Indicator values (IndVal) mimic ranking by $F$-ratio in real-world vegetation data

O. Wildi$^{1,2}$ and E. Feldmeyer-Christe$^{1,3}$

$^1$WSL Swiss Federal Institute for Forest, Snow and Landscape Research, 8903 Birmensdorf, Switzerland.  
$^2$Corresponding author, e-mail: otto.wildi@wsl.ch, Fax: +41 44 739 22 15, http://www.wsl.ch/info/mitarbeiterlde/wildi  
$^3$E-mail: elizabeth.feldmeyer@wsl.ch, Fax: +41 44 739 22 15, http://www.wsl.ch/info/mitarbeiterlde/feldmeye

Keywords: Clustering, Null model, Mantel correlation, Rank correlation.

Abstract: The identification of differentiating species, also termed indicator species, is a key issue in striving for group pattern in vegetation samples. With this objective in mind Jancey (1979) proposed species ranking on a variance ratio ($F$-ratio) criterion, the mechanism of analysis of variance. Almost 20 years later Dufrené and Legendre (1997) presented their indicator value analysis with the same objective in mind but based on different reasoning. This raises the question if (i) the results of the two approaches are equivalent or (ii), if one of the measures of performance is superior in predicting specific properties of ecosystems, such as site conditions, biodiversity, succession or other. Because the outcome of both methods strictly relies on the strength of vegetation pattern reflected by the data sets used as well as the quality of classifications, we compare results from a small and also a large real-world data set and we evaluate the effect of the number of groups involved when clustering sites. In a subsequent step, the ranking of indicator species identified by either of the methods is compared with a ranking obtained by correlating species with measured environmental factors. The results confirm that the outcome of ranking by maximum indicator value (IndVal) according to Dufrené and Legendre (1997) is very similar to the ranking devised by Jancey (1979). Rank correlation reaches a maximum of $r > 0.95$ when the data set is large and group number in clustering is low. In our examples Jancey’s method is generally more closely related to the environmental predictive power of species, outperforming IndVal when applied to continuous variables measured in the field. We conclude that the potential of Jancey’s method is generally underrated. As expected, the agreement of results between the two approaches depends on the strength of similarity patterns inherent in the data sets analysed. The method of Dufrené and Legendre (1997) is well adapted to issues of phytosociology where classification is frequently based on expert knowledge. If the resolving power of species is used as a surrogate for patterns and processes of plant-environment systems, then ranking by variance ratio may be the more promising approach.

Introduction

Indicator values were introduced by Dufrené and Legendre (1997) to measure the uniqueness of occurrences of species within a given vegetation type. The indicator value (IndVal) measures the probability of a site belonging to a specific site group $G$ (a vegetation type, for example) in case a species $S$ has been encountered (De Cáceres et al. 2012), $A=p(G|S)$. The quantity $A$ is also termed ‘specificity’ or ‘positive predictive value’ of species $S$. To further adapt the indicator value to practical requirements it is multiplied by a probability, $B$, expressing the chance that the species can be detected in the field. The success of the method depends on properties of a given vegetation sample, that is, a species by site table with species performance scores as elements and a classification of the sites. The indicator value is an attribute of any one species and in practice it can serve reducing species lists deemed too long for practical purposes such as in monitoring or mapping (De Cáceres et al. 2012). In such applications, the emphasis is not on the magnitude of an indicator variable but rather its rank order within a usually long species list.

