la SMU, la caliza activa en el perfil y la unidad litoestratigráfica; y en negativo la conductividad eléctrica medida en el extracto de la pasta de saturación más alta en el perfil, la relación entre concentración y porcentaje de potasio intercambiable en el perfil y las clases texturales en el epipedon (Tabla 5-6).

La categorización de la distribución en 5 clases mediante la clasificación jerárquica da lugar a agrupaciones con rangos muy diferenciados, siendo la clase 1 la que reúne los valores más elevados y la clase 5 los más bajos. La superficie correspondiente a cada una de ellas tampoco son homogéneas, si bien aproximadamente el 50% del viñedo se concentra entre las clases 1 y mayoritariamente la 2 (Figura 5-7).

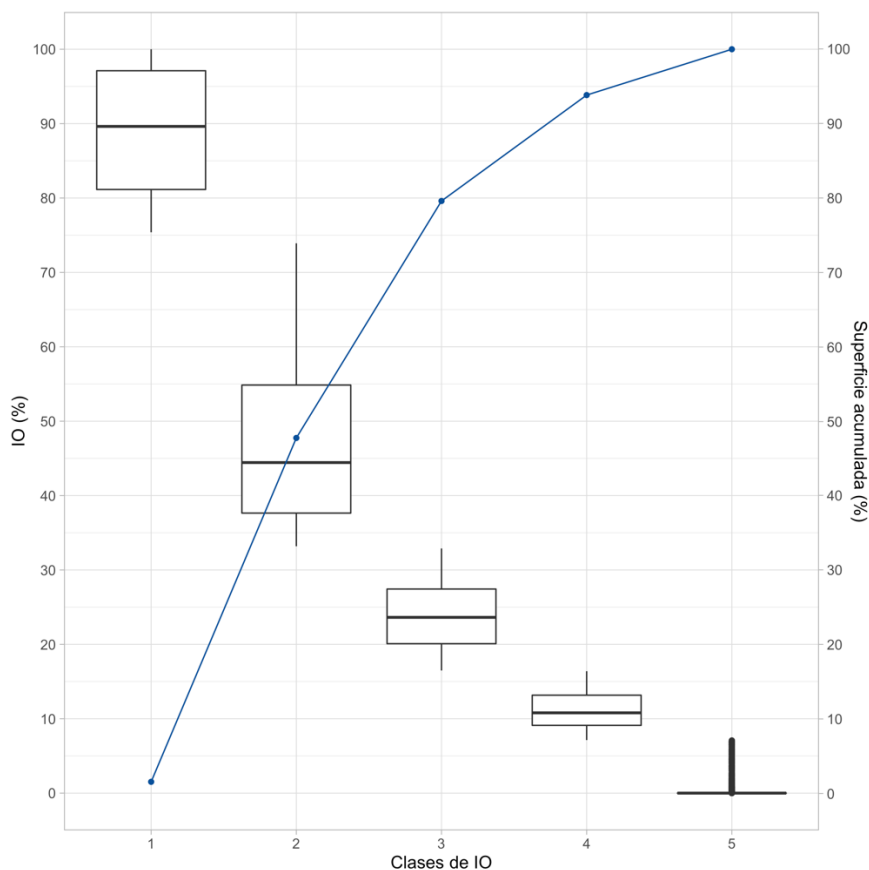


Figura 5-7. Distribución del Índice de Ocupación en la región del proyecto según las 5 clases definidas (eje principal) y la superficie abarcada por cada clase (eje secundario). IO: Índice de Ocupación.

