

Ecology, biodiversity and mining: science solving the challenges

L Mucina *Harry Butler Institute, Murdoch University, Australia*

JL Tsakalos *The University of Western Australia, Australia*

PD Macintyre *The University of Western Australia, Australia*

Abstract

Mining has an impact on the ecology and biodiversity of an area, and this impact is required to be mitigated under regulatory requirements. The mitigation is achieved by applying measures, deeply rooted in scientific knowledge, of the functioning of the impacted (and rehabilitated) ecosystems and the biota supported by these ecosystems. This paper focusses on selected issues of the pre-mining ecology and biodiversity surveys and the scientific basis of regulatory requirements. Using Western Australian experience and drawing on comparisons from other countries, this paper addresses two challenges. The first is the classification of vegetation of mining tenements to inform on the variability of vegetation using a series of data-analytical experiments. The paper reveals flaws in the current practise aimed at recognition and description of plant communities in vegetation surveys. Concrete steps serving and enhancing the clarity and plausibility of regulatory tools (guidance manuals) are identified. It also advises on what we should do if we find discrepancies between the regulatory expectations and the level they reflect current scientific knowledge. The second is the comparative and spatial aspects of the identification of communities of conservation interest. The paper further identifies the missing vital information of the vegetation mapping procedures and briefly analyses the need for tools assisting assessment of conservation value of the vegetation classifications, including the building of a centralised vegetation database, formulation of vegetation and habitat classification systems, and construction of scientifically sound and ecologically informative vegetation maps at various scales of complexity.

Keywords: *classification of vegetation, data analysis, post-mining rehabilitation, vegetation mapping*

1 Introduction

Mining might have a very negative impact on the environment. The environment is, among other things, a subject of the scientific discipline of ecology. Therefore, mitigation of the impacts of mining on the environment requires science for advice on what to do and how to do it right.

Mining activities, including mine closure and all its facets that are of importance to the mining company and society at large, are a subject of government regulations aimed at maximising the positive outcome of the mining for economic and social development and minimising the negative impact on the environment. This delicate balance is far from simple and it requires scientific input.

Naturally, post-mining rehabilitation—a vital part of mine closure procedures—should be based on knowledge of the fundamental processes of vegetation dynamics underpinning the assisted revegetation. Ultimately, however, successful rehabilitation outcomes also depend on our knowledge of the structure and dynamics of ecosystems that were originally impacted. Therefore, we shall also investigate, review, and demonstrate how science should contribute to gathering this knowledge, appearing indispensable for designing of the post-mining rehabilitation. By drawing on our experience in working in the field of ecology and ecological restoration in Western Australia, we shall address two challenges:

1. Addressing the classification of the natural (native) vegetation of the proposed mining tenement before mining, focussing on the selection of numerical-analytical tools to handle the collected vegetation data.
2. Addressing the spatial aspects of the identification of communities of conservation interest.

2 Challenge 1: classifying native vegetation before mining

2.1 Motivation: relevance to mine closure

As a part of a two-tier process of flora and vegetation surveying of the proposed mining tenement, mining companies are obliged to perform a survey of the biota; identifying which species of plants and animals are found there and what kind of plant communities can be distinguished on the tenement. Plant communities inform (in terrestrial conditions) about the habitat conditions on the tenement or, in other words, about the types of ecosystems supported by the current environmental matrix (geological substrate, soils, hydrology, and large-scale disturbance factors such as fire). This survey is needed to formulate the mine closure planning as the mining company needs to know which plant communities were occurring on the mining tenement before the (surface) mining activities had major impact on them. Recognition and understanding of the past vegetation patterns is a vital provision for reasonable rehabilitation or reconstruction of the vegetation cover after mining.

2.2 Environmental Protection Authority instructions

Focussing on vegetation survey only, Environmental Protection Authority (EPA) guidance documents (Environmental Protection Authority [EPA] 2004, 2016) instruct how to collect the field vegetation (plot) data, how to handle (analyse) the data, and how to present the results of the survey.

Focussing on the challenge in hand, EPA 2004 prescribes:

- Consideration of disturbance, focussing on the number of native species compared to weed species, the proportion of native species present compared with that expected in an intact plant community of the same type, and on the condition of the described units of vegetation.
- To define vegetation categories where the scope is large—in terms of number of vegetation types, area, and/or multiple locations—a form of multivariate analysis of the data is likely to be warranted. At intermediate scales, or in the absence of the resources for such analysis, a site/species matrix may be used to group sites on the basis of like suites of the most common/indicator species.
- Multivariate analysis should include, as a minimum, presence/absence data and perennial species.
- A table of the area of each vegetation type including the percentage affected by the project area (both at the locality and in the region) and the categories of vegetation and their environmental values (which may include near pristine, unique, limited extent, extensively cleared, and significant flora present).
- General observations and other qualitative information on the site.

The latest technical guidance (EPA 2016, pp. 18–19) advises:

“When comparing quadrat data with data from another survey, it is recommended that only species presence/absence data be used, as variation in cover estimates may significantly affect the analysis. Ensuring that all taxon names are representative of the same nomenclature is critical for analyses involving multiple data sources (data reconciliation). When interpreting the results of analysis from multiple data sources, consideration should also be given to the influence of differences in survey effort, timing of survey, seasonal conditions and disturbance history.”

“Multivariate comparative (cluster) analysis should be performed on a species-by-site matrix to measure the similarity between sites based on the presence or absence of species. A clearly legible dendrogram (Figure 3) should be produced to illustrate the similarities between the vegetation units that have been identified.”

2.3 Interpretation of the Environmental Protection Authority instruction

The researchers in charge of the vegetation surveys are usually contracted environmental consulting companies. These are the ones who face a daunting task: to make the right methodical choices and use appropriate tools to collect, collate, and analyse the vegetation data. The EPA instructions are, however, very vague in terms of choice of the selection of multivariate techniques and therefore, the service providers have a free hand to shape their choice of tools. We argue that this freedom also opens doors to many risks. Further, the poorly informed advice (for more, see Section 2.4.2 to use presence/absence data allows corner cutting that results in the production of poor classification outcomes.

2.4 Risk assessment

Selection of right numerical (quantitative) tools assisting the task at hand—the objective (or better: formalised and quantitative) definition of plant communities—is critical. There are hundreds of ways to construct the numerical apparatus to analyse the vegetation data. The construction of such apparatus, in the case of a multivariate-based classification technique, entails selection of data transformation, selection of resemblance measure, selection of classification algorithm (clustering, divisive technique, intuitive non-formalised classification), and selection of tools to analyse the numerical result, i.e. analysing the dendrograms, for instance. For the technical terminology, we refer to Legendre & Legendre (2012).

A combination of transformation, resemblance, and classification algorithms defines so-called data-analytical combination (DAC). There are several tens of ways to transform the data (Podani 2000), there are hundreds of resemblances (or ways to describe the similarity between objects), and there are also many ways to classify the data (McCune & Mefford 1999: PC-ORD; Belbin 2003: PATN; Legendre & Legendre 2012; Clarke & Gorley 2015: Primer; Wildi 2017: dave; Oksanen et al. 2019: vegan). When different DACs are applied on the same dataset, different, and often indeed very disparate, classifications result. We wish to demonstrate this issue using two data-analytical experiments.

2.4.1 Experiment 1: classification of the same dataset using different data-analytical combination

The choice of multivariate analytical tools is not trivial and requires knowledge of the inner workings of the techniques themselves and understanding the analytical model, as well as the nature of the data used in the analysis. Issues such as having quantitative (measured or assessed) data, the species diversity in the sampled relevés, the effective size of the plots, their spatial distribution in the landscape, etc., are all parameters which matter in the choice of the tools.

We have generated a simulated relevé data matrix comprising 12 relevés/plots (we use the terms interchangeably) sharing 23 species. The quantitative importance values using scale 0–9 were scored for each species to create a vegetation relevé matrix of continuous nature: as one species is fading out along the gradient spanning relevés 1 to 12, the species might be fading out (having lower importance values) in the opposite direction (Table 1).

Table 1 Simulated vegetation matrix composed of 12 plots/relevés and 23 species (A through W) showing continuous nature of vegetation. The values in the cells of the matrix are importance values (scale: 0–9; 0 means absence of the species) of each species in each plot

Species	Plot											
	1	2	3	4	5	6	7	8	9	10	11	12
A	3	2	1	0	0	0	0	0	0	0	0	0
B	4	3	2	1	0	0	0	0	0	0	0	0
C	5	4	3	2	1	0	0	0	0	0	0	0
D	6	5	4	3	2	1	0	0	0	0	0	0
E	7	6	5	4	3	2	1	0	0	0	0	0
F	8	7	6	5	4	3	2	1	0	0	0	0
G	9	8	7	6	5	4	3	2	1	0	0	0
H	1	9	8	7	6	5	4	3	2	1	0	0
I	2	1	9	8	7	6	5	4	3	2	1	0
J	1	2	3	4	5	6	7	8	9	1	2	3
K	1	2	3	4	5	6	7	8	9	1	2	3
L	1	2	3	4	5	6	7	8	9	1	2	3
M	1	2	3	4	5	6	7	8	9	1	2	3
N	1	2	3	4	5	6	7	8	9	1	2	3
O	0	1	2	3	4	5	6	7	8	9	1	2
P	0	0	1	2	3	4	5	6	7	8	9	1
Q	0	0	0	1	2	3	4	5	6	7	8	9
R	0	0	0	0	1	2	3	4	5	6	7	8
S	0	0	0	0	0	1	2	3	4	5	6	7
T	0	0	0	0	0	0	1	2	3	4	5	6
U	0	0	0	0	0	0	0	1	2	3	4	5
V	0	0	0	0	0	0	0	0	1	2	3	4
W	0	0	0	0	0	0	0	0	0	1	2	3

We then classified the same dataset using six different DACs, combining two resemblances (Euclidean distance and chord distance), and three clustering techniques (average link, single link, and Ward's method). Although there are clear (optically) similarities between dendrograms using the different resemblances using the same clustering, there is also clear variance between the shapes of dendrograms using the same resemblance but different clustering technique (compare the first row comprising three dendrograms in Figure 1).

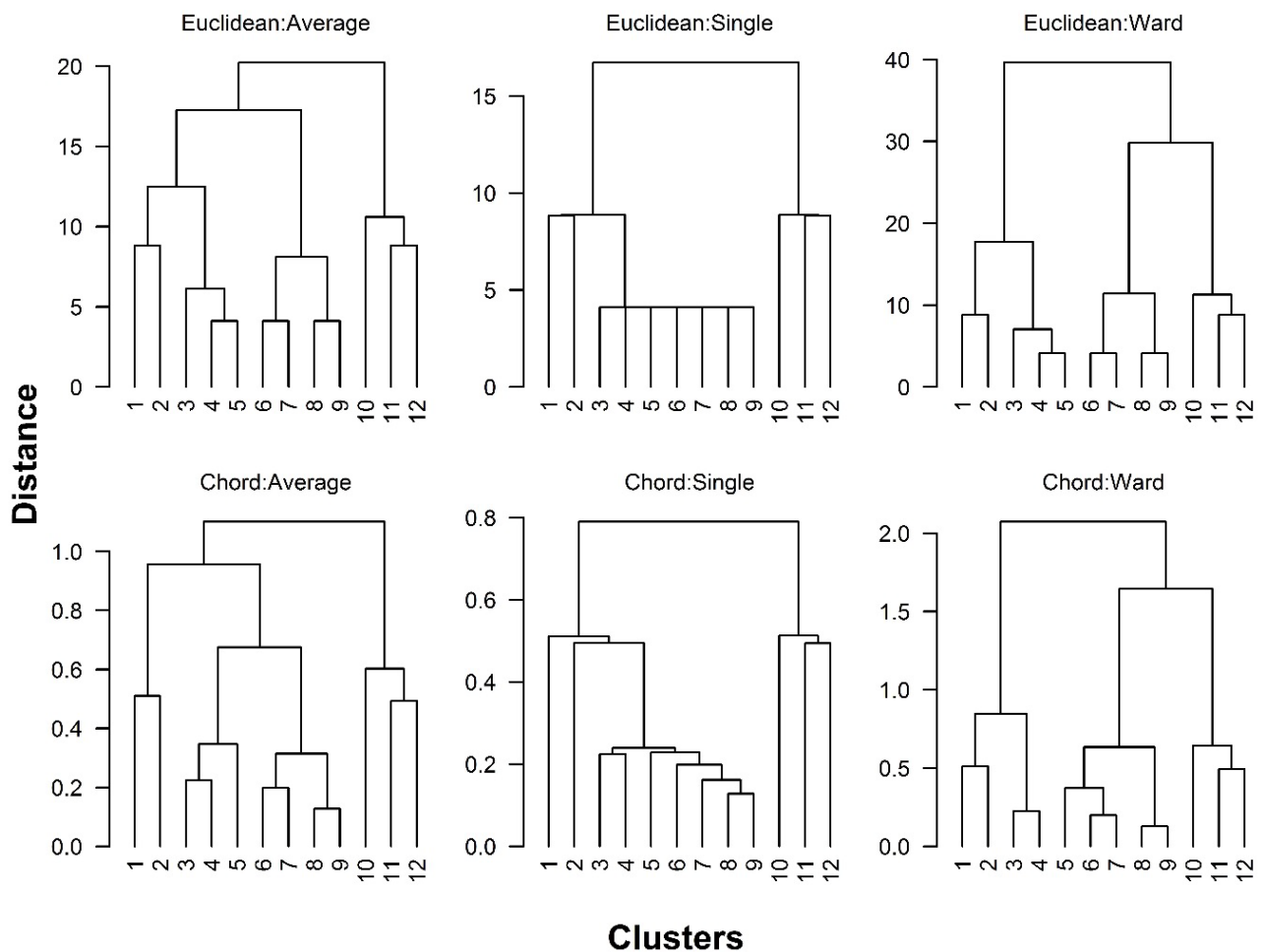


Figure 1 Resulting dendrograms of the classification of the Table 1 matrix using three clustering techniques (single link, average link, Ward's method), in combination with two resemblances (distances in this instance), namely Euclidean distance and chord distance (normalised chord distance). For technical details on the rationale underpinning each clustering and the resemblances used, see Legendre & Legendre (2012). All six dendrograms show the fusion hierarchy which is driven by the resemblance between objects (12 relevés) and clusters (as they grow by successive addition of the relevés). The y-axis shows the distance at which clusters have been fusing. The calculations were performed using the 'hclust' function from the base R package (R Core Team 2019)

Which one is correct? All of them are, in their own way. A quick optical assessment reveals that all dendrograms have a unique fusion pattern; in other words, they are all different. Looking at the differences in the distance level at the level of final fusions, one can easily see that the choice of the resemblance did not play such a big role. The average link clustering suggests the existence of perhaps three clusters, the single link clustering suggests two clusters, while Ward's method again would suggest three clusters (not identical with those of average link solution though).

The challenge faced is how to determine the optimal number of clusters (presumed plant communities) in such classification using an independent, robust procedure. EPA guidelines do not give advice on this issue, equally as they do not warn that various analyses would bring different results. Not surprisingly, yet still very wrongly, researchers (and almost invariably the way used in environmental consultant reports) use 'optical judgement' – a subjective decision based on shape of the classification dendrogram. In the best case, they might have applied some sort of formalised procedure to identify the optimal number of clusters, but they fail to inform the consumers about how that was done (Figure 2).

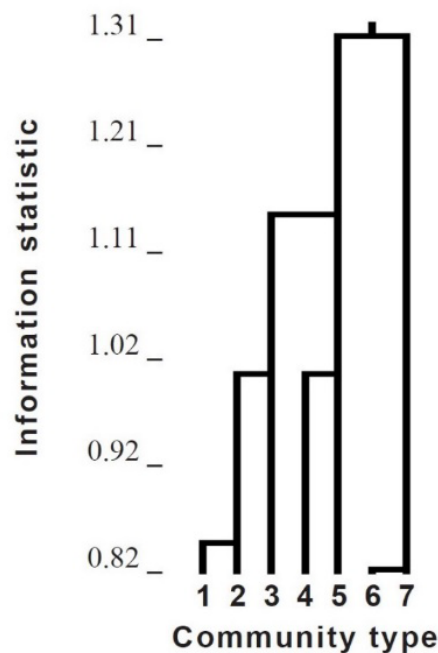


Figure 2 Dendrogram (with suggested seven clusters/community types) based on 54 relevés (Gibson 2004). Although the author has indicated the choice of the resemblance (Czekanowski's index) and clustering (unweighted pair group method with arithmetic mean), the author has failed to explain the meaning of the 'information statistics', and hence how he arrived at the seven-cluster solution

There are many ways to identify the optimal number of clusters and some are widely used in vegetation ecology (Tichý et al. 2010: OptimClass; Caliński & Harabasz 1974: C-H index). Here we analysed the six dendrograms using the C-H index (Figure 3). These analyses suggest that an optimal number in four of six cases is nine clusters; the other two cases (both using the single link clustering) suggest as many as 11 clusters.

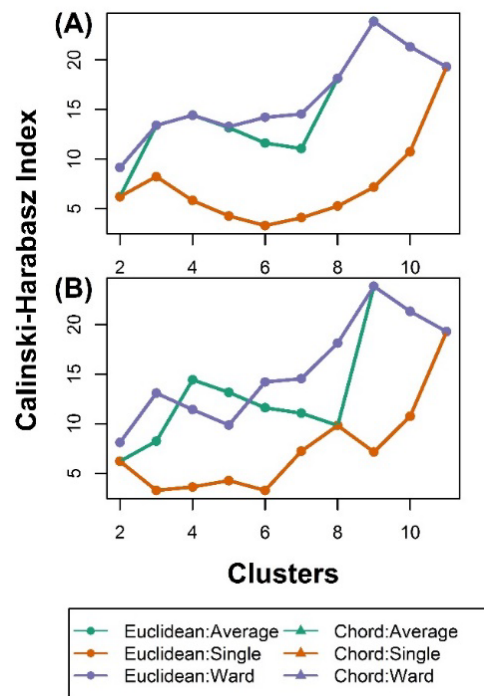


Figure 3 Calinski-Harabasz analysis identified the optimal number of clusters for six dendrograms generated by three data clustering techniques (Average link, Single link, Ward's method) and two resemblances: (A) Euclidean distance; (B) chord distance. The highest score of the C-H indicates that two (orange) classifications (single link) suggest 11 clusters as the optimal solution while the rest of the classifications peak at nine clusters. A solution involving three or four clusters appears as the second best for average link and Ward's method, respectively. The calculations were performed using the calinhara function coded in the fpc R package (Hennig 2015)

2.4.2 Experiment 2: effect of data transformation on classification outcome

EPA (2016) guidance advice to focus on presence/absence (P/A or 1/0) data argues that 'variation in cover estimates may influence the analysis'. Indeed, there is a variance in estimates if more than one surveyor is involved, but this variance can be mitigated by the choice of an appropriate data transformation. Applying data flattening, all quantitative data are transformed into 1/0 or P/A data is the crudest way which disregards the nature of plant community per se. Many vegetation types are dominated by one or few species and the less species-rich these communities are, the more pronounced the dominants appear, as a rule. Dominant species control the functioning of the plant community (Grime 1987), hence, dominance matters.

To demonstrate this effect, we used the single link classification technique and three transformations options involving no transformation, logarithmic transformation, and P/A transformation which causes data flattening (i.e. all quantitative values are transformed to assume the value of 1). We use another simulated data matrix (Table 2) where we have, on purpose, introduced some dominant and subdominant species.

Table 2 Simulated vegetation matrix composed of 12 plots/relevés and 23 species (A through W), incorporating some dominants (value 5) and subdominants (value 3). The values in the cells of the matrix are importance values (scale: 0–9; the blank cells indicate 0 values) of each species in each plot

Species	Plot											
	7	11	3	9	6	5	8	10	4	12	1	2
A			5						5		5	5
B			3			1			3		1	1
C			2		1	1			2		1	1
D	1		2		1	1			2		2	1
E	5		2		5	5	5		1		2	1
F	3		2	1	1	1	3		1		2	1
G	2		1	1	1	1	1	1	1		2	1
H	2	1	1	1	1	1	1	1	1		1	1
I	2	5	1	5	1	1	1	5	1	5	1	1
J	2	3	1	1	2	1	1	1	1	3	1	2
K	1	2	1	1	2	1	1	1	2	1	1	2
L	1	2	1	1	2	1	1	2	2	1	1	2
M	1	2	1	2	2	1	1	2	2	2	1	2
N	1	2	1	2	1	2	1	2	2	2	1	1
O	1	1	1	2	1	2	1	2	1	2	1	1
P	1	1	1	2	1	2	2	2	1	2		1
Q	1	1	1	1	1	2	2	2	1	2		
R	1	1		1	1	2	2	1	1	1		
S	1	1		1	1	2	2	1		1		
T	1	1		1	1		2	1		1		
U	1	1		1			1	1		1		
V		1		1			1	1		1		
W		1		1				1		1		

The results of this clustering exercise are summarised in Figure 4. The results indicate that no transformation (i.e. preservation of the quantitative values and hence also dominance), as well as moderate transformation by application of logarithm of the original values, produces similar results. The P/A transformation has, however, major impact on the analysis as it is producing a very different outcome. Flattening of the data destroyed the original data structure and created an artefact.

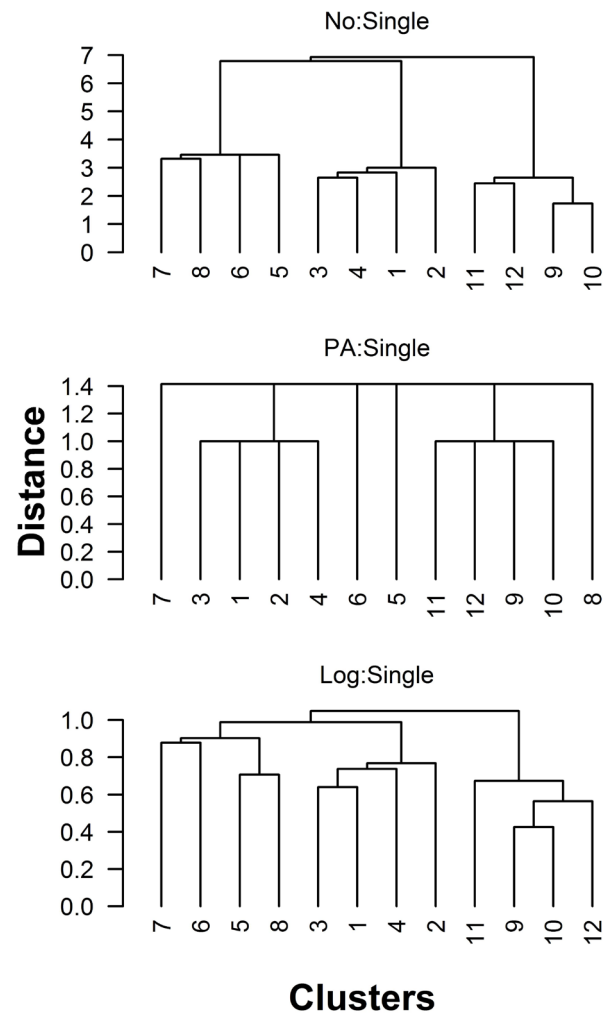


Figure 4 Dendrograms of three clustering analyses using the same clustering algorithm (single link), demonstrating the effect of the data transformation (in top-down order: no transformation applied; P/A=presence/absence (= flattening of data to values 1/0; logarithmic transformation. The calculations were performed using the 'hclust' function from the base R software (R Core Team 2019)

2.5 Conclusion

Our experiments revealed risks which cannot be underestimated. Focussing only on one (often ill-advised way of analysing data) can lead to wrong conclusions when seeking resolution of challenge 1. The emerging problems, common flaws, and solutions are summarised in Table 3.

