Appendix S3. The artificial data set

Methods

The artificial data set comprised plant species abundances on 2 orthogonal gradients of soil resources and disturbance intensity, respectively, and the following traits: canopy height (CH), specific leaf area (SLA), leaf nitrogen content (LNC), seed mass (SM), onset of flowering date (Onset), terminal velocity (TV). Each of the two orthogonal gradients consisted of a sequence of 16 numbers, offering $16x_{16} = 256$ combinations ("plots"). For each environmental gradient, niches for 50 species were constructed with a Gaussian response curve with the formula $z=c \exp[-0.5(x-u)^2/t^2]$ where z is the abundance value, x is the resources or disturbances vector, c is the species' maximum abundance (set to 70%), u its optimum (niche centre), and t its tolerance (niche breath). To ensure that the species had slightly different niche positions, breaths, and abundances, the final optimum u_i was a normal random number with mean u and sd=1. Likewise, t_i was a normal random number with mean=1.5 and sd=1. The individual abundances z_i were random deviates of a Poisson distribution with mean=z which were separately calculated for each environmental vector. The final abundances were the product of each combination of the two vectors to produce the plots x species table. Likewise, the plots by environment table was the expanded combination of the resource and disturbance vector. To ensure that the environmental space was not completely regular, we did the calculations for 16x16=256 plots and then applied a random stratified sampling of 144 plots and associated species' abundances over the whole gradient, divided in 6 strata. Stratification was done to prevent the random sampling from just picking plots from the upper or lower part of the gradients.

To construct the species by trait table, we used the optima u_i of each species on the strong resources and disturbance gradients, denoted here as "resopt" and "disopt". Canopy height

decreased non-linearly with disturbance intensity, using CH=150+(-0.48*distopt)+(-0.48*disto

 $0.48*(distopt^2)$. Onset of flowering date decreased with both disturbance intensity and resources, with Onset=220+(-9*disopt)+(-2*resopt). That is, plants became smaller when more disturbed and onset of flowering was earlier, even more so, when resources were abundant. SLA increased with earlier onset, because plants need a higher growth rate, indicated by higher SLA, to attain a given height when the number of growing days till onset of flowering is reduced. The formula was SLA=50+(-0.25*onset)+(0.1*CH). LNC increased with resources and SLA, using LNC=10+(2.5*resopt)+(0.5*SLA]. Seed mass was positively correlated with canopy height (Moles et al. 2004) and LNC, using SM=0.5+(0.02*CH)+(0.04*LNC). The final trait values for each species were normal random numbers with means given by these formulas. Finally, TV had a random distribution. This trait was included to assess the method's ability to distinguish non-responsive from responsive traits. Formulas were created to give biologically meaningful values for each trait, and in the case of SM to conform to log-transformed values.



Fig. S3.1 The two orthogonal gradients disturbance and resources, each with values ranging from 1 to 16. The species' optima were simulated to occur either in the corners (dark grey) and in the centre (strong gradient), in the intermediate positions (grey) and in the centre (intermediate gradient), or in the centre alone (light grey, no gradient). A fourth simulation distributed the species randomly on the gradient plane.

To test whether the different methods were sensitive to gradient length, we constructed the niches and associated trait values for three independent scenarios. In the first scenario, niches of ten species each were placed at the corners (2,2; 2,15; 15,2; 15,15) and the centre (8) of the disturbance-resource template, producing two strong orthogonal gradients in both abiotic and functional terms (Fig. S3.1). In the second scenario, the species were placed at the intermediate positions between centre and endpoints (4,4; 4,12; 12,4; 12,12) as well as the centre itself (8), producing intermediate gradients. The third scenario was characterized by absent gradients, i.e. all species were distributed around the centre (8). The assumption behind the first scenario was that strong environmental differences selected 50 functionally different species forming five well separated functional groups in the four corners and the centre of resourcedisturbance template. The third scenario corresponded to the analysis of a single community type, e.g. meadows with homogenous soil resources and functional community composition. A fourth scenario consisted of random distributions of species optima and random abundances. The optima used to calculate the trait values for this scenario were those from the strong gradients scenario, so that functionally very different species were randomly distributed along the two gradients. In this scenario, we assumed early stages of a succession on bare soil, when no species sorting has yet taken place.

We expected that TV was not responsive to either resources or disturbances, and, due to the trait-trait correlations, either canopy height or onset were replaced by SLA, seed mass or/and LNC as responsive trait combinations or vice versa. We further assumed that traitenvironment relationships would be stronger and classification of functional groups more consistent across all repetitions when gradients are strong and collapse when gradients are absent. Trait-environment relationships should be insignificant in the random scenario. We ran 200 repetitions for each scenario and method and used the mean and standard deviation of different output metrics produced by the methods to compare their performance in the scenarios.

Most methods produced a single result per scenario and repetition. Cluster Regression however offered a range of different trait combinations that may have yielded similar responses to the environmental gradients. Of these, one best performing trait combination was retained in each repetition. This trait combination had (i) the highest R² averaged over all clusters modelled on the environmental gradients, (ii) the highest number of clusters, indicating the resolution of the trait dissimilarity in functional groups, and (iii) the highest cophenetic correlation index, indicating the stability of the clusters. All calculations were done in R.