Entre ambas clases reúnen 332 UHM que se caracterizan por distribuirse sobre 81 SMU, entre las que es posible destacar M02, M12, M17 y M73, que son monotáficas. Estas a su vez contienen 102 STU diferentes, donde predominan las series (J03, J04, J07 y J08) y especialmente J03 que también aparece recogida en la clase superior del IC. La mayoría se clasifican como Calcixerept típico, Haploxeralf cálcico, Xerorthent típico y Xerofluvent y que coinciden en gran medida con las clases de suelo identificadas como más favorables. Además, se identifican los Xerofluvent relacionados con zonas de fondos de valle. Respecto a las propiedades de suelos señaladas de mayor relevancia, en estas series predominan valores de caliza activa en el perfil medios y altos, entre 6 y 15%; y de concentración y porcentaje de potasio intercambiable en el epipedon medios también, entre 0,15 y 0,6 cmolc/kg y entre 2 y 12% respectivamente. La conductividad eléctrica medida en el extracto de la pasta de saturación más alta en el perfil alcanza valores medios y se sitúa entre 3 y 5 dS/m. Por último, presentan principalmente clases texturas de USDA franco arcillosas sobre franca, franca sobre franca limosas y arcillo limosas sobre arcillosas.

Tabla 5-6. Modelo de Índice de Ocupación a partir de Índices de Calidad parciales de los parámetros considerados en la zonificación del terroir por regresión lineal con selección de variables en etapas por adición según el Criterio de Información de Akaike y con validación cruzada.

Coefficientes	Estimación	Error estándar	t valor	Pr(> t )	
(Intercepto)	-47.832	11.172	-4.282	1.95E-05	***
Pendiente	40.647	5.627	7.224	7.30E-13	***
Zona climática	37.132	2.922	12.709	< 2e-16	***
SMU	20.309	8.253	2.461	0.01395	*
Caliza activa del perfil	12.85	5.499	2.337	0.01956	*
Unidad Litoestratigráfica	9.608	2.587	3.714	0.00021	***
CEe más alta	-7.396	3.865	-1.913	0.05584	.
Relación K / PKI	-18.498	7.476	-2.474	0.01343	*
Clases texturales del perfil	-20.35	6.392	-3.184	0.00148	**
<i>R<sup>2</sup>: 0.1421; F: 39.21 sobre 8 y 1894 grados de libertad; p-valor: &lt; 2.2e-16</i>					

CEe: conductividad eléctrica medida en el extracto de la pasta de saturación más alta en el perfil; K: concentración de Potasio intercambiable; PKI: porcentaje de Potasio intercambiable.

Estas UHM se sitúan sobre 16 tipos de litoestratigrafías diferentes, si bien la mayoría se concentra sobre aluviales de fondos de valle menores y modelados de los afluentes, que se extienden por el 22% de la región del proyecto. Este hecho guarda sentido con algunas de las clases de suelos de mayor presencia en la región del proyecto mencionadas. Por último, las UHM se localizan principalmente en las zonas climáticas 3A y 4A, como ocurría en la clase 1 según el IC.

## 4 Conclusiones

El empleo del mapa de suelos disgregado como zonificación de suelo en la nueva zonificación de la DO Campo de Borja mediante la metodología de la ZIT frente a la zonificación original basada en el mapa de suelos convencional ha dado lugar a las siguientes conclusiones:

- i) Hay una mayor identificación de UHM en la región del proyecto, que se quintuplican en número
- ii) Existe una mayor diferenciación, que se refleja en la cuantificación a través del IC con un rango más extenso.
- iii) El aumento en la detección de superficie de zonas con mayor potencialidad según al IC dentro de áreas uniformes con menor valoración originalmente es apreciable, aun

existiendo un alto grado de consistencia entre ambos mapas. Este hecho se aprecia en los flujos existentes entre clases en la matriz de confusión de la intersección entre las UHM categorizadas de los dos mapas. El análisis se sustenta en el IC, que cuenta con parámetros identificados como significativos para entender la distribución de viñedo en la región de trabajo y el grado de ocupación de las UHM.

Estos elementos formaban parte de los objetivos marcados en la nueva zonificación y son consecuencia directa de las mejoras del mapa del suelo disgregado en términos de detalle cartográfico y categórico mediante técnicas de cartografía de suelos digital. Todo ello validaría la metodología desarrollada de disgregación del mapa de suelo y verificaría los beneficios de su implementación en la zonificación integrada.