Table 3 Summary of problems, bad responses, and solutions in the field of selecting data-analytical tools to deliver on challenge 1

Problem	Bad response to the problem	Solutions
There are many ways to analyse data	Lack of reporting on selected analytical procedures	EPA guidance should require detailed reporting of how the data have been collated, prepared, and stored) before the analyses are then transformed, and which resemblances and classification tools have been chosen
	Unjustified selection of analytical tools	EPA guidance should require reporting on a plausible, scientifically sound, justification of the choice of each of the data-analytical tools
	Wrong selection of analytical tools	Seek the advice of experts or re-start self-education
How to identify clusters in hierarchical analyses	Using visual rather than analytical tools to define clusters	If visual approaches are used, then the procedure should be formalised (describe steps logically and convincingly); otherwise use freely available formal (data-analytical) tools to perform the task

In general terms, it is the black-box approach to the use of data-analytical tools which reflects a concerning lack of expertise, or at best, simple complacency on the side of those tasked with the delivery of expertise to the client. The EPA must change its policy and demand more detailed and robust scientific reasoning underpinning the choice of the tools. Admittedly, robust scientific reasoning, coupled with higher quality analyses, comes at a price. Until that happens, mining companies should implement their own, stringent criteria to secure the quality of the advice received via additional demands on the service providers, careful selection of those undertaking the analysis possessing needed skills, and independent audit of the reports.

3 Challenge 2: identification of communities of conservation interest

3.1 Motivation: relevance to mine closure

Vegetation of a region or landscape is composed of plant communities; distinct and recognisable units, characterised by unique species composition, links to the environment, and occurring in a distribution area that reflects the environmental conditions, as well as many other, mainly evolutionary and historical factors. Some of the communities are rare (have a limited distribution) or are even unique to a region, and others may form the vegetation matrix of extensive regions. Some communities can be home to many endemic species (species occurring in a limited, geographically and/or ecologically well-defined area) and some can be composed of species having a broad distribution. Some communities are species-rich or support species of conservation interest, i.e. Western Australian kwongan scrub, and others can be naturally species-poor, i.e. Mulga shrubland, or depauperate due to intensive human influence. Those which are rare or unique, or supporting species of conservation interest, can fit into the category of a threatened ecological community (TEC). Pre-mining surveys are requested to detect such communities and to establish if the potential mining tenement has already declared TECs or priority ecological communities (PEC) present, or there are some communities that are unique and appear irreplaceable.

3.2 Environmental Protection Authority instructions

Section 3.2 of the EPA technical guidance (EPA 2016, pp. 3–4) on vegetation states:

“A search of EPBC Act List of Threatened Ecological Communities and Parks and Wildlife’s threatened and priority ecological communities’ databases is essential for all surveys. A wide search area encompassing the survey area should be selected for database queries.

The following sources should be used to gather information about threatened ecological communities (TECs) and priority ecological communities (PECs):

- EPBC Act list of TECs;
- Parks and Wildlife's threatened and priority ecological community list (Parks and Wildlife 2015a, 2015b);
- Reports that contain the original descriptions of particular communities;
- Nomination or listing descriptions of the TEC or PEC, if available from Parks and Wildlife or the Commonwealth;
- Recovery plans and other reports containing information on the preferred habitats and distributions of TECs that can be checked against the study area; and
- Survey reports or references in the region or locality (refer to Appendix A for a list of selected flora and vegetation reports). Desktop study may also identify vegetation units that are not currently listed as TECs or PECs but may have significance for other reasons. Any information likely to be useful during field survey to identify the characteristics and habitat of significant communities should be recorded."

Section 8.4 of the EPA technical guidance (EPA 2016, p. 23) on defining TECs and PECs elaborates further on how to identify and describe the TECs and PECs and, among other instructions, it stipulates:

"Analysis of whether vegetation units described in the survey area represent TECs or PECs is a significant part of the vegetation classification process for EIA."

"Where TECs or PECs are defined at a local scale, the objective of the vegetation classification will be to determine whether and where these occur within the survey area. Where the TEC or PEC is defined at a regional scale (association, alliance, complex, system or broader), vegetation sub-units of the listed ecological community should be identified so that the overall impact on the biodiversity values of the TEC or PEC can be evaluated."

"Where TECs or PECs have been described from quadrat-derived data, similarity should be determined by comparing data from the survey quadrats with data from the survey in which the TEC or PEC was identified. These data are often available in the literature or directly from Parks and Wildlife."

3.3 Interpretation of the instructions

The EPA recognises that the identification of some vegetation types is of great conservation importance. Recognition of some communities as such needs consultation of existing data (plot databases, regional studies, maps). The presence of TECs and PECs and other putative community types of conservation importance should be established by consulting existing information.

3.4 Risk assessment

Service providers tasked at finding out if the mining tenement has communities of conservation interest are required to use comparative methods. Specifically, they are asked to classify and map vegetation and compare those mapping units to similar vegetation in the area to establish if vegetation types are unique, rare, or of pre-defined list of already recognised TECs and PECs (Department of Biodiversity, Conservation and Attractions 2019) of conservation interest.

There are two major problems with this approach as it implies that:

1. The comparative information (databases) are available.
2. The plant communities of interest (in the investigated area and beyond) have been mapped.

The impact of these two problems are analysed in the following two examples.

