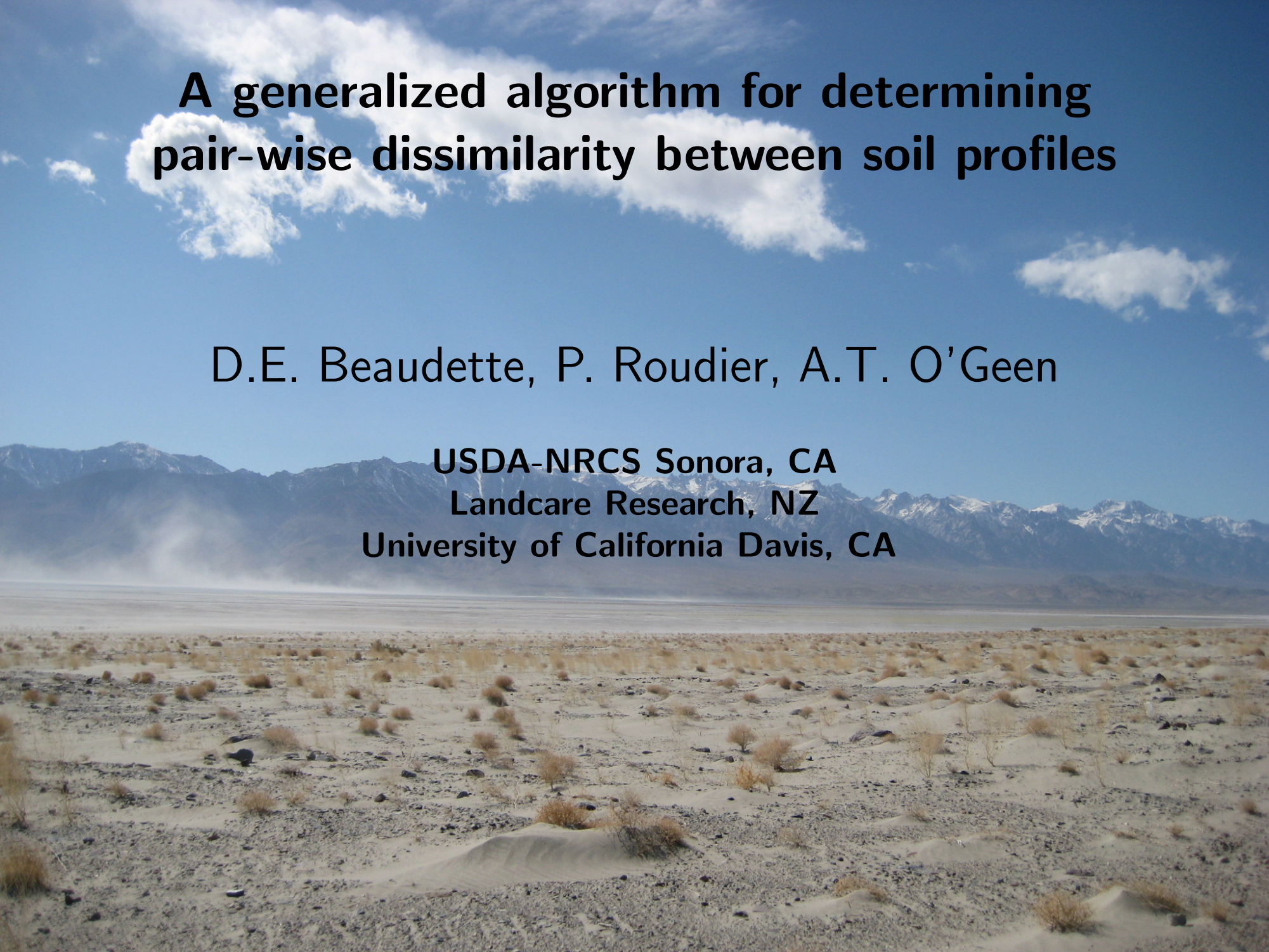
A generalized algorithm for determining pair-wise dissimilarity between soil profiles

D.E. Beaudette, P. Roudier, A.T. O'Geen

USDA-NRCS Sonora, CA
Landcare Research, NZ
University of California Davis, CA



Quantitative (pair-wise) Comparison of Soils

(a)

