Integrating environmental variables by multivariate ordination enables the reliable estimation of mineland rehabilitation status

Markus Gastauer[1]*, Cecílio Frois Caldeira[1], Sílvio Junio Ramos[1], Leonardo Carreira Trevelin[1], Rodolfo Jaffé[1], Guilherme Oliveira[1], Mabel Patricia Ortiz Vera[1,2], Eder Pires[1], Flavia Louzeiro de Aguiar Santiago[3], Marco Aurélio Carbone Carneiro[3], Felipe Tadashi Asoa Coelho[4], Rosilene Silva[4], Pedro Walfir M. Souza-Filho[1], José-Oswaldo Siqueira[1]

[1]Instituto Tecnológico Vale, Rua Boaventura da Silva, 955, Umarizal, Belém, Pará, Brazil

[2]Universidade Federal de Pará, Programa de Pós-Graduação em Genética e Biologia Molecular, Rua Augusto Corrêa, 01, Guamá, Belém, Pará, Brazil

[3]Universidade Federal de Lavras, Departamento de Ciências do Solo, Lavras, MG, Brasil

[4]Vale - Diretoria de Ferrosos Centro Oeste, Rua Cabral, 1555, Centro, Corumbá, Mato Grosso do Sul, Brazil

*Corresponding author. Email: markus.gastauer@itv.org, Tel.: +55 91 3213 5554

Author contributions

MG developed this study with important contributions from CFC, SJR, FTAC, RS, PWMSF and JOS; LCT and RJ carried out the statistical analysis; GO, MPOV and EP conducted the metagenomic analysis; MACC and FLAS provided the biochemical soil attributes; and MG and CFC led the writing of the manuscript. All of the authors contributed critically to the drafts and gave final approval for publication.

Acknowledgements

This work was funded by Vale (Canga Plant & Soil to PWMSF and Canga Plant Genomics to GO), and received additional support by CNPq to SJR (305831/2016-0), RJ (301616/2017-5), GO (402756/2018-5, 307479/2016-1), PWMSF (306430/2013-5) and JOS (303580/2013-5), by CAPES to GO (88887.130628/2016-00), and by CABANA to GO, who was funded by the BBSRC under the Global Challenges Research Fund (GCRF) Growing Research Capability call, contract number BB/P027849/1 (www.ukri.org/research/global-challenges-research-fund/funded-projects/). We are grateful for the valuable comments of four anonymous reviewers.

Highlights

- A multivariate estimation of the environmental status of rehabilitating areas is presented
- The method integrates processes, diversity and structure in a single estimation
- Rehabilitation status is successfully estimated after iron ore mining in Corumbá
- The approach enables tests for biases regarding single variables or variable groups
- Defining indicators reduces the number of variables needed for further assessments

Abstract

Despite the wide variety of variables commonly applied to measure different aspects of rehabilitation, the assessment and subsequent definition of indicators of environmental rehabilitation status are not simple tasks. The main challenges are comparing rehabilitated sites with target ecosystems as well as integrating individual environmental and eventually collinear variables into a single tractable measure of the state of a system before effective indicators that track rehabilitation may be modeled. For that, a consensus is lacking regarding which and how many variables need to be surveyed. Our approach considered ecological processes, vegetation structure, and community diversity from nonrehabilitated, rehabilitating and reference sites. We applied this approach to a curated set of 32 environmental variables retrieved from nonrevegetated, rehabilitating and reference sites associated with iron ore mines from the Urucum Massif, Mato Grosso do Sul, Brazil. By integrating variables from a single attribute or the entire set of variables into a single estimation of rehabilitation status, the proposed multivariate approach is straightforward and able to adequately address collinearity among variables. The proposed approach allows for the identification of biases towards single variables, surveys or analyses, which is necessary to rank environmental variables regarding their importance to the assessment. Furthermore, we show that bootstrapping permitted the detection of the minimum number of environmental variables necessary to achieve reliable estimations of the rehabilitation status. Finally, we show that the proposed variable integration enables the definition of environmental indicators for more comprehensive monitoring of mineland rehabilitation. Thus, the proposed multivariate ordination represents a powerful tool to outline the benefits of rehabilitating sites for the maintenance of biodiversity and ecosystem functions and services provided that sufficient environmental variables related to ecological processes, diversity and vegetation structure are gathered from nonrehabilitated, rehabilitating and reference study sites. By identifying deviations from predicted rehabilitation trajectories and providing assessments for environmental agencies, this proposed multivariate ordination increases the effectiveness of (mineland) rehabilitation.

Keywords: environmental monitoring; ecological processes; functional diversity; environmental indicators; primers for environmental rehabilitation; Urucum Massif

1. Introduction

Decision making as well as timely and adequate interventions in complex systems require indicators for tracking the states of a system. The environmental rehabilitation of mining areas, which is intended to restitute ecosystems and their functionality to as close as possible to premining levels to reduce the overall impacts of mining operations (Perring et al., 2013), is affected by a myriad of forces operating at multiple spatiotemporal scales (Arroyo-Rodríguez et al., 2017; Ren et al., 2017). The monitoring of rehabilitation success and the definition of effective indicators of rehabilitation status are necessary for monitoring whether rehabilitation goals are achieved (Gastauer et al., 2018a; Latawiec and Agol, 2016). Moreover, monitoring can guarantee institutional tractability of rehabilitation activities (Lamb et al., 2015) and may produce positive feedback for environmental rehabilitation practices (Lechner et al., 2018). In this context, we define rehabilitation status as the portion of environmental advances achieved at rehabilitating sites in relation to the overall trajectory necessary to restitute the biodiversity and ecosystem functioning of nonrevegetated sites to predisturbance levels. Despite their importance, the development of such metrics is challenging, as there is no consensus about which and how many variables to survey and how to integrate different variables into a single tractable measure of the state of a system (Gastauer et al., 2018b; Kollmann et al., 2016; Perring et al., 2015).

In 2004, the Society of Ecological Restoration (SER, 2004) postulated nine primers to evaluate environmental rehabilitation success. From that list, Ruiz-Jaen and Aide (2005a) recommended the inclusion of two variables from each of the following attributes to achieve a reliable evaluation

Peer-reviewed version available at Journal of Environmental Management 2020; doi:10.1016/j.jenvman.2019.109894

of revegetation success: vegetation structure, diversity of restituted communities and ecological processes. Whereas vegetation structure may comprise vegetation cover, tree density, and the basal area of trees, diversity may be captured as the taxonomic, phylogenetic and functional diversity of different taxonomic groups (Brancalion and Holl, 2016; Cadotte et al., 2015). Additionally, the inclusion of species composition in restoration success measures makes these measures resilient to compositional failures (Reid, 2015). Ecological processes encompass intrinsic ecosystem characteristics that maintain the integrity of an ecosystem (Birkhofer et al., 2015), i.e., the mechanisms by which ecosystems provide ecosystem services, including decomposition, (primary) production, nutrient and water cycling, nutrient and energy fluxes, community dynamics, plant and animal successions, energy flow, species interactions, the movements of organisms and natural disturbance regimes (Bennett et al., 2009).

Although most rehabilitation status assessments comprise surrogates for all three attributes proposed by Ruiz-Jaen and Aide (Wortley et al., 2013, but see Gatica-Saavedra et al., 2017), the number of different ecosystem characteristics surveyed to monitor rehabilitation activities differs significantly among available studies (e.g., Derhé et al., 2016; Kollmann et al., 2016; Vickers et al., 2012). Questions about how to integrate different variables into a single measure of rehabilitation status are rarely addressed in the literature (Gastauer et al., 2018b; McDonald et al., 2016). Instead, most studies derive rehabilitation status from individual variables without any form of integration (e.g., Jiao et al., 2012; Monie et al., 2013; Sasaki et al., 2018) or define rehabilitation status arbitrarily as a single variable, such as the species richness of the restituted communities (e.g., Londe et al., 2017); however, different variables may estimate rehabilitation status to different magnitudes. Not integrating different environmental variables in a single conclusive figure impedes the subsequent modeling of suitable indicators for tracking the rehabilitation status, statistical testing for the number of environmental variables necessary for reliable estimations and testing for biases regarding certain variables. Additive approaches that define rehabilitation status as the mean score of different variables (Ruiz-Jaén and Aide, 2005b) are restricted to uncorrelated variable sets (Rosenfield and Müller, 2018). As far as we know, multivariate ordination approaches, which are common in community ecology and other ecological fields (e.g., Dray et al., 2012), have rarely been applied to compute rehabilitation status (but see Ruiz-Jaén and Aide, 2005b), even though they may represent effective tools for

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 September 2019 doi:10.20944/preprints201909.0206.v1

Peer-reviewed version available at Journal of Environmental Management 2020; doi:10.1016/j.jenvman.2019.109894

addressing collinearity, sample sufficiency and the importance of single variables to achieve reliable estimations of the rehabilitation state.

In order to overcome these challenges, we propose here a multivariate approach to integrate different environmental variables in a single estimate of mineland rehabilitation status. We develop this approach using a curated set of 32 environmental variables, including surrogates for ecological processes, community diversity and vegetation structure retrieved from nonrevegetated, rehabilitating and natural reference sites associated with iron mines from the Urucum Massif, Mato Grosso do Sul, Brazil. To outline whether all variables are equally important, we checked for biases towards single variables or groups of variables. Additionally, we examined the sufficiency of the sampling effort, i.e., the number of variables used for the assessment, before we identified the variables or the combination of variables that were possible surrogates for (indicators of) the postmining rehabilitation status in the region.

2. Methods

2.1. Rehabilitation status estimation

We define the rehabilitation status as the environmental modifications achieved by rehabilitating sites in relation to both undisturbed reference and nonrevegetated sites, being 0 in nonrevegetated sites and 100, when ecological processes, vegetation structure and community diversity correspond to levels achieved in undisturbed reference sites. For single variables, the status is computed as shown in equation (1).

$$Rehabilitationstatus = \left(1 - \frac{\Delta Reh - Ref}{\Delta NR - Ref}\right) * 100 \qquad (1)$$

Δ*Reh-Ref* is the difference between the scores achieved by rehabilitating a site and the mean scores from reference sites; Δ*NR-Ref* is the mean difference between nonrevegetated and reference sites.

To integrate available environmental variables in a single estimation, we propose a multivariate ordination method to remove collinearity among variables before we compute the quotient between the rehabilitation-reference site distances and the mean nonrevegetated-reference site

Peer-reviewed version available at *Journal of Environmental Management* **2020**; doi:10.1016/j.jenvman.2019.109894

distance. Because environmental variable sets necessary for the estimation of rehabilitation status may include continuous (e.g., estimation of biomass or enzyme activities) as well as count data (e.g., observed and estimated species richness), the computation of principal components based on Euclidean distances among the plots is not appropriate (Ramette, 2007). Instead, we compute Gower distances among sites based on the variable set, which are able to address variables of different natures (e.g., 'daisy' function in the cluster package, Maechler et al., 2018). Based on this distance matrix, we conduct principal coordinate analyses using the 'cmdscale' function. We then compute the mean Euclidean distances between the coordinates of each plot and the centroid coordinates from all the reference plots among all the principal coordinates with positive eigenvalues. Then, the quotient of the mean plot-reference site distance and the mean nonrehabilitated reference distance represents the portion of environmental advances still necessary to achieve predisturbance levels (in relation to the overall trajectory necessary). Subtracting this value from 1 returns the portion of environmental success already achieved by the rehabilitation trajectory, i.e., the environmental rehabilitation status. Thus, the mean rehabilitation status of all nonrehabilitated sites will be 0, and the mean score of all reference plots will equal to 100%; the rehabilitated sites receive scores relative to the reference study sites.

2.2. Verifying the influences of the variables and the minimum number of variables for the estimation

To check how single variables or different combinations of variables (e.g., ecological processes, vegetation structure or community diversity) influence assessments of rehabilitation status, the rehabilitation status is estimated without the use of focus variable(s). For each reduced variable set, the estimation computed for the entire set of variables is then fitted with the estimation from the reduced set. For each comparison, the minimum and maximum residuals from the models are interpreted as a measure of the magnitude of deviations that might occur when dropping single variables or variable combinations. The minimum and maximum residuals indicate the extent to which the values of single sites might be over- or underestimated when the rehabilitation status is computed based on a reduced set of variables. The coefficient of determination (R^2), although inappropriate for model selection, indicates the predictability of the complete rehabilitation status with a reduced set of variables. The larger the R^2 is, the lower the

Peer-reviewed version available at *Journal of Environmental Management* **2020**; doi:10.1016/j.jenvman.2019.109894

rehabilitation status differences are that are computed between the complete and reduced sets of variables.

To outline the minimum number of variables for a reliable estimation of rehabilitation status, the number of variables for the estimation is bootstrapped. For each number of variables, the rehabilitation status is estimated 1,000 times based on randomly selected variable combinations. As above, the analysis of minimum and maximum residuals, as well as the coefficients of determination, enables to outline over- or underestimation as well as predictability when the rehabilitation status is computed based on a reduced set of variables.

2.3. Identification of indicators for the estimation of rehabilitation status

To determine which variables or variable combinations used for estimation are the most appropriate indicators of environmental rehabilitation status, global models with the estimated rehabilitation status as the response variable and all independent environmental variables are built. The Akaike information criterion (AIC) is used to select the variable combinations that best fit the estimation (Symonds and Moussalli, 2011). For that, all of the models containing noncollinear variables ($r < 0.6$) using the 'dredge' function (MuMIn package) should be compared (Bartón, 2015). All of the models with ΔAICc values smaller than 2 in relation to the best model are considered equally parsimonious.

2.4. Case study

To test our method, we analyzed a rehabilitating chronosequence after open-cast iron ore mining situated between 19º10'-19º13' S and 57º33'-57º37' W in the Urucum Massif, Corumbá and Ladário municipalities, Mato Grosso do Sul, midwestern Brazil (Gastauer et al., 2019). The climate of the region is tropical warm (Aw in the Köppen classification) and is characterized by dry winters and rainy summers, with a mean annual temperature of 25°C and a mean annual precipitation amount of 1,070 mm. A mosaic of seasonal deciduous and semideciduous forests, as well as different savanna formations, characterize the vegetation of the region. While forests cover the slopes and the neighborhoods of water courses, different savanna formations, ranging from arborized physiognomies to treeless grasslands (open savanna formation), may be found

Peer-reviewed version available at *Journal of Environmental Management* **2020**; doi:10.1016/j.jenvman.2019.109894

on the upper sections of the massif. Iron ore reserves are covered by *canga* vegetation, an open savanna formation composed of a characteristic set of species (Takahasi, 2010).

Open-cast iron ore mining occurs after vegetation suppression and the removal of the topsoil layers. For rehabilitation, the mined soils are limed and fertilized, and biomass from the suppressed areas is added. Plants rescued from the suppressed areas and seedlings of native species produced in tree nurseries are planted for rehabilitation. Additionally, seed mixtures of native species collected from the vegetation remnants are applied. If required, further activities, such as the replantation of seedlings, the further application of seeds, and the combating of alien invasive species, are executed.

We sampled three bare soil, i.e., nonrevegetated, seven rehabilitating (two-, three- and six-year-old stands), and three reference sites (Fig. 1). Because minelands were originally covered by *canga* vegetation and environmental rehabilitation after iron ore extraction is intended to restitute this specific formation, we placed undisturbed reference sites in remnants of this unique vegetation type. Bare soil or nonrevegetated sites were minelands immediately before beginning of rehabilitation activities. At each study site (Fig. 1), three 10×10 m plots were installed in homogeneous, representative vegetation without signs of external disturbances, comprising a total of 39 sampling plots. This sample design guaranteed sufficient revegetation site samples (Supplementary Material 1). Within the plots, the vegetation and soils were sampled to obtain a total of 32 variables related to the three attributes proposed by Ruiz-Jaen & Aide (2005a) (Table 1). For comments on variable selection and the methodological details, please see Supplementary Material 2.

Unless otherwise stated, all of the statistical analyses were carried out in the R environment (R Core Team, 2018). Before estimating the rehabilitation status, we built a correlation matrix between all of the variables using the ‘corrplot’ function. Then, we outlined the differences in all 32 variables among the nonrevegetated, rehabilitating and reference plots using Kruskal-Wallis tests followed by post hoc Dunn tests.

Then, we estimated the rehabilitation status for the entire set of variables and each attribute proposed by Ruiz-Jaen and Aide (2005a), i.e., ecological processes, vegetation structure and diversity. To outline the influences of single variables or groups of variables and the minimum number of variables for a reliable estimation, we fitted mixed-effect models using the ‘lmer’

Peer-reviewed version available at Journal of Environmental Management 2020; doi:10.1016/j.jenvman.2019.109894

function in the lme4 package (Bates et al., 2014). The random effect was the study site, as the rehabilitation activities and abiotic stand properties may differ between sites, and the arrangement of the three plots within small study sites did not guarantee complete sample independence. We checked the dependent variable against truncated distribution (Greene, 2012) and evaluating for endogeneity (Wooldridge, 2009), before we computed the coefficient of determination (R^2) for the mixed-effect models using the 'r.squaredGLMM' function in the MuMIn package (Johnson, 2014). To model effective indicators for more comprehensive monitoring, we built global mixed-effect models with the estimated rehabilitation status as the response variable, 32 independent environmental variables, and study site as the random effect. We considered only the models with two or fewer variables to avoid overfitting.

3. Results

The Spearman correlation matrix revealed that the different variables regarding the ecological processes, vegetation structure and community diversity were, in most cases, positively correlated with each other (Supplementary Material 3), so additive approaches could not be applied to estimate rehabilitation status.

As expected, most of the variables derived from vegetation were higher in the reference than in the rehabilitating sites, and the nonrehabilitated sites showed the lowest scores (Table 1). The percentage of zoochoric species in the rehabilitating sites was not different from the percentage detected in the reference sites. Additionally, the biochemical soil attributes (variables related to C and N cycles and enzyme activities) showed no difference between the nonrehabilitated and rehabilitating sites, but the values were higher in the reference sites than in the other two types of sites. The microorganism richness was lower in the rehabilitating than in both the nonrehabilitated and reference study sites; the Shannon diversity of microorganisms exhibited no differences. Thus, single variables (except the variables with mean values in the rehabilitated sites outside the interval given by the mean scores from the nonrehabilitated and reference study sites) were used to estimate the mean environmental rehabilitation status after iron mining in the Urucum Massif. The status was estimated to be between 0.7 and 93.23% in relation to the reference sites (Table 1).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 September 2019 doi:10.20944/preprints201909.0206.v1

Peer-reviewed version available at Journal of Environmental Management 2020; doi:10.1016/j.jenvman.2019.109894

Principal coordinate analyses showed significant differences among nonrehabilitated, rehabilitating and reference study sites for the entire variable set as well as for variables related to ecological processes, vegetation structure and diversity (Supplementary Material 4, 5).

The nonrehabilitated sites received rehabilitation scores less than 10% in all four comparisons, and these scores were derived from the distances among the plots within PCoA space (Fig. 2). In all of the comparisons, the reference sites achieved scores higher than 0.9. Considering the variables regarding only the ecological processes, the estimated rehabilitation status of the rehabilitated study sites varied between 6.74 and 52.88% (mean 32.10%). The diversity attributes indicated a higher rehabilitation status for the rehabilitating study sites than further comparisons, ranging from 67.71 to 87.71% (mean 78.54%). For most of the rehabilitating sites, 60 to 80% of the vegetation structure was rehabilitated, although site 2 showed values less than 50%. The environmental status ranged between 50.46 and 73.90% (mean 64.39%, Fig. 2) when all 32 variables were used.

With coefficients of determination greater than 0.99 and absolute maximum or minimum residual values, which are a measure of the over- or underestimation of single plots, less than five, the withdrawal of single variables showed only a nonsignificant influence on the rehabilitation status estimation (Supplementary Material 6). The withdrawal of attributes, entire surveys or analyses from the rehabilitation status estimation showed only marginal effects in most cases ($R^2 > 0.98$). However, due to the low coefficient of determination (~0.95), i.e., a large influence on the outcomes of the estimation, the vegetation surveys were indispensable for the reliable estimation of rehabilitation status (Fig. 3).

Randomly increasing the number of variables used to estimate the rehabilitation status increased the strength of the mean correlation with the estimation of the rehabilitation status using all of the variables; a random subset of 12 variables was necessary to achieve sufficient prediction power (mean R^2 greater than 0.95, Fig. 4).

Two models were selected after fitting the rehabilitation status estimation by using different combinations of all 32 variables (considering only models with two or fewer variables). Both models included vegetation cover plus either RaoQ, a measure of functional diversity, or phylogenetic diversity (Table 2). Both models showed R^2 values greater than 0.99; high

Peer-reviewed version available at *Journal of Environmental Management* **2020**; doi:10.1016/j.jenvman.2019.109894

correlation coefficients were not due to truncated regressions or endogeneity of the dependent variables (Electronic Supplementary Materials 7 and 8).

4. Discussion

The integration of 32 environmental variables in a unique estimation of rehabilitation status by the multivariate ordination method proposed in this study was straightforward because the principal coordinate analysis based on Gower distances represents an adequate way to integrate different types (continuous and count data) of collinear environmental variables into a single estimation of rehabilitation status. Estimations can be carried out for different attributes, e.g., as proposed by Ruiz and Jaen (2005a), or for the entire set of variables. The multivariate ordination can integrate variables with scores from rehabilitating sites that are outside the intervals detected in nonrevegetated and reference sites (some biological or biochemical soil properties in our case), indicating deterioration or over-rehabilitation in analyses that consider single variables individually. Furthermore, the proposed procedure enables the evaluation for biases against single variables or groups of variables and allows for issues regarding which variables are most important for environmental monitoring to be addressed. The proposed approach also provides a statistical tool (bootstrapping) to outline the minimum number of variables required for a reliable assessment and permits the derivation of useful indicators for more comprehensive monitoring of rehabilitation activities. By producing practical feedback, i.e., identifying deviations from predicted rehabilitation trajectories (e.g., along time series or chronosequences) and providing reliable assessments, the proposed method is thus able to contribute to more effective rehabilitation, given that sufficient environmental variables from nonrehabilitated and reference study sites are available.

The rehabilitation outcomes of our case study were highly variable from the perspective of single variables, indicating deterioration (e.g., richness of microorganism communities, glomalin contents), no rehabilitation (further biochemical soil attributes) or restitution of up to 80% of predisturbance levels (functional diversity measures). Nonetheless, the multivariate estimation of rehabilitation status showed low bias towards single variables, attributes, surveys and analysis. The high coefficients of determination and low minimum and maximum residuals between the estimation computed from the entire set of variables and the computations lacking single variables

Peer-reviewed version available at *Journal of Environmental Management* **2020**; doi:10.1016/j.jenvman.2019.109894

indicate that the estimation is not biased by any single variable. Concerning different analyses and surveys, the exclusion of only the entire vegetation survey caused the inaccurate prediction of the rehabilitation status when the entire set of variables was used for the estimation. The importance of the vegetation surveys was expected, as vegetation structure was one of the three attributes proposed by Ruiz-Jaen and Aide (2005a), rendering the attribute mandatory for rehabilitation status assessments.

Furthermore, the bootstrapping of the variables showed estimation stability in our case (mean correlation coefficients > 0.95) when more than 12 variables (30% of available variables) were present in the analysis. As the estimation became reliable with a much smaller subset of variables than those applied, we can attest sample sufficiency here, but highlight that the minimum number of variables for a reliable assessment in our case is larger than that postulated by Ruiz-Jaen and Aide (2005a). Additionally, this phenomenon and the low influence of the variable groups indicate that the inclusion of additional analyses that are not yet available but intended to capture a more holistic picture of environmental status may influence the overall estimation only marginally as well, highlighting the reliability of the produced estimation.

The selection of appropriate reference sites is critical for the reliable estimation of rehabilitation status. Reference sites from ecosystems that will not or cannot be achieved by rehabilitation activities due to abiotic or biotic changes in the system (e.g., Suding et al., 2004) reduce the quality of the estimation and may produce meaningless results, further influencing downstream analyses such as the identification of useful rehabilitation status indicators. Unpredictable rehabilitation outcomes, including the emergence of novel ecosystems in rehabilitated minelands (Gastauer et al., 2018a), may require the use of multiple reference ecosystems to minimize the risk of relating rehabilitation outcomes to inappropriate reference systems in further assessments (Palmer et al., 2016).

In our rehabilitation case study, the mean rehabilitation status of the open savanna formations after iron mining in the Urucum Massif amounted to approximately 60%, indicating the return of ecosystem services in the examined minelands. Diversity indicated a higher rehabilitation status than indicated by structural parameters or ecological processes, although the structure was expected to restore more quickly than processes and diversity (Laughlin et al., 2017; Suganuma and Durigan, 2014). In particular, the basal area and biomass scores in the rehabilitating sites

Peer-reviewed version available at *Journal of Environmental Management* **2020**; doi:10.1016/j.jenvman.2019.109894

were far from the reference site values; thus, long periods are required for the rehabilitation of these attributes (Chazdon et al., 2016). Furthermore, some of the variables, including soil organic matter, in the rehabilitating sites did not differ from those in the nonrehabilitated sites, although they were expected to increase after the beginning of rehabilitation (Yuan et al., 2018). High temperatures and low soil humidity may contribute to the rapid decomposition of applied biomass; additionally, short periods of vegetation establishment might be insufficient for an increase in soil organic matter. The lack of organic matter may reduce microbial richness, activity and diversity, including mutualistic, glomalin-producing mycorrhiza necessary for the establishment of certain plant species (Singh et al., 2016). Thus, further rehabilitation activities in these areas should focus on increasing the soil organic matter content to enhance the actual rehabilitation status by increasing microbial activity.

For the presented case from the Corumbá Massif, the suitability of vegetation cover plus the functional or phylogenetic diversity were the best indicators of the environmental rehabilitation status in the region. Together, they predict environmental rehabilitation status with sufficient precision (> 99%). The importance of phylogenetic and functional diversity for environmental assessments as the linkage between traditional taxonomic measures of biodiversity and ecosystem functioning has previously been assumed (e.g., Díaz-García et al., 2017; Gastauer et al., 2018b). Thus, prospective or more spatially comprehensive assessments of rehabilitation status after iron mining in the Urucum Massif may focus on a much smaller subset of variables. Application of the proposed methodology to further (mineland) rehabilitation projects will contribute to more effective monitoring of environmental rehabilitation activities.

5. Conclusions

Addressing collinearity, the proposed multivariate ordination approach is a straightforward method to integrate different environmental variables in a unique estimation of rehabilitation status, given that environmental variables regarding ecological processes, vegetation structure, and diversity from chronosequences including nonrehabilitated, rehabilitating and undisturbed reference sites are available. For our Corumbá study case, the proposed method attributes that, at average, about two thirds of environmental rehabilitation has already been concluded, although

Peer-reviewed version available at *Journal of Environmental Management* **2020**; doi:10.1016/j.jenvman.2019.109894

reference levels have not yet been achieved. To enhance the actual rehabilitation status, further rehabilitation activities should focus on increasing the soil organic matter content.

By testing for biases against single variables or variable groups or bootstrapping the number of variables, the proposed approach provides statistical support to answer key questions about which and how many variables are necessary for a reliable estimation of rehabilitation status. Additionally, the estimation permits the prospection of useful indicators able to track the status of the rehabilitated systems to facilitate further monitoring activities. Thus, the methodological approach presented here provides a practical framework to comply with environmental legislation regarding the monitoring of mineland rehabilitation activities. The statistical tools to define effective indicators for the overall rehabilitation status contribute to more effective monitoring of future (mineland) rehabilitation projects. However, the importance of single environmental variables, the minimum number of variables necessary to achieve a consistent estimation and the indicators that are effective for comprehensive monitoring may differ among cases, indicating the need for further research on these topics.

References

Arroyo-Rodríguez, V., Melo, F.P.L., Martínez-Ramos, M., Bongers, F., Chazdon, R.L., Meave, J.A., Norden, N., Santos, B.A., Leal, I.R., Tabarelli, M., 2017. Multiple successional pathways in human-modified tropical landscapes: new insights from forest succession, forest fragmentation and landscape ecology research. Biol. Rev. 92, 326–340. https://doi.org/10.1111/brv.12231

Bartón, K., 2015. MuMIn. R package version 1.13.4.

Bates, D., Meachler, M., Bolker, B., Walker, S., 2014. Lme4: linear mixed-effects models using Eigen and S4. R package version 1.1-7.

Bennett, A.F., Haslem, A., Cheal, D.C., Clarke, M.F., Jones, R.N., Koehn, J.D., Lake, P.S., Lumsden, L.F., Lunt, I.D., Mackey, B.G., Nally, R. Mac, Menkhorst, P.W., New, T.R., Newell, G.R., O'Hara, T., Quinn, G.P., Radford, J.Q., Robinson, D., Watson, J.E.M., Yen, A.L., 2009. Ecological processes: A key element in strategies for nature conservation.

Ecol. Manag. Restor. 10, 192–199. https://doi.org/10.1111/j.1442-8903.2009.00489.x

Birkhofer, K., Diehl, E., Andersson, J., Ekroos, J., Früh-Müller, A., Machnikowski, F., Mader, V.L., Nilsson, L., Sasaki, K., Rundlöf, M., Wolters, V., Smith, H.G., 2015. Ecosystem services—current challenges and opportunities for ecological research. Front. Ecol. Evol. 2, 1–12. https://doi.org/10.3389/fevo.2014.00087

Brancalion, P.H.S., Holl, K.D., 2016. Functional composition trajectory: A resolution to the debate between Suganuma, Durigan, and Reid. Restor. Ecol. 24, 1–3. https://doi.org/10.1111/rec.12312

Cadotte, M.W., Arnillas, C.A., Livingstone, S.W., Yasui, S.L.E., 2015. Predicting communities from functional traits. Trends Ecol. Evol. 30, 510–511. https://doi.org/10.1016/j.tree.2015.07.001