3.4.1 Problem complex 1: availability of the vegetation (plot) data

In this regard, the advice from Section 3.2 of the EPA (2016) to consult the existing regional survey reports (and published sources too) is naïve and contra-productive. The problems with these sources are manifold:

1. Most of the Western Australian vegetation-ecological knowledge is buried in poorly-available reports and these are not readily available due to all sorts of administrative reasons. In this space, the recent initiative of the Western Australian Biodiversity Institute aimed at creating a digital library of reports pertinent to approved mining activities is welcome, but not poised to provide a general solution. The Department of Biodiversity, Conservation and Attractions also provides, through NatureMap (<https://naturemap.dpaw.wa.gov.au/>) access to many reports (Gibson 2018), but obviously not all of relevance.
2. There is no centralised database of vegetation plots/relevés publicly available to all service providers that would provide material to compare their vegetation systems with other pools of data. We are aware that some service providers maintain their ad hoc databases, but these are invariably incomplete and electronically hardly compatible.
3. Equally, a serious problem (if such reports become available) is the rampant lack of comparability between vegetation classification systems defined for regions of limited extent, published or not. In simple language, the plant communities defined in one report are (as a rule) very poorly comparable with those described in other reports because of different (and often poorly described) methods used to define such communities and for disparate nomenclature used to name them. As an example, we wish to refer to three disparate vegetation classification systems coined for one single small region as featured in a critical report by Mucina (2009).
4. There is neither a general vegetation classification system nor a standardised habitat system of value available across Western Australia to provide the platform for comparability of the regional studies.

3.4.1.1 Solutions

1. Further efforts should be made to make all reports available to the public, which, de facto, owns that knowledge (bar operationally sensitive and classified information).
2. The vegetation plot data should be captured in a centralised vegetation database embedded within a vegetation information system. Western Australia can learn from Europe where almost every nation runs a centralised database, linked by an exchange platform (Chytrý et al. 2016: European Vegetation Archie), using a unified database platform (Hennekens & Schaminée 2001: Turboveg/TvWin). It has been a mammoth task for the European Vegetation Survey Working Group of the International Association for Vegetation Science (<http://euroveg.org/>) to arrive at that point, but Western Australia should be far less challenging than a concoction of 50 sovereign states, often having their own agendas.
3. Because EPA (2016) does not outline the data collection and analysis procedures, comparability between various vegetation systems would always stay compromised. We all need more specific, scientifically sound instructions that would entail, at the least, a demand that the authors of the reports (and publications) would provide detail on the numerical-analytical methods and the reasons for their choices to allow informed judgement on the usefulness of the reported material in hand.
4. Having vegetation data centralised in a database, metanalysis (or re-analysis) of the data is possible. We need general classifications schemes, for starters, for selected vegetation types (kwongan, mulga, tropical woodlands, etc.). These schemes would provide the building stones of a general classification system for the state; a system that would form the basis of the vegetation mapping legend (see Section 3.4.1). Europe has such a system (Mucina et al. 2016) because of historically important recognition that a building of a vegetation system is possible and needed. This system informs the Habitat Directory of the European Union (European Commission 2007; Janssen et al.

2016) which is a major legislative tool in the space of nature management and use of renewable resources. We do not need to adopt the same scientific language (vegetation unit categories) as Europe is using. We need to develop a conceptual framework which is scientifically sound and underpinned by data. The vegetation system developed for South Africa (Mucina & Rutherford 2006) serves as an example. Western Australian ecology, despite efforts of legendary John Beard, (for more, see Section 3.4.1) has failed to deliver in this space so far.

3.4.2 Problem complex 2: mapping instructions and quality of the vegetation mapping sources

Plant communities occur in nature in the form of patches and these occupy space, hence, those communities are mappable (can be expressed in a graphical spatial model called 'map'). In the context of vegetation surveys, EPA (2016, p. 24) advises the following:

“The results of a survey should be mapped to present information such as survey effort, distribution of vegetation units, significant vegetation and populations of significant flora in the survey area and provide complementary information to determine the significance of impact. All maps should be legible, with an explanatory title and current information. Aerial photography should be the base layer for most maps with the subject of the map overlaid with transparent colours and labelled features. Colours of features and/or shapes of point symbols should be readily distinguishable from one another. The colours or textures used to indicate recurring features (e.g. impact footprint) should be consistent for all maps within the survey report. See Figure 4 for an example map showing minimum features. The scale of maps will vary depending on the size of the survey area, spatial heterogeneity of vegetation and amount of information that needs to be displayed.”

3.4.2.1 Problems

1. EPA advice on how to map is limited to using aerial photography. This advice is, in times when useful satellite imagery (e.g. Landsat) is free, unacceptable.
2. Besides the need to map vegetation in the regions under specific surveys, to perform the task of establishing if the communities are of conservation interest, one needs a series of modern vegetation maps reflecting various scales of the landscape complexity. For a long time, Western Australian ecology and biodiversity science have been using the maps produced by John S Beard (for the sources, see Beard et al. 2014). These maps use vegetation structural criteria (vegetation layering, dominants), but also other salient features like ecology (soils, geology, hydrology) and at the fine levels of the mapping units also floristic composition. This mapping was done in times where the only accessible remote-sensed data were aerial photos. The world of vegetation mapping has long moved on and embraced tools such as the use of vegetation plot data and formalised classifications, satellite imagery interpretation, and machine-learning based predictive modelling (Mucina & Daniel 2013; Macintyre et al. 2018).
3. Vegetation maps are not an ornament of service providers' reports; they are models of the vegetation of an area and they inform (if created using honest and robust scientific approaches) on the ecological fabric and even evolutionary history of the region. They are indispensable at any level when spatial information on the distribution of biota (vegetation in terrestrial environment in the first place) is required. It is not surprising therefore, that in countries which acknowledge the gravity of serious environmental challenges (land-use change, urbanisation, disturbance caused by agriculture and mining), vegetation maps are used profusely. In South Africa, for instance, the National Vegetation Map (Mucina & Rutherford 2016) became a legal standard, ranging from environmental consulting to informing local, regional, and national policy decisions.
4. However legendary, the maps of John Beard (summarised in electronic format by Beard et al. 2013) are outdated and imprecise. Western Australia needs a new vegetation mapping program and here we offer solutions to remedy the current unsatisfactory situation.

As an example, let us consider the same area mapped by four different mapping schemes (Figure 5).

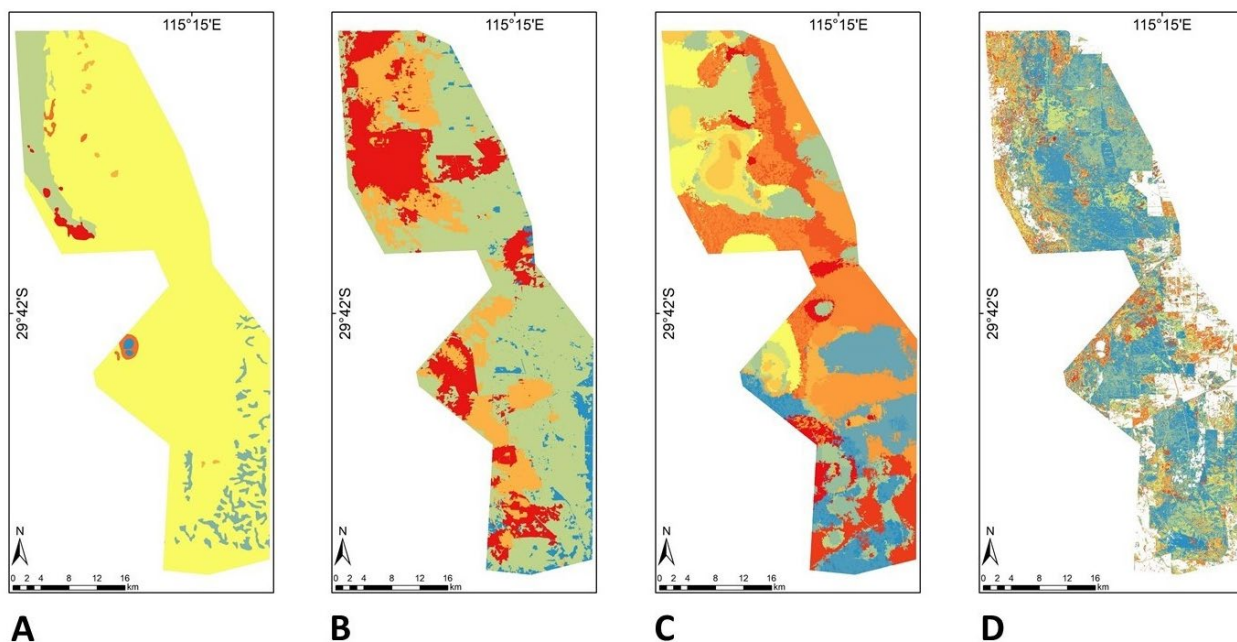


Figure 5 Four examples of vegetation mapping across the same area. The clip is showing an area of approximately 30 x 80 km in the surrounds of Eneabba, Western Australia. A) vegetation map of Western Australia (Beard et al. 2013); B) and C) potential natural vegetation predictions based on both coarse (B) and fine (C) classifications (Macintyre et al. 2018); and (D) supervised classification (same classification as C) of real, extant vegetation using multispectral imagery

As mapped by Beard, this area is predominately considered to be a single unit of low heath, scattered low woodlands, and non-vegetated patches. Compare this with image B, showing a potential natural vegetation (PNV; Mucina 2010; Macintyre et al. 2018) map of the coarse vegetation units derived using floristic criteria. In the latter map, the units align primarily within differences in the geological substrate as they show (broadly) kwongan on sand, kwongan on laterite, wet kwongan, and non-kwongan scrub. As in Beard's mapping, these broad units provide context in both the type of vegetation (kwongan), with additional information about their ecological positioning, with the advantage of the ability to predict these boundaries at essentially any scale (subject to adequate resolution of spatial data). Taking this to the level of fine-scale communities derived on floristic basis, i.e. the finest scale of classification, map C shows the predicted patterns to be significantly more complex than can be inferred by the regional reference (map A). The maps of PNV can have limitations. For example, in the production of maps B and C (Macintyre et al. 2018), it was shown that at the coarse scale, the predictions were robust, while at the finer scales, additional factors of ecological redundancy made these predictions unreliable. In contrast, map D shows a supervised per-pixel classification of the vegetation of the area using multi-temporal, remotely sensed data from the Sentinel 2 platform (equivalent to ~ 1:12,500 scale) which returned a highly reliable classification. While ecologically the complexity may reduce the accuracy, this applies to maps A, B, and C (independently of actual method). The ability of freely available, high frequency, remotely sensed data offers a serious alternative to methods used today. While all examples contain pros and cons, for example, the PNV can predict patterns over disturbed or cleared areas, while the satellite classifications are restricted to extant units and vegetation only, they contain one singular advantage which is crucial moving forward; that of independent and assessable accuracy. Whereas traditional field intensive mapping exercises, where the boundaries are identified through a combination of photointerpretation and ground truthing conducted by two people independently, would likely result in discrepancies (either minor or not), these alternatives are completely reproducible. The same data combined with the same recorded methods will result in an identical map. Additionally, while we could not tell which of the maps produced on the ground were more accurate (if either), the alternate methods

allow quantitative accuracy assessments which allow for greater confidence in their use by parties that have never been in that area.

3.4.2.2 Solutions

1. EPA guidance should include advice on how to make vegetation at various mapping scales reflect the needs of the clients and to reflect the current scientific knowledge and level of technology available.
2. In the context of the mine closure and all associated mining procedures, we suggest that all concerned parties including the Western Australian government and its relevant institutions, the mining sector, the environmental consulting world, and scientific experts embedded within universities and research institutes, join forces and get the job done. Western Australia needs an inventory of its vegetation treasures and needs tools that would inform processes of the mining activities. 2,529,875 km² our state occupies a huge chunk of land but the price tag of such an exercise can be reasonable and defensible in considering the benefits for decades to come. Good science is crucial for achieving this goal.

Acknowledgement

This work has been supported by the ARC Linkage Grant LP150100339 as well as the Iluka Chair in Vegetation Science and Biogeography at The University of Western Australia (School of Biological Sciences) and Murdoch University (Harry Butler Institute). We thank Dr Mark Dobrowolski for comments and advice.

References

- Beard JS, Beeston, GR, Harvey JM, Hopkins, AJM & Shepherd, DP 2014, 'The vegetation of Western Australia at the scale 1:3,000,000 scale', 2nd edn, *Conservation Science Western Australia*, vol. 9, issue 1, pp. 1–152.
- Belbin, L 2003, *PATN. A Revised USER's Guide*, Blatant Fabrications, Bonnet Hill.
- Caliński, T & Harabasz, J 1974, 'A dendrite method for cluster analysis', *Communications in Statistics*, vol. 3, pp. 1–27.
- Chytrý, M, Hennekens, SM, Jiménez-Alfaro, B, Knollová, I, Dengler, J, Jansen, F ... Yamalov, S 2016, 'European Vegetation Archive (EVA): an integrated database of European vegetation plots', *Applied Vegetation Science*, vol. 19, iss. 1, pp. 173–180.
- Clarke, KR & Gorley, RN 2015, *PRIMER v7: User Manual/Tutorial*, PRIMER-E Plymouth, Plymouth.
- Department of Biodiversity, Conservation and Attractions 2019, *Threatened Species and Communities*, Government of Western Australia, <https://www.dpaw.wa.gov.au/plants-and-animals/threatened-species-and-communities>.
- Environmental Protection Authority 2004, *Guidance Statement No. 51, Terrestrial Flora and Vegetation Surveys for Environmental Impact Assessment in Western Australia*, Government of Western Australia, Perth.
- Environmental Protection Authority 2016, *Technical Guidance - Flora and Vegetation Surveys for Environmental Impact Assessment*, Government of Western Australia, Perth.
- European Commission 2007, *EUNIS Habitat Classification*, European Commission, Bruxelles, <http://www.eea.europa.eu/themes/biodiversity/eunis/eunis-habitat-classification>
- Gibson, N 2004, 'Flora and Vegetation of the Eastern Goldfields Ranges: Part 6. Mt Manning Range', *Journal of the Royal Society of Western Australia*, vol. 87, pp. 35–47.
- Gibson, N 2018, 'Availability of vegetation plot data in Western Australia: a reply to Gellie et al.', *Phytocoenologia*, vol. 48, pp. 321–324.
- Grime, JP 1987, 'Dominant and Subordinate Components of Plant Communities: Implications for Succession, Stability and Diversity', in AJ Gray & MJ Crawley (eds), *Colonization, Succession and Stability*, Blackwell Scientific Publications, Oxford.
- Hennekens, SM & Schaminée, JHJ 2001, 'Turboveg, a comprehensive database management system for vegetation data', *Journal of Vegetation Science*, vol. 12, pp. 589–591.
- Hennig, C 2015, *FPC: Flexible Procedures for Clustering*, R package version 2.1-10, computer software, <https://CRAN.R-project.org/package=fpc>
- Janssen, JAM, Rodwell, JS, Garcia Criado, M, Gubbay, S, Haynes, T, Nieto, A ... Valachovič, M 2016, *European Red List of Habitats. Part 2. Terrestrial and Freshwater Habitats*, European Union, Bruxelles.
- Legendre, P & Legendre, LF 2012, *Numerical Ecology*, 3rd edn, Elsevier, Amsterdam.
- Macintyre, PD, Van Niekerk, A, Dobrowolski, MP, Tsakalos, JL & Mucina, L 2018, 'Impact of ecological redundancy on the performance of machine learning classifiers in vegetation mapping', *Ecology and Evolution*, vol. 8, pp. 6728–6737.
- McCune, B & Mefford, MJ 1999, *PC-ORD. Multivariate Analysis of Ecological Data. Version 4*, MjM Software Design, Gleneden Beach.
- Mucina L 2009, *Jack Hills Iron Ore Project Area: Auditing of the Conflicting Vegetation Classifications and Vegetation Mapping of Plant Communities in Reports by Mattiske Consulting (2006) and Ecologia Environment (2009)*, Department of Environmental and Aquatic Sciences, Curtin University of Technology, Perth. <https://library.dbca.wa.gov.au/static/FullTextFiles/070359.pdf>

- Mucina, L 2010, 'Floristic-phytosociological approach, potential natural vegetation, and survival of prejudice', *Lazaroa*, vol. 31, pp. 7–16.
- Mucina, L, Bültmann, H, Dierssen, K, Theurillat, J-P, Raus, T, Čarni, ... Tichý, L 2016, 'Vegetation of Europe: hierarchical floristic classification system of vascular plant, bryophyte, lichen, and algal communities', *Applied Vegetation Science*, vol. 19, pp. 3–264.
- Mucina, L & Daniel, G 2013, *Vegetation Mapping in the Northern Kimberley, Western Australia*, Curtin University, Perth.
- Mucina, L & Rutherford, MC 2016, *The Vegetation of South Africa, Lesotho and Swaziland*, South African National Biodiversity Institute, Pretoria.
- Oksanen, J, Blanchet, FG, Friendly, M, Kindt, R, Legendre, P, McGlinn, D, ... Wagner H 2019, *Vegan: Community Ecology Package*, R Package version 2.5-4, computer software, <https://CRAN.R-project.org/package=vegan>
- Parks and Wildlife 2015a, *Priority Ecological Communities for Western Australia, Version 22, 6 June 2015*, Department of Parks and Wildlife, Perth.
- Parks and Wildlife 2015b, *List of Threatened Ecological Communities Endorsed by the Minister for the Environment, 25 June 2015*, Department of Parks and Wildlife, Perth.
- Podani, J 2000, *Introduction to the Exploration of Multivariate Biological Data*, Backhuys Publishers, Leiden.
- R Core Team 2019, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, <https://www.R-project.org/>
- Tichý, L, Chytrý, M, Hájek, M, Talbot, SS & Botta-Dukát, Z 2010, 'OptimClass: using species-to-cluster fidelity to determine the optimal partition in classification of ecological communities', *Journal of Vegetation Science*, vol. 21, pp. 287–299.
- Wildi, O 2017, *Dave: Functions for "Data Analysis in Vegetation Ecology"*, R package version 2.0, computer software, <https://CRAN.R-project.org/package=dave>