(b)

(c)

"is **a** more like **b**, as compared to **c**?"

ideally transcending horizonation and description style

Numerical Soil Classification

essentially: an evaluation of "distance" in property space

Numerical Soil Classification

essentially: an evaluation of "distance" in property space

---

reminder: those who ignore the past are doomed to re-implement it– poorly

Numerical Soil Classification

essentially: an evaluation of "distance" in property space

reminder: those who ignore the past are doomed to re-implement it– poorly

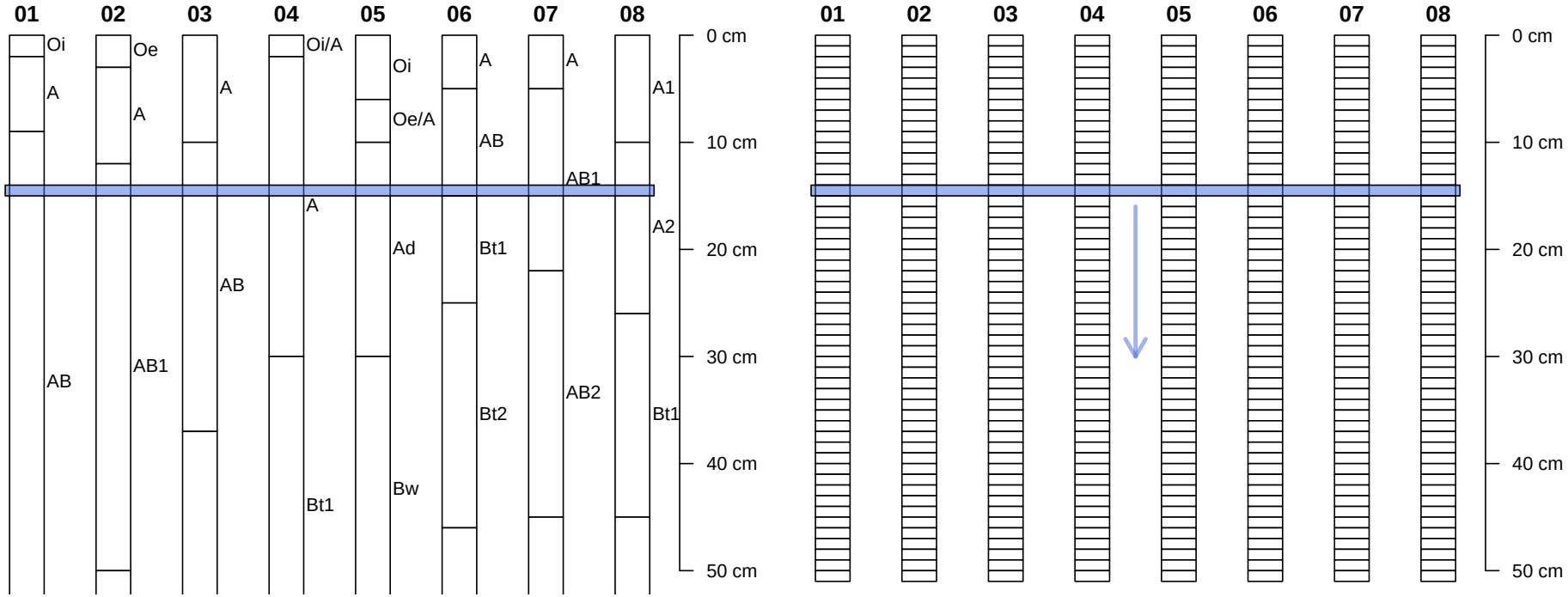
Examples

- ordination of soil properties (Hole and Hironaka, 1960)
- horizon "matching" between profiles (Rayner, 1966)
- depth-intervals, depth func. coeff., transition mat. (Moore et al., 1972)
- allocation to "reference horizons" (King and Girard, 1998)
- k-means, several "distance" metrics (Carre and Jacobson, 2009)

Issues, Assumptions, Limitations

- soil depth not always parameterized
- reference profiles required
- profile-scale (aggregation) vs. hz-scale properties
- algorithm complexity $\leftrightarrow$ parsimony
- allocation vs. pair-wise dissimilarity
- distance metric selection & continuous vs. categorical variables