With just the same objectives in mind Jancey (1979) proposed what he called ‘species ordering on a variance criterion’. This is simply analysis of variance (AOV) of individual species, based on a classified vegetation sample, identical with the requirements for calculating IndVal. The difference lies in the use of a well-established linear model yielding a variance ratio, $F=V_g/V_w$, where $V_g$ is between and $V_w$ is within group variance. This is interpreted as a measure for the resolving power of species, whereas a practice-oriented reasoning is inherent in the IndVal method. As noted by Jancey (1979) the $F$-ratio may have an $F$-distribution allowing statistical testing. Jancey proposed its use as a ranking criterion. His main purpose was identical with the same later suggested for IndVal: ‘using some subset of the total number of species present in the area.’ He also suggested considering combinations of site groups, when species are indicative for two or more groups rather than one group only: ‘A partitioning of the ranking into all pairwise site comparisons overcomes this problem.’ Although this idea has not been used hitherto De Cáceres et al. (2010) suggest this for IndVal. They not only consider pairs of site groups but any possible combination of groups. Unlike in the original concept (Dufrené and Legendre 1997) where each group is compared with all remaining
and the maximum IndVal taken as a final result, this extension takes into account up to $2^k$ combinations of k groups. Almost simultaneously Podani and Csányi (2010) addressed the same issue by proposing an extension to IndVal in which the number of groups involved is taken into account. This brings it closer to Jancey’s analysis of variance where variance within all groups is used and sample size as well as group number are affecting error probability of the $F$-ratio by considering the degrees of freedom. Yet another possible extension of indicator analysis is in accounting for species combinations proposed by De Cáceres et al. (2012). This differs from Jancey (1979) who explicitly abstains from species combinations considering this problem a task of ordination.

Wildi (1989) used Jancey’s ranking to distinguish differentiating species individually from indifferent ones in the context of traditional ordering of vegetation tables (Mueller-Dombois and Ellenberg 1974) where it proved to be rather satisfying. This raises the question if ranking by IndVal in this context would be an alternative to variance ranking, a question addressed in this paper. A case in point is a small artificial example of a data set with 21 species shown in Wildi (2013) where the two ranking methods show roughly similar outcome. To illustrate the issue a very small artificial example shown in Table 1 confirms the striking similarities of the results despite obvious difference in the magnitude of the indices. The proportions (and hence the ranks) of indicator values and the variance ratios, respectively, are almost identical. The same holds for the error probabilities, although those from the $F$-values refer to the variance ratio of all groups simultaneously. In case of IndVal, the error probabilities are the outcome of a randomization test and this concerns the group only for which the value is highest.

From a formal point of view the two approaches strive for selection of single variables reflecting best a classification of a multivariate sample. Plant ecologists have identified various questions where this is of potential merit. Already Jancey (1979) promotes ranking as a tool to ‘achieve greater efficiency’ in sampling. McGeoch and Chown (1998) argue that species should obtain priority reflecting best environmental quality. And in fact, De Cáceres et al. (2012) see indicator species (i) reflecting not only the biotic but also the abiotic environment, (ii) to be useful in detecting environmental change and even (iii) predicting diversity of other species or taxa within the sampling frame. Ranking does of course not answer such questions directly. It fulfils the expectations only in case where there is a relationship between the biotic space (the species, for example) and the environmental factors, the diversity of vegetation types or the succession pattern in case of change. Change may concern the state of the natural system, but the same happens when applying an alternative classification to the same body of data. To reveal what variance ratios or indicator values really express one has to analyse a system where not only the biotic, but also the abiotic or even the spatial and temporal components are known and available for independent testing.

### Table 1. An artificial example demonstrating ranking by IndVal (error probability derived in a randomization test) and by $F$-value (error probability according to analysis of variance).

<table>
<thead>
<tr>
<th>Relevé</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>IndVal</th>
<th>Error $p$</th>
<th>$F$-ratio</th>
<th>Error $p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group no.</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td></td>
<td>1.000</td>
<td>0.040</td>
</tr>
<tr>
<td>Species 1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>0.548</td>
<td>0.049</td>
<td>16.092</td>
</tr>
<tr>
<td>Species 2</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td></td>
<td>0.500</td>
<td>0.178</td>
<td>13.290</td>
</tr>
</tbody>
</table>

In this paper we try to answer two questions. First, how big is the difference between the variance-ratio method (species ranking by variance) and the IndVal concept in real-world examples? Second, is there a relationship between the results of these two methods and the correlation of species with environmental factors? We answer these using two real-world data sets, published in Wildi (2013).

### Methods and data sets

The quality of classification used for the two ranking methods is of course crucial and it may be of external origin (e.g., when reflecting a sampling design) or internal (e.g., from clustering of relevés). In the present case we use clustering, offering freedom in the selection of group number. For classification of relevés Ward’s method is used (Legendre and Legendre 2012, Wildi 2013) and the resulting dendrograms are cut progressively to trace the effect of varying group number. Rank orders delivered by the two approaches are compared by the Spearman correlation coefficient. Technically, we first transformed the $F$-ratio and IndVal scores into a rank scale and then applied the Kendall correlation coefficient yielding the same result as when using the Spearman coefficient directly. The resulting confidence limits, however, strictly apply to the ranks only rather than the original performance scores. Computations always start with two groups. Group number is then incremented until correlation of the two ranking methods drops considerably.

In a subsequent step the question is posed if there is a relationship of rank order and the correlation of the species with measured environmental factors. The selection of these factors is based on previous analyses of the data sets and priority is given to those explaining best vegetation composition. Then, species are ranked again based on fit to environment and this rank order is compared with ranking by variance ratio and IndVal, respectively.

We choose Mantel correlation (Borcard and Legendre 2012) to correlate single species with external factors. Like any other linear method, such as redundancy analysis (Borcard et al. 2011) or non-parametric multivariate analysis of variance (Anderson 2001) this can handle low species frequencies and it provides an error probability measure based on permutation. Even though non-linear methods, as implemented in generalised linear models (GLM’s) or generalised additive models (GAM’s) (Guisan and Zimmermann 2000, Elith et al. 2006, Elith et al. 2008) would be favoured, these do not work for low-frequency species.
The two data sets used are included in the R package ‘dave’ (Wildi 2013). The first, R object ‘sveg’, encompasses 63 acidic subalpine wetland relevés near Rothenthurm, Switzerland, with 119 species as variables (Wildi 1977). The sampling design is systematic with a grid width of 10 m by 10 m and a plot size of 1 m². Braun-Blanquet scores are used as a rank scale for cover-abundance. The second, ‘ws200’, comprises 726 Alpine forest relevés with 1157 species (Wohlgenuth et al. 2008). This study is also based on a systematic grid of 4 km width covering the whole of Switzerland. Plot size is 200 m² and cover-abundance also relies on the Braun-Blanquet scale. In both samples there are various site measurements available to choose among. We restrict the computations to the variables offering best correlations, pH in one, altitude in the other case. All computations are derived in R version 2.15.2 (R development core team 2012).

Results

The first example, using the ‘sveg’ vegetation data, demonstrates the outcome in case of a fairly small data set (Figure 1). Rank correlation reaches a maximum already at the level of four groups with \( r = 0.914 \pm 0.025 \), then decreasing slowly but continuously. A high agreement with \( r = 0.832 \pm 0.084 \) still persists when using 7 groups where mean groups size is 9 relevés. The agreement of the methods vanishes completely near 20 groups, a number far to high for the sample size of \( n = 63 \).

In the second example the number of species is now 1157 describing 726 relevés. The agreement between the two ranking methods is stunning, as confirmed by Figure 2. Correlation is highest when distinguishing 6 groups, \( r = 0.984 \pm 0.002 \). When choosing 20 groups it still reaches \( 0.914 \pm 0.009 \). Even for 30 groups we get \( r = 0.837 \pm 0.018 \). This gives rise to the supposition that the two ranking methods strongly converge when applied to large data sets.

Then, the question is answered whether the indices, IndVal and F-ratio respectively, also reflect the response of species to real environmental factors. Mantel correlation delivers the required response for all species and this has to be correlated with rankings from different classifications. The ‘sveg’ data set encompasses a pronounced floristic and environmental gradient that promises a strong response. The first step is in establishing a species ranking based on Mantel correlation. Using 5 relevé groups this is shown in Table 2 for the 20 first ranks out of 119. This is compared with the two different floristic rankings. Not surprisingly the methods differ in result. But there is also some agreement. The 20 species ranked first by Mantel correlation all reside in the first third of F-ranking and roughly in the first half of IndVal ranks. Hence, some relationship useful for ecological interpretation can be expected.

