Geological Magazine

www.cambridge.org/geo

Multispectral satellite imaging improves detection of large individual fossils

Elena Ghezzo[1], Matteo Massironi[2] and Edward B. Davis[1,3]

[1]Department of Environmental Sciences, Informatics and Statistics, Ca'Foscari University of Venice, Venezia, Via Torino 155, 30170, Italy; [2]Department of Geosciences, University of Padua, Padova, Italy and [3]Department of Earth Sciences, University of Oregon, Eugene, USA

Original Article

Cite this article: Ghezzo E, Massironi M, and Davis EB (2023) Multispectral satellite imaging improves detection of large individual fossils. *Geological Magazine* **160**: 535–544. https://doi.org/10.1017/S001675682200108X

Received: 12 April 2022
Revised: 28 September 2022
Accepted: 30 September 2022
First published online: 28 November 2022

Keywords:
remote sensing; cluster classification; Petrified Forest National Park; Triassic fossil logs; mapping fossils

Author for correspondence:
Elena Ghezzo,
Email: elena.ghezzo@unive.it

Abstract

Palaeontological field surveys in remote regions are a challenge, because of uncertainty in finding new specimens, high transportation costs, risks for the crew and a long time commitment. The effort can be facilitated by using high-resolution satellite imagery. Here we present a new opportunity to investigate remote fossil localities in detail, mapping the optical signature of individual fossils. We explain a practical workflow for detecting fossils using remote-sensing platforms and cluster algorithms. We tested the method within the Petrified Forest National Park, where fossil logs are sparse in a large area with mixed lithologies. We ran both unsupervised and supervised classifications, obtaining the best estimations for the presence of fossil logs using the likelihood and spectral angle mapper algorithms. We recognized general constraints and described logical and physical pros and cons of each estimated map. We also explained how the outcomes should be critically evaluated with consistent accuracy tests. Instead of searching for fossiliferous outcrops, our method targets single fossil specimens (or highly condensed accumulations), obtaining a significant increase in potential efficiency and effectiveness of field surveys. When repeatedly applied to the same region over time, it could also be useful for monitoring palaeontological heritage localities. Most importantly, the method here described is feasible, easily applicable to both fossil logs and bones, and represents a step towards standard best practices for applying remote sensing in the palaeontological field.

1. Introduction

The use of GIS and remote-sensing tools are exponentially increasing the effectiveness of palaeontological surveys, diminishing costly field time and uncertain travel in remote areas. Fifteen years of analyses, focused on the recognition of potential fossiliferous outcrops, have all anticipated the capacity to search for single fossils in remote-sensing data. That opportunity has arrived. Here, we show it is possible to detect individual fossils exposed on the surface using multispectral satellite imagery of the Petrified Forest National Park (Arizona, USA).

An early structured preliminary analysis of the field in palaeontology was proposed by Oheim (2007) for the Two Medicine Formation in Montana (Campanian Stage, Upper Cretaceous). Her model was refined with successive field trips, giving a rudimentary guide for the best understanding of the region. Oheim (2007) had access to several geographic data layers, i.e. geological maps, land cover dataset, elevation dataset and roads, because her study area is located in the United States, which is well covered by different kinds of remote-sensing data. A first attempt to work around a regional shortage of spatial data was addressed by Malakhov *et al.* (2009). They approached the Lower Syrdarya Uplift in southern Kazakhstan (lower Bostobe Formation, Upper Cretaceous) using satellite images from Landsat 7 ETM+. They ran unsupervised (ISODATA) and supervised (spectral angle mapper, SAM) cluster classifications of the images, demonstrating that classified images could be very useful for determining likely areas for new fossil localities prior to prospecting on the ground.

Anemone *et al.* (2011*b*) and Emerson & Anemone (2012) analysed the Paleocene–Eocene deposits of the Great Divide Basin of southwestern Wyoming using the same medium-resolution satellite images (Landsat 7 ETM+). Even though their approach was limited by the data they used, they can be considered pioneers in the field of satellite images applied to palaeontological remote surveys. They attempted to identify productive fossiliferous deposits by creating an artificial neural network (ANN) with the intent to identify differences in pixel reflectance. The model correctly classified most of the potential fossil localities, but it was affected by overprediction (Emerson *et al.* 2015), proving that machine learning tools can be used for a coarse refining of a region to be examined before on-the-ground surveys. B. Bommersbach (unpub. Master's thesis, Western Michigan Univ., 2014) and Emerson *et al.* (2015) tried to improve the previous results using an object-oriented classification (GEOBIA) and QuickBird and Landsat 8 OLI satellite images, obtaining a moderate success; nevertheless, they admitted this method has limits and potential issues related to, among others, the effective possibility

© The Author(s), 2022. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

of distinguishing the spectral signature of productive outcrops from other objects on the scene, the exposure and slope of the outcrops, and biased verification on the ground. This study was also the first to test commercial satellite images for palaeontological work, with an outstanding resolution of 0.61 m per pixel in panchromatic in QuickBird (released by the DigitalGlobe corporation).

Conroy *et al.* (2012) applied maximum likelihood classification to the Eocene deposits of the Uinta Basin in Utah, creating a model for potential new fossil localities. They confirmed the map reliability with post-2005 field surveys. Thus, in this case the trained (supervised) classification had a positive outcome, and it can be potentially tested in other fossiliferous contexts. Two years later, Conroy (2014) investigated the use of unsupervised classification in the same region with unknown *a priori* knowledge of fossil occurrences. The results were promising because most of the tested palaeontological localities fell into the regions classified as fossiliferous, and the simple classification created a predictive map that refined the potential areas for future surveys to 5 % of the entire region. At the same time, the ISODATA classification did not reach 100 % inclusion of previously known fossil localities, and some of the outcrops were misclassified. Consequently, the ISODATA classification was demonstrated as not conclusive, and it should be implemented only with other cluster analyses.

Burk (2014) studied the Lower Cretaceous Cedar Mountain Formation in Utah through Landsat 8 OLI/TIRS images. He concluded that the obtained classified map is not accurate enough to be considered alone for palaeontological surveys. Physical and environmental information, such as slope degree and aspect, should be applied to refine the model. Wills *et al.* (2018) matched georeferenced information extracted from heterogeneous sources, including Landsat 8 OLI, to infer locality distribution in the Upper Triassic – Lower Jurassic Elliott Formation in South Africa. They ran advanced statistics to objectively consider the results, stressing again the importance of creating spatial predictive models before planning palaeontological surveys. Finally, a recent study by d'Oliveira Coelho *et al.* (2021) applied unsupervised (*k*-means) classification on a portion of the East African Rift System (Urema Rift, Mozambique) reaching 84.6 % accuracy in fossil site detection. In recent years, researchers have begun pointing their attention to extraterrestrial ichnofossil signs on other planets, such as Mars (McKeown *et al.* 2009, 2011; Noe Dobrea *et al.* 2010, 2017; Baucon *et al.* 2021), giving a new motivation to using satellite images and spatial analyses to detect life.