Pair-wise dissimilarity along depth-slices (Moore et al, 1972)

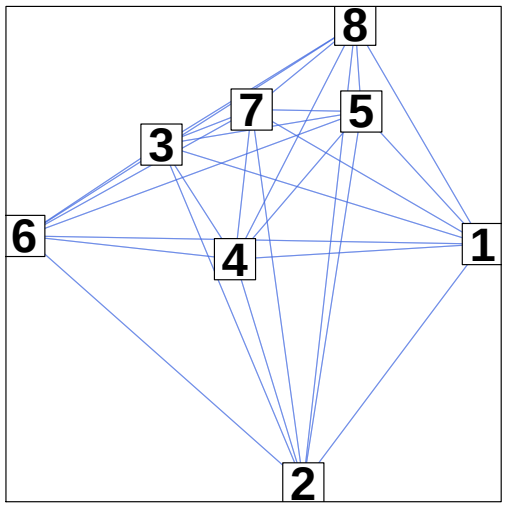


soil properties at slice 15

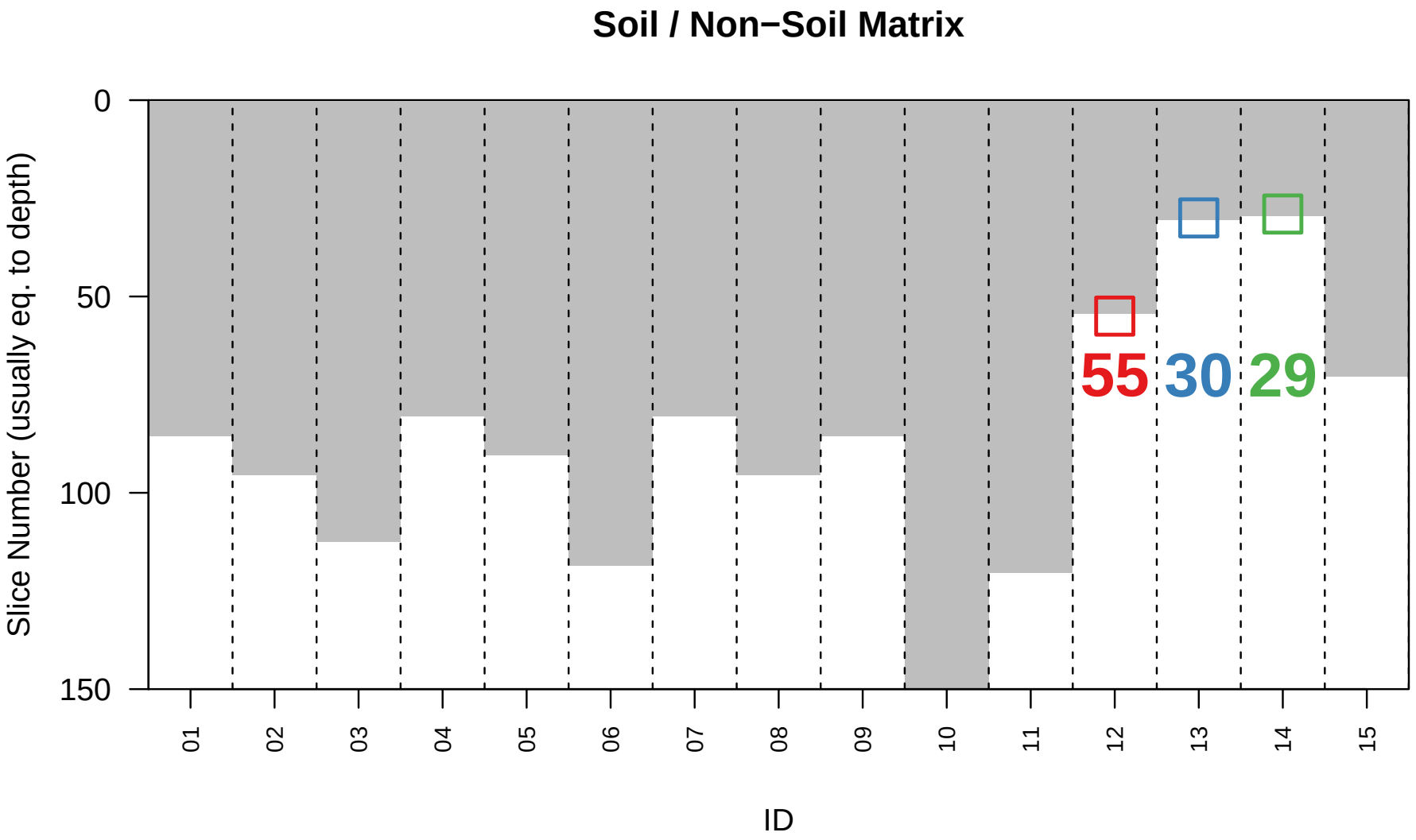
	clay	vcs	ln_tc	cec	A
1	6.4	21.5	-1.5	3.8	8.1
2	10.6	17.6	-1.5	5.9	4.9
3	8.8	10.5	-0.2	7.5	6.0
4	9.1	12.6	-0.8	5.6	5.9
5	6.8	16.2	-0.5	5.0	8.2
6	17.9	11.0	-0.1	9.4	5.2
7	7.0	11.7	-0.4	4.7	6.3
8	6.3	17.0	-0.1	4.7	7.9