The full outcome is shown in Figure 3. For F-ratio ranking the highest correlation is \( r = 0.776 \pm 0.083 \), for the IndVal ranking this is \( r = 0.664 \pm 0.114 \). Whereas for F-ratio ranking correlation drops just slowly towards \( r = 0.6 \) for 20 groups, the same done with IndVal vanishes quickly with increasing group number. Hence, some agreement exists in either case, but only when group number is low.

---

**Table 2.** The 20 species (out of 119) in data set ‘sveg’ ranked highest by Mantel correlation with pH in peat. Compared with rankings by F-ratio and Indval, respectively, both based on 5 groups.

<table>
<thead>
<tr>
<th>Rank (Mantel)</th>
<th>Species</th>
<th>Mantel r</th>
<th>Rank (F-ratio)</th>
<th>Rank (Indval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Frigates violaceae</td>
<td>0.517</td>
<td>5</td>
<td>21</td>
</tr>
<tr>
<td>2</td>
<td>Carex elata</td>
<td>0.492</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>3</td>
<td>Anemone montana</td>
<td>0.476</td>
<td>4</td>
<td>14</td>
</tr>
<tr>
<td>4</td>
<td>Sphagnum magellanicum</td>
<td>0.477</td>
<td>13</td>
<td>18</td>
</tr>
<tr>
<td>5</td>
<td>Vaccinium oxyccocos</td>
<td>0.430</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>Aster latifolium</td>
<td>0.439</td>
<td>25</td>
<td>24</td>
</tr>
<tr>
<td>7</td>
<td>Juncus effusus</td>
<td>0.408</td>
<td>15</td>
<td>40</td>
</tr>
<tr>
<td>8</td>
<td>Carex pulicaris</td>
<td>0.401</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>9</td>
<td>Sphagnum recurvatum</td>
<td>0.399</td>
<td>12</td>
<td>26</td>
</tr>
<tr>
<td>10</td>
<td>Linum crassifolium</td>
<td>0.377</td>
<td>29</td>
<td>28</td>
</tr>
<tr>
<td>11</td>
<td>Carex scirpoides</td>
<td>0.342</td>
<td>0</td>
<td>31</td>
</tr>
<tr>
<td>12</td>
<td>Ledebouria gracilis</td>
<td>0.332</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>13</td>
<td>Tribulus patience</td>
<td>0.320</td>
<td>20</td>
<td>33</td>
</tr>
<tr>
<td>14</td>
<td>Carex echinata</td>
<td>0.318</td>
<td>11</td>
<td>47</td>
</tr>
<tr>
<td>15</td>
<td>Eriophorum plumbeum</td>
<td>0.316</td>
<td>26</td>
<td>23</td>
</tr>
<tr>
<td>16</td>
<td>Polystichum pumilum</td>
<td>0.314</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>17</td>
<td>Sphagnum rubrum</td>
<td>0.306</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>18</td>
<td>Meadiommum drummonse</td>
<td>0.301</td>
<td>33</td>
<td>54</td>
</tr>
<tr>
<td>19</td>
<td>Cirsium disynodeses</td>
<td>0.300</td>
<td>22</td>
<td>59</td>
</tr>
<tr>
<td>20</td>
<td>Epipactis palustris</td>
<td>0.264</td>
<td>28</td>
<td>27</td>
</tr>
</tbody>
</table>

---

**Figure 1.** Correlation of species rankings by IndVal and F-ratio respectively for classifications with group number ranging from 2 through 20. The ‘sveg’ vegetation data set is used (119 species ranked).

**Figure 2.** Correlation of species rankings by IndVal and F-ratio respectively for classifications with group number ranging from 2 through 30. The ‘ws200’ vegetation data set is used (1157 species ranked).
The resulting pattern when considering a large sample is shown in Figure 4 with 726 Alpine forest relevés and elevation as the explaining site factor. Correlations of rankings are near $r = 0.5$, again those of $F$-ranking being somewhat higher than the same of IndVal ranking. An increase of group number again results in a drop of correlation, but not as pronounced as in the smaller example shown in Figure 3.