In sum, most of the research activities described above obtained coarse classifications of large portions of fossiliferous regions, progressively testing more efficient cluster classification algorithms on medium- and high-resolution satellite images (Landsat 7 ETM+, Landsat 8, QuickBird) combined with digital elevation models (DEM) and land cover data, i.e. geological map constraints, slope and vegetation indexes (normalized difference vegetation index, NDVI), the presence of streets and artefacts, etc. The goal of much of this work was to map outcrops with a high potential for fossil presence to limit funding and time, as well as find ways to improve safety for the crew and the effectiveness of field logistics. All those analyses anticipated the improvement of satellite spatial and spectral resolution.

In this paper, we deliver on that expectation as we describe the rationale and workflow for the detection of individual fossils exposed on the surface using multispectral satellite imagery, testing the resulting maps on the Crystal Forest walk trail, as part of the Petrified Forest National Park (Arizona, USA). This workflow is generally applicable, flexible in terms of data sources, as it is applicable to both multispectral and hyperspectral imagery, uses a scalar approach in progressive spatial resolution and takes advantage of satellites as well as atmospheric aircraft (drones or airplanes). More importantly, it is adaptable to diverse kinds of fossil localities with remains of exposed fossils, from fossil logs to marine mammals to dinosaurs.

2. Logical workflow

Searching for fossils in remote-sensing data requires careful consideration of three factors: the mineral composition of the fossils, the physical properties of the region to be surveyed and the availability of spectral imagery. Each factor is intimately related to the other two, and one factor affects the characteristics and outcomes of the other two. First, because we are considering only optical, multispectral imagery, fossils must be exposed on the surface. The size of the exposed part (or the exposed association, in the case of closely packed fossils) should cover at least one pixel of the multispectral imagery, or be larger and occupy multiple continuous pixels. Smaller objects can create issues related to pixel mixing. The pixel size depends on the aircraft and the camera used for recording. Because drones, airplanes and satellites fly at increasing altitudes, an increasing portion of land surface is available for each vehicle, and the ground resolution decreases accordingly. Drones can have a resolution of less than a centimetre, while the most common satellites have a resolution of tens of metres. For palaeontological surveys, the minimal required resolution cannot be larger than 2 m, for a covered area of 4 m^2. In this paper, we restricted our analysis to this extreme limit using WorldView-2 (WV2) imagery to stress the minimal potential of the method. A higher resolution should only improve the outcomes explained here.

For this workflow to be successful, the way fossils respond to solar radiation needs to diverge enough from the reflected light of the surrounding area to be detected in the imagery as a different object in the scene. The greater the divergence in spectral reflectance between fossils and the sedimentary matrix in which they are embedded, the easier it will be to distinguish these fossils from the background. The number of bands in the multispectral imagery can be increased or reduced in accordance with this divergence. Imagery from areas with low slope values and lacking vegetation cover are preferred: the former prevent problems related to shadowing and variance of reflected light, and the latter ensure that fossils are visible on the ground. Finally, radiation scattering in the atmosphere creates random variation in the quantity of light that reaches the camera. The scattering is caused by floating particles of dust and water, so imagery recorded on wind-free days with minimal cloud cover are preferred. In the case of deserts, wind usually blows in the afternoon, so morning imagery can minimize the effects of wind, even though an early morning image can be affected by the low angle of the Sun over the horizon.

In their analysis of satellite imagery in Kazakhstan, Malakhov *et al.* (2009) introduced three main reasons to utilize satellite images in palaeontological surveys: (1) the investigated area is far distant from any kind of facilities, (2) fossil remains were previously reported, and (3) fossils are concentrated in small areas scattered over the field. Because our attempt entails finding single fossils instead of outcrops, it is also important to pay attention to the optical properties of the fossils and matrix: there must be a great enough contrast to be able to detect fossils in the imagery.

Once all those points are clarified, and the best potential imagery is chosen, the workflow can focus on the cluster classification of the scene. In the procedure synthesized in Figure 1, we

https://doi.org/10.1017/S001675682200108X Published online by Cambridge University Press

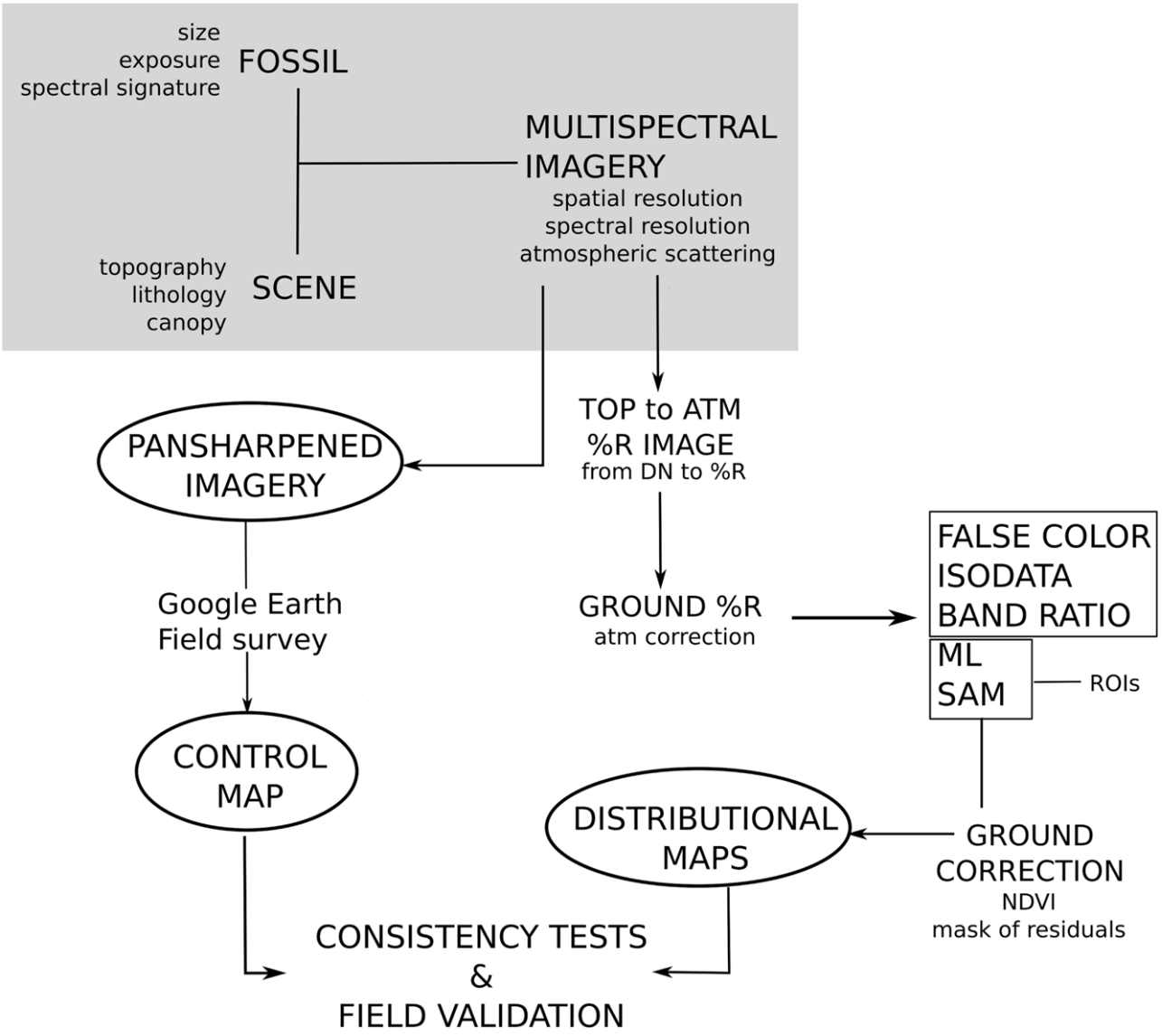


Fig. 1. Logic workflow of the analyses explained in this paper. The area in grey indicates steps to consider before processing the data. Ovals indicate imagery returned as outputs. Uppercase notes not in ovals indicate algorithms and statistics applied to the imagery. Unsupervised and supervised classifications are grouped in white rectangles. ATM – atmosphere; DN – digital numbers; %R – reflectance; ML – maximum likelihood; NDVI – normalized difference vegetation index; ROIs – regions of interest; SAM – spectral angle mapper.

suggest proceeding through these steps: The WV2 multispectral imagery should be pansharpened and compared to other available sources (exemplified in the schema as Google Earth and a field survey). The pansharpened imagery loses numerical homogeneity throughout the scene, so it should not be used for further spectral analyses. The original multispectral imagery must be corrected and converted to the ground reflectance (details in online Supplementary Material S1). Because the way images are shown on computer screens is the result of matching among three bands (usually red, blue and green), changing the spectral bands to be shown returns information about what is most reflective in a particular band in comparison to the other objects in the scene. The same result can be obtained with the ratio between divergent bands, producing a grey scale map of values corresponding to the increasing signal for one band versus the other. This first simple comparison between bands cannot be considered exhaustive for fossil detection, but it is a good starting point to improve knowledge of the region, and diagnose the spectral characteristics of the targeted fossils in comparison to the background sedimentary rock. After the user becomes confident with the features in the scene, algorithms for cluster classification can be applied. If the fossils are well recognizable, an ISODATA analysis should reach the goal. In the ISODATA unsupervised classification, the algorithm works independently from any predefined clusterization made by the user, and it can be applied where there is a very poor knowledge of the objects in the scene. It extracts evenly distributed classes and iteratively includes the other pixels in the scene in one of the classes according to their minimum distance. The process is iterated multiple times until a threshold or a maximum number of repetitions is reached (Tou & Gonzalez, 1974). Unfortunately, because fossils are frequently very similar to the matrix in terms of their colour, the ISODATA is often not able to correctly classify each of them separately. Nevertheless, it can narrow the area where fossils are potentially exposed, highlighting differences in grain size, shape and texture, and the result can be used to mask successive analyses.

https://doi.org/10.1017/S001675682200108X Published online by Cambridge University Press

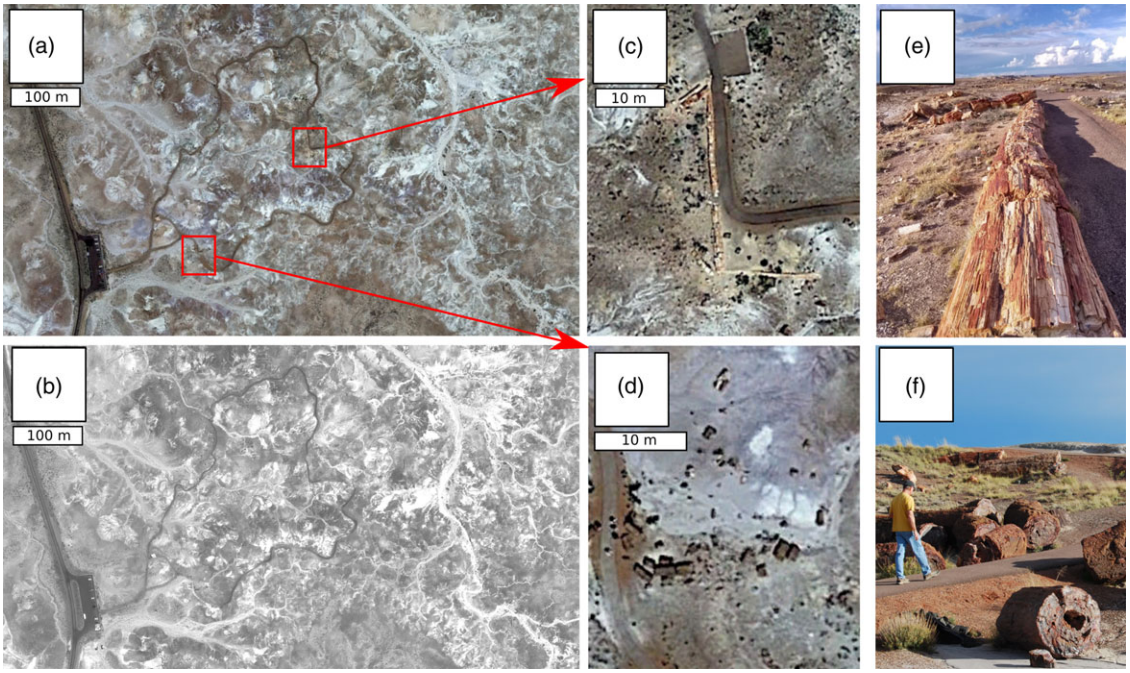


Fig. 2. (Colour online) (a) The Crystal Forest recorded by Google Earth on 25 May 2019 (Map data: © 2019 Google) and (b) the same scene in the WV2 image in panchromatic collected on 21 August 2014 with 50 cm of pixel resolution. Squared regions are magnified in (c) and (d), corresponding to type-1 and type-2 fossil logs (photo from the ground in (e) and (f), respectively).

In cases where the ISODATA does not work, we propose two other approaches to classifying the pixels: maximum likelihood (ML) classification and spectral angle mapper (SAM). These two are supervised approaches; they force the pixels to be classified according to spectral differences derived from objects recognized in the scene by the user. That is, ML and SAM are trained by the user on known specimens, and the resulting rule set can be applied to unknown imagery to identify potential fossils. The two algorithms can be affected by subjectivity because the recognition of training pixels is made by a person. The previous false colour and band ratio analyses have been proposed to minimize this effect. Once the maps of potential fossil distribution have been obtained, and residuals (shadows, anthropic artefacts, vegetation, etc.) are removed, they can be used to plan more efficient, safer field work, where researchers can compare the true presence of fossils with the outcomes of the estimated maps for future improvement of remote palaeontological mapping.

The method here described is the basic approach for single fossil detection in remote regions using multispectral imagery. It would certainly be enormously improved if applied to higher resolution and hyperspectral images. Unfortunately, satellites with hyperspectral cameras are still not available with a ground resolution of 1 m or less, and unmanned aircraft with a hyperspectral camera as a payload are not a practical solution, because they have prohibitive costs and the data must be recorded *in loco*, limiting the spatial extent of analyses and the overall advantage in pre-planning field work. As a consequence, the proposed method is the only one currently available to ameliorate costs and constraints of field work, improving effectiveness. Another technology that is coming to prominence for surface analysis is the active data collector available as synthetic aperture radar (SAR). TerraSAR-X, COSMO-SkyMed and RADARSAT-2 satellites provide radar imagery of the Earth's surface up to 1 m resolution, so they are comparable to the ground resolution of the WorldView fleet. Nevertheless, because radar and optical imagery require different approaches to analyses and corrections of the scene, they have not been considered in this paper.