pair-wise dissimilarity at slice 15

	1	2	3	4	5	6	7
2	42						
3	68	46					
4	51	29	19				
5	30	46	40	30			
6	96	59	30	45	68		
7	48	45	20	16	22	48	
8	32	50	39	37	11	64	24

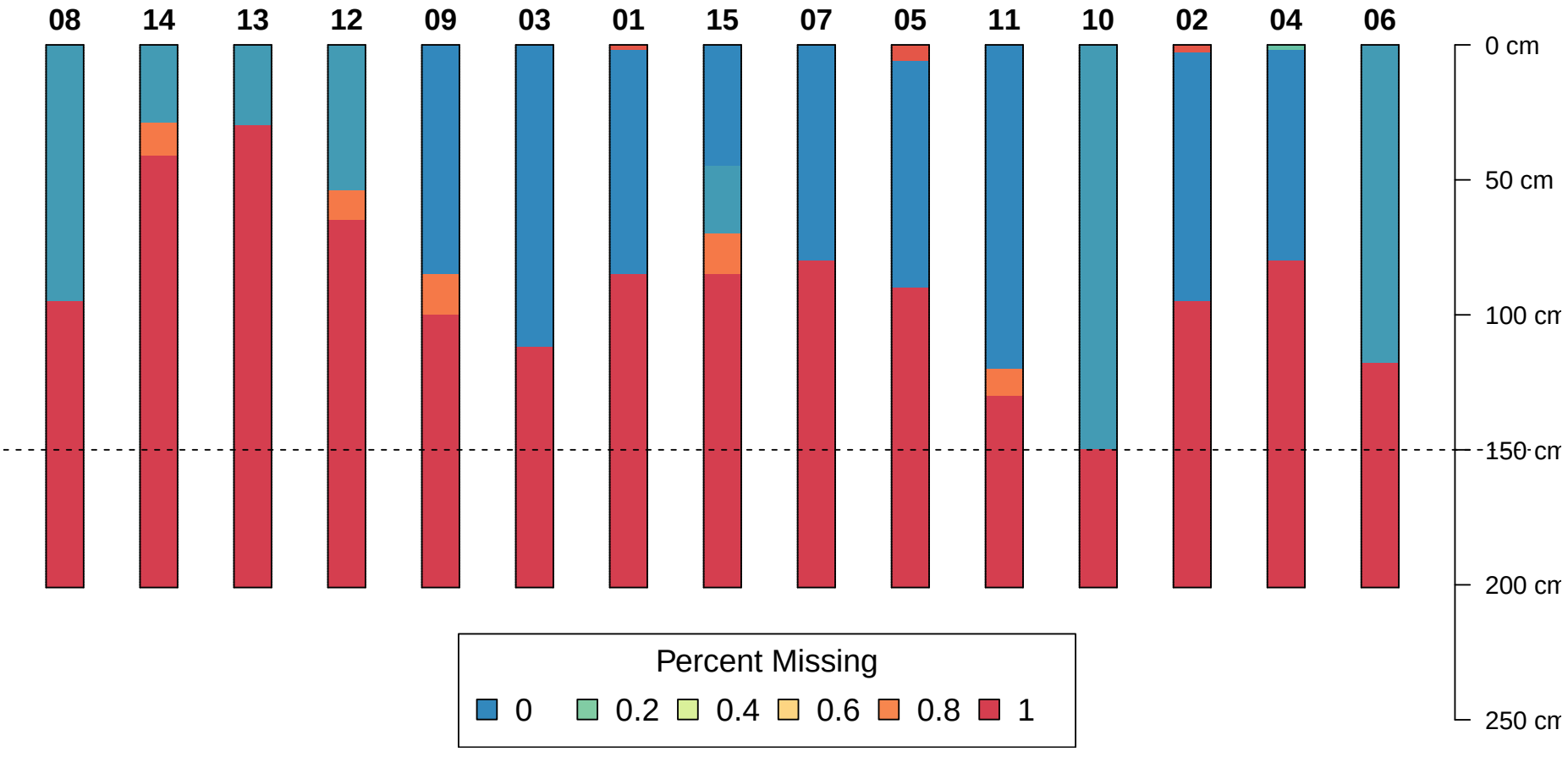