**Discussion**

Using $F$-ratio or IndVal for ranking species in classified samples is mainly motivated by practical considerations, as mentioned by Jancey (1979) and, for instance, by De Cáceres et al. (2010). The resulting values are a means to reduce the number of species if deemed helpful. Both methods yield weights that allow ordering by ranks. The variance ratio proposed by Jancey (1979) may have $F$-distribution, but in the case of relevé by species tables this will rarely be the case: it will often not be suited for statistical testing due to skewed distributions of species scores. In the IndVal method there is a randomization test suggested (Dufrène and Legendre 1997). In its original version as many indicator values per species result as there are relevé groups involved, and for each of these groups there is a corresponding error probability. Ranking as suggested above, however, will only rely on the largest indicator value, representing a fraction of the results only. Despite this constraint the present case study suggests that both methods yield useful results. Even more surprising is the high agreement of the results with increasing sample size. Or, vice-versa, disagreement in small samples seems to be caused mainly by random noise. In the present examples the rank orders achieved also reflect the abiotic environment, as hypothesized by De Cáceres et al. (2012). The origin of this correlation is in the agreement of patterns in the species community data, $Y$, and the same of the environmental data, $E$ (Dray et al. 2012). In ecological modelling $Y$ and $E$ serve the computation of a fitted community table, $F$, and a residual community, $R$. If variation in $F$ is sufficiently large, then there is considerable relationship between community and environmental data. It is only in this case that species ranking as shown above can have some ecological meaning. If there is no relationship, as for instance when using a ‘null model’ for $Y$ (De Bello 2012), then species ranking cannot have any practical purpose. The same is true for all other purposes mentioned by De Cáceres et al. (2012), such as predicting diversity or succession, for example.

Another important restriction is the dependence of the results on the classification used. As mentioned by Jancey (1979) a typical application for a reduced species list would be in vegetation mapping. However, once vegetation starts changing in time, classification is likely to obliterate and the resolving power of high-ranked species vanishes. Nevertheless, ranking may serve the assessment of change, but only if the classification reflects a time process. A typical example is shown by Feldmeyer-Christe et al. (2011) with a succession from bare peat towards bog vegetation. If groups used in ranking of their data set represent states of succession, then the emergence of any one species is a signal of succession progress.

An obvious limitation of IndVal is in omitting the issue of degrees of freedom, namely group number, group sizes and sample size, parameters analysis of variance relies on. In the examples shown here this manifests in decreasing correlations when group number in clustering is increased (Figure 3, for instance). Although we did not test this, the extensions suggested by Podani and Csányi (2010) may reconcile this situation. These authors also mention that IndVal can be used only if scores are metric or ranks, but not presence-absence data. The $F$-ratio does not carry this restriction. Furthermore, it can equally be applied to metric data without zero scores. Hence, Jancey’s ranking also works in the context of environmental variables rather than just species scores (see Wildi [2013] for an example). Because environmental variables often rely on different scales and have different distributions some transformation may be needed prior to ranking.

De Cáceres et al. (2012) suggest including species combinations into the computation of indicator values. They also show that pairs of species can have higher resolving power.
than the individual species in case they overlap partly. The most critical issue in their method is in selecting ‘candidate species’ first (that is, an already reduced set of species) and forming useful species combinations in the sequel. This delivers the outset for indicator value analysis: a data table where not only the sites are classified, but also the species. Such a change of the setup of the data could potentially also be used for variance ranking, although no example is known to us so far.

No matter which method is used, there is yet another shortcoming in any kind of species ranking: redundancy in species lists. Indicator values do not give us any information about the role of species in n-dimensional space. This is what ranking by orthogonal components does, devised by Orłoci (1973, 1978). In this method redundant species drop off the list, which shrinks tremendously in case of high redundancy. In the future the ideal ranking method may therefore be a combination of indicator value analysis and ranking by orthogonal components.

Conclusions

Our results suggest that ranking by F-ratio as devised by Jancey (1979) is more appropriate in our case study than IndVal analysis. The results are less vulnerable to the number of groups chosen when clustering sites and the resulting ranking of species may even more closely reflect environmental conditions. We therefore conclude that the ranking by F-ratio is generally underrated in the literature.

References


Received April 29, 2013
Revised July 9, 2013
Accepted July 10, 2013