Chazdon, R.L., Broadbent, E.N., Rozendaal, D.M.A., Bongers, F., Zambrano, A.M.A., Aide, T.M., Balvanera, P., Becknell, J.M., Boukili, V., Brancalion, P.H.S., Craven, D., Almeida-Cortez, J.S., Cabral, G.A.L., de Jong, B., Denslow, J.S., Dent, D.H., DeWalt, S.J., Dupuy, J.M., Durán, S.M., Espírito-Santo, M.M., Fandino, M.C., César, R.G., Hall, J.S., Hernández-Stefanoni, J.L., Jakovac, C.C., Junqueira, A.B., Kennard, D., Letcher, S.G., Lohbeck, M., Martínez-Ramos, M., Massoca, P., Meave, J.A., Mesquita, R., Mora, F., Muñoz, R., Muscarella, R., Nunes, Y.R.F., Ochoa-Gaona, S., Orihuela-Belmonte, E., Peña-Claros, M., Pérez-García, E.A., Piotto, D., Powers, J.S., Rodríguez-Velazquez, J., Romero-Pérez, I.E., Ruíz, J., Saldarriaga, J.G., Sanchez-Azofeifa, A., Schwartz, N.B., Steininger, M.K., Swenson, N.G., Uriarte, M., van Breugel, M., van der Wal, H., Veloso, M.D.M., Vester, H., Vieira, I.C.G., Bentos, T.V., Williamson, G.B., Poorter, L., 2016. Carbon sequestration potential of second-growth forest regeneration in the Latin American tropics. Sci. Adv. 2, e1501639. https://doi.org/10.1126/sciadv.1501639

Cristescu, R.H., Frère, C., Banks, P.B., 2012. A review of fauna in mine rehabilitation in Australia: Current state and future directions. Biol. Conserv. 149, 60–72. https://doi.org/10.1016/j.biocon.2012.02.003

Derhé, M.A., Murphy, H., Monteith, G., Menéndez, R., 2016. Measuring the success of reforestation for restoring biodiversity and ecosystem functioning. J. Appl. Ecol. 53, 1714–1724. https://doi.org/10.1111/1365-2664.12728

Peer-reviewed version available at *Journal of Environmental Management* **2020**; doi:10.1016/j.jenvman.2019.109894

Díaz-García, J.M., Pineda, E., López-Barrera, F., Moreno, C.E., 2017. Amphibian species and functional diversity as indicators of restoration success in tropical montane forest. Biodivers. Conserv. 26, 2569–2589. https://doi.org/10.1007/s10531-017-1372-2

Dray, S., Pélissier, R., Couteron, P., Fortin, M.-J., Legendre, P., Peres-Neto, P.R., Bellier, E., Bivand, R., Blanchet, F.G., De Cáceres, M., Dufour, A.-B., Heegaard, E., Jombart, T., Munoz, F., Oksanen, J., Thioulouse, J., Wagner, H.H., 2012. Community ecology in the age of multivariate multiscale spatial analysis. Ecol. Monogr. 82, 257–275. https://doi.org/10.1890/11-1183.1

Gastauer, M., Souza Filho, P.W.M., Ramos, S.J., Caldeira, C.F., Silva, J.R., Siqueira, J.O., Furtini Neto, A.E., 2018a. Mine land rehabilitation in Brazil: Goals and techniques in the context of legal requirements. Ambio. https://doi.org/10.1007/s13280-018-1053-8

Gastauer, M., Silva, J.R., Caldeira Junior, C.F., Ramos, S.J., Souza Filho, P.W.M., Furtini Neto, A.E., Siqueira, J.O., 2018b. Mine land rehabilitation: Modern ecological approaches for more sustainable mining. J. Clean. Prod. 172, 1409–1422. https://doi.org/10.1016/j.jclepro.2017.10.223

Gastauer, M., Vera, M.P.O., de Souza, K.P., Pires, E.S., Alves, R., Caldeira, C.F., Ramos, S.J., Oliveira, G., 2019. A metagenomic survey of soil microbial communities along a rehabilitation chronosequence after iron ore mining. Sci. Data 6, 190008.

Gatica-Saavedra, P., Echeverría, C., Nelson, C.R., 2017. Ecological indicators for assessing ecological success of forest restoration: a world review. Restor. Ecol. 25, 850–857. https://doi.org/10.1111/rec.12586

Greene, W.H., 2012. Econometric Analysis. Pearson, Upper Saddle River.

Jiao, J., Zhang, Z., Bai, W., Jia, Y., Wang, N., 2012. Assessing the Ecological Success of Restoration by Afforestation on the Chinese Loess Plateau. Restor. Ecol. 20, 240–249. https://doi.org/10.1111/j.1526-100X.2010.00756.x

Johansen, K., Erskine, P.D., McCabe, M.F., 2019. Using Unmanned Aerial Vehicles to assess the rehabilitation performance of open cut coal mines. J. Clean. Prod. 209, 819–833. https://doi.org/https://doi.org/10.1016/j.jclepro.2018.10.287

Johnson, P.C.D., 2014. Extension of Nakagawa & Schielzeth's R2GLMM to random slopes models. Methods Ecol. Evol. 5, 944–946. https://doi.org/10.1111/2041-210X.12225

Peer-reviewed version available at *Journal of Environmental Management* **2020**; doi:10.1016/j.jenvman.2019.109894

Kollmann, J., Meyer, S.T., Bateman, R., Conradi, T., Gossner, M.M., de Souza Mendonça, M., Fernandes, G.W., Hermann, J.M., Koch, C., Müller, S.C., Oki, Y., Overbeck, G.E., Paterno, G.B., Rosenfield, M.F., Toma, T.S.P., Weisser, W.W., 2016. Integrating ecosystem functions into restoration ecology—recent advances and future directions. Restor. Ecol. 24, 722–730. https://doi.org/10.1111/rec.12422

Lamb, D., Erskine, P.D., Fletcher, A., 2015. Widening gap between expectations and practice in Australian minesite rehabilitation. Ecol. Manag. Restor. 16, 186–195. https://doi.org/10.1111/emr.12179

Latawiec, A., Agol, D., 2016. Sustainability Indicators in Practice. https://doi.org/https://doi.org/

Laughlin, D.C., Strahan, R.T., Moore, M.M., Fulé, P.Z., Huffman, D.W., Covington, W.W., 2017. The hierarchy of predictability in ecological restoration: are vegetation structure and functional diversity more predictable than community composition? J. Appl. Ecol. 54, 1058–1069. https://doi.org/10.1111/1365-2664.12935

Lechner, A.M., Arnold, S., McCaffrey, N.B., Gordon, A., Erskine, P.D., Gillespie, M.J., Mulligan, D.R., 2018. Applying modern ecological methods for monitoring and modelling mine rehabilitation success. From Start to Finish – a Life-of-mine Perspect. 109–116.

Londe, V., Sousa, H.C.D., Kozovits, A.R., 2017. Key plant indicators for monitoring areas undergoing restoration: A case study at the Das Velhas River, southeast Brazil. Ecol. Eng. 103, 191–197. https://doi.org/10.1016/j.ecoleng.2017.04.012

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K., 2018. cluster: Cluster Analysis Basics and Extensions. R package version 2.0.7-1.

McDonald, T., Jonson, J., Dixon, K.W., 2016. National standards for the practice of ecological restoration in Australia. Restor. Ecol. 24, S4–S32. https://doi.org/10.1111/rec.12359

Monie, K., Florentine, S., Palmer, G., 2013. Recruitment and functionality traits as bioindicators of ecological restoration success in the Lurg hills district, Victoria, Australia. Ecol. Process. 2, 1–11. https://doi.org/10.1186/2192-1709-2-27

Palmer, M.A., Zedler, J.B., Falk, D.A., 2016. Ecological Theory and Restoration Ecology BT - Foundations of Restoration Ecology, in: Palmer, M.A., Zedler, J.B., Falk, D.A. (Eds.), . Island Press/Center for Resource Economics, Washington, DC, pp. 3–26. https://doi.org/10.5822/978-1-61091-698-1_1

Peer-reviewed version available at Journal of Environmental Management 2020; doi:10.1016/j.jenvman.2019.109894

Perring, M.P., Standish, R.J., Hobbs, R.J., 2013. Incorporating novelty and novel ecosystems into restoration planning and practice in the 21st century. Ecol. Process. 2, 1–8. https://doi.org/10.1186/2192-1709-2-18

Perring, M.P., Standish, R.J., Price, J.N., Craig, M.D., Erickson, T.E., Ruthrof, K.X., Whiteley, A.S., Valentine, L.E., Hobbs, R.J., 2015. Advances in restoration ecology: Rising to the challenges of the coming decades. Ecosphere 6. https://doi.org/10.1890/ES15-00121.1

R Core Team, 2018. R: A language and environment for statistical computing.

Ramette, A., 2007. Multivariate analyses in microbial ecology. FEMS Microbiol. Ecol. 62, 142–160. https://doi.org/10.1111/j.1574-6941.2007.00375.x

Reid, L.J., 2015. Indicators of success should be sensitive to compositional failures: reply to Suganuma and Durigan. Restor. Ecol. 23, 519–520. https://doi.org/10.1111/rec.12254

Ren, Y., Lü, Y., Fu, B., Zhang, K., 2017. Biodiversity and Ecosystem Functional Enhancement by Forest Restoration: A Meta-analysis in China. L. Degrad. Dev. 28, 2062–2073. https://doi.org/10.1002/ldr.2728

Rosenfield, M.F., Müller, S.C., 2018. Assessing ecosystem functioning in forests undergoing restoration. Restor. Ecol. 0. https://doi.org/10.1111/rec.12828

Ruiz-Jaen, M.C., Aide, T.M., 2005a. Restoration Success: How Is It Being Measured? Restor. Ecol. 13, 569–577. https://doi.org/10.1111/j.1526-100X.2005.00072.x

Ruiz-Jaén, M.C., Aide, T.M., 2005b. Vegetation structure, species diversity, and ecosystem processes as measures of restoration success. For. Ecol. Manage. 218, 159–173. https://doi.org/10.1016/j.foreco.2005.07.008

Sasaki, T., Ishii, H., Morimoto, Y., 2018. Evaluating restoration success of a 40-year-old urban forest in reference to mature natural forest. Urban For. Urban Green. 32, 123–132. https://doi.org/https://doi.org/10.1016/j.ufug.2018.04.008

Singh, A.K., Rai, A., Singh, N., 2016. Effect of long term land use systems on fractions of glomalin and soil organic carbon in the Indo-Gangetic plain. Geoderma 277, 41–50. https://doi.org/https://doi.org/10.1016/j.geoderma.2016.05.004

Society for Ecological Restoration International Science & Policy Working Group, S., 2004. The SER International primer on ecological restoration 2, 206–207. https://doi.org/S34

Suding, K.N., Gross, K.L., Houseman, G.R., 2004. Alternative states and positive feedbacks in

Peer-reviewed version available at *Journal of Environmental Management* **2020**; doi:10.1016/j.jenvman.2019.109894

restoration ecology. Trends Ecol. Evol. 19, 46–53. https://doi.org/https://doi.org/10.1016/j.tree.2003.10.005

Suganuma, M.S., Durigan, G., 2014. Indicators of restoration success in riparian tropical forests using multiple reference ecosystems. Restor. Ecol. 23, 238–251. https://doi.org/10.1111/rec.12168

Symonds, M.R.E., Moussalli, A., 2011. A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. Behav. Ecol. Sociobiol. 65, 13–21. https://doi.org/10.1007/s00265-010-1037-6

Takahasi, A., 2010. Ecologia da vegetação em bancadas lateríticas em Corumbá. Universidade de São Paulo. https://doi.org/10.11606/T.41.2010.tde-30092010-160712

Vickers, H., Gillespie, M., Gravina, A., 2012. Assessing the development of rehabilitated grasslands on post-mined landforms in north west Queensland, Australia. Agric. Ecosyst. Environ. 163, 72–84. https://doi.org/10.1016/j.agee.2012.05.024

Wooldridge, J.M., 2009. Introductory Econometrics: A Modern Approach, Fourth ed. ed. South-Western.

Wortley, L., Hero, J.-M., Howes, M., 2013. Evaluating Ecological Restoration Success: A Review of the Literature. Restor. Ecol. 21, 537–543. https://doi.org/10.1111/rec.12028

Yuan, Y., Zhao, Z., Li, X., Wang, Y., Bai, Z., 2018. Characteristics of labile organic carbon fractions in reclaimed mine soils: Evidence from three reclaimed forests in the Pingshuo opencast coal mine, China. Sci. Total Environ. 613–614, 1196–1206. https://doi.org/https://doi.org/10.1016/j.scitotenv.2017.09.170

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 September 2019 doi:10.20944/preprints201909.0206.v1

Peer-reviewed version available at *Journal of Environmental Management* **2020**; doi:10.1016/j.jenvman.2019.109894

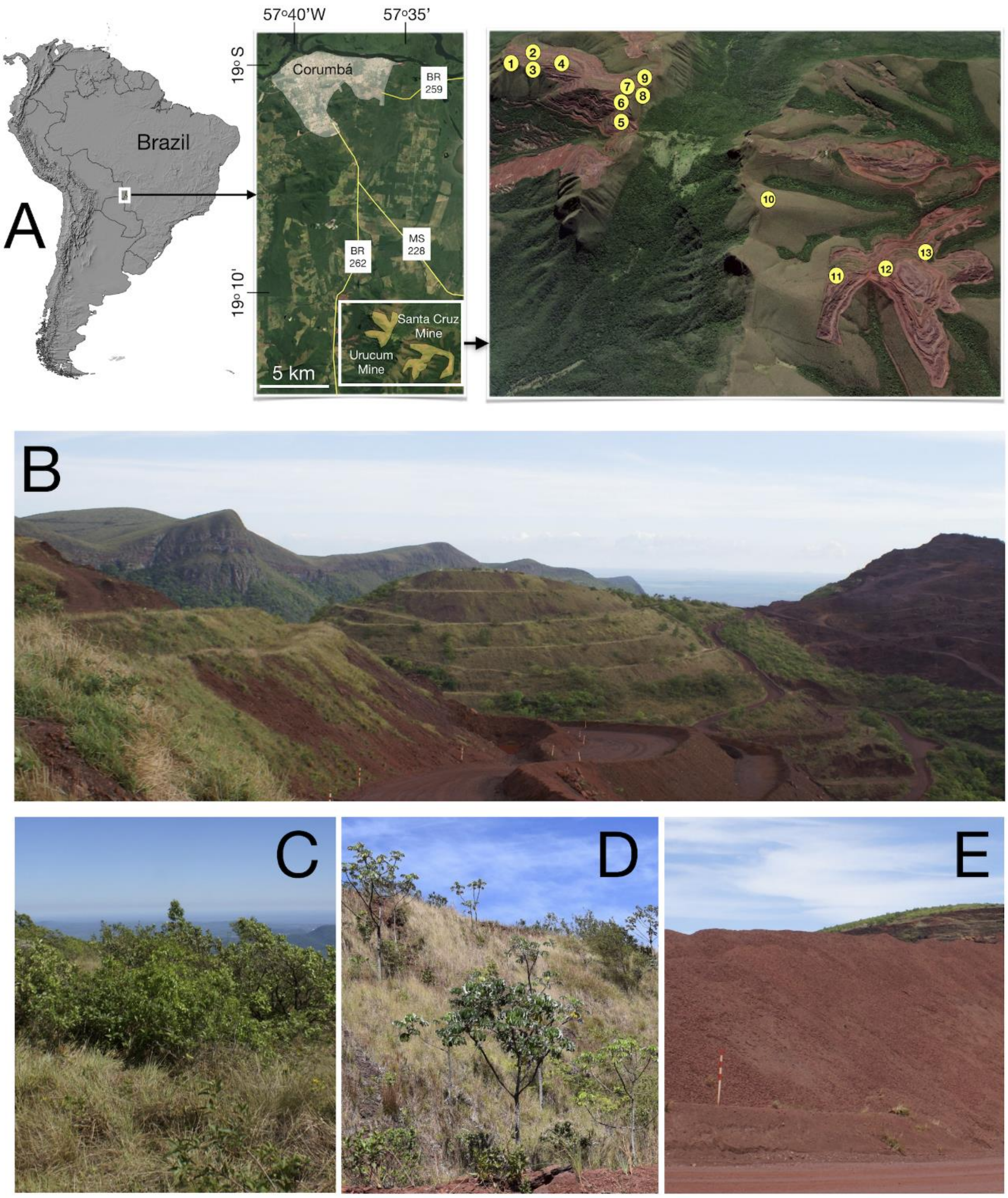

Fig. 1. Map of the study sites (A) and the characteristic vegetation mosaic found in the Urucum Massif, Corumbá, Brazil (B). The general landscape shows open savanna formations covering the mountaintops in the background of B and forest formations on the hillsides, the environmental rehabilitation areas (mountains in the center and on the left portion of the photo) and the nonrehabilitated mining sites (mountain on the left). C-E provide details of premining open savanna formations (sites 1, 9, 10), rehabilitating sites (2, 4, 6, 7, 8, 12, 13) and nonrehabilitated benches (further sites).

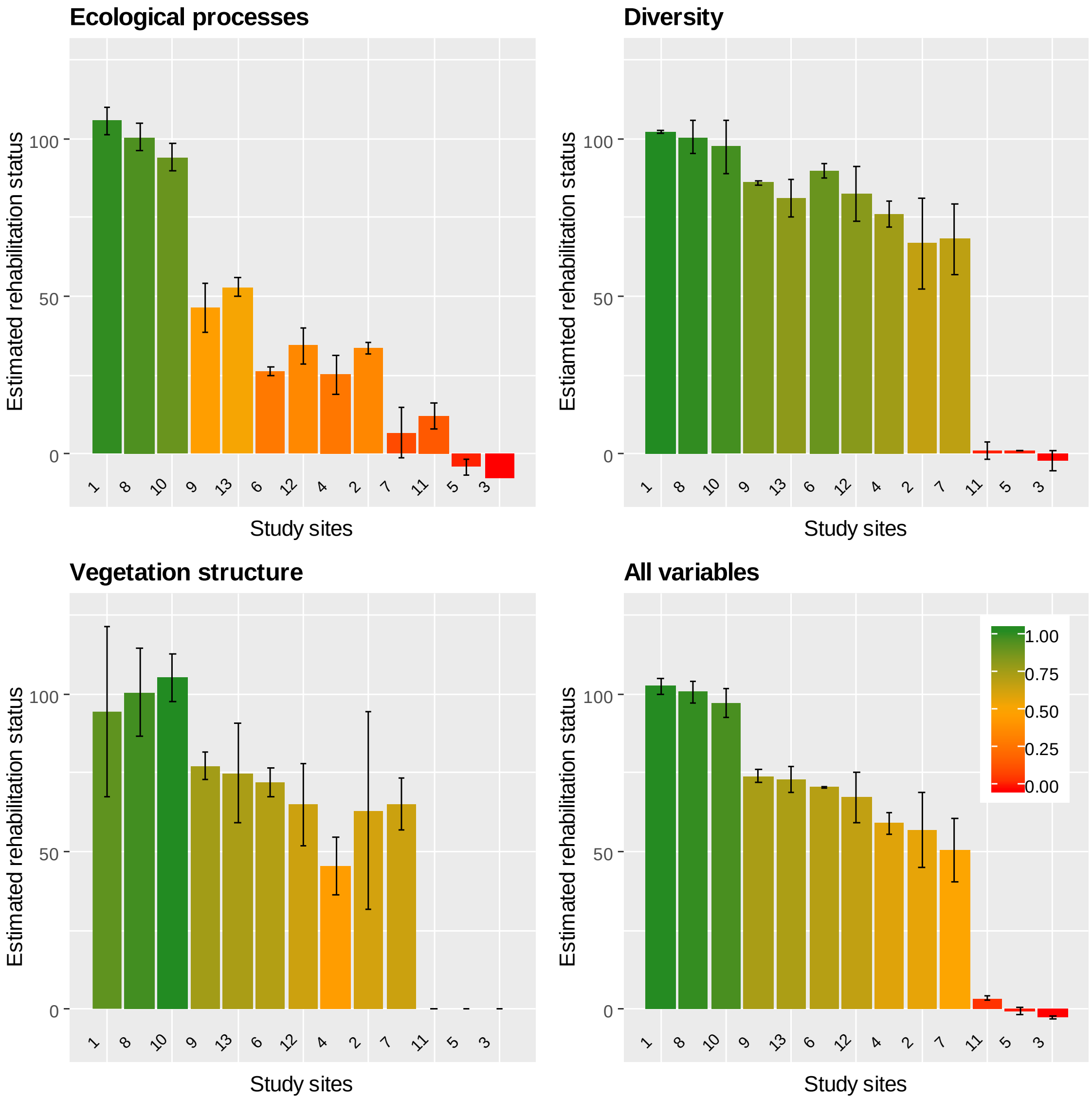

Fig. 2. Mean estimated rehabilitation status (error bars indicate standard deviations) based on proxies of the ecological processes, diversity, and vegetation structure and the entire set of variables computed by the multivariate approach for reference sites (study sites 1, 9, 10), nonrehabilitated benches (sites 3, 5, 11) and rehabilitating sites (further sites). The colors change from red to green with increasing rehabilitation status. For the geographic locations of the study sites, see Fig. 1.

Peer-reviewed version available at Journal of Environmental Management 2020; doi:10.1016/j.jenvman.2019.109894

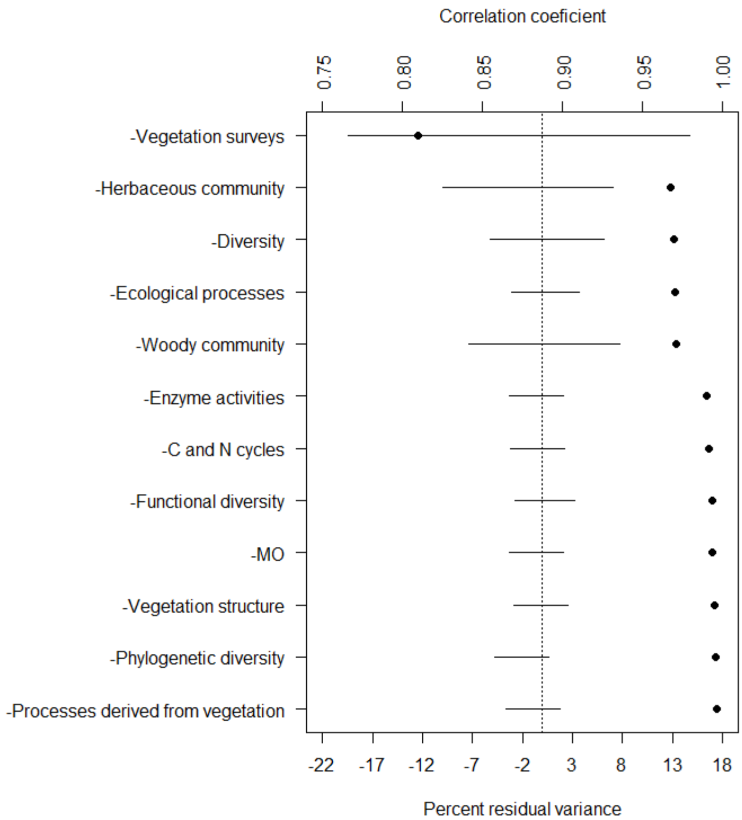

Fig. 3. Influence of single attributes, surveys or analyses on the outcomes of the rehabilitation status estimation. The percent residual variances (lines) and the coefficients of determination (points) fitting the complete estimation by the estimations computed from smaller variable sets using mixed-effect models are shown. See the methods for details.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 September 2019 doi:10.20944/preprints201909.0206.v1

Peer-reviewed version available at *Journal of Environmental Management* **2020**; doi:10.1016/j.jenvman.2019.109894

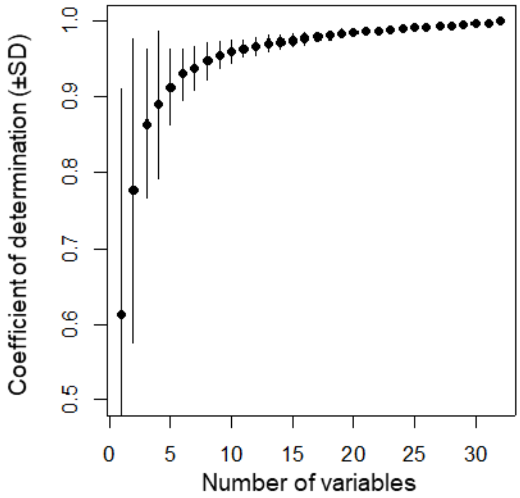

Fig. 4. Mean coefficients of determination (black points) and their standard deviations (lines) between 1,000 randomizations based on a reduced number of variables and the rehabilitation status estimation from the entire set of variables. See the methods for details.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 September 2019 doi:10.20944/preprints201909.0206.v1

Peer-reviewed version available at *Journal of Environmental Management* **2020**; doi:10.1016/j.jenvman.2019.109894

Table 1. Variable mean values (± SD) from the nonrehabilitated sites (NR), the rehabilitated sites (Reh) and the reference sites (Ref), as well as the mean scores of the rehabilitated sites scaled between the values achieved in the reference and nonrehabilitated sites (REH[%]), used for the rehabilitation status estimation after iron ore mining in the Urucum Massif. Different letters indicate significant differences according to a post hoc Dunn test ($p < 0.05$).

Attribute	Analysis/Surveys	Environmental variable	Acronym	NR	Reh	Ref	REH [%]*
Ecological processes	N and C cycles	Soil organic carbon content	Corg	3.3 ±2.23b	3.09 ±2.15b	10.86 ±8.44a	-
		Soil basal respiration	Respiration	10.25 ±3.53b	12.19 ±2.88b	25.3 ±2.37a	12.87 ±19.13
		Metabolic quotient	qCO2	0.04 ±0.01	0.04 ±0.01	0.04 ±0.02	-
		Nitrogen in the microbial biomass	N_BM	1.86 ±1.61b	1.25 ±1.51b	7.38 ±4.41a	-
		Carbon in the microbial biomass	C_BM	254.94 ±38.49b	337.16 ±116.38b	777.65 ±492.21a	15.73 ±22.26
	Enzyme activity	Urease activity	Urease	3.34 ±0.58b	3.91 ±3.28b	19.17 ±8.26a	3.62 ±20.72
		β-glucosidase activity	β_glucosidase	23 ±4.56b	24.32 ±6.04b	210.62 ±45.21a	0.70 ±3.22
		Hydrolysis of the fluorescein diacetate	FDA	90.84 ±35.64b	158.51 ±87.56b	291.99 ±44.09a	33.64 ±43.53
		Phosphatase activity	Phosphatase	209.35 ±118.25b	309.94 ±271.67b	2375.03 ±406.53a	4.65 ±12.54
	Glomalin	Easily extractible glomalin content	EEG	2.27 ±0.87b	1.77 ±1.09b	3.26 ±0.74a	-
		Total glomalin content	TG	11.24 ±10.94b	9.33 ±7.68b	21.32 ±8.87a	-
	Vegetation	Percentage of zoochorous species	%Zchory	0 ±0b	57.37 ±8.29a	61.54 ±4.84a	93.23 ±13.48
		Percentage of species with fleshy fruits	%Fleshy	0 ±0b	4.2 ±6.22b	27.2 ±8.87a	15.44 ±22.88
		Aboveground plant biomass	AGB	0 ±0c	0.31 ±0.26b	0.92 ±0.32a	33.87 ±28.17
Vegetation structure	Vegetation	Tree density	Tree Density	0 ±0c	10.29 ±8.33b	17.56 ±5.39a	58.59 ±47.42
		Vegetation cover	Vegetation Cover	0 ±0c	15.04 ±13.67b	89.89 ±5.83a	65.24 ±13.32
		Basal area	Basal area	0 ±0c	0.01 ±0.01b	0.05 ±0.04a	15.48 ±15.72
Diversity	Vegetation	Observed species richness of the woody community	S_woody	0 ±0c	2.05 ±0.97b	4 ±1.5a	51.19 ±24.34
		Estimated species richness of the woody community	Chao_woody	0 ±0c	2.12 ±1.07b	4.19 ±1.59a	50.59 ±25.57
		Phylogenetic diversity of the woody community	PD_woody	0 ±0c	253.67 ±89.71b	434.39 ±147.19a	58.4 ±20.65

Peer-reviewed version available at Journal of Environmental Management 2020; doi:10.1016/j.jenvman.2019.109894

	Observed species richness of the herbaceous community	S_herbaceous	0 ±0c	11.62 ±2.48b	16.44 ±2.74a	70.66 ±15.08
	Estimated species richness of the herbaceous community	Chao_herbaceous	0 ±0c	15.73 ±4.92b	23.05 ±5.2a	68.26 ±21.36
	Phylogenetic diversity of the herbaceous community	PD_herbaceous	0 ±0c	731.72 ±147.56b	1273.41 ±166.59a	57.46 ±11.62
	Species richness of the entire community	S_total	0 ±0c	12.19 ±2.91b	17.89 ±2.57a	68.15 ±16.26
	Phylogenetic diversity of the entire community	PD_total	0 ±0c	791.39 ±172.93b	1417.63 ±179.45a	55.82 ±12.2
	Functional richness (entire community)	FRic	0 ±0c	7.86 ±2.01b	14.11 ±2.03a	55.68 ±14.22
	Rao's entropy (entire community)	RaoQ	0 ±0c	0.16 ±0.02b	0.2 ±0.01a	78.66 ±12.33
	Functional dispersion (entire community)	FDis	0 ±0c	0.38 ±0.04b	0.44 ±0.02a	85.72 ±9.14
	Functional evenness (entire community)	FEve	0 ±0c	0.57 ±0.08b	0.74 ±0.05a	76.72 ±10.46
	Mean dissimilarity of the reference sites	Dissimilarity	1 ±0a	0.67 ±0.05b	0.37 ±0.08c	51.94 ±8.23
Microorganisms	Richness of the microorganism communities	S_MO	1553.44 ±287.24a	1284.9 ±232.9b	1516.86 ±193.02a	-
	Shannon diversity of the microorganism communities	DIV_MO	5.13 ±0.35	5.12 ±0.25	5.27 ±0.17	-

*Missing values indicate that rehabilitation success estimation is not feasible because the scores achieved by the rehabilitating sites are outside the interval given between nonrehabilitated and reference study sites.

Peer-reviewed version available at *Journal of Environmental Management* **2020**; doi:10.1016/j.jenvman.2019.109894

Table 2. Best models best fitting the estimation of the rehabilitation status. AIC is the Akaike information criterion, and R^2 is the coefficient of determination of the model, which is listed here to indicate the predictive power of the model for the entire rehabilitation status estimation and not as a criterion for model selection. All models are mixed-effect models.

Model	Variable	Estimate (±SE)	t-value	p-value	AIC	Weight	R^2
1	Intercept	0.578 ±0.017	34.649	< 0.001	-107.23	0.571	0.9908
	Vegetation Cover	0.171 ±0.024	7.160	< 0.001			
	RaoQ	0.183 ±0.024	7.714	< 0.001			
2	Intercept	0.578 ±0.013	45.128	< 0.001	-106.09	0.340	0.9906
	Vegetation Cover	0.179 ±0.024	7.234	< 0.001			
	PD	0.179 ±0.024	7.318	< 0.001			