To cut the feet to fit the shoes:

Forcing unsupervised cluster analysis to reproduce results of subjective vegetation classification

David Zelený & Ching-Feng Li

Masaryk University
Czech Republic
Zelený & Li: To cut the feet to fit the shoes ...
Terminology

Methods for creating vegetation classification scheme

- **expert-based classification (“subjective”)**
  - hand sorting of vegetation relevés
  - based on expert knowledge and experience

- **unsupervised numerical classification (“objective”)**
  - cluster analysis (incl. TWINSPAN)
  - based on selected numerical algorithm

Methods for reproduction of existing classification

- **supervised classification**
  - expert based or numerical (e.g. COCKTAIL)
How does good vegetation classification looks like?

• based on ecologically meaningful concept
• easily being observed in the field
• simple, or at least not overly complicated
• reproducible on new vegetation records

• good classification is like a good story – it’s not only about list of diagnostic species, but also about other features, like habitat quality, disturbance regime, origin, historical development, management etc.
• but it should not be just a story – it must be based on real data and real patterns (concept-driven vs data-driven classification)
Why is cluster supposed to be better than expert-based?

... numerical cluster analysis is considered to be

- objective
- quantitative
- reproducible
- convenient and fast
- consistent
Cluster analysis: application of consistent method on inconsistent data
Numerical reproduction of existing vegetation associations

Data:

- Czech National Phytosociological Database
- 10 associations, 4 alliances
- classification based on the *Vegetation of the Czech Republic 4. Forest and scrub vegetation* (editor M. Chytrý)
- 40 randomly selected relevés per association
- 400 relevés in total

<table>
<thead>
<tr>
<th>Alliance</th>
<th>Association</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quercion pubescenti-petrae</td>
<td>(1) Sorbo torminalis-Quercetum</td>
</tr>
<tr>
<td>Carpinion betuli</td>
<td>(2) Melico pictae-Quercetum roboris</td>
</tr>
<tr>
<td>Melico pictae-Quercetum roboris</td>
<td>(3) Galio sylvatici-Carpinetum betuli</td>
</tr>
<tr>
<td>Carici pilosae-Carpinetum betuli</td>
<td>(4) Carici pilosae-Carpinetum betuli</td>
</tr>
<tr>
<td>Acer-Tilietum</td>
<td>(5) Aceri-Tilietum</td>
</tr>
<tr>
<td>Mercuriali perennis-Fraxinetum excelsioris</td>
<td>(6) Mercuriali perennis-Fraxinetum excelsioris</td>
</tr>
<tr>
<td>Arunco sylvestris-Aceretum pseudoplatani</td>
<td>(7) Arunco sylvestris-Aceretum pseudoplatani</td>
</tr>
<tr>
<td>Alnion incanae</td>
<td>(8) Alnetum incanae</td>
</tr>
<tr>
<td>Carici remotae-Fraxinetum excelsioris</td>
<td>(9) Carici remotae-Fraxinetum excelsioris</td>
</tr>
<tr>
<td>Stellario nemorum-Alnetum glutinosae</td>
<td>(10) Stellario nemorum-Alnetum glutinosae</td>
</tr>
</tbody>
</table>
Numerical reproduction of existing vegetation associations

**Analysis**

- cluster analysis and (modified) TWINSPAN

- cluster analysis: 5 distances x 5 clustering algorithms
  - distances: Bray-Curtis, Canberra, Euclidean, Kulczynsky, Manhattan
  - clustering algorithms: average linkage, beta flexible, complete linkage, single linkage and Ward

- comparison with subjective classification using Goodman-Kruskal's lambda
Numerical reproduction of existing vegetation associations

Results: How good are the methods in reproducing all clusters?

(sorting of methods reflects how well they reproduce overall classification)
Numerical reproduction of existing vegetation associations

Results:
How good are the methods in reproducing individual associations?

Zelený & Li: To cut the feet to fit the shoes ...
Conclusions

• it’s time to say explicitly that unsupervised numerical classification is **not better** than expert-based, “subjective” one

• human brain has extraordinary ability to sense the signal in noisy data and to combine complex information

• “numerical effort” is better invested into detailed and formalized description of distinguished vegetation units to make them reproducible, instead of “cutting the feet to fit the shoes”

• unconstrained cluster analysis is a good servant, but a bad master – it’s useful as an exploratory tool, but not for creating classification schema
Thank you for your attention!