Finally, advanced algorithms based on machine learning (i.e. neural network analysis and object-oriented analysis; Anemone *et al.* 2011*a*) are available to improve the outcome of fossil detection. This procedure requires many images as a training set, and this is not typically the case for those fossil localities where ground characteristics vary a lot throughout the scene, and where there is a very limited knowledge of fossils exposed on the ground. Consequently, the machine learning approach is still restricted as a promising tool in this topic.

3. Test case: mapping fossil logs at Petrified Forest National Park

We tested the proposed method at the Crystal Forest (Fig. 2) within the Petrified Forest National Park, using a selected high-resolution multispectral image from the WV2 satellite (released by Digital Globe Inc.).

The imagery has a ground resolution of 50 cm in the panchromatic and 2 m in the 8-bands around visible light, spanning from ultraviolet (UV) light to the near infrared (NIR) (metadata in the details in online Supplementary Material S1).

The Petrified Forest National Park has one of the most outstanding geological records of the Upper Triassic. The stratigraphy is characterized by alternating alluvial and lacustrine deposits which are visible over the whole *c.* 500 km^2 of the national park (Loughney *et al.* 2011). The sediments, accumulated in different redox depositional environments and climatic conditions, recorded the environmental changes as differences in colours of the exposed stratigraphic units, visible from satellite images.

We selected an area within the Crystal Forest, in the southern portion of the national park. Geologically, it is part of the Upper Triassic Chinle Formation, dated to the boundary between the Adamanian and Revueltian vertebrate biozones (219–213 Ma;

https://doi.org/10.1017/S001675682200108X Published online by Cambridge University Press

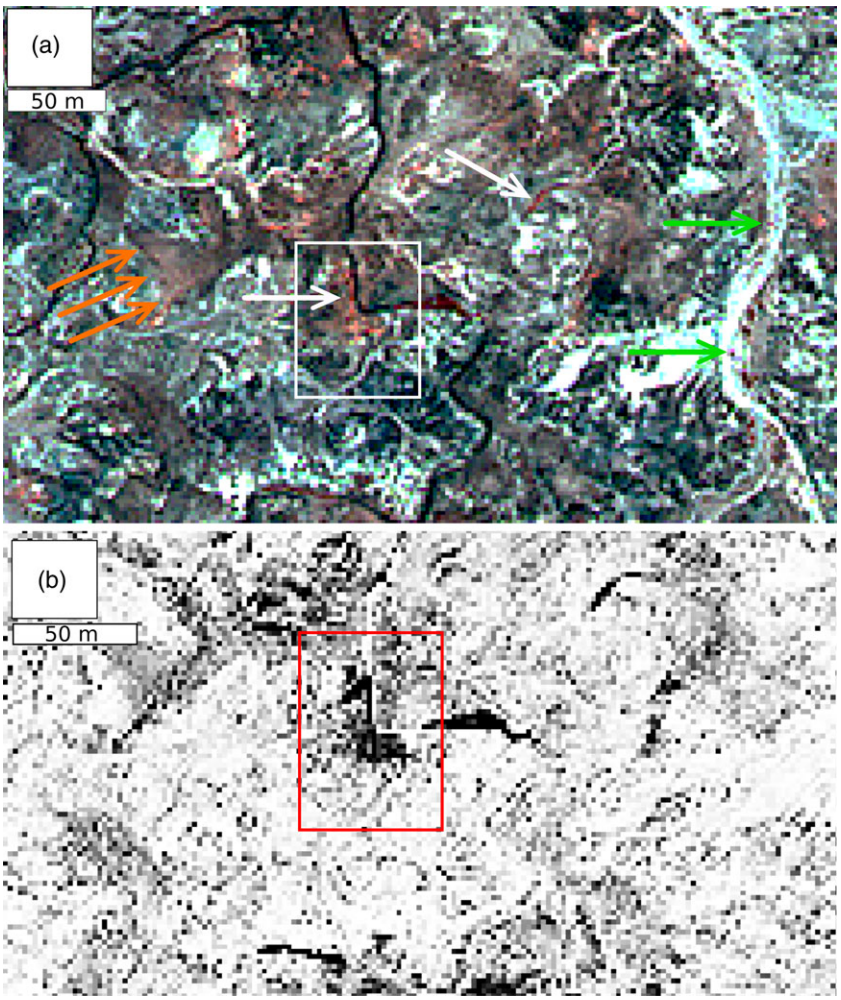


Fig. 3. (Colour online) (a) Multispectral imagery of the Crystal Forest, zoomed in on the region with large fossil logs. The false colour image 8-2-1 has been selected to emphasize the high values in the NIR (in bright red, white arrows) related to fossil presence in comparison to the rest of the scene. Arrows point to fossil logs (in white), vegetation (in green) and soil (in orange). (b) Blue/Red Edge band ratio (minimum values in black, maximum values in white; see online Supplementary Material S1 for details). Squares correspond to the region indicated in Figure 4.

Martz *et al.* 2012). During Late Triassic time, this locality was part of western equatorial Pangaea, palaeogeographically located at *c.* 10° N of the equator and 1500–2000 m above sea level (Steiner & Lucas, 2000; Golonka, 2007; Nordt *et al.* 2015). Palaeosols indicate that the existing ecozones were warm temperate or subtropical, never firmly tropical (Nordt *et al.* 2015). Conifers are represented by the genus *Araucarioxylon* (Ash & Creber, 2000) and are represented by flat-lying fossil wood logs deposited in channels and overbanks (Jiang *et al.* 2018). The environment was warm and humid, with a progressive trend towards a semi-humid and arid climate (Nordt *et al.* 2015). Seasonal monsoon and aridity, occasional flooding and rapid sedimentation facilitated transport and silicification of plants (A. S. Gottesfeld, unpub. M.S. thesis, Univ. Arizona, 1971; Sigleo, 1979; Nordt *et al.* 2015; Jiang *et al.* 2018).

Satellite images had been previously applied to the national park, but none of them attempted to map the position of single fossil logs. Mickus & Johnson (2001) mapped sedimentary and volcanic units within the park and the surrounding area, while Jimenez (2011) tested the method described by Malakhov *et al.* (2009) on an extensive surface within the national park, finding a good match between his classified Landsat 7 ETM+ image and the USGS geological map of the region.

We recognized two main kinds of preserved fossil logs at the Crystal Forest, according to Jiang *et al.* (2018). Type-1 fossil logs consist of mostly reddish logs, preserved as long specimens lying on the surface (Fig. 2c, e). Type-2 fossil logs are brownish, and frequently broken in multiple slices, partially dismembered and dislocated on the ground (Fig. 2d, f). Type-1 and type-2 fossil logs show opposite characteristics in terms of preservation of cellular structures (cell walls) and Fe content, with the former showing a low degree of preservation and higher mineral density (Mustoe & Acosta, 2016; Jiang *et al.* 2018).

3.a. Results

The scalar analysis of the Crystal Forest started with the evaluation of the scene and the selected satellite imagery (see metadata in the online Supplementary Material S1). Comparison between the extracted image from Google Earth Pro and both the reflectance (%R) and the pansharpened imagery shows, as expected, that the best superficial details are visible on Google Earth. Nonetheless, the way Google Earth treats its data creates aberrant pixel values, preventing further analyses of the scene using such bands in the images.

The first consideration of the Crystal Forest through multispectral imagery as continuous %R values already shows that near infrared bands (NIR1 and NIR2) are intense for fossil logs with respect to the surrounding areas and in relation to UV and blue bands (emphasized in Fig. 3a). This result has been confirmed by the bands ratio analyses, i.e. the mathematical ratio between values in each aligned pixel of two different bands, where the

https://doi.org/10.1017/S001675682200108X Published online by Cambridge University Press

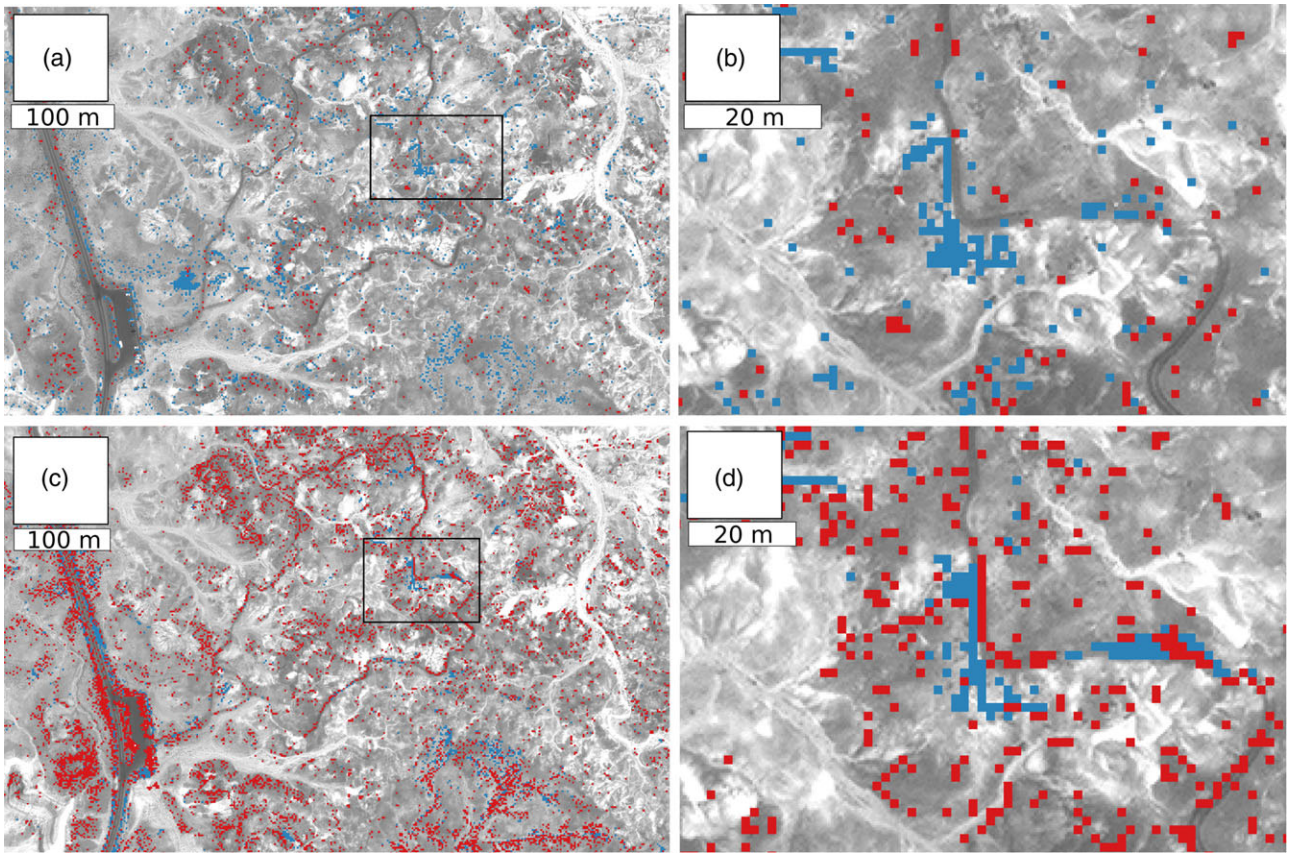


Fig. 4. (Colour online) (a, c) ML and SAM classifications of the Crystal Forest, respectively. (b, d) Magnification of main target area (outlined by rectangles in (a) and (c)). Clusters correspond to the two types of fossil logs we considered as end-members (type-1 in light blue and type-2 in red).

long-wave bands are both less absorbed (in this case, we used the Red/NIR1 bands and the Blue/RedEdge bands; see online Supplementary Material S1 and zoomed in Fig. 3b). In the latter, differences between extreme bands allow us to distinguish fossil logs, streambeds and lowland regions. Unfortunately, high values of NIR are also returned by modern vegetation, recognizable as scattered pixels along the streambed edges, and by scattered vegetation and un-reworked soil (Fig. 3a, b). Therefore, successive steps in the analyses (consisting of clustering and filtering) are needed to attempt the extraction of signals related specifically to fossil logs.

Both the band ratio and the unsupervised classification (ISODATA) could not recognize fossil features on the ground at the Crystal Forest (results in the online Supplementary Material S1). The spectral signatures were mixed and too similar throughout the scene to be discerned by an unsupervised analysis. The resulting images show a confetti pattern, a signal of a chaotic distribution of values.

Therefore, we hypothesized that a better response should be obtained when the algorithm was trained and directed, with previous knowledge of the objects in the scene. We selected five regions of interest (ROIs), two of them corresponding to potential fossil logs (type-1 and type-2). The other selected reference regions consist of human artefacts, stream beds and stable un-reworked soil.

With a preliminary evaluation, the ML classification returned the most plausible map (Fig. 4a) while the SAM was affected by the affinity among spectral signatures of the matrix and fossils at least for the type-2 logs (Fig. 4c). It resulted in a very close angle between the vectors of each pixel and the end-members, grouping too many of them together.

Both supervised classifications performed well on the recognition of the type-1 petrified logs (our main target), even though the mapped features partially do not match to each other (Fig. 4b, d).

3.b. Accuracy assessment

3.b.1. Confusion matrix

To assess the success of our classification of the Crystal Forest, we compared our estimation maps with both observations in the field and a digitized map of the distribution of fossil logs, the latter obtained manually using Google Earth Pro (details in online Supplementary Material S1).

It is worth stressing that the manual mapping procedure was highly time-consuming (>3 work days), and it was affected by under detection and subjective misclassification, so it was checked multiple times to correct the polygons to represent only real fossil logs. The multispectral procedure, on the other hand, allowed us to process the spatial data independently from human judgement, and starting from deeper information about the ground (multispectral layers). It was timesaving, but the saved time was at the cost of a lower resolution.