Soil Depth: comparisons between "soil" and "not soil"



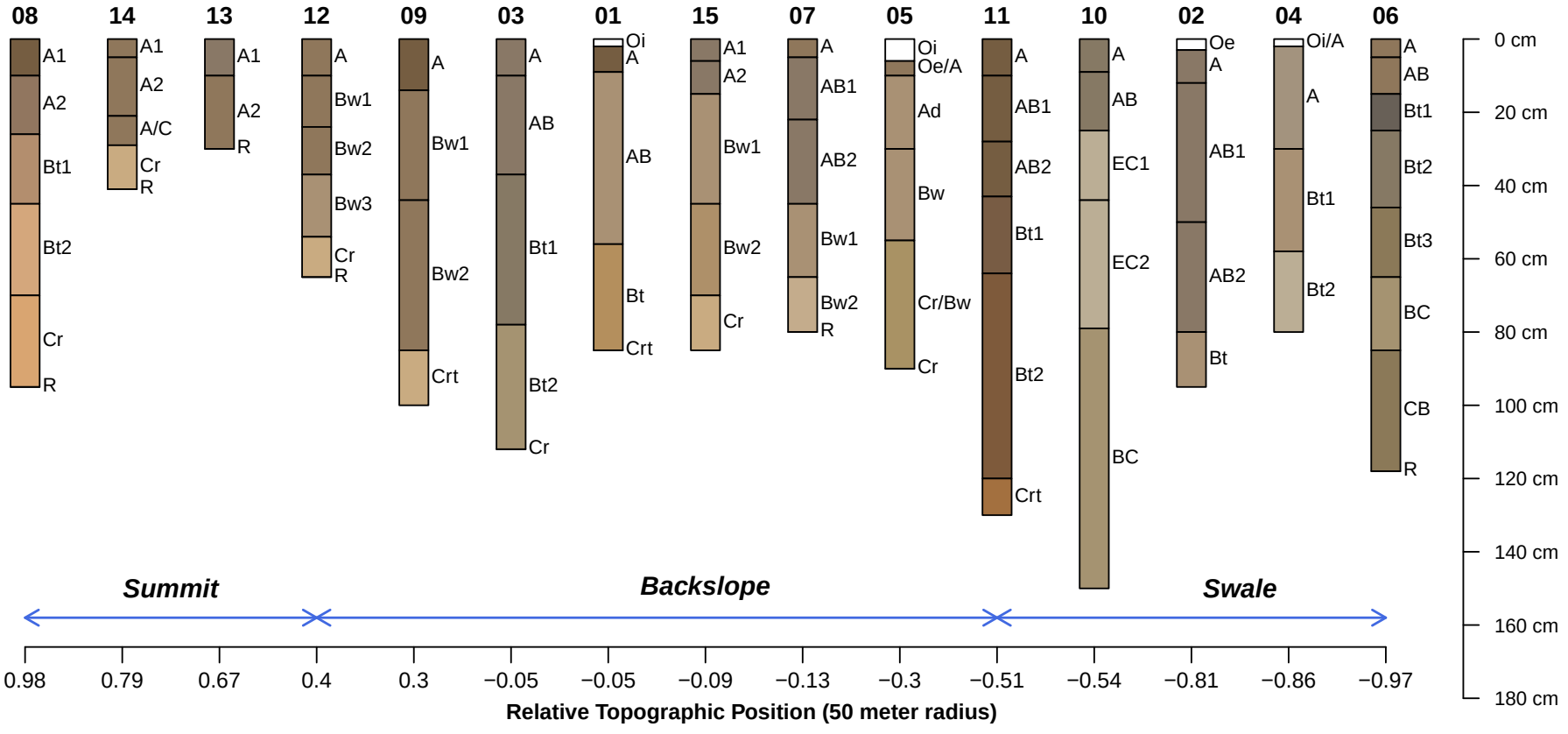
Pair-wise dissimilarities are only accumulated to the *deepest* of two profiles.