## **5 Bibliografía**

- Fregoni, Mario., Schuster, Danny., Paoletti, Andrea., 2003. Terroir, zonazione, viticoltura : trattato internazionale. Phytoline, Verona.
- Gómez-Miguel, V.D., 2011. Terroir: Parte III, in: Atlas das Castas da Península Ibérica: História, Terroir, Ampelografia. Dinalivro, p. 56.
- Gómez-Miguel, V.D., Sotés, V., 1992. Metodología y primeros resultados para la zonificación vitícola de la denominación de origen Ribera del Duero. Presented at the XX World Congress on Vine and Wine, p. 2:20.
- Gómez-Miguel, V.D., Sotés, V., 2002. Delimitation of terroir in the AOC Rueda and Toro (Castilla y León, Spain). Presented at the IVe Symposium International Zonage Vitivinicole, Avignon.
- Gómez-Miguel, V.D., Sotés, V., 2004. Zonificación del Terroir: Estudio de Suelos y Ordenación del Cultivo de la Vid en la DO Somontano. Universidad Politécnica de Madrid (UPM), Madrid.
- Gómez-Miguel, V.D., Sotés, V., 2006. Zonificación del Terroir: Estudio de Suelos y Ordenación del Cultivo de la Vid en la DO Cigales. Universidad Politécnica de Madrid (UPM), Madrid.
- Gómez-Miguel, V.D., Sotés, V., 2010. Zonificación del Terroir: Estudio de Suelos y Ordenación del Cultivo de la Vid en la DO Arribes. Universidad Politécnica de Madrid (UPM), Madrid.
- Gómez-Miguel, V.D., Sotés, V., 2011. Zonificación del Terroir: Estudio de Suelos y Ordenación del Cultivo de la Vid en la RD Douro. Universidad Politécnica de Madrid (UPM), Madrid.
- Gómez-Miguel, V.D., Sotés, V., 2015. Zonificación del Terroir: Estudio de Suelos y Ordenación del Cultivo de la Vid en la DO Campo de Borja (Zaragoza). Universidad Politécnica de Madrid (UPM), Madrid.
- Gómez Sánchez, P.J., 1995. Desarrollo de una metodología edafoclimática para Zonificación Vitícola: aplicación a la D.O. Ribera del Duero. Universidad Politécnica de Madrid, Madrid.
- Kuhn, M., Johnson, K., 2013. Applied Predictive Modeling. Springer New York, New York, NY. <https://doi.org/10.1007/978-1-4614-6849-3>

OIV, 2012. Resolución 423/2012: OIV Guidelines for vitiviniculture zoning methodologies on a soil and climate level.

Soil Survey Division Staff, 2014. Keys to Soil Taxonomy, 12th Edition. ed. USDA-Natural Resources Conservation Service, Washington, DC.

Vaudour, E., Costantini, E., Jones, G.V., Mocali, S., 2015. An overview of the recent approaches to terroir functional modelling, footprinting and zoning. SOIL 1, 287–312. <https://doi.org/10.5194/soil-1-287-2015>

## GENERAL CONCLUSIONS

The "disaggregation" of polytaxic cartographic units from conventional maps by disclosing of the non-explicit soil distribution pattern is an active line of action within the Digital Soil Mapping framework with positive results and growing interest.

It is based on quantitative soil-landscape models that are built using classification methods whose results are strongly influenced by the category assignment and sampling scheme during the definition of the training sets. Variations in most recent applications have focused on improving these two key aspects

In this regard, the proposed improvements in terms of expert-knowledge rules and geoforms patterns to potentially allocate categories as well as sampling limits, distribution, and size, have given rise to improved accuracy values on their implementations. Further research on the comparison between strategies and parameters is deemed necessary to identify performers and in which conditions.