A confusion matrix was created comparing the rasterized control map with the estimation maps of fossil log distribution obtained with ML and SAM classification (Fig. 5, see details in online Supplementary Material S2). The matrix covered 156 375 pixels for a total of 0.6255 km^2 (close to 117 American football

https://doi.org/10.1017/S001675682200108X Published online by Cambridge University Press

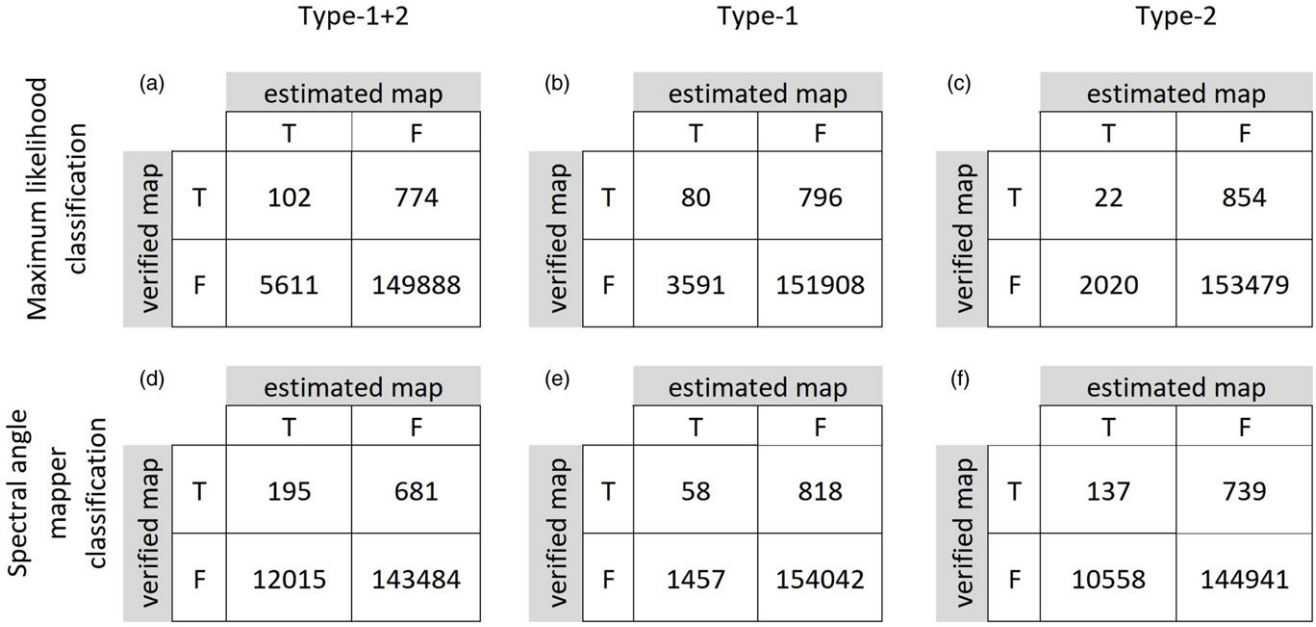


Fig. 5. Tables in the top row are the confusion matrices of the maximum likelihood classification. The bottom row shows the confusion matrices of the spectral angle classification, for (a, d) type-1+2, (b, e) type-1 and (c, f) type-2 fossil classes, respectively (see online Supplementary Material S2 for more details).

fields, or about 60 soccer fields). A total of 876 pixels were true fossil logs in the control map. Of them, 11.64 % were recognized as true fossils by the ML map for type-1 and type-2 fossil logs together (9.13 % and 2.51 % for each type, respectively) corresponding to only 1.79 % of the total number of positive potential fossils pixels recognized with this method. The percentage of truly detected fossils goes up to 22.26 % when the SAM classification is considered (6.62 % and 15.64 % for each type, respectively), but also in this case it represents only 1.60 % of the total number of potential fossil pixels detected. When we consider the two types of fossil logs separately, the ML classification performed better in the recognition of the type-1 fossil logs, matching 1/10 of the true fossil logs in the control map (9.13 %), while SAM worked better in the recognition of the type-2, with 1/6 of the matching pixels (15.64 %).

Both classifications had a high number of false positive pixels, overpredicting the presence of fossils on the ground, but they recognized close to 100 % of the true negative pixels, performing well in eliminating those pixels not suitable for fossil detection.

The power of our classifiers was tested using canonical metrics (Fielding & Bell, 1997) and the *z*-test (Davis *et al.* 2014). All classifications returned an overall diagnostic power close to 100 % (0.99) indicating a high degree of confidence for checking areas without fossil logs, and above 0.92 for the correct classification rate, a signal that the algorithms correctly classified regions matching the control map.

Observed matching between the control map and the estimated maps ('o' in online Supplementary Material S2) resulted always in a sum lower than the randomly expected one ('e' in online Supplementary Material S2), giving a negative value and a non-significant result for the *z*-test. This result is a consequence of the overestimated presence of fossil logs (false positives). As a consequence, all classifications failed to reduce the number of false positives (that lead to overestimation).

Sensitivity (the ratio of the number of true positives from the classification over the number of actual positives from the known dataset; Fielding & Bell, 1997) and specificity (the ratio of the number of true negatives from the classification over the actual negatives from the known dataset; Fielding & Bell, 1997) provided information about the correct classification of true and false fossil logs on the scene. Because of the high number of pixels with no fossil logs on the scene in the control map, specificity can be aligned to the accuracy of the result, recognizing close to 100 % of all regions with no fossils.

Sensitivity shows a minimum value of 0.03 for the type-2 ML classification, and the best value of 0.22 for the SAM algorithm applied to both type-1 and 2 fossil logs. If we assume that a person walking across the scene should return a 0.0056 probability of finding true fossil logs on the surface (derived from the ratio of pixels corresponding to true fossils versus the count of pixels of the whole surface: our null hypothesis), the SAM classification returned a map representing a large improvement of our chances of finding fossil logs, even when we estimate a 95 % confidence interval (0.19–0.24; calculated using the webtool https://www.medcalc.org/calc/diagnostic_test.php).

The classification is still missing a large portion of fossil logs exposed on the ground (it detected less than one third of the true fossil logs) but it improves by 38 times the efficiency of field work compared to uninformed searching.

3.b.2. Polygon matching

To compare not just areas but the actual number of fossil logs detected in the control map to our estimated maps of potential fossil distribution (type-1 and type-2 together), we converted the two estimated maps to a vector layer. Also, because it was assumed that fossil logs are large enough to be visible in the proximity of any walking path chosen by a hypothetical field crew, we added a 2 m buffer zone to the positive pixels to create these polygons.

The outcome from the SAM classification worked better than the ML classification, recognizing one third more fossil logs than the latter. In fact, 135 out of 325 polygons of the control map were touched (intersected) by the SAM map, corresponding to 41.5 % of

https://doi.org/10.1017/S001675682200108X Published online by Cambridge University Press

the real fossil logs detected though Google Earth, whereas the percentage dropped to 27.1 % (88 polygons) for the ML map. The buffer increased the percentages to 64.3 % and 50.8 % of the detected polygons, respectively.

The matching ratio enormously increased when we considered only polygons in the Google Earth-based map bigger than 4 m^2, corresponding to the minimum spatial resolution of a pixel in the original multispectral imagery. The output returned a percentage of 54.5 % and 83.6 % of matching features for the ML and SAM classifications, respectively. It reached 63.6 % and up to 92.7 % of positive detections in the buffered ML and SAM maps, respectively.

In the count of matching objects, we used a control map with just positive polygons. As a consequence, we could calculate only a few metrics out of the confusion matrix (Fielding & Bell, 1997). As expected, the confusion matrices, populated using the number of polygons of the control map versus each estimated map, retraced the general outcomes, with an increasing sensitivity from the ML to the buffered SAM, and a lowering of the rate of false negatives, with close to 100 % estimation of true negatives.

It can be argued that the positive result in this comparison is mostly a consequence of overclassified positive potential fossils. Nevertheless, the ML and SAM classifications matched the control map 44 % and 67.7 % of the time, i.e. 357.8 and 540.3 m^2, respectively. In comparison, the null hypothesis of a random distribution of positive fossil predictions should be able to detect 0.128 % of the true fossils, corresponding to the ratio between the area of true fossil logs in the control map versus the total area of the Crystal Forest, i.e. 1.01 m^2 of true positives by random chance. Both ML and SAM enormously exceeded this limit, as they classified as positive 3.6 % and 7.9 % of the total surface, respectively.

We acknowledge that the positive detection of polygons <4 m^2 was both a result of a low degree of pixel mixing for those fossil logs close to the lowest pixel resolution of our original imagery and random overestimation of true fossils. In any case, it did not falsify our main results.

4. Discussion

In this paper, we propose a standard workflow that should be adopted when satellite images are analysed for any kind of palaeontological pre-survey planning. We recognized three critical constraints: (1) the size, exposure and spectral signature divergence of the fossils we want to target, (2) the spatial and spectral resolution of the satellite imagery, and atmospheric scattering, and (3) the topography, lithology and vegetation canopy of the study scene (Fig. 6). When each constraint is correctly accommodated, the method can be applied to any kind of field work to map the potential presence of new fossils, from dinosaurs to large mammals to fossil logs. As satellite technology improves, resolution will increase and allow the identification of progressively smaller specimens exposed at the surface.

As explained in the introduction, the idea of leveraging satellite images to enhance palaeontological surveys is not new, and this is particularly true for the Petrified Forest National Park (Mickus & Johnson, 2001; Jiménez, 2011). We tested an improved approach, focusing our attention on the recognition of single fossil logs exposed on the surface. The best outcomes came from the ML classification and SAM on selected 8-band imagery of the Crystal Forest area of the Petrified Forest National Park. Our results show that the method is feasible and promising: a planned survey can focus on an area that is 27 or 12 times smaller than the entire

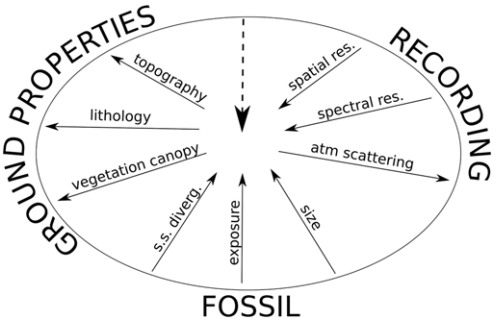


Fig. 6. Schema of constraints, grouped between evaluation of fossil properties, ground features and recording potentials. Arrows point to the direction of each factor: the more the complexity/quality/quantity of the factor is in itself, the more that factor pushes towards the centre or the outside of the circle. Quality of the output improves from the edge to the centre (dashed arrow), being the point with the best potential for single fossil detection (s.s. diverg. – spectral signature divergence; atm – atmospheric; res. – resolution).

Crystal Forest, to find one eighth or one quarter of the true fossil logs, using the ML and SAM cluster classifications, respectively. Our results show that unsupervised classification is not sufficient to discern fossils from matrix at the Crystal Forest, reconfirming what was already stated by Conroy (2014); the game changed when we recognized target objects in the scene, pushing the algorithm to cluster pixels together according to our identifications. Consequently, we can confirm that any preliminary knowledge of the ground (such as slope, lithology, vegetation coverage, etc.) should improve the performance of such cluster classifications, and that the multispectral data can be replaced by a less informative data source, such as Google Earth or geological maps for the outcrops, to train the algorithm and still obtain positive results.

Both cluster classifications were affected by overprediction of fossil presence (false positives) because the signal of the fossil logs is similar to residual vegetation and stable soil. Scattered vegetation can be ignored when the pixels are located in the proximity of streambeds or when they correspond to large, smoothed areas in the false colour map.

Finally, we evaluated the correspondence between real fossil logs on the surface (control map) to our estimated maps of type-1 and type-2 fossil logs. The confusion matrix on the raster layers did not completely describe the matching between true fossil logs and the estimated maps. This comparison is based on pixel counting instead of single features; so it can be misleading, giving a general result of overestimation of fossil logs. In fact, the conversion from the vector to raster layer of the control map drove the comparison to overrepresenting non-matching pixels, doubling the negative effect of unmatching pixels. Even so, the confusion matrix points to a sensible reduction of the area to be surveyed to detect an important portion of the fossil logs.

When we converted the ML and SAM prediction maps to vector layers of predicted polygon logs, we were able to test their success by matching polygons instead of pixels. We find this comparison to be more objective and realistic. We could verify that the SAM classification performed better than the ML. About half the fossil logs were detected (27.1 % and 41.5 %, respectively), a percentage that doubled (54.5 % and 83.6 %, respectively) when only fossil logs larger than the minimum spatial resolution of the multispectral imagery (4 m^2) were considered. It exceeded 90 % when we

https://doi.org/10.1017/S001675682200108X Published online by Cambridge University Press

buffered the regions of the estimated maps, considering a visibility of 2 m around the working path of a hypothetical palaeontological field crew.

5. Conclusion

In this paper, we described the workflow needed to find single fossils from satellite images. The method allows researchers to reduce the negative effect of a paucity of field data, improving the effectiveness of expeditions in remote regions, as scientists can plan daily exploration using computer-generated predictive maps to choose priority localities for prospecting. Previous studies using remote sensing for field palaeontology were focused on searching for outcrops with good potential for palaeontological survey. Our method pushes the field forward, stressing the potential of a new generation of satellites to detect single fossils exposed on the surface. It should not only become a common practice that allows better planning of field expeditions, but also allow remote monitoring of fossil resources in public lands such as National Parks, as well as remote localities of cultural heritage (i.e. UNESCO sites).

We tested the hypothesis on a large area within the Petrified Forest National Park. Our results demonstrate that, even though an overestimation of fossil sites still persists, we could reduce the area to be surveyed and improve the potential of finding fossils within a restricted path, in comparison to a traditional walking survey on the surface. Also, we critically evaluated methods for testing outcome maps for remote-sensing-based predictive palaeontology. We suggest that a more realistic assessment can be obtained comparing a confusion matrix of vector maps of fossils as real or predicted polygons instead of comparison of predicted pixel values in a raster.

In conclusion, the proposed workflow is limited by three main constraints, related to (1) the characteristics of the targeted fossils (size, exposure, texture, mineral composition), (2) the scene (topography, lithology, canopy), and (3) the used recording tool (spectral and spatial resolution of the satellite device). Each constraint contains multiple factors which must be evaluated before running the analyses. Because none of the factors is related to a specific geological timeframe, or is limited by evolutionary clades, or by geology, our method is extremely flexible, applicable and adaptable to nearly any kind of palaeontological field research.

Supplementary material. To view supplementary material for this article, please visit https://doi.org/10.1017/S001675682200108X

Acknowledgements. The authors are grateful to the Petrified Forest National Park team, in particular to William (Bill) G. Parker and Adam Marsh, for introducing E.G. to the park after the SVP Annual Meeting (2018). We are grateful to Robert Anemone and Bas Van de Schootbrugge for the positive suggestions to improve this manuscript. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Action, Global Fellowship (E.G., grant number 785821).

