Application of artificially intelligent systems for the identification of discrete fossiliferous levels

David M Martín-Perea [Corresp., 1, 2, 3] , **Lloyd A Courtenay** [4] , **M. Soledad Domingo** [5] , **Jorge Morales** [1]

[1] Palaeobiology Department, Museo Nacional de Ciencias Naturales - CSIC, Madrid, Madrid, Spain

[2] Geodynamics, Stratigraphy and Palaeontology Department, Universidad Complutense de Madrid, Madrid, Madrid, Spain

[3] Institute of Evolution in Africa, Madrid, Madrid, Spain

[4] Department of Cartographic and Land Engineering, Higher Polytechnic School of Avila, University of Salamanca, Avila, Avila, Spain

[5] Sciences, Social Sciences and Mathematics Department, Universidad Complutense de Madrid, Madrid, Madrid, Spain

Corresponding Author: David M Martín-Perea
Email address: davidmam@ucm.es

The separation of discrete fossiliferous levels within an archaeological or palaeontological site with no clear stratigraphic horizons has historically been carried out using qualitative approaches, relying on two-dimensional transversal and longitudinal projection planes. Analyses of this type, however, can often be conditioned by subjectivity based on the perspective of the analyst. This study presents a novel use of Machine Learning algorithms for pattern recognition techniques in the automated separation and identification of fossiliferous levels. This approach can be divided into three main steps including; (1) unsupervised Machine Learning for density based clustering, (2) expert-in-the-loop Collaborative Intelligence Learning for the integration of geological data followed by (3) supervised learning for the final fine-tuning of fossiliferous level models. For evaluation of these techniques, this method was tested in two Late Miocene sites of the Batallones Butte paleontological complex (Madrid, Spain). Here we show Machine Learning analyses to be a valuable tool for the processing of spatial data in an efficient and quantitative manner, successfully identifying the presence of discrete fossiliferous levels in both Batallones-3 and Batallones-10. Three discrete fossiliferous levels have been identified in Batallones-3, whereas another three have been differentiated in Batallones-10.

1 # Application of Artificially Intelligent Systems for the
2 # identification of discrete fossiliferous levels

4 David M. Martín-Perea[1, 2, 3], Lloyd A. Courtenay[4], M. Soledad Domingo[5], Jorge Morales[1]

6 [1]Palaeobiology Department, Museo Nacional de Ciencias Naturales - CSIC, Madrid, Spain
7 [2]Geodynamics, Stratigraphy and Palaeontology Department, Universidad Complutense de
8 Madrid, Madrid, Spain
9 [3]Institute of Evolution in Africa, Madrid, Madrid, Spain
10 [4]Department of Cartographic and Land Engineering, Higher Polytechnic School of Avila,
11 University of Salamanca, Avila, Spain
12 [5]Sciences, Social Sciences and Mathematics Department, Universidad Complutense de Madrid,
13 Madrid, Spain

15 Corresponding Author:
16 David Manuel Martín-Perea

18 Email address: davidmam@ucm.es

20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45

46
47
48
49

50 ## Abstract

51

52 The separation of discrete fossiliferous levels within an archaeological or palaeontological site
53 with no clear stratigraphic horizons has historically been carried out using qualitative
54 approaches, relying on two-dimensional transversal and longitudinal projection planes. Analyses
55 of this type, however, can often be conditioned by subjectivity based on the perspective of the
56 analyst. This study presents a novel use of Machine Learning algorithms for pattern recognition
57 techniques in the automated separation and identification of fossiliferous levels. This approach
58 can be divided into three main steps including; (1) unsupervised Machine Learning for density
59 based clustering, (2) expert-in-the-loop Collaborative Intelligence Learning for the integration of
60 geological data followed by (3) supervised learning for the final fine-tuning of fossiliferous level
61 models. For evaluation of these techniques, this method was tested in two Late Miocene sites of
62 the Batallones Butte paleontological complex (Madrid, Spain). Here we show Machine Learning
63 analyses to be a valuable tool for the processing of spatial data in an efficient and quantitative
64 manner, successfully identifying the presence of discrete fossiliferous levels in both Batallones-3
65 and Batallones-10. Three discrete fossiliferous levels have been identified in Batallones-3,
66 whereas another three have been differentiated in Batallones-10.

67

68 ## Introduction

69

70 The Batallones Butte, located 30 km to the south of Madrid (Spain, Fig.1A) and 1 km to the east
71 of Valdemoro (Fig.1B), is home to nine Late Miocene palaeontological sites (named Batallones-
72 1, Batallones-2, and so on). These sites hypothetically correspond to hourglass-shaped cavities
73 with upper openings (Fig.1C), formed as a consequence of pseudokarstic processes (Pozo et al.,
74 2004; Calvo et al., 2013), where mammals became trapped (Domingo et al., 2011; Domingo et
75 al., 2012; Domingo et al., 2013a). The upper part of this hour-glass structure is formed by
76 deposits where mammalian herbivore bones predominate, whereas the lower part of the structure
77 is overwhelmingly dominated by carnivoran remains (more than 90% of the fossils). Both
78 assemblages contain abundant, diverse and well-preserved remains (Domingo et al., 2013a).
79 These cavities were located in a landscape composed by woodland with patches of wooded
80 grassland (Domingo et al., 2013b; Domingo et al., 2016)

81

82 The Batallones Butte sites have a late Vallesian age (ca. 9.1 Ma; early Late Miocene), based on
83 the faunal association (Morales et al., 1992; Morales et al., 2004; Domingo et al., 2007; Morales
84 et al., 2008; López-Antoñanzas et al., 2010; Gómez Cano et al., 2011). However, micromammal
85 studies have shown that site formation was not synchronic between sites, with Batallones-10
86 being older than Batallones-1, which in turn is older than Batallones-3 (López-Antoñanzas et al.,
87 2010).

88

89 The decimeter to centimeter-scale separation of discrete fossiliferous levels within the same
90 lithostratigraphic unit is common in Cenozoic sites (Canals et al., 2003; Uribelarrea et al., 2014;
91 Sañudo et al., 2016; Gravina et al., 2018; Martín-Perea et al., 2019), but tend to be qualitative,

92 using two-dimensional transversal and longitudinal projection planes made in continuous strips
93 (Canals et al., 2003). Using similar methods, Martín-Escorza & Morales (2005) preliminarily
94 inferred discrete fossiliferous levels at Batallones-1 using spatial data from the first three years of
95 the excavation (2,273 remains, 17.19% of the total Batallones-1 excavated sample).
96
97 Palaeoecological, palaeoenvironmental and taphonomical studies carried out at Batallones Butte
98 (Antón & Morales, 2000; Domingo et al., 2011; Domingo et al., 2013a; Domingo et al., 2013b;
99 Domingo et al., 2016) have treated the upper and lower part of the hourglass-shaped cavities
100 separately based on their different depth and their different taxonomical composition (herbivores
101 vs. carnivores, respectively). Due to the challenges posed by the structure of the deposits that
102 filled the Batallones Butte cavities, such as lateral facies changes, deformations, depositional
103 asymmetries, collapse structures, slickensides and/or local tilting typical of cave deposits (Calvo
104 et al., 2013), geological sub-levels have not been identified within each of these parts. However,
105 the question still remains on whether or not discrete fossiliferous levels are found embedded in
106 the apparently geologically continuous deposits.
107
108 The pioneering introduction of Artificially Intelligent Algorithms (AIAs) in fields of archaeology
109 and palaeoanthropology has revolutionized numerous sub-disciplines such as those related with
110 genetic sequencing (Mondal et al., 2018), site and object detection (Anemone et al., 2011;
111 Conroy et al., 2012; Emerson and Anemone, 2012; Emerson et al., 2015; Benhabiles and Tabia,
112 2016; Block et al., 2016; Wills et al., 2017; Anemone and Conroy, 2018; Caspari & Crespo,
113 2019; Verschoof-van der Vaart and Lambers, 2019), physical anthropology (Bewes et al., 2019),
114 biomechanics (Püschel et al., 2018) restoration (Derech et al., 2018; Hermoza and Sipian, 2018),
115 as well as taphonomy (Arriaza and Domínguez-Rodrigo, 2016; Domínguez-Rodrigo, 2018;
116 Egeland et al., 2018; Byeon et al., 2019; Courtenay et al., 2019; Moclán et al., 2019).
117
118 AIAs, including those trained through Machine Learning and Deep Learning (ML & DL)
119 techniques, present numerous possibilities for the processing of highly complex and noisy data
120 sets (Bishop, 2006). Combined with their versatility, in many cases ML and DL approaches have
121 outperformed human experts in a multitude of specialized tasks. Here we present a means of
122 utilizing AIAs for the detection of patterns in 3D spatial data, establishing a method to
123 objectively and efficiently detect discrete levels that human experts may have mistakenly
124 overlooked. This study includes both unsupervised and supervised ML, joined through a hybrid
125 Human-AI collaborative approach in order to establish a final model that can be used to describe
126 the palaeontostratigraphic nature of each site.
127
128 The main goal of this study is to establish whether or not ML analysis can be used to
129 quantitatively identify discrete fossiliferous levels in palaeontological and archaeological sites
130 based on the study of the spatial distribution of fossils. Specifically, this method has been applied
131 and tested in two Batallones Butte sites (Batallones-3 and Batallones-10) in order to determine if
132 the deposits are homogeneous or heterogeneous with discrete fossiliferous levels.
133

Materials and Methods

135
136 The site of Batallones-3 has undergone systematic field excavations between the years 2001 and
137 2017, whereas excavations at Batallones-10 started in 2007 and are ongoing. Batallones-3

138 corresponds to the carnivoran-rich lower part of the hour-glass structure previously mentioned. It
139 is not clear whether an upper part ever existed for this fossil site, but, if it existed, it was
140 destroyed by slope-erosion processes (this site is located in the slope of the Batallones Butte). In
141 turn, Batallones-10 corresponds to the herbivore-rich upper part of the hour-glass structure and,
142 if present, will have a lower part, yet to be excavated. At both sites, standard fossil vertebrate
143 excavation protocols were followed in the extraction of palaeontological remains (Eberth et al.,
144 2007). On-site documentation of excavated remains has primarily consisted of taxonomical and
145 anatomical identification as well as in depth documentation of the spatial distribution of remains.
146 Spatial data obtained thus include standard x, y and z coordinates according to their position
147 within a grid, as well as the trend and plunge of elongated fossil remains. Additional data, for
148 future taphonomical studies, were also collected concerning the degree of articulation between
149 anatomical elements, the overall preservation of the remains and the restoration techniques used.
150
151 For the purpose of this study, the data used for both supervised and unsupervised ML
152 applications consisted solely in the spatial distribution of remains, including their x, y and z
153 coordinates for in-depth 3D analysis. Manageable slices were extracted, similar to the approach
154 proposed by Canals et al. (2003), 2 for each site across the entire length of the excavated area. In
155 the case of Batallones-10, one 2-metre-wide slice was taken across the x and one across the y
156 axis, searching for representative areas with the greatest object densities to capture a global
157 vision of the entire site. In the case of Batallones-3, located in a domically-shaped cave with a
158 debris cone in the middle under an inferred opening, two 2-metre-wide slices were obtained
159 along the site's x axis, one on each side of the debris cone.
160
161 Once slices were selected, the coordinates of each of the remains were imported into the free R
162 software, x64 v.3.5.1 (www.rproject.org, Core-Team, 2018), for further analysis.
163
164 **Unsupervised Machine Learning**
165
166 For initial detection of hidden patterns among the levels of Batallones Butte site, an unsupervised
167 density based clustering algorithm was used. Unsupervised ML algorithms are highly efficient
168 AIAs that are exceptional for their use in tasks such as pattern recognition, anomaly detection,
169 noise and dimensionality reduction, feature engineering as well as generative modelling (Patel,
170 2019). While unsupervised learning performs poorly in specifically defined tasks, their greatest
171 advantage can be seen through their flexibility when applied to versatile datasets for general
172 feature extraction. A popular application of unsupervised algorithms lies in clustering tasks, a
173 component of pattern recognition that is considered useful for processing and searching for
174 hidden trends in highly noisy datasets.
175 A non-parametric Density-Based Spatial Clustering of Applications with Noise (DBSCAN)
176 algorithm from the "fpc" R package was used for pattern recognition in 3D spatial data.
177 Proposed as a means of overcoming clustering issues where groupings of samples are not
178 straightforward (Ester et al., 1996), DBSCAN is highly efficient at detecting patterns in arbitrary
179 and noisily distributed samples (Sander et al., 1998; Schubert et al., 2017).
180
181 DBSCAN performs through establishing areas of significant point densities, thus providing an
182 efficient means of finding non-parametric patterns among noisy data sets while easily detecting
183 outliers in low-density regions. This is especially useful when used for pattern recognition in

184   spatial studies, where traditional clustering models such as partitioning and hierarchical
185   algorithms tend to focus groupings around a centroid or subsamples in a convex manner. This
186   frequently creates conflict in real-life circumstances.
187
188   DBSCAN has proven efficient when applied to 2D, 3D and/or higher dimensional feature spaces
189   (Ester et al., 1996; Sander et al., 1998; Schubert et al., 2017). The algorithm works by separating
190   points within the cluster (i.e. *core points*) from those on the border of the cluster (i.e. *border*
191   *points*). The differentiation between the two is established mathematically depending on the
192   reachability of neighbouring points (q) from a core point (p). To define this, DBSCAN requires
193   two main hyperparameters, the ε value defining the neighbourhood (a.k.a. *eps*) of a point, and the
194   minimum number of points (MinPts) that are required to form the cluster (i.e. the density of
195   points). From here DBSCAN defines the neighbourhood of p within dataset D as $N_{Eps}(p) = \{ q \in$
196    $D \mid \text{dist}(p,q) \leq \varepsilon \}$, adjusting for border points and the exclusion of noise through $p \in N_{Eps}(q)$ and
197    $|N_{Eps}(q)| \geq$ MinPts (Ester et al., 1996). The definition of *density-reachable* points for clustering is
198   thus established if the points can be linked in a chain, whereby $p_1, \ldots, p_n = q, p_n = p$ ensuring p_{i+1}
199   is directly reachable through density from p_i. Through this definition, border points may not be
200   density reachable if they do not fulfill the core point condition $|N_{Eps}(q)| \geq$ MinPts. The final
201   components that define point density for clustering are established through *density-connected*
202   points, whereby a point o is density-reachable from both p and q, essentially connecting both
203   border and cluster points.
204
205   Once density has been established within the sample, the assigning of these points into cluster C
206   requires the fulfilment of conditions; (1) $\forall$ p, q: if $p \in C$ and q fulfill the definition of being
207   density reachable, then q $\in$ C, while (2) $\forall$ p, q $\in C$: p must be density-connected to q. These
208   defined conditions are named *Maximality* and *Connectivity*, respectively (Ester et al., 1996). In
209   accordance with the aforementioned conditions, the final definition of noise can be logically
210   established through $\{p \in D \mid \forall \text{ i: p} \notin C_i\}$.
211
212   DBSCAN was thus trained on the entire dataset in an unsupervised manner, establishing an
213   average MinPts value between 3 and 5. These values were intuitively chosen depending on slices
214   with tightly packed densities (MinPts $\approx$ 5) and slices with a lower number of remains (MinPts $\approx$
215   3). This abides by general rules recommended by numerous authors (Sander et al., 1998;
216   Schubert et al., 2017), who define an approximate optimum MinPts = 2 * n° dimensions in the
217   dataset which is then adjusted depending on the complexity of point distributions and the
218   analyst's knowledge of the domain under consideration (Sander et al., 1998; Schubert et al.,
219   2017). ε values were established in accordance with both the correspondent MinPts values and
220   the slices under study. For optimization of this hyperparameter, k-distance graphs were plotted
221   according to the nearest neighbor (Ester et al., 1996), employing the "elbow" technique to find a
222   ε value according to both the MinPts parameter and the actual dataset (Thorndike, 1953; Patel et
223   al., 2009; Satopa, 2017). The use of this heuristic, as provided by the "dbscan" R package, can
224   then be adjusted to find the smallest ε value possible that can define a final model for training.
225   DBSCAN then establishes clusters within the dataset, separating noise into a separate "0"
226   cluster, and creating convex hulls through the use of border points q.
227
228   **Collaborative Intelligence Learning**
229

230 Considering the currently unquantifiable nature of numerous geological features that ML
231 algorithms may be unable to detect, a human-in-the-loop collaborative strategy was proposed for
232 this study (Kamar, 2016; Holzinger, 2016; Dellermann et al., 2019). The objectives of including
233 this hybrid intelligence strategy meant that human interaction could complement the strengths of
234 the pattern detection algorithms (Simard et al., 2017; Dellermann et al., 2019), whereby creating
235 a bridge between both unsupervised and supervised techniques for model creation.
236
237 The Expert-in-the-Loop (EitL) approach adopted here consisted in the revision of the clustered
238 dataset by considering possible underlying geological components that could be generating noise
239 undetectable by the model. Components of this nature could easily include disconformities,
240 geological faults, erosive surfaces, uneven sedimentation or even sedimentation of geological
241 elements such as large boulders that could be separating point-densities that would essentially
242 belong to the same layer.
243
244 The strategies employed consequently used EitL (in this case geologist-in-the-loop) interaction
245 to manually correct clustered patterns, assessing which of the clustered groups were geologically
246 separated and assigning these clusters to a new labelled layer. This was performed as objectively
247 as possible through complimenting palaeontological data routinely collected from the site during
248 excavation and evaluating the nature of each of DBSCAN's groupings. EitL thus benefits
249 through *Machine Teaching* (MT) in as much as the expert knowledge is used for troubleshooting
250 and debugging (Dellerman et al., 2019:7; also known as a *sense-making* approach). Moreover,
251 the amount of human input for models was monitored in a *collective* manner, using numerous
252 *domain experts* to ensure accuracy (Dellerman et al., 2019:7). In cases where clusters seemed
253 dubious or could not be clearly defined by either the EitL participants or the MT algorithm, these
254 were stripped of their labels and included in the noise (cluster "0") group for objective
255 classification by the trained ML algorithms in the supervised phase of the system. Once revised
256 by the EitL participants, each cluster group was assigned a new label that could be used to train
257 supervised algorithms in the following part of the workflow.
258
259 **Supervised Learning**
260
261 Once spatial data points had been passed through the DBSCAN algorithm for initial pattern
262 detection and then corrected by geologist-in-the-loop interventions, supervised algorithms were
263 employed to define the final fossiliferous level models and classify those points in cluster group
264 "0".
265
266 Considering the amount of data available, no bootstrapping procedures were deemed necessary
267 prior to supervised training. Each algorithm was thus trained using 70:30% [*train:test*] splits
268 using k-fold cross-validation ($k = 10$) in order to ensure the model could efficiently adjust its
269 weights. While some authors propose the use of Spatial Cross-Validation (SCV) for studies
270 regarding geographic data (Miller, 2004; Brenning, 2012; Pohjankukka et al., 2017; Lovelace et
271 al., 2019), these algorithms were not seen to produce any significant change to the quality of
272 results. Additionally, hyperparameter optimization was performed using a random search loop
273 function programmed in R; establishing the best hyperparameter values for each model using a
274 combination of random values until finding the optimum settings (Bergstra and Bengio, 2012).

275 These loop algorithms ran for 50 iterations and were then extrapolated and used for the final
276 classification models.
277
278 Two primary supervised ML algorithms were used for the final fine tuning of the fossiliferous
279 level models.
280

281 ● **Support Vector Machines (SVM)**. SVMs map out input vectors into a non-linear high
282   dimensional feature space, using hyperplanes to calculate the degree of separation
283   between samples (Cortes and Vapnik, 1995). In order to overcome traditional limitations
284   imposed by linearity, a *kernel* function is used to define the feature space (Bishop, 2006).
285   The constructed hyperplane can thus be used as a discriminant classifier decision surface
286   which uses maximized margin or decision boundaries to reduce chances of overfitting
287   (Cortes and Vapnik, 1995; Bishop, 2006). The consequent hyperplane can then be used to
288   plot the maximized margin separations between samples and can provide a visual means
289   of dividing strata, in an efficient and objectively computed way. For SVM applications
290   the "e1071" R package was used.
291

292 ● **Random Forest (RF)**. RF can be described as a robust and highly complex form of
293   decision tree which has proven useful in the past for the processing of spatio-temporal
294   data (Hengl et al., 2018). The RF algorithms use small random numbers of data set
295   variables rather than the whole dataset, constructing an independent decision tree with
296   each subsampling (Breiman et al., 1984; Breiman, 2001). The random variable selection
297   is additionally performed using bootstrap aggregation, using a technique frequently
298   referred to as out-of-bag observations. RF is then able to calculate the number of
299   iterations needed to minimize the out-of-bag error. Once the number of trees has been
300   selected, the algorithm averages results to produce a robust classification model
301   (Breiman, 2001). For RF applications the "caret" R package was used.
302

303 Model evaluations were performed following standardized ML protocol, evaluating the Kappa
304 (κ), Sensitivity, Specificity and Balanced Accuracy values obtained through confusion matrices
305 (Kuhn and Johnson, 2013; Lantz, 2013). Sensitivity, Specificity and Balanced Accuracy are
306 values derived from evaluations of Type I and Type II statistical errors in proportion with the rest
307 of the calculated confusion matrix. These values are presented as numbers between 0 (poor) and
308 1 (high performing classifiers). κ values are a further statistical adjustment of the accuracy metric
309 that measure model agreement relative to what would be expected by chance (Kuhn, 2008).
310 These values are similarly presented as numbers between 0 and 1 although negative values can
311 (although rarely) occur. Further evaluation of this statistic uses 0.8 as the threshold between poor
312 and powerful classification models. Finally, loss metrics were recorded to evaluate the predictive
313 power of each model. For this, the Mean Squared Error (MSE) metric was employed:

314
$$\frac{1}{n}\sum_{i=1}^{n}\left(t_i - p_i\right)^2$$

315 taking into consideration the target (t_i) and predicted (p_i) values for each classification. MSE
316 values are further interpreted considering AIAs of this type are trained to reduce the error

317   produced in CV sets, therefore the smaller the MSE value the more powerful the predictive
318   model. Model evaluation was performed using the "caret" R package.
319

320   **Fine Tuning of Fossiliferous Levels**
321

322   Once the optimal models were defined, these were used to classify any of the points separated as
323   noise by DBSCAN as well as the clusters that the EitL participants could not objectively define
324   as a part of either fossiliferous level (so as to avoid subjectivity). In order to empirically class
325   any point as a member of a fossiliferous level, the trained SVM and RF algorithms were used to
326   predict the class label of each point. A threshold of 80% security was established as acceptable.
327   Any points that could not be classed by the model with over an 80% predictive decision
328   boundary were thus rejected and therefore not classified into any level in the final fossiliferous
329   levels models.
330

Results

331

332

333   **Unsupervised Machine Learning clustering**
334

335   DBSCAN clustering proved highly efficient at quickly processing each slice, grouping fossil
336   remains into abundant clusters in both Batallones-3 (Fig. 2) and Batallones-10 (Fig. 3).
337   DBSCAN results in most cases present very noisy profiles, with the detection of large numbers
338   of tightly packed clusters. This is a result of the irregular density patterns detected in the data and
339   probably results from numerous geological and palaeontological disconformities that the
340   algorithm is unable to quantify in some areas. Nevertheless, profiles become less noisy towards
341   the extremities of each slice, helping the differentiation between levels, identifying at least 3
342   different fossiliferous levels.
343

344   On average, DBSCAN identified 166 points (5.6%) as noise in Batallones-10 and 62 (7.76%) in
345   Batallones-3, with the highest levels of noise appearing in Batallones-10 x axis slice (n = 141)
346   and Batallones-3 right slice (n = 56). For Batallones-3 right slice, this can be attributed to certain
347   areas of the profile having little separation between the layers (Fig. 2C). Nevertheless, excluding
348   noise, most clustered groups seem to follow a trend or pattern that highlights a separation
349   between different fossiliferous levels (see next section).
350

351   **Collaborative Intelligence Learning**
352

353   While DBSCAN results produced a large number of clusters, detailed analysis through the EitL
354   process identified numerous groupings that belonged to the same fossiliferous level. After
355   careful evaluation, each of these separations could be recognized as products of minute or large
356   disconformities. These can be attributed to the cavity's asymmetrical geomorphology at
357   Batallones-10, visible in the x axis slice (Fig. 4A) or to a fallen large carbonate block in
358   Batallones-10, as seen in the y axis slice (Fig. 4B). This highlights the importance of the
359   geologist-in-the-loop hybrid intelligence system for clarification and unification of clusters.
360

361   In other cases, numerous examples can be seen where depositional processes have placed
362   fossiliferous levels closer together such as the sedimentary onlap between levels II and III in

363   Batallones-10 x axis slice (Fig. 4A), and Batallones-3 right slice (Fig. 5B). Nevertheless,
364   considering the algorithms' effectiveness processing 3D data, DBSCAN is still able to detect
365   density patterns across the slice's x, y and z axes, thus separating the different levels.
366
367   **Supervised Learning**
368
369   Once processed and cleaned by both DBSCAN and the EitL specialists, databases were
370   channeled into supervised algorithms for training. In all cases, each of the databases generated by
371   unsupervised learning algorithms proved to be highly efficient for the training of their supervised
372   counterparts. Both SVM and RF obtained exceptional results with over 90% accuracy when
373   differentiating between groups and small MSE values on all accounts (Table 1). While SVM
374   proved to have greater balanced accuracy on training sets, RF in general obtained optimal MSE
375   values in testing, especially in the case of Batallones-10. Interestingly Batallones-3 (both left and
376   right slices) seems to create the most amount of confusion, nevertheless on all accounts κ values
377   remain above the acceptable threshold.
378
379   Using AIAs to clean the final profiles and de-noise the datasets, both SVM and RF proved to be
380   confident classifiers when making predictions. In most cases, over half the points denoted as
381   noise by the previous phases were successfully classified and assigned into a
382   palaeontostratigraphic level (Table 2). For Batallones-10 y axis slice and Batallones-3 left slice,
383   RF outperformed SVM with greater MSE values and success in classifying indeterminable
384   points. Nevertheless, in either case, RF tended to have greater decision making capabilities on
385   some individual points than SVM, even though SVM obtained greater overall MSE in others.
386
387   Upon further evaluation of model performance, SVM maximized margin decision boundaries
388   (dotted line in Fig. 6) are seen to expand or contract closer to the identified sedimentary onlap
389   areas (in both Batallones-3 and Batallones-10) while an increase in MSE is observed for these
390   particular areas as well (Table 2).
391
392   **Fine-Tuned Models**
393
394   Once the data had been processed by each of the phases of the workflow, a final fine-tuned
395   model of the fossiliferous levels was created.
396
397   The fine-tuned model produced for Batallones-3 shows three discrete fossiliferous levels on
398   either side (left and right) of the debris cone (Fig. 7). The left profile (Fig.7A) is laterally less
399   extensive (~3 m) than the right profile (Fig. 7B), which extends over 5 m. All identified levels in
400   both profiles dip from the debris cone outwards, with dip angles decreasing towards the
401   outermost parts of the cavity and towards the top of the infilling, with Level 3 dipping only
402   slightly in Batallones-3 left profile (Fig. 7A) and nearly horizontal in the outer part of the
403   Batallones-3 right profile (Fig. 7B).
404
405   Batallones-10 fine-tuned model shows three different discrete fossiliferous levels (Fig.8). A first
406   level (Level I), dips northward (Fig.8A) and towards the East and West dips towards the central
407   part of the cavity, adapting to the cave limits (Fig.8B). Level II conformably overlies Level I in
408   the southern section, dipping towards the North, and in the northern, eastern and western cavity

409   limits dips towards the central part of the cavity (Fig.8A and 8B). Finally, Level III is practically
410   horizontal, with gentle folding in the outermost limits of the cavity (Fig.8A and 8B).
411
412   **Discussion**
413
414   The methodology proposed in this study, based on artificially intelligent systems, has been
415   deemed effective in quantitatively identifying discrete fossiliferous levels in palaeontological and
416   archaeological sites based on the study of the spatial distribution of fossils. This new, more
417   quantitative approach could be considered a valuable substitute for previous qualitative
418   techniques that have been applied in the past.
419
420   Evaluating variable importance through testing for mean decrease in node impurity for RF
421   algorithms (Louppe et al., 2013) explains how a boost in the weight of the third dimension of
422   each slice helps support the differentiation between levels (Fig. 9). Batallones-3, for example,
423   shows a shift in the value of the x axis between the left-hand and right-hand sides of the debris
424   cone, having more weight in the final decision making capabilities of RF than the y dimension.
425   Data of this nature thus supports the need for 3D over 2D analysis in order to ensure an efficient
426   detection of patterns that 2D data may not be able to reveal.
427
428   Fine-tuned models produced for Batallones-3, in the lower part of a hypothetical hourglass-
429   shaped cavity, clearly show three discrete levels on either side of the debris cone (Fig. 7). Cave
430   asymmetry, as explained in Calvo et al. (2013), is clearly visible in these profiles, with greater
431   cave development towards the Northeast, in Batallones-3 right profile (Fig. 7B). Batallones-10
432   fine-tuned models, in the upper part of the hourglass-shaped cavities, show another three discrete
433   fossiliferous levels (Fig. 8). Similar geological structures were observed by Calvo et al. (2013) in
434   Batallones-9, located 50 m to the north of Batallones-10.
435
436   These newly discovered levels in both the lower levels (Batallones-3) and upper levels
437   (Batallones-10) of the hourglass-shaped cavities of Batallones Butte site support the preliminary
438   inferences made by Martín-Escorza & Morales (2005) about the possibility of discrete
439   fossiliferous levels within Batallones-1. Careful revision should therefore be carried out
440   considering previous palaeoecological, palaeoenvironmental and taphonomical analyses. As of
441   this point, future studies at Batallones-3 and Batallones-10 should consider each of these
442   individual discrete levels, instead of for the whole upper or lower part of the cavity.
443
444   As previously explained, Batallones-3 deposits and fossils are located in the lower part of the
445   cavity and the fossil site is considered exhausted as the encasing rock was already found, for this
446   reason, the three identified levels can be considered definitive. In the case of Batallones-10, the
447   encasing rock has not yet been found and, therefore fossil material recovered in future
448   excavations may reveal additional discrete and separable fossiliferous layers. Thus, the increase
449   in the number of levels at Batallones-10 cannot be discarded.
450
451   Future studies should analyze the embedded faunal assemblages separately for each of the
452   described levels rather than the entire upper or lower part of the cavity, paying special attention
453   to micromammal remains in order to establish chronological intra-site relationships, such as
454   those inter-site relationships described by Lopez-Antoñanzas et al. (2010). These detailed

455 micromammal studies would greatly refine the geochronological framework of the Batallones
456 Butte sites.
457
458 The overall workflow of this hybrid intelligence system has proven effective in detecting the
459 presence of 3 levels in both Batallones-10 and Batallones-3. While in some slices the detection
460 of these levels is clearer than in others, the use of two representative, thick and potent slices for
461 each site helps empirically support their identification. Through the implementation of
462 unsupervised algorithms for the initial detection of patterns, expert (geologist)-in-the-loop
463 interactions for sense making and the final tuning of profiles using supervised algorithms, the
464 overall system can be used to find areas where identification of fossiliferous levels is clearer.
465 This can then be extrapolated for the supervised classification of the rest of the site.
466
467 Moreover, while some isolated points suspiciously appear to be classified into one layer and can
468 be arguably considered out of place, future efforts should try to combine more data for the fine
469 tuning of teaching and training processes. These could include taxonomic or taphonomic
470 variables (Brain, 1981; Domínguez-Rodrigo et al., 2018), data regarding object density
471 (Kreutzer, 1992; Lam et al., 1998; 1999; 2003), weight and size (Bunn et al., 1980; Brain, 1981;
472 Bunn, 1987; Bunn and Pickering, 2010; Domínguez-Rodrigo et al., 2018), as well as orientation
473 and plunge patterns (Woodcock, 1977; Woodcock and Naylor ,1983; Lenoble and Betran, 2004;
474 Domínguez-Rodrigo and García-Pérez, 2013).
475
476 While models and systems without humans-in-the-loop would be an optimal solution for the
477 future (Holzinger, 2016), numerous geological factors that need extensive investigation for
478 modelling are required before this can be achieved. Likewise, a number of underlying
479 depositional and post-depositional processes may be generating confusion that machines are
480 unable to understand or process for the time being. Such is the case for the confusion generated
481 amongst algorithms caused by sedimentary onlaps, causing fossiliferous levels to lie closer to
482 each other, at both Batallones-3 and Batallones-10. This may, however, be solved by more
483 complex AIAs and the use of Deep Machine Teaching systems, such as clustering AIAs using
484 auto-encoders (Xie et al., 2016; Guo et al., 2017; Mrabah et al., 2019; Yang et al., 2019), or
485 those used for reinforcement learning tasks (Lake et al., 2014; Mnih et al., 2015; Holzinger,
486 2016; Simard et al., 2017). Efforts should therefore be made to investigate the effects of these
487 numerous geological components on pattern detection algorithms.
488

489 **Conclusions**
490
491 This study presents a novel use of Artificially Intelligent Systems for the quantitative
492 identification of discrete fossiliferous levels in palaeontological and archaeological sites based
493 solely on the study of the spatial three-dimensional distribution of fossil remains. These results
494 have been able to reveal new discoveries in the Batallones Butte site inner structure, including
495 three discrete levels at both Batallones-3 and Batallones-10.
496
497 The two lowermost levels of Batallones-3, Level I and Level II, dip outwards from the debris
498 cone towards the outer limits of the cave, becoming progressively more horizontal, whereas the
499 uppermost level, Level III is sub-horizontal. On the other hand, Batallones-10's lowermost Level
500 I dips northward into the cavity and adapts laterally to the cave limits whereas Levels II and III

501 also show northern dips but are more influenced by cave limits and dip towards the center of the
502 cavity in the outermost areas.
503
504 Through these discoveries, AIAs have been proved to be a highly efficient and objective means
505 of detecting spatial patterns in palaeontological sites. Unsupervised and unsupervised techniques
506 can greatly reduce (but not eliminate) the amount of subjectivity in discrete fossiliferous level
507 identification. The possibilities provided by the combination of these algorithms and expert-in-
508 the-loop systems are multiple, however it is important to point out that in order for these methods
509 to be widely used, a need for further experimentation and investigation is essential. This should
510 include an increase in the data used to feed these AIAs. For example, the incorporation of
511 taphonomic features of the remains in the analyzed fossil may provide an even finer tuning of
512 these models.
513

Acknowledgements

515

516 We would like to thank the TIDOP Group from the Department of Cartographic and Land
517 Engineering of the Higher Polytechnics School of Avila, University of Salamanca, especially
518 Diego González-Aguilera and Miguel-Ángel Maté-González for their support. We are very
519 grateful to Robert Anemone, João Coelho and an anonymous reviewer for their insightful
520 comments which helped greatly improve this paper.

521

References

523

524 Anemone RL, Emerson, CW, Conroy GC. 2011. Finding fossils in new ways: An artificial
525 neural network approach to predicting the location of productive fossil localities. *Evolutionary*
526 *Anthropology,* 20: 169-180.
527
528 Anemone RL, Conroy CG. 2018. *New Geospatial Approaches to the Anthropological Sciences*.
529 University of New Mexico Press/SAR Press, Alburquerque, New Mexico.
530
531 Antón M, Morales J. 2000. Inferencias paleoecológicas de la asociación de carnívoros del
532 yacimiento de Cerro Batallones. In: Morales J, ed. *Patrimonio Paleontológico de la Comunidad*
533 *de Madrid. Madrid: Arqueología, Paleontología y Etnografía*. Serie de la Conserjería de
534 Educación Comunidad de Madrid, 6, 190-201.
535
536 Arriaza MC, Domínguez-Rodrigo M. 2016. When Felids and Hominins ruled at Olduvai Gorge:
537 a Machine Learning Analysis of the Skeletal Profiles of the Non-Anthropogenic Bed I Sites.
538 *Quaternary Science reviews,* 139: 43-52.
539
540 Benhabiles H, Tabia H. 2016. Convolutional Neural Network for pottery retrieval. *Journal of*
541 *Electronic Imaging*. DOI: 10.1117/1.jei.26.1.011005
542
543 Bergstra J, Bengio, Y. 2012. Random Search for Hyper-Parameter Optimization, *Journal of*
544 *Machine Learning Research*, 13: 281-305.
545

546 Bewes J, Low A, Morphett A, Pate FD, Henneberg M. 2019. Artificial Intelligence for Sex
547 Determination of Skeletal Remains: Application of a Deep Learning Artificial Neural Network to
548 human skulls. *Journal of Forensic and Legal Medicine*, 62: 40-43.
549
550 Bishop C. 2006. *Pattern Recognition and Machine Learning*. Springer: Singapore.
551
552 Block S, Saltré F, Rodríguez-Rey M, Fordham DA, Unkel I, Bradshaw CJA. 2016. Where to Dig
553 for Fossils: Combining Climate-Envelope, Taphonomy and Discovery Models. *PLoS ONE,*
554 11(3): e0151090.
555
556 Breiman L, Friedman JH, Stone CJ, Olshen RA. 1984. *Classification and Regression Trees*.
557 London: Chapman and Hall.
558
559 Breiman L. 2001. Random Forests. *Machine Learning*, 45: 5-32.
560
561 Brenning A. 2012. Spatial Cross-Validation and Bootstrap for the Assessment of Prediction
562 Rules in Remote Sensing. *The R Package Sperrorest IEEE*: 5372-5375. DOI:
563 10.1109/IGARSS.2012.6352393
564
565 Brain CK. 1981. *The Hunters or the Hunted? An Introduction to African Cave Taphonomy*.
566 Chicago: The University of Chicago Press.
567
568 Bunn HT, Harris JWK, Isaac G, Kaufulu Z, Kroll E, Schick K, Toth N, Behrensmeyer A. 1980.
569 FxJj50: an Early Pleistocene Site in Northern Kenya. *World Archaeology*, 12(2): 109-136.
570
571 Bunn HT. 1987. Patterns of Skeletal Representation and Hominid Subsistence Activities at
572 Olduvai Gorge, Tanzania, and Koobi Fora, Kenya. *Journal of Human Evolution*, 15: 673-690.
573
574 Bunn HT, Pickering TR. 2010. Methodological Recommendations for Ungulate Mortality
575 Analyses in Paleoanthropology. *Quaternary Research*, 74: 388-394.
576
577 Byeon W, Domínguez-Rodrigo M, Arampatzis G, Baquedano E, Yravedra J, Maté-González
578 MA, Koumoutsakos P. 2019. Automated identification and deep classification of cut marks on
579 bones and its palaeonthropological implications. *Journal of Computation Science*, 32: 36-43.
580 DOI: 10.1016/j.jocs.2019.02.005
581
582 Calvo JP, Pozo M, Silva PG, Morales J. 2013. Pattern of sedimentary infilling of fossil mammal
583 traps formed in pseudokarst at Cerro de los Batallones, Madrid Basin, central Spain.
584 *Sedimentology*, 60: 1681-1708.
585
586 Canals A, Vallverdú J, Carbonell E. 2003. New archaeo-stratigraphic data for the TD6 level in
587 relation to Homo antecesor (Lower Pleistocene) at the site of Atapuerca, North-Central Spain.
588 *Geoarchaeology. An International Journal*, 18 (5): 481-504.
589
590 Caspari G, Crespo P. 2019 Convolutional Neural Networks for archaeological site detection -
591 Finding "princely" tombs. *Journal of Archaeological Science*. DOI: 10.1016/j.jas.2019.104998

592

593   Conroy GC, Emerson CW, Anemone RL, Townsend KEB. 2012. Let your fingers do the
594   walking: a simple spectral signature model for "remote" fossil prospecting. *Journal of Human*
595   *Evolution,* 63: 79-84.

596

597   Core R Team. 2018. A Language and Environment for Statistical Computing. R Foundation for
598   Statistical Computing, Vienna. *Available at http://www.R-project.org/*
599   Cortes C, Vapnik V. 1995. Support-Vector Networks. *Machine Learning*, 20: 273-29.
600   Courtenay LA, Yravedra J, Huguet R, Aramendi J, Maté-González MA, González-Aguilera D,
601   Arriaza MC. 2019 Combining Machine Learning algorithms and Geometric Morphometrics: a
602   study of carnivore tooth marks. *Palaeogeography, Palaeoclimatology, Palaeoecology.* 522:28-
603   29.

604

605   Dellermann D, Calma A, Lipusch N, Weber T, Weigel S, Ebel P. 2019. The Future of Human-AI
606   Collaboration: a taxonomy of design knowledge for hybrid intelligence systems. Hawaii
607   International Conference on System Sciences.

608

609   Derech N, Tal A, Shimshoni H. 2018. Solving archaeological puzzles. *axXiv*: 1812.10553

610

611   Domingo MS, Alberdi MT, Azanza B. 2007. A new quantitative biochronological ordination for
612   the Upper Neogene mammalian localities of Spain. *Palaeogeography, Palaeoclimatology,*
613   *Palaeoecology*, 255: 361-376.

614

615   Domingo MS, Domingo L, Sánchez IR, Alberdi MT, Azanza B, Morales J. 2011. New insights
616   on the taphonomy of the exceptional mammalian fossil sites of Cerro de los Batallones (Late
617   Miocene, Spain) based on rare earth element geochemistry. *Palaios*, 26: 55-65.

618

619   Domingo MS, Alberdi MT, Azanza B, Morales J. 2012. Mortality patterns and skeletal physical
620   condition of the carnivorans from the Miocene assemblage of Batallones-1 (Madrid Basin,
621   Spain). *Neues Jahrbuch für Geologie und Paläontologie – Abhandlungen*, 265: 131-145.

622

623   Domingo MS, Alberdi MT, Azanza B, Silva PG, Morales J. 2013a. Origin of an assemblage
624   massively dominated by carnivorans from the Miocene of Spain. *PLoS One*, 8: e63046.

625

626   Domingo MS, Domingo L, Badgley C, Sanisidro O, Morales J. 2013b. Resource partitioning
627   among top predators in a Miocene food web. *Proceedings of the Royal Society B*, 280:
628   20122138.

629

630   Domingo MS, Domingo L, Abella J, Valenciano A, Badgley C, Morales J. 2016. Feeding
631   ecology and habitat preferences of top predators from two Miocene carnivore-rich assemblages.
632   *Paleobiology*, 42: 489-507.

633

634   Domínguez-Rodrigo M. (2018) Successful classification of experimental Bone Surface
635   Modifications (BSM) through Machine Learning. A solution to the controversial use of BSM in
636   Palaeoanthropology. *Archaeological and Anthropological Sciences*. DOI: 10.1007/s12520-018-
637   0684-9

638

639 Domínguez-Rodrigo M, Cobo-Sáncez L, Yravedra J, Uribelarrea D, Arraiza C, Organista E,
640 Baquedano E. 2018. Fluvial Spatial Taphonomy: A New Method for the Study of Post-
641 Depositional Processes. *Archaeological and Anthropological Sciences*, 10: 1769-1789.
642

643 Dunmore CJ, Wollny G, Skinner MM. 2018. MIA-Clustering: a novel for segmentation of
644 palaeontological material. *PeerJ*. DOI: 10.7717/peerj.4374
645

646 Eberth DA, Rogers RR, Fiorillo AR. 2007. A practical approach to the study of bonebeds. In:
647 Rogers RR, Eberth DA, Fiorillo AR, eds. *Bonebeds: Genesis, analysis and paleobiological*
648 *significance.* The University of Chicago Press, Chicago, 265-331.
649

650 Egeland CP, Domínguez-Rodrigo M, Pickering TR, Menter CG, Heaton JL. 2018. Hominin
651 skeletal part abundances and claims of deliberate disposal of corpses in the Middle Pleistocene.
652 *PNAS*. DOI: 10.1073/pnas.1718678115
653

654 Emerson CW, Anemone RL. 2012. An artificial neural network approach to identifying
655 mammalian fossil localities in the Great Divide Basin, Wyoming. *Remote Sensing Letters*, 3:
656 453-460.
657

658 Emerson CW, Bommersbach B, Nachman B, Anemone RL. 2015. An object-oriented approach
659 to extracting productive fossil localities from remotely sensed imagery. *Remote Sensing,* 7
660

661 Ester M, Kriegel HP, Sander J, Xu X. 1996. A density-based algorithm for discovering clusters
662 in large spatial databases with noise. *Proceedings of the 2nd International Conference on*
663 *Knowledge Discovery and Data Mining*.
664

665 Gómez Cano AR, Hernández Fernández M, Álvarez-Sierra MA. 2011. Biogeographic
666 provincialism in rodent faunas from the Iberoccitanian Region (southwestern Europe) generates
667 severe diachrony within the Mammalian Neogene (MN) biochronologic scale during the Late
668 Miocene. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 307: 193-204.
669

670 Gravina B, Bachellerie F, Caux S, Discamps E, Faivre JP, Galland A, Michel A, Teyssandier N,
671 Bordes JG. 2018. No reliable evidence for a Neanderthal-Châtelperronian association at La
672 Roche-à-Pierrot, Saint-Césaire. *Scientific Reports*, 8: 15134.
673

674 Guo X, Liu X, Zhu E, Yin J. 2017. Deep Clustering with Convolutional Autoencoders.
675 *International Conference on Neural Information Processing*. DOI: 10.1007/978-3-319-70096-
676 0_39
677

678 Hengl T, Nussbaum M, Wright MN, Heuvelink GBM, Gräler B. 2018. Random Forest as a
679 Generic Framework for Predictive Modeling of Spatial and Spatio-Temporal Variables. *PeerJ*, 6:
680 e5518 DOI: 10.7717/peerj.5518
681

682 Hermoza R, Sipiarn I. 2018. 3D Reconstruction of Incomplete Archaeological Objects using a
683 Generative Adversarial Network. *Proceedings of Computer Graphics International*, arXiv:
684 1711.06363v2
685
686 Holzinger A. 2016. Interactive machine learning for health informatics: when do we need the
687 human in the loop? *Brain Informatics*, 3: 119-131.
688
689 Kamar E. 2016. Directions in hybrid intelligence: Complementing AI systems with human
690 intelligence. *IJCAI International Joint Conference on Artificial Intelligence*.
691
692 Kreutzer LA. 1992. Bison and Deer Bone Mineral Densities: Comparisons and Implications for
693 the Interpretation of Archaeological Faunas. *Journal of Archaeological Science*, 19: 271-294.
694
695 Kuhn M. 2008. Building Predictive Models in R Using the Caret Package. *Journal of Statistical*
696 *Software*, 28(5): 1-26.
697 Kuhn M, Johnson K. 2013. *Applied Predictive Modelling*. Heidelberg: Springer.
698 Lake BM, Ullman TD, Tenenbaum JB Gershman SJ. 2014. Building Machines that Learn and
699 Think Like People. *Behavioral and Brain Sciences*, 40, e253. DOI:
700 10.1017/S0140525X16001837
701
702 Lam YM, Chen X, Marean CW, Frey CJ. 1998. Bone Density and Long Bone Representation in
703 Archaeological Faunas: Comparing Results from CT and Photon Densitometry. *Journal of*
704 *Archaeological Science*, 25(6): 559-570.
705
706 Lam YM, Chen X, Pearson OM. 1999. Intertaxonomic Variability in Patterns of Bone Density
707 and the Differential Representation of Bovid, Cervid, and Equid Elements in the Archaeological
708 Record. *American Antiquity,* 64(2): 343-362.
709
710 Lam YM, Pearson OM, Marean CW, Chen X. 2003. Bone Density Studies in Zooarchaeology.
711 *Journal of Archaeological Science*, 30(12): 1701-1708.
712 Lantz B. 2013. *Machine Learning with R*. Birmingham: Packt Publishing.
713
714 Lenoble A, Bertran P. 2004. Fabric of Palaeolithic Levels: Methods and Implications for Site
715 Formation Processes. *Journal of Archaeological Science*, 31: 457-469.
716
717 López-Antoñanzas R, Peláez-Campomanes P, Álvarez-Sierra MA, García-Paredes I. 2010. New
718 species of Hispanomys (Rodentina, Cricetodontinae) from the Upper Miocene of Batallones
719 (Madrid, Spain). *Zoological Journal of the Linnean Society*, 160: 725-747.
720
721 Louppe G, Wehenkel L, Sutera A, Geurts P. 2013. Understanding Variable Importance in Forest
722 of Randomized Trees, *Neural Information Processing Systems Proceedings*.
723
724 Lovelace R, Nowosad J, Muenchow J. 2019. *Geocomputation with R*. CRC Press
725

726 Martín Escorza C, Morales J. 2005. Estructura interna del yacimiento mioceno de Batallones
727 (Madrid, España). *XVI Reunión Bienal de la Real Sociedad Española de Historia Natural*: 109-
728 111.
729
730 Martín-Perea DM, Fesharaki O, Rey Samper JJ, Arroyo X, Uribelarrea D, Cobo-Sánchez L,
731 Baquedano E, Mabulla A, Domínguez-Rodrigo M. 2019. Mineral assemblages and low energy
732 sedimentary processes in the FLK-Zinj, DS, PTK and AMK complex palaeolandscape (Olduvai
733 Gorge, Tanzania). *Quaternary International*, 526: 15-25.
734
735 Miller H. 2004. Tobler's First Law and Spatial Analysis. *Annals of the Association of American
736 Geographers*, 94(2).
737
738 Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, Graves A, Riedmiller
739 M, Fidjeland AK, Ostrovski G, Petersen S, Beattie C, Sadik A, Antonoglou I, King H, Kumaran
740 D, Wierstra D, Legg S, Hassabis D. 2015. Human-level control through deep reinforcement
741 learning. *Nature Letters*. DOI: 10.1038/nature14236
742
743 Moclán A, Domínguez-Rodrigo M, Yravedra J. 2019. Classifying bone breakage patterns: an
744 experimental analysis of fracture planes to discern between Hominin and carnivore activity using
745 Machine Learning (ML) algorithms. *Anthropological and Archaeological Sciences*. DOI:
746 10.1007/s12520-019-00815-6
747
748 Mondal M, Bertranpetit J, Lao O. 2018. Approximate Bayesian Computation with Deep
749 Learning supports a third archaic introgression in Asia and Oceania. *Nature Communications*.
750 DOI: 10.1038/s41467-018-08089-7
751
752 Morales J, Capitán J, Calvo JP, Sesé C. 1992. Nuevo yacimiento de vertebrados del Mioceno
753 Superior al Sur de Madrid (Cerro de los Batallones, Torrejón de Velasco). *Geogaceta*, 12: 77-80.
754
755 Morales J, Alcalá L, Álvarez-Sierra MA, Antón M, Azanza B, Calvo JP, Carrasco P, Fraile S,
756 García-Paredes I, Gómez E, Hernández Fernández M, Merino L, van der Meulen A, Martín-
757 Escorza C, Montoya P, Nieto M, Peigné S, Pérez B, Peláez-Campomanes P, Pozo M, Quiralte V,
758 Salesa MJ, Sánchez IM, Sánchez-Marco A, Silva PG, Soria MD, Turner A. 2004. Paleontología
759 del Sistema de yacimientos de mamíferos miocenos del Cerros de los Batallones, Cuenca de
760 Madrid. *Geogaceta*, 35: 139-142.
761
762 Morales J, Pozo M, Silva PG, Domingo MS, López-Antoñanzas R, Álvarez-Sierra MA, Antón
763 M, Martín-Escorza C, Quiralte V, Salesa MJ, Sánchez IM, Azanza B, Calvo JP, Carrasco P,
764 García-Paredes I, Knoll F, Merino L, van der Meulen A, Montoya P, Peigné S, Peláez-
765 Campomanes P, Sánchez-Marco A, Turner A, Abella J, Alcalde GM, Andrés M, Demiguel D,
766 Cantalapiedra JL, Fraile S, García Yelo B, Gómez Cano AR, López Guerrero P, Oliver Pérez A,
767 Siliceo G. 2008. El sistema de yacimientos de mamíferos miocenos del Cerro de los Batallones,
768 Cuenca de Madrid: estado actual y perspectivas. *Publicaciones del Seminario de Paleontología
769 de Zaragoza*, 8: 41-117.
770

771 Mrabah N, Khan NM, Ksantini R, Lachiri Z. 2019. Deep Clustering with a Dynamic
772 Autoencoder. *arXiv*: 1901.07752v4.
773
774 Patel AA. 2019. *Hands-On Unsupervised Learning with Python: How to Build Applied Machine*
775 *Learning Solutions from Unlabeled Data*. Tokyo: O'Reilly.
776
777 Pozo M, Calvo JP, Silva PG, Morales J, Peláez-Campomanes P, Nieto M. 2004. Geología del
778 sistema de yacimientos de mamíferos miocenos del Cerro de los Batallones, Cuenca de Madrid.
779 *Geogaceta*, 35: 143-146.
780
781 Püschel TA, Marcé-Nogué J, Gladman JT, Bobe R, Sellers WI. 2018. Inferring locomotor
782 behaviours in Miocene New World monkeys using finite element analysis, geometric
783 morphometrics and machine-learning classification techniques applied to talar morphology.
784 *Journal of the Royal Society Interface,* 15(146): 20180520.
785
786 Sander J, Ester M, Kriegel HP, Xu X. 1998. Density-based clustering in spatial databases: the
787 algorithm GDBSCAN and its applications. *Data Mining and Knowledge Discovery*, 2: 169-194.
788
789 Sañudo P, Blasco R, Fernández Peris J. 2016. Site formation dynamics and human occupations at
790 Bolomor Cave (Valencia, Spain): An archaeolstratigraphic analysis of levels I to XII (100-200
791 ka). *Quaternary International*, 417: 94-104.
792
793 Satopa V, Albrecht J, Irwin D, Raghavan B. 2017. Finding a "Kneedle in a Haystack: Detecting
794 Knee Points in System Behavior, Python PYPI. *Available at https://github.com/arkevi/kneed.git*
795
796 Schubert E, Sander J, Ester M, Kriegel HP, Xu X. 2017. DBSCAN revisited: Why and how you
797 should (still) use DBSCAN. *ACM Transactions on Database Systems*, 42(3) DOI:
798 10.1145/3068335
799
800 Simard PY, Saleema A, Chickering DM, Pelton AE, Ghorashi S, Meek C, Ramos G, Suh J,
801 Verwey J, Wang M, Wernsing J. 2017. Machine Teaching. A New Paradigm for Building
802 Machine Learning Systems. *arXiv*: 1707.06742v3
803
804 Thorndike RL. 1953. Who Belongs in the Family? *Psychometrika*, 48(4), 1267-276.
805
806 Uribelarrea D, Domínguez-Rodrigo M, Pérez-González A, Vegas Salamanca J, Baquedano E,
807 Mabulla A, Musiba C, Barboni D, Cobo-Sánchez L. 2014. Geo-archaeological and geometrically
808 corrected reconstruction of the 1.84 Ma FLK Zinj paleolandscape at Olduvai Gorge, Tanzania.
809 *Quaternary International*, 322: 7-31.
810
811 Verschoof-van der Vaart W, Lambers K. 2019. Learning to Look at LiDAR: the use of R-CNN
812 in the Automated Detection of Archaeological Objects in LiDAR Data from the Netherlands.
813 *Journal of Computer Applications in Archaeology*, 2(1): 31-40.
814
815 Wills S, Choiniere JN, Barrett PM. 2018. Predictive modelling of fossil-bearing locality
816 distributions in the Elliot Formation (Upper Triassic – Lower Jurassic), South Africa, using a

817   combined multivariate and spatial statistical analysis of present-day environmental data.
818   *Palaeogeography, Palaeoclimatology, Palaeoecology,* 489: 186-197.
819
820   Woodcock NH. 1977. Specification of Fabric Shapes using the Eigenvalue Method. *Geological*
821   *Society of America Bulletin*, 88: 1231-1236.
822
823   Woodcock, N.H.; Naylor, M.A. 1983. Randomness Testing in Three Dimensional Orientation
824   Data. *Journal of Structural Geology*, 5(5): 539-548.
825
826   Xie J, Girshick R, Farhadi A. 2016. Unsupervised Deep Embedding for Clustering Analysis.
827   *Proceedings of the 33rd International Conference on Machine Learning*. arXiv: 1511.06335
828
829   Yang X, Deng C, Zheng F, Yan J, Liu W. 2019. Deep Spectral Clustering using Dual
830   Autoencoder Network. *arXiv*: 1904.13113

Figure 1

Geographical and geological background.

(A) Map of Spain showing the location of the cities of Madrid and Valdemoro. (B) Detailed map of the situation of the Batallones Butte sites Batallones-3 and Batallones-10. (C) Hypothetical geomorphology and geology of Batallones Butte cavities, with an herbivore-dominated upper part and a carnivore-dominated lower cavity (modified from Calvo et al., 2013).

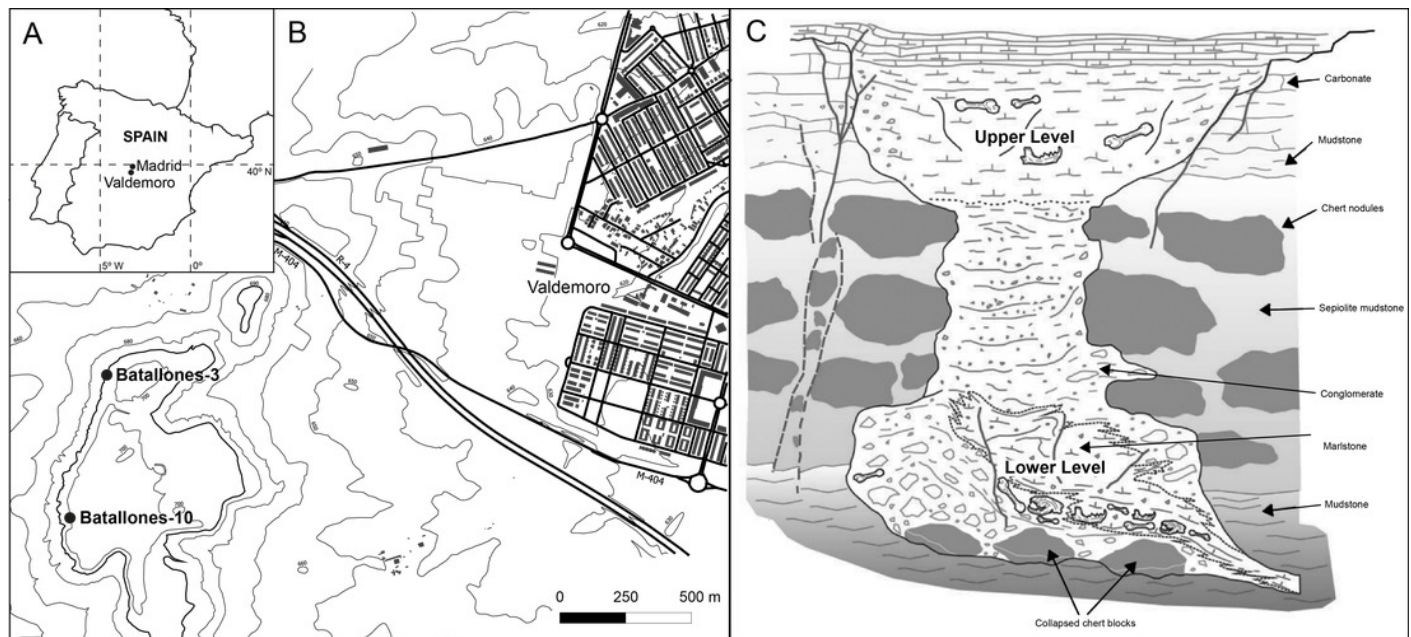



Figure 2

DBSCAN clustering for the studied slices at Batallones-3, each point representing a fossil remain. Z=depth.

(A) Batallones-3 grid and slice orientation. (B) Batallones-3 left slice clusters (MinPts = 3, ε = 11.5). (C) Batallones-3 right slice clusters (MinPts = 5, ε = 15).

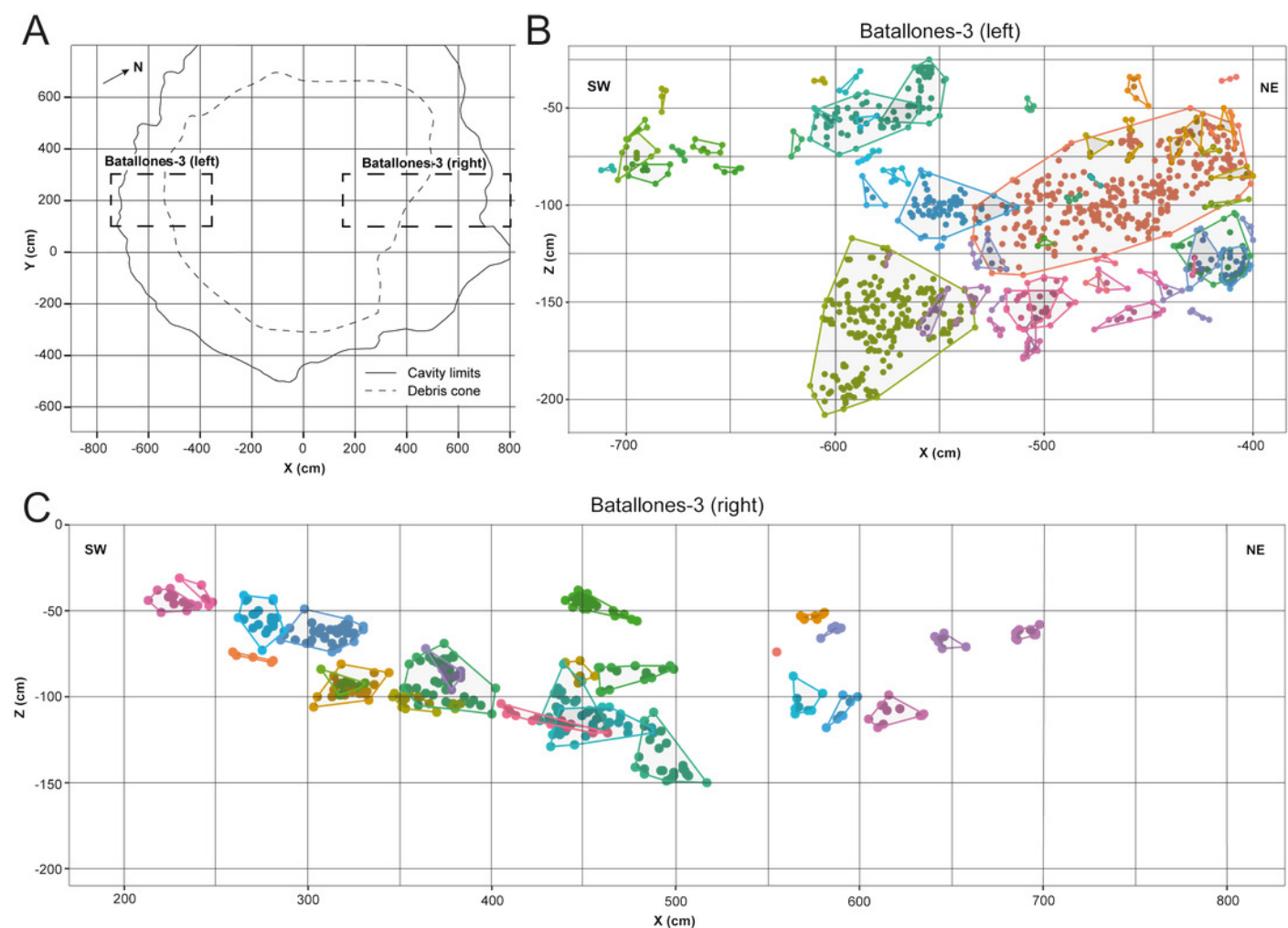



Figure 3

DBSCAN clustering for the studied slices at Batallones-10, each point representing a fossil remain. Z=depth.

(A) Batallones-10 grid and slice orientation. (B) Batallones-10 x axis slice clusters (MinPts = 5, ε = 27.5). (C) Batallones-10 y axis slice clusters (MinPts = 5, ε = 25).

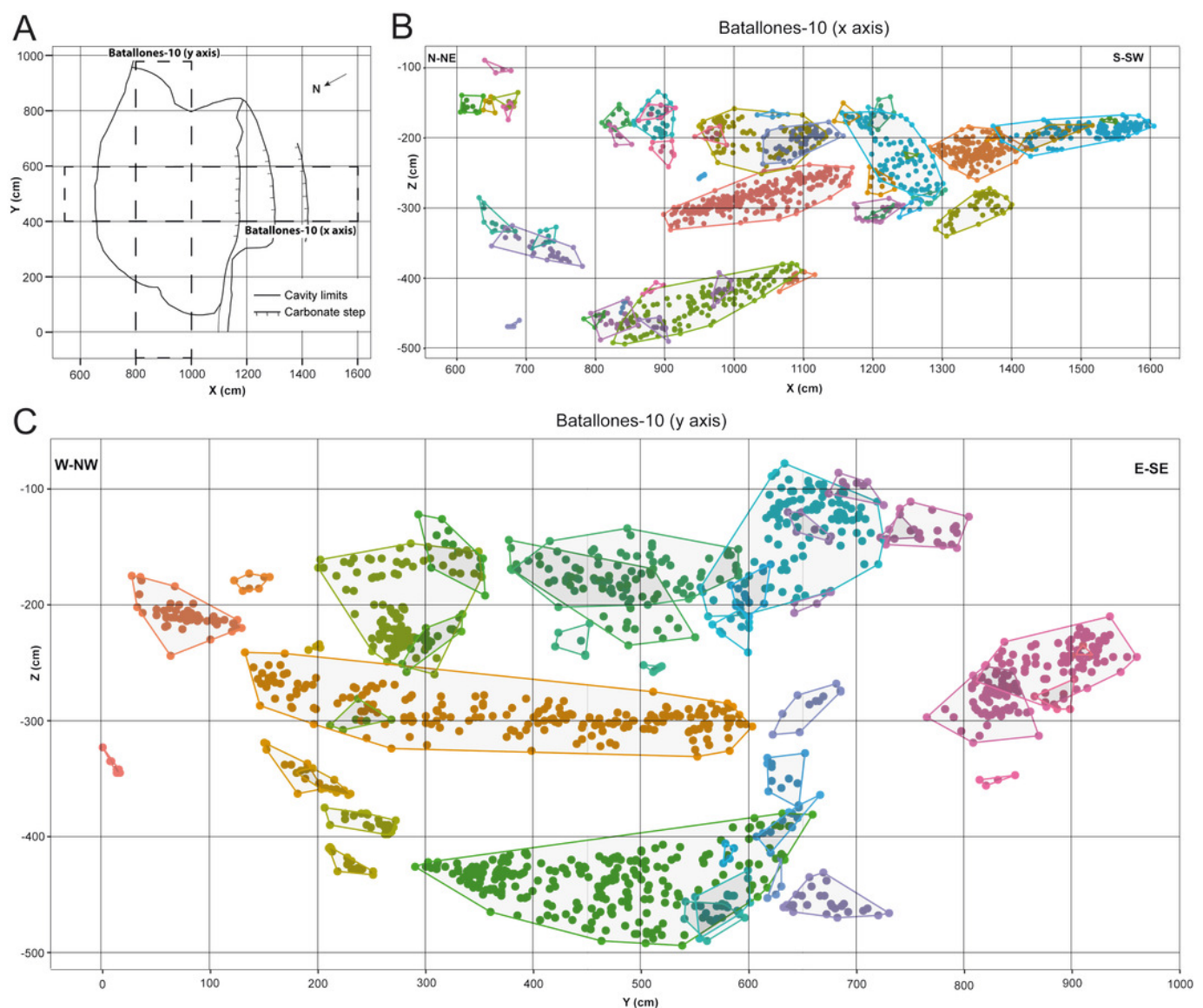



Figure 4

DBSCAN clustering and geological interpretation. In grey, site limits.

(A) Batallones-10 x axis slice clusters and interpretation. (B) Batallones-10 y axis slice clusters and interpretation.

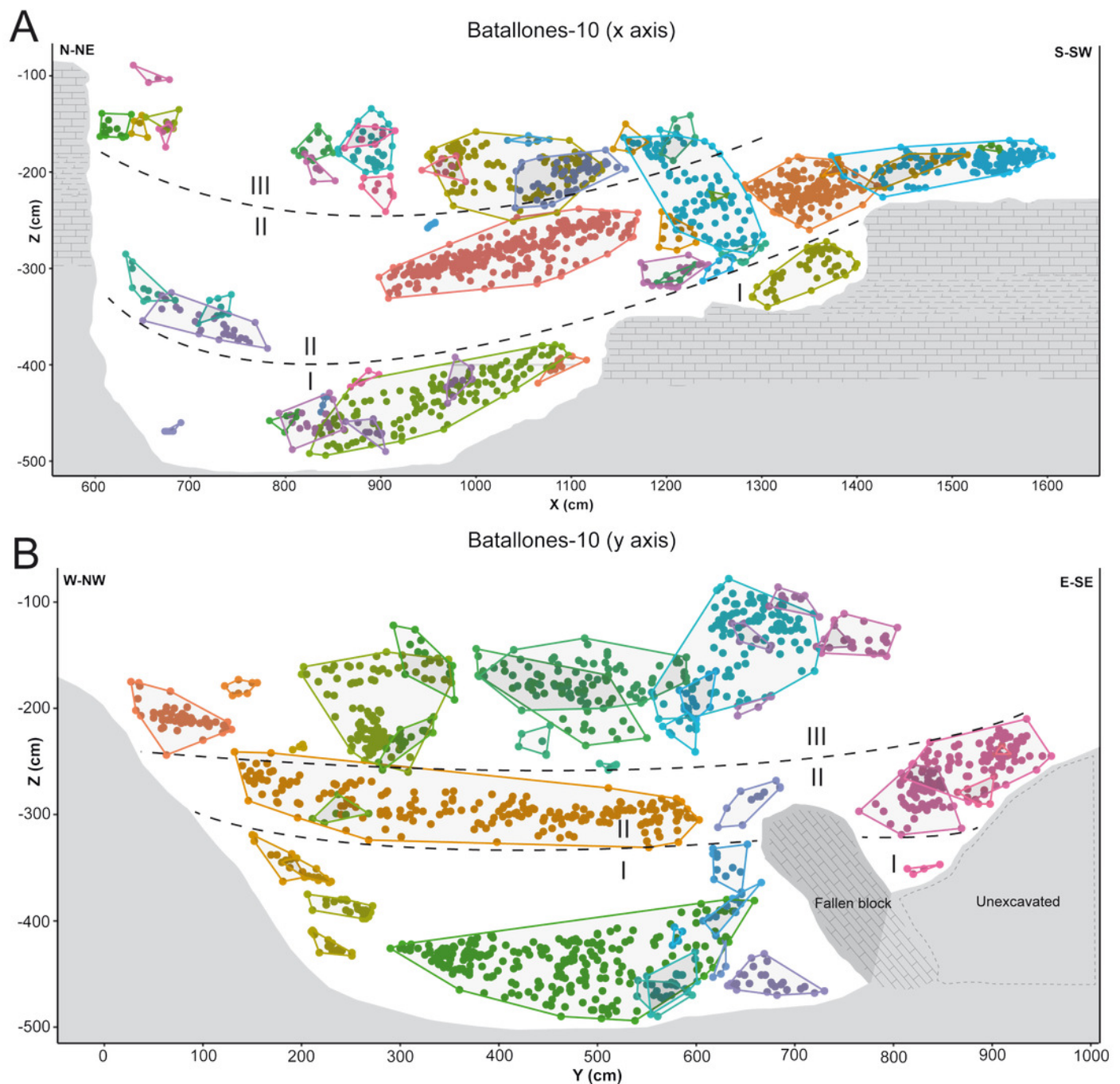



Figure 5

DBSCAN clustering and geological interpretation. In grey, site limits.

(A) Batallones-3 left slice clusters and interpretation. (B) Batallones-3 right slice clusters and interpretation.

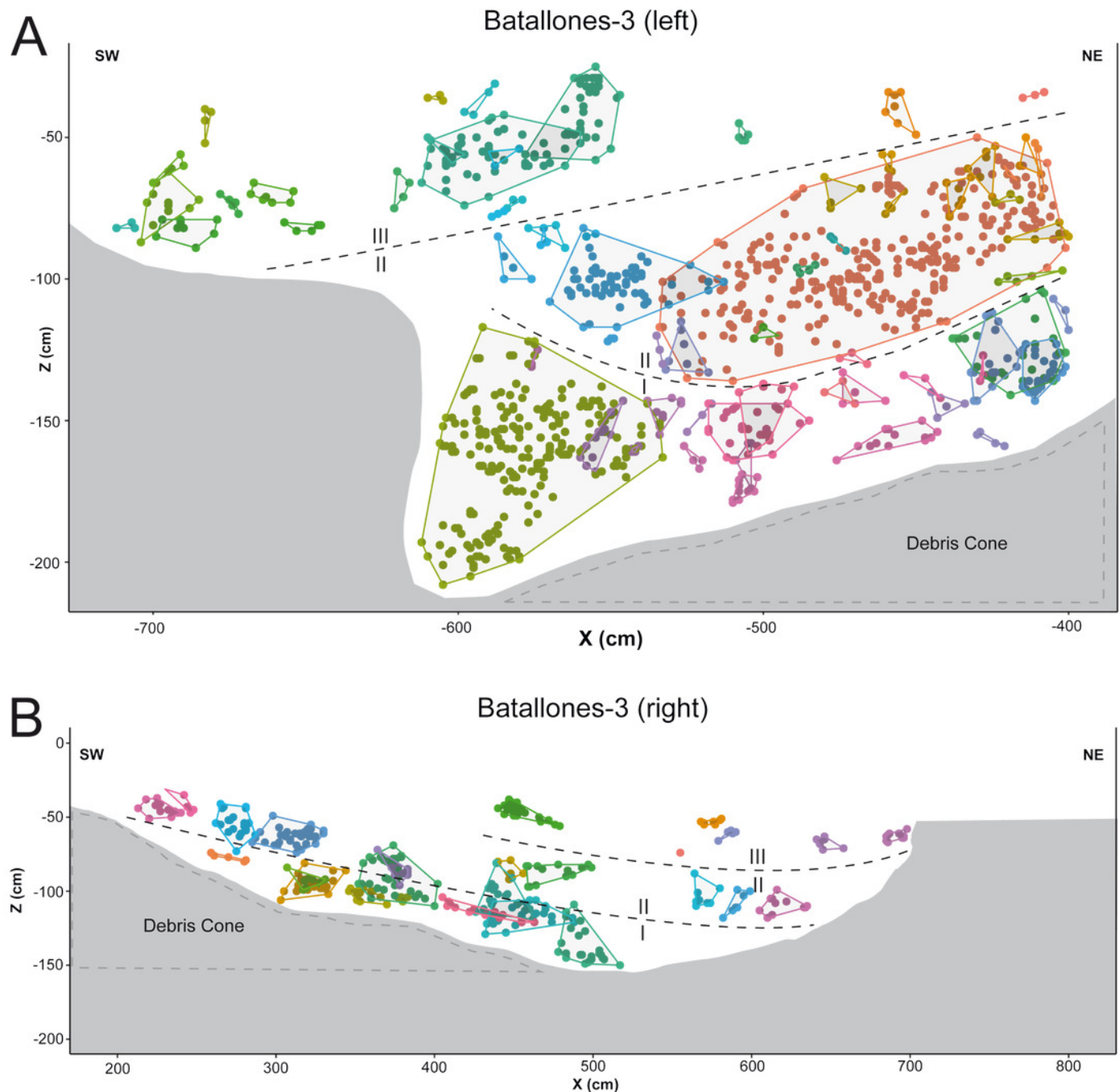



Figure 6

Examples of SVM maximized margins (dotted) and decision boundaries (solid line), plotted against calculations of material densities. Density values are reported as the number of points per square unit (cm), per quadrat.

(A) Batallones-10 y axis slice. (B) Batallones-3 right slice.

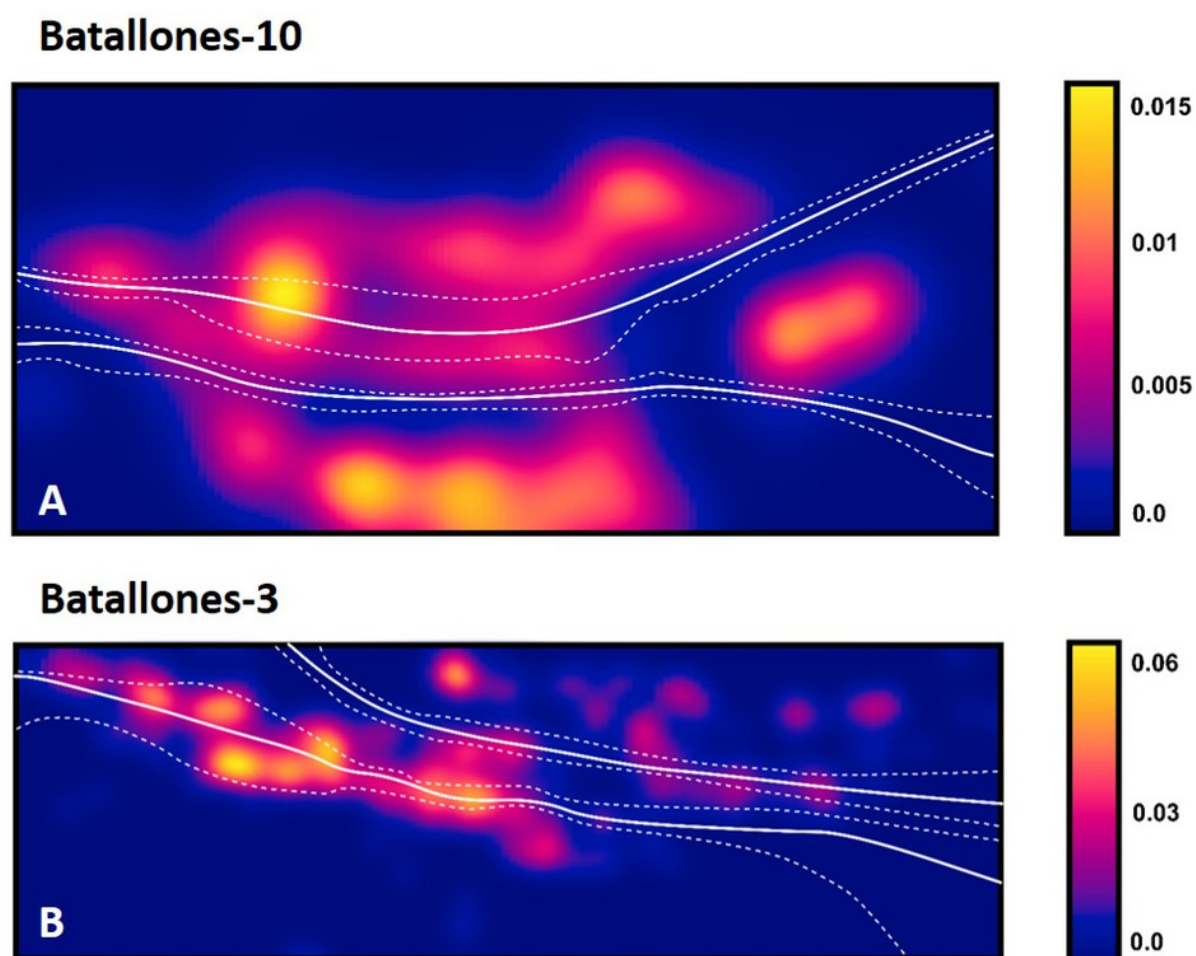



Figure 7

Fine-tuned fossiliferous level models.

(A) Batallones-3 left slice. (B) Batallones-3 left slice associated 2-metre-wide perpendicular slice. (C) Batallones-3 right slice. (D) Batallones-3 right slice associated 2-metre-wide perpendicular slice.

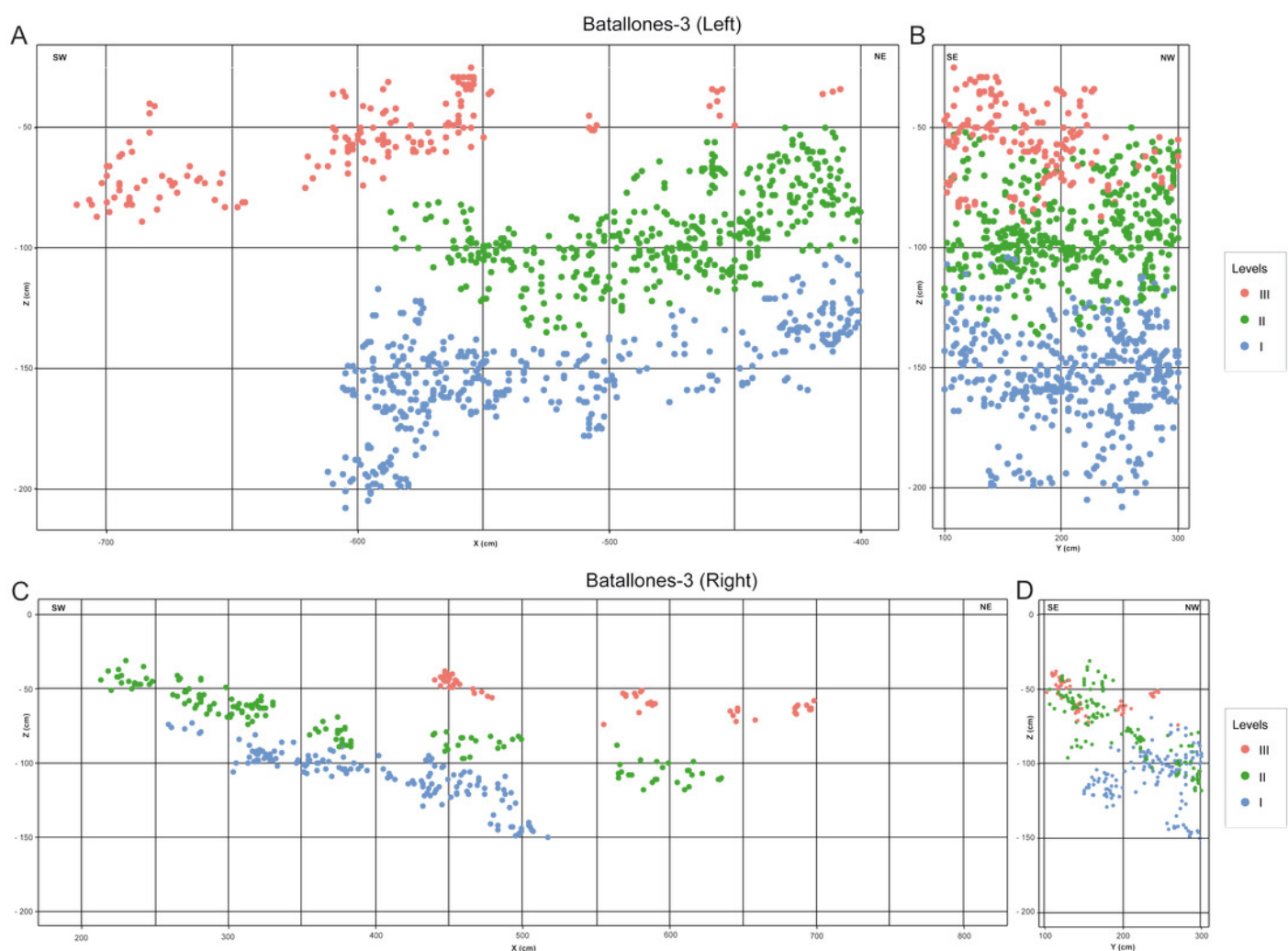



Figure 8

Fine-tuned fossiliferous level models.

(A) Batallones-10 x axis slice. (B) Batallones-10 x axis slice associated 2-metre-wide perpendicular slice. (C) Batallones-10 y axis slice. (D) Batallones-10 y axis slice associated 2-metre-wide perpendicular slice.

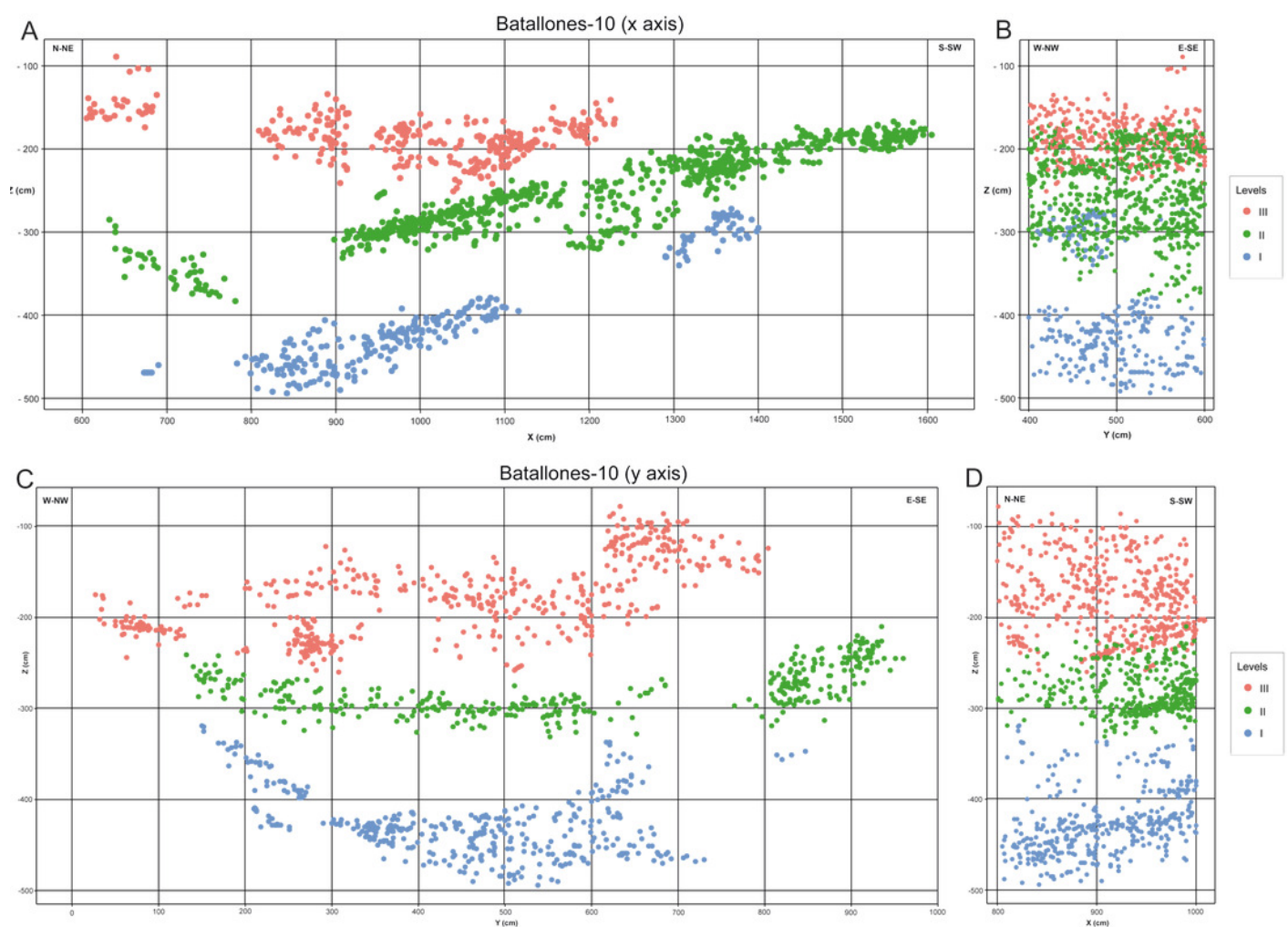
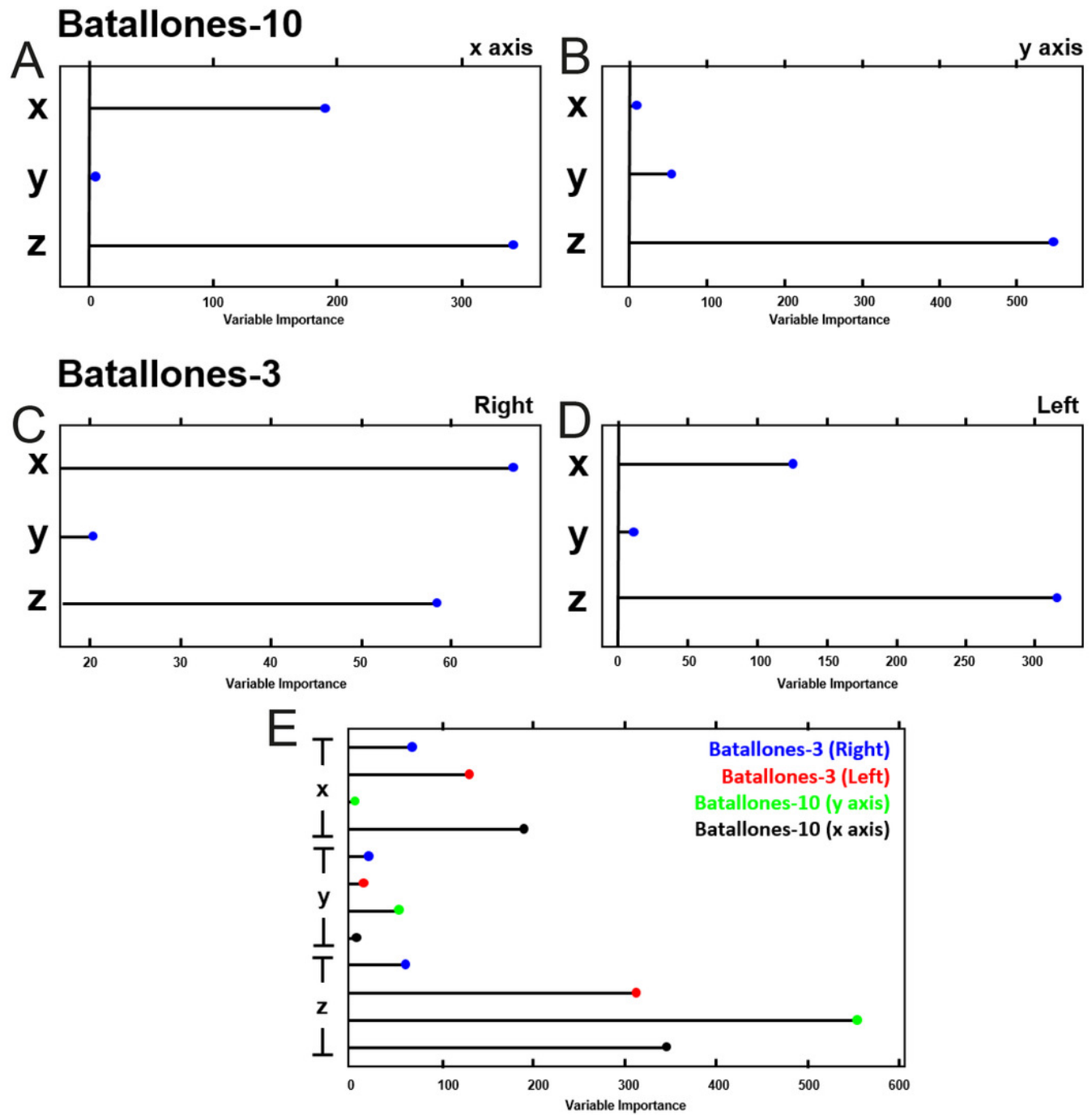



Figure 9

Variable Importance Plots for the RF model.

Each panel represents the weight each variable has on the decision making capabilities of the RF model, indicating the importance in some cases of the use of 3D data for stratigraphic model definition. (A) Batallones-10 x slice. (B) Batallones-10 y slice. (C) Batallones-3 right slice. (D) Batallones-3 left slice. (E) Comparison of all slices.

Batallones-10

Batallones-3

Evaluation metrics of each ML algorithm on test sets, as derived from confusion matrices in Supplementary File 1. Confidence Intervals (CI) are established for the accuracy metric.

MSE = Mean squared error.

	Random Forest				**Support Vector Machine**			
	Batallones-10		Batallones-3		Batallones-10		Batallones-3	
	X axis	Y axis	Left	Right	X axis	Y axis	Left	Right
Kappa	0.99	0.99	0.97	0.97	1.00	1.00	0.99	1.00
Lower CI*	0.98	0.98	0.93	0.93	0.99	0.99	0.96	0.96
Accuracy	0.99	0.99	0.98	0.98	1.00	1.00	1.00	1.00
Upper CI*	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Sensitivity	0.99	1.00	0.98	0.98	1.00	1.00	1.00	1.00
Specificity	1.00	1.00	0.99	0.99	1.00	1.00	1.00	1.00
MSE	7.73E-06	3.39E-05	1.15E-03	1.15E-03	7.76E-05	6.84E-04	6.07E-05	9.37E-04

Results obtained from ML algorithm for the classification of points considered indeterminable after both MT and EitL intervention.

MSE values represent the confidence of classifications.

		Batallones-10 X axis	Batallones-10 Y axis	Batallones-3 Left	Batallones-3 Right
SVM	MSE	3.45E-03	1.21E-02	NA	7.11E-03
	Determinable	113	15	0	45
	Indeterminable	28	10	6	11
RF	MSE	4.01E-03	6.17E-05	3.18E-02	8.88E-03
	Determinable	106	23	4	39
	Indeterminable	35	2	2	17