Under these considerations, a new methodology for the disaggregation of polytaxic Soil Map Units of conventional soil maps has been developed. It relies on the detection of soil homogeneous areas with no prior definition using the unsupervised CLARA classification technique with Mahalanobis distance within a space of selected covariates and the expert knowledge-driven correlation of the areas with the Soil Taxonomy Units. It is an alternative approach to predominant methods based on supervised classification techniques that is suited to large datasets and that reduces the effect of redundancy by considering the correlation between covariates.

The new methodology successfully implemented in the DO Campo de Borja with a positive result. Altogether, it resulted in a remarkable increase of identified SMUs (threefold) and their delineations (+ 1.223%), while reducing their average-size area (-93%) and the average number of STU per SMU, so most of the new SMU are monotaxic or single-component. Such effects reflect its ability to detect soil homogeneous areas. Given these characteristics, its application might be geared towards regions where the potential location of STU within SMUs is deemed unclear, and furthermore, it has a potential to support new two-stage strategies.

A strategy to refine the map delineations for optimum publication criteria was considered beneficial insofar as it has little influence on the Overall Accuracy values and, on the contrary,

achieves balance in the map texture, reducing the IMR by 71%. However, these accuracy values were not particularly high. Further improvements could be achieved by incorporating lithological or soil related covariates or by validation through new sets of random observations.

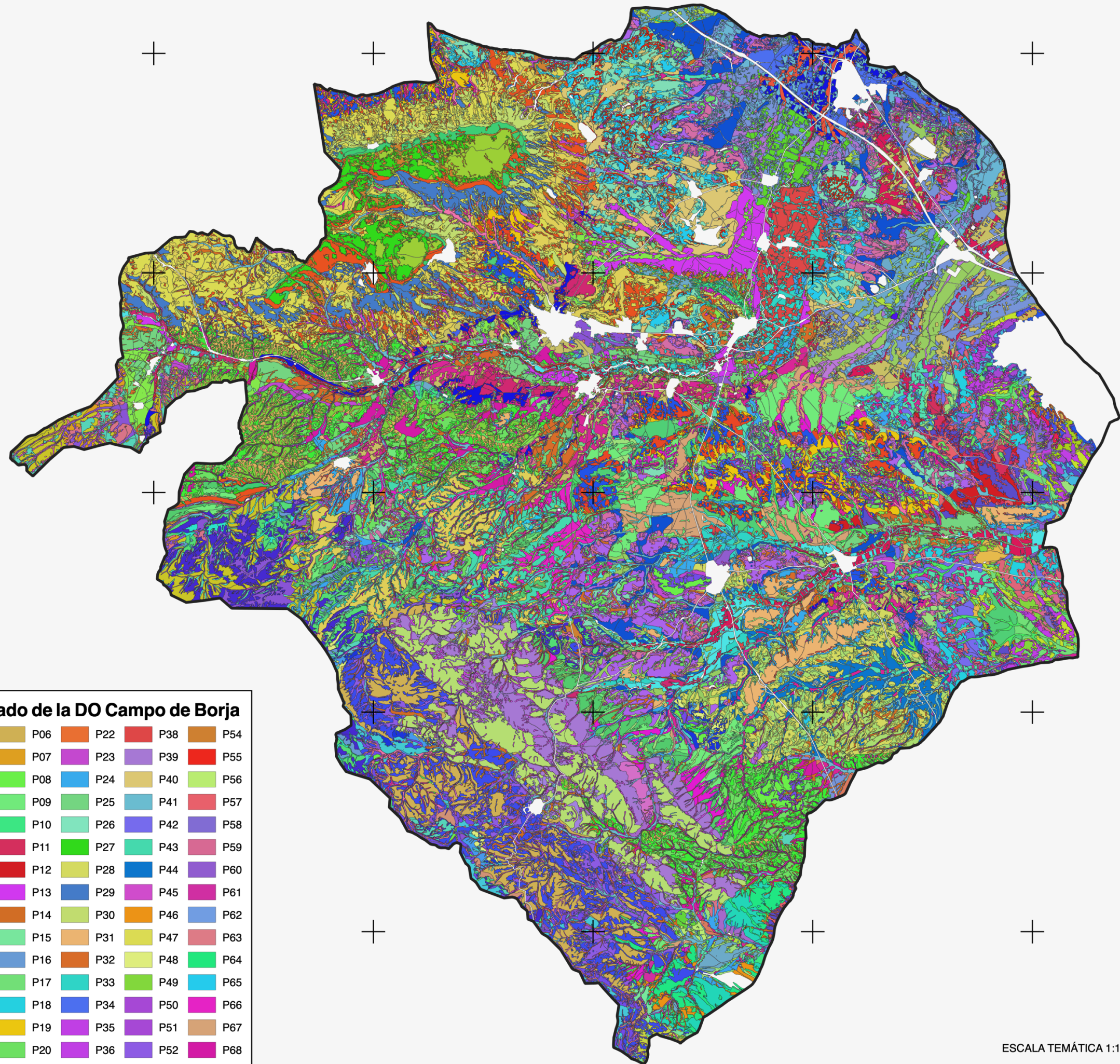
The utilisation of the disaggregated soil map in terroir zoning of the DO Campo de Borja by the Integrated Terroir Zoning methodology compared to the original one based on the conventional soil map has resulted in a better identification of Homogeneous Terroir Units within the project region and a greater differentiation. These facts emerge from the increase in the number of map units and the expansion of the range within the index distribution that quantifies their quality. Both aspects have led to an improvement in detecting areas with higher potential quality within uniform areas that were originally undervalued. The quality assessment is based on parameters identified as meaningful to understand the distribution of vineyards in the region and the occupation extent of the Homogeneous Terroir Units.