Conflict of interest. None.

References

Anemone RL, Conroy GC and Emerson CW (2011a) GIS and paleoanthropology: incorporating new approaches from the geospatial sciences in the analysis of Primate and human evolution. *Yearbook of Physical Anthropology* **54**, 19–46.

Anemone R, Emerson C and Conroy G (2011b) Finding fossils in new ways: an artificial neural network approach to predicting the location of productive fossil localities. *Evolutionary Anthropology* **20**, 169–80.

Ash SR and Creber GT (2000) The Late Triassic *Araucarioxylon arizonicum* trees of the Petrified Forest National Park, Arizona, USA. *Palaeontology* **43**, 15–28.

Baucon A, Neto de Carvalho C, Briguglio A, Piazza M and Felletti F (2021) A predictive model for the ichnological suitability of the Jezero crater, Mars: searching for fossilized traces of life-substrate interactions in the 2020 Rover Mission Landing Site. *PeerJ* **9**, e11784. doi: 10.7717/peerj.11784.

Burk D (2014) A fossil locality predictive model for the Early Cretaceous Cedar Mountain Formation, Utah, USA. *Geological Society of America Abstracts with Programs* **47**, 681.

Conroy GC (2014) Walking back the cat: unsupervised classification as an aid in 'remote' fossil prospecting. *Evolutionary Anthropology* **23**, 172–6.

Conroy GC, Emerson CW, Anemone RL and Townsend KEB (2012) Let your fingers do the walking: a simple spectral signature model for 'remote' fossil prospecting. *Journal of Human Evolution* **63**, 79–84.

Davis EB, McGuire JL and Orcutt JD (2014) Ecological niche models of mammalian glacial refugia show consistent bias. *Ecography* **37**, 1133–8.

D'Oliveira Coelho J, Anemone RL and Carvalho S (2021) Unsupervised learning of satellite images enhances discovery of late Miocene fossil sites in the Urema Rift, Gorongosa, Mozambique. *PeerJ* **9**, e11573. doi: 10.7717/peerj.11573.

Emerson CW and Anemone RL (2012) An artificial neural network-based approach to identifying mammalian fossil localities in the Great Divide Basin, Wyoming. *Remote Sensing Letters* **3**, 453–60.

Emerson C, Bommersbach B, Nachman B and Anemone R (2015) An object-oriented approach to extracting productive fossil localities from remotely sensed imagery. *Remote Sensing* **7**, 16555–70. doi: 10.3390/rs71215848.

Fielding AH and Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* **24**, 38–49.

Golonka J (2007) Late Triassic and Early Jurassic palaeogeography of the world. *Palaeogeography, Palaeoclimatology, Palaeoecology* **244**, 297–307.

Jiang H, Lee CT and Parker WG (2018) Trace elements and U–Pb ages in petrified wood as indicators of paleohydrologic events. *Chemical Geology* **493**, 266–80.

Jiménez A (2011) Research project in course Geol 4048. Department of Geology, University of Puerto Rico at Mayaguez, 10 pp. http://gers.uprm.edu/geol6225/pdfs/a_jimenez.pdf.

Loughney KM, Fastovsky DE and Parker WG (2011) Vertebrate fossil preservation in blue paleosols from the Petrified Forest National Park, Arizona, with implications for vertebrate biostratigraphy in the Chinle Formation. *Palaios* **26**, 700–19. doi: 10.2110/palo.2011.p11-017r.

Malakhov DV, Dyke GJ and King C (2009) Remote sensing applied to paleontology: exploration of Upper Cretaceous sediments in Kazakhstan for potential fossil sites. *Paleontologia Electronica* **12**, 1–10.

Martz JW, Parker WG, Skinner L, Raucci JJ, Umhoefer P and Blakey RC (2012) Geologic Map of Petrified Forest National Park, Arizona. Contributed Map CM-12-A, 1 map sheet, map scale 1:50,000. Tuscon: Arizona Geological Survey, 18 pp.

McKeown NK, Bishop JL and Noe Dobrea EZ (2011) A comparative lab-field-aerial spectral study of the Painted Desert, AZ: implications for CRISM data interpretation. *American Geophysical Union, Fall Meeting, San Francisco, California, December 2011*, Abstract if.P43C-1695.

McKeown NK, Bishop JL, Noe Dobrea EZ, Ehlmann BL, Parente M, Mustard JF, Murchie SL, Swayze GA, Bibring JP and Silver EA (2009) Characterization of phyllosilicates observed in the central Mawrth Vallis region, Mars, their potential formational processes, and implications for past climate. *Journal of Geophysical Research* **114**, E00D10. doi: 10.1029/2008JE003301.

Mickus K and Johnson E (2001) Mapping sedimentary and volcanic units within and surrounding Petrified Forest National Park, Arizona, using Landsat-5 and SPOT panchromatic data. *International Journal of Remote Sensing* **22**, 1919–35.

Mustoe G and Acosta M (2016) Origin of petrified wood color. *Geosciences* **6**, 1–25.

Noe Dobrea EZ, Bishop JL, McKeown NJ, Fu R, Rossi CM, Michalski JR, Heinlein C, Hanus V, Poulet F, Mustard RJF, Murchie S, McEwen AS,

Swayze G, Bibring J-P, Malaret E and Hash C (2010) Mineralogy and stratigraphy of phyllosilicate-bearing and dark mantling units in the greater Mawrth Vallis/west Arabia Terra area: constraints of geological origin. *Journal of Geophysical Research* **115**, E00D19. doi: 10.1029/2009JE003351.

Noe Dobrea EZ, McAdam AC, Freissinet C, Franz H, Belmahdi I, Hamersley MR, Archer R, Stoker PWG, Glavin DP and Calef F (2017) Mechanism for preservation of organics in jarosite at the Painted Desert. *Astrobiology Science Conference 2017 (LPI Contrib. No. 1965)*, 3356.

Nordt L, Archley S and Dworkin S (2015) Collapse of the Late Triassic megamonsoon in the western equatorial Pangea, present-day American Southwest. *Geological Society of America Bulletin* **127**, 1798–815.

Oheim KB (2007) Fossil site prediction using geographic information systems (GIS) and suitability analysis: the Two Medicine Formation, MT, a test case. *Palaeogeography, Palaeoclimatology, Palaeoecology* **251**, 354–65.

Sigleo AC (1979) Geochemistry of silicified wood and associated sediments, Petrified Forest National Park, Arizona. *Chemical Geology* **26**, 151–63.

Steiner MB and Lucas SG (2000) Paleomagnetism of the Late Triassic Petrified Forest Formation, Chinle Group, western United States: further evidence of "large" rotation of the Colorado Plateau. *Journal of Geophysical Research, Series B, Solid Earth and Planets* **105**, 25791–808.

Tou JT and Gonzalez RC (1974) *Pattern Recognition Principles*. Reading, MA: Addison-Wesley Publishing Company.

Wills S, Choiniere JN and Barrett PM (2018) Predictive modelling of fossil-bearing locality distribution in the Elliot Formation (Upper Triassic–Lower Jurassic), South Africa, using a combined multivariate and spatial statistical analyses of present-day environmental data. *Palaeogeography, Palaeoclimatology, Palaeoecology* **489**, 186–97.

https://doi.org/10.1017/S001675682200108X Published online by Cambridge University Press