Results

Consistency of trait-environment relationships

The total variance in species or trait composition explained by the environmental gradients in the ordination-based methods and the average R² of the clusters modelled by Cluster Regression indicated that the "strong" and "intermediate gradients" scenarios could be well captured by all methods (Fig. S3.2, S3.3). As intended, the explained variation of the "random" scenario was always close to zero. For the "absent gradients" scenario, the species-based methods RLQ, double CCA and OMI-GAM hardly managed to explain any variance, whereas the community-based CWM-RDA still revealed some variance, though less than for the stronger gradients. Likewise, RDA-mRDA and Cluster Regression yielded explained variance in some repetitions. For Cluster Regression this was the case when a single cluster contained all species, because their traits were all similar and they were all located at the centre of the environmental template.



Fig. S3.2 Total variance explained (sum of eigenvalues) as a measure of total traitsenvironment relationships. Ran: "random" scenario; Abs: "absent gradients" scenario; Int: "intermediate gradients" scenario; Str: "strong gradients" scenario.



Fig. S3.3 Left: Sum of eigenvalues of the OMI as a measure of species-environment relationships. Centre: Percentage of variance explained by the tree of RDA-mRegTree. Right: Average R² of modelled clusters.

Consistency in individual trait responses

In the "strong" and "intermediate gradients" scenarios, LNC mainly responded to resources, whereas canopy height, SLA and onset of flowering responded to disturbance and TV was random and therefore non-responsive. CWM-RDA and RLQ reproduced these relationships very well for the "strong" and "intermediate gradient" scenarios, with axis 1 representing the disturbance and axis 2 the resources gradient (Fig. S3.4). Variation was remarkably low across all repetitions. In the "absent gradients" scenario, these relationships were still visible but less strong and with higher variance. OMI-GAM produced traits weights that allow assessing the relevance of each trait for the explanation of species distributions on the environmental gradients. Correlation of axis 1 and 2 to either resources or disturbances shifted strongly between repetitions so that trait weights showed similar responses to both axes. Onset, LNC and SLA had higher weights than the other traits in the first three scenarios. Like OMI-GAM, associations of the axes to the two environmental gradients shifted between the repetitions in double CCA. This led to a high variance in trait responses while the averages were similar to those of RLQ and CWM-RDA. The random trait TV had almost no correlation with the axes in RLQ and double CCA, and very minor correlations in OMI-GAM and CWM-RDA.

For Cluster Regression and RDA-mRegTree, summaries of trait correlations to ordination axes could not be computed. Instead, we used the proportion of mentions of the traits in the trait combinations with the best clusters in terms of model fit in the regression trees, respectively, across all repetitions for each scenario (Fig. S3.5). RDA-mRegTree gave higher relevance to the resources gradient because the traits associated with this gradient, i.e. LNC, canopy height and onset, appeared most often in the trees. In contrast, Cluster Regression favoured traits that were associated with both gradients, such as seed mass, SLA, onset. Both methods correctly discarded the random trait TV.

CWM-RDA and Cluster Regression did not produce any relationships for the "random" scenario. OMI-GAM, RDA-mRegTree, and double CCA yielded minor responses which however did not differ between the traits, including TV. In contrast, RLQ produced responses with high variance between the repetitions.





Fig. S3.4 Correlation of environmental gradients (row A) and traits (row B) to ordination axes in the four scenarios, shown for methods A, D-F. Abbreviations of scenarios see Fig. S3.1.



Fig. S3.5. Proportion of mentions of traits in classifications over all repetitions in four scenarios: Left: trait combinations produced by Cluster Regression. Right: regression trees produced by RDA-mRegTree. Abbreviations of scenarios see Fig. S3.1.

Consistency in the classification of functional groups

We assessed the consistency of functional group classifications across all repetitions for the "intermediate" and "strong gradients" scenarios by using (i) the frequency of number of groups per repetition (Fig. S3.6), and (ii) the stability of the assignment of species to groups among repetitions (Fig. S3.7). This was done for all methods except CWM-RDA which could not produce functional groups. To quantify stability, we used the corrected Rand index, which compares classification agreement ('crand index'; R package 'e1071'). RLQ, dCCA and OMI-GAM showed lower diversity in group numbers and higher agreement in species assigned to functional groups than RDA-mRegTree and Cluster Regression. This was not surprising because the former methods used all traits to define the functional groups whereas the latter methods selected only responsive traits. This introduced an additional source of variation because the selection of traits could change between repetitions depending on differences

in niche distribution and associated trait values introduced by the random component. Groups produced by RLQ, double CCA and OMI-GAM were more stable in the "strong" than in the "intermediate gradients" scenario.



Fig. S3.6 Frequency of number of groups identified in 200 repetitions.



Fig. S3.7 Stability of the assignment of species to groups among repetitions according to the corrected Rand index.

References: see main paper