These insights are directly derived from the improvements of the disaggregated soil map in terms of cartographic and categorical detail via digital soil mapping techniques. Ultimately, the proposed methodology for disaggregating conventional soil maps would be validated, while the applicability of the resulting maps and the benefits of their implementation into the integrated zoning of the terroir would be verified.



**ANEXO I MAPA DEL INVENTARIO DEL RECURSO SUELO  
DISGREGADO DE LA DENOMINACIÓN DE ORIGEN  
CAMPO DE BORJA**





**Inventario del Recurso Suelo disgregado de la DO Campo de Borja**

SMU	M16	M32	M48	M64	P06	P22	P38	P54
M01	M17	M33	M49	M65	P07	P23	P39	P55
M02	M18	M34	M50	M66	P08	P24	P40	P56
M03	M19	M35	M51	M67	P09	P25	P41	P57
M04	M20	M36	M52	M68	P10	P26	P42	P58
M05	M21	M37	M53	M69	P11	P27	P43	P59
M06	M22	M38	M54	M70	P12	P28	P44	P60
M07	M23	M39	M55	M71	P13	P29	P45	P61
M08	M24	M40	M56	M72	P14	P30	P46	P62
M09	M25	M41	M57	M73	P15	P31	P47	P63
M10	M26	M42	M58	M74	P16	P32	P48	P64
M11	M27	M43	M59	P01	P17	P33	P49	P65
M12	M28	M44	M60	P02	P18	P34	P50	P66
M13	M29	M45	M61	P03	P19	P35	P51	P67
M14	M30	M46	M62	P04	P20	P36	P52	P68
M15	M31	M47	M63	P05	P21	P37	P53	P69

ESCALA TEMÁTICA 1:10.000

Autor de la información temática: Alberto Lázaro-López (UBU). 2021  
 Basado en el IRS convencional (Gómez-Miguel, et al., 2014)



Tabla de contenido de las unidades cartográficas de suelos del mapa de suelos disgregado de la Denominación de Origen Campo de Borja en términos de unidades taxonómicas de suelos.