Warning: missing data will bias results!



Real data are often sprinkled with missing values $\rightarrow$ D(6, NA) = NA.
Estimate or constrain the dissimilarity calculation to a chunk of non-missing data.

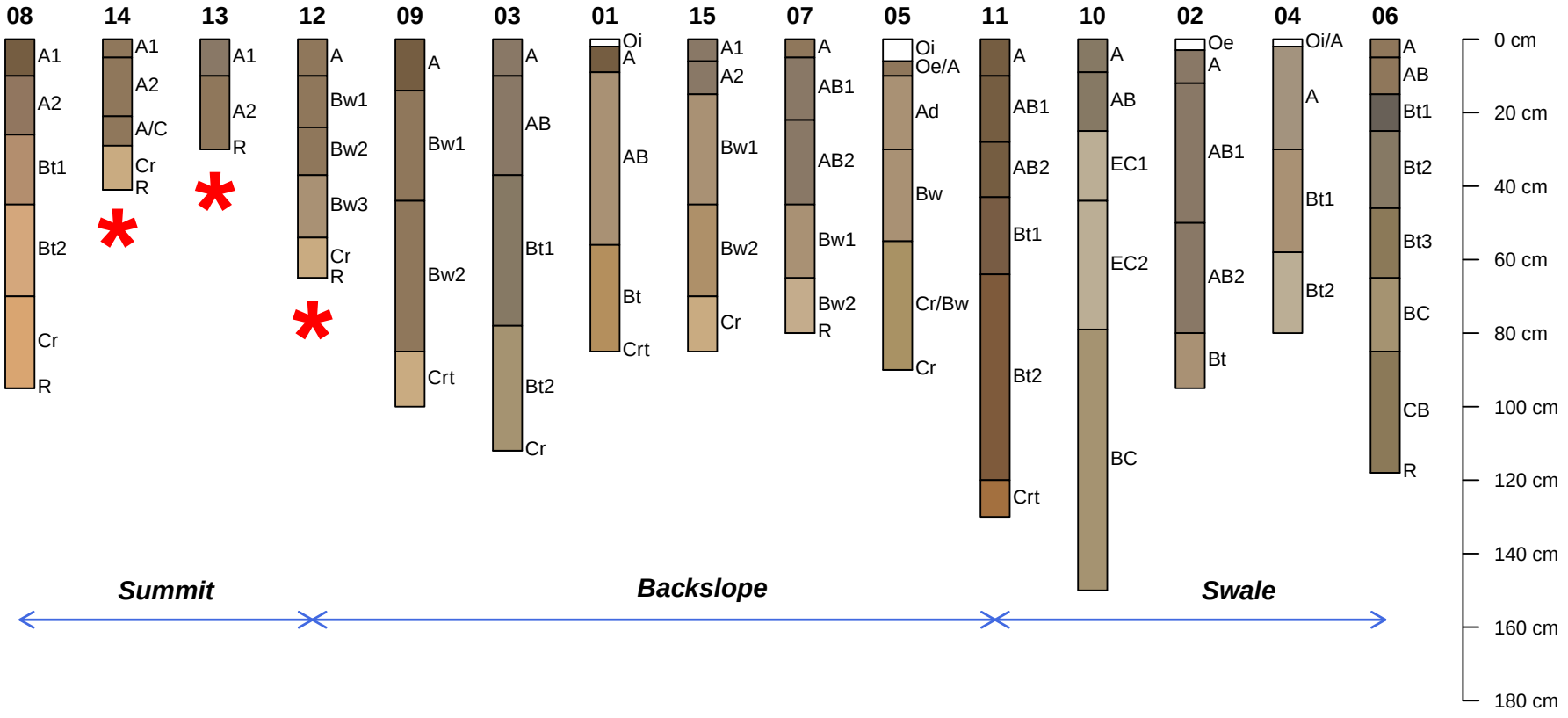
Example: Residual Soils formed on Granite