SMU	STU 1ª	STU 1ª (%)	STU 2ª	STU 2ª (%)	STU 3ª	STU 3ª (%)	STU 4ª	STU 4ª (%)
M01	HxaC1N6	100						
M02	CxpT4F	100						
M03	HxaC1F	100						
M04	CxpT1N4	100						
M05	HxaT1N7	100						
M06	XfnT1F	100						
M07	CxpT7F	100						
M08	HxaC1C4	100						
M09	XonT1C6	100						
M10	CxpP2F	100						
M11	CxpT1F	100						
M12	CxpT1FV	100						
M13	CxpT1M2	100						
M14	CxpT1N2	100						
M15	CxpT1N7	100						
M16	CxpT1P5	100						
M17	CxpT4Y2	100						
M18	CxpT4Y5	100						
M19	CxpT5F	100						
M20	CxpT8F	100						
M21	HxaC1C9	100						
M22	HxaC1P6	100						
M23	HxaC1T	100						
M24	HxaC2F	100						
M25	HxaC2Y5	100						
M26	HxaC3F	100						
M27	HxaT1F	100						
M28	XfnT1FV	100						
M29	XfnT3F	100						
M30	XonL1M2	100						
M31	XonL1R5	100						
M32	XonT1M2	100						
M33	CxpP1F	100						
M34	CxpT1C9	100						
M35	CxpT1G	100						
M36	CxpT1M9	100						
M37	CxpT2F	100						
M38	CxpT2Y7	100						
M39	CxpT3C9	100						
M40	CxpT3F	100						
M41	CxpT3T	100						
M42	CxpT3Y2	100						
M43	CxpT3Y7	100						
M44	CxpT5Y5	100						
M45	EapT1FV	100						
M46	HxaC1FY	100						
M47	HxaC1G	100						
M48	HxaC1P1	100						

SMU	STU 1 <sup>a</sup>	STU 1 <sup>a</sup> (%)	STU 2 <sup>a</sup>	STU 2 <sup>a</sup> (%)	STU 3 <sup>a</sup>	STU 3 <sup>a</sup> (%)	STU 4 <sup>a</sup>	STU 4 <sup>a</sup> (%)
M49	HxaC1Y7	100						
M50	HxaC2Y2	100						
M51	HxaT1FV	100						
M52	HxaT2M2	100						
M53	PxaC1C9	100						
M54	PxaC1P1	100						
M55	PxaC1T	100						
M56	XfnT1C9	100						
M57	XfnT1N4	100						
M58	XfnT2N4	100						
M59	XfnT2Y2	100						
M60	XfnT3FV	100						
M61	XfnT4FY	100						
M62	XfnT1F	100						
M63	XonT1FV	100						
M64	XonT1R5	100						
M65	XonT1Y5	100						
M66	XonT1Y9	100						
M67	XonT2C4	100						
M68	XonT2F	100						
M69	XonT2M2	100						
M70	XonT2Y5	100						
M71	XonT3FY	100						
M72	XonT4C9	100						
M73	XonT4Y2	100						
M74	XonT6C9	100						
P01	HxaC1N4	65	XonT1C6	35				
P02	CxpT1M9	70	HxaT2M2	30				
P03	CxpT1P1	85					XonT2M2, CxpT2P1	15
P04	CxpT2T	95					CxpT5T	5
P05	HxaC1C9	75	XfnT1C9	25				
P06	HxaT1M2	55	CxpT1M2	45				
P07	HxaT2N7	70	XonT1C6	30				
P08	XonT1R5	55	XonL1R5	35			HxaT2N7	10
P09	CxpP1T	80					HxaC2T, CxpTA1T, HxaT2T	20
P10	CxpP2T	90					XonT3T	10
P11	CxpT1C6	90					CxpT2C6	10
P12	CxpT1C6	85					CxpT2C6, XfnT1C6	15
P13	CxpT1F	90					CxpT4F	10
P14	CxpT1FV	85					HxaA1FV, HxaG1FV	15
P15	CxpT1M2	80	PxaT1M2	20				
P16	CxpT1N4	75	XfnT2N4	25				
P17	CxpT2C4	95					CxpT4C4	5
P18	CxpT2Y7	90					PxaP1Y2	10
P19	CxpT3C9	85					CxpT2C9, XonT2C9	15
P20	CxpT3Y2	60	CxpT4Y2	35			XfnT3Y2	5
P21	CxpT3Y7	80					CxpP1Y7, PxaP1Y7, XfnT1Y7	20
P22	CxpT4C9	60	PxaT1C9	40				
P23	CxpT4F	90					XonT1FV	10

SMU	STU 1 <sup>a</sup>	STU 1 <sup>a</sup> (%)	STU 2 <sup>a</sup>	STU 2 <sup>a</sup> (%)	STU 3 <sup>a</sup>	STU 3 <sup>a</sup> (%)	STU 4 <sup>a</sup>	STU 4 <sup>a</sup> (%)
P24	CxpT5F	80					CxpT7F, CxpP2F	20
P25	CxpT5F	90					CxpT4F	10
P26	XfnT2F	50	CxpT5F	50				
P27	CxpT5Y5	95					CxpP2F	5
P28	CxpT5Y5	95					CxpT3Y5, HxaC1Y5	5
P29	CxpT5Y5	90					XfnT1Y5	10
P30	CxpT5Y5	90					XonT2Y5	10
P31	CxpT6F	85					CxpT7F, CxpP2F	15
P32	CxpT8F	80	CxpT4F	20				
P33	CxpT8F	75					CxpT4F, XfnT2F	25
P34	HxaC3F	50	EapT1FV	50				
P35	HxaC1C4	90					XfnT1C6	10
P36	HxaC1C4	90					XonT3C4	10
P37	HxaC1F	80					XfnT3F, HxaC4F, CxpP2F	20
P38	HxaC1F	90					XonT1F	10
P39	HxaC1P1	60	HxaT1P1	40				
P40	HxaC1Y2	95					XonT1Y2	5
P41	HxaC2FV	85	HxaT2FV	15				
P42	HxaC2Y2	80	XfnT3Y2	20				
P43	HxaC2Y5	90					CxpT1Y5, RxaC1F	10
P44	HxaC3Y5	95					HxaT1Y5	5
P45	HxaT1N7	80	XonT1C6	20				
P46	HxaT1P4	65	PxaC1P1	35				
P47	HxaT1Y5	85	HxmC1Y5	15				
P48	HxaT2M2	90					HxaC1P6	10
P49	PxaC1C9	70	HxaC1C4	30				
P50	XfnTA1Y2	50	PxaP1Y2	50				
P51	PxaT1C9	85	PxaC1C9	15				
P52	PxaT1M2	50	HxaT3M2	25	HxaT4M2	25		
P53	XfnT1F	85	HxaT1T	15				
P54	XfnT1Y2	80	RxaC1F	20				
P55	XfnT2F	50	XfnT2FV	50				
P56	XonL1R5	70	HxaT2M2	30				
P57	XonT1C6	90					XonT2C6	10
P58	XonT1C9	80	HxaC1C4	20				
P59	XonT1C9	60	XonT7C9	30			PxaT1C9	10
P60	XonT1G	75					XonT2G, CxpT2G	25
P61	XonT1M2	75	HxaC1P1	25				
P62	XonT1R5	75	HxaT2M2	25				
P63	XonT1R5	90					HxaT2N7	10
P64	XonT2M2	80	HxaT2M2	20				
P65	XonT2Y2	80					PxaT1Y2, XonL1Y2	20
P66	XonT2Y5	90					RxaC1F	10
P67	XonT3FY	90					XfnT1T	10
P68	XonT3Y7	80	RxaP1Y7	20				
P69	XonT4C9	85	XonT5C9	15				





**ANEXO II MAPA DE LA NUEVA ZONIFICACIÓN  
INTEGRADA DEL TERROIR DE LA DENOMINACIÓN DE  
ORIGEN CAMPO DE BORJA**