- summit: Lithic Haploxerolls, Typic Haploxerolls, Mollic Haploxeralfs
- backslope: Typic Haploxerepts, Ultic Haploxerepts
- swale: Oxyaquic Haploxerolls, Typic Argixerolls

slice-wise comparison via: clay, VCS, CEC, pH

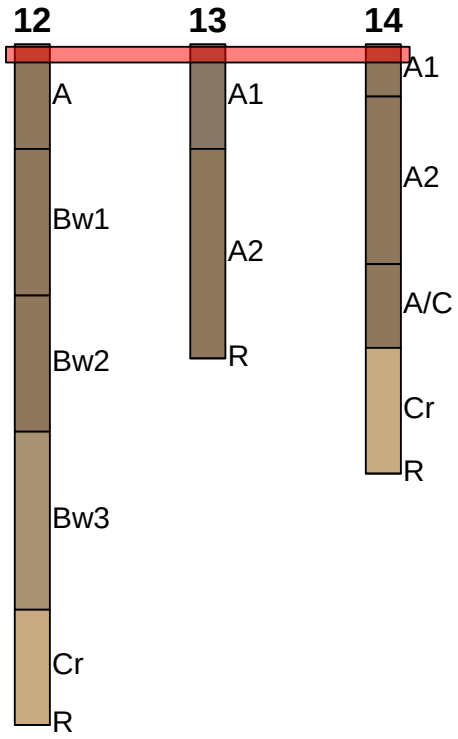
Example: Algorithm Applied to Subset

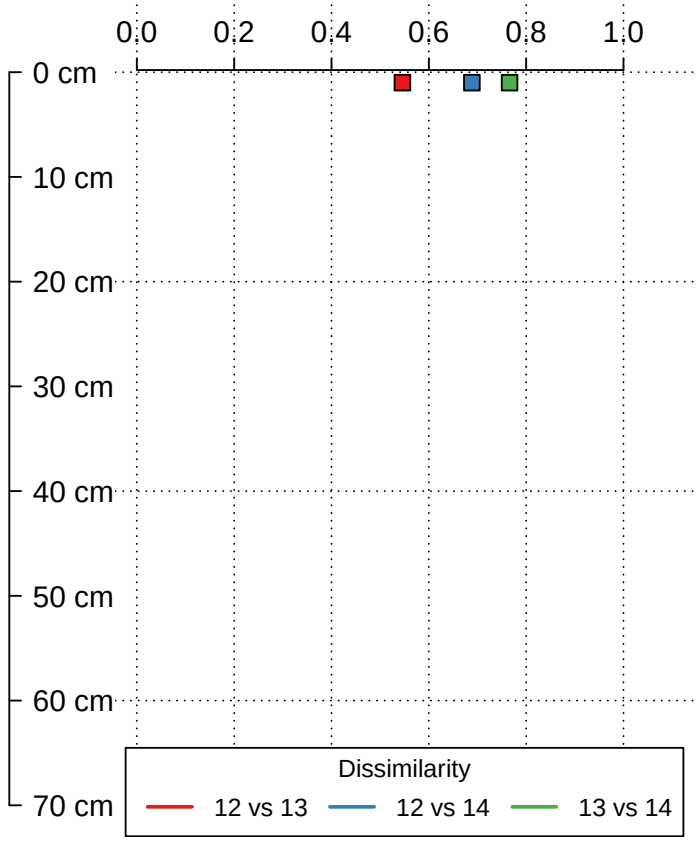


- summit: Lithic Haploxerolls, Typic Haploxerolls, Mollic Haploxeralfs
- backslope: Typic Haploxerepts, Ultic Haploxerepts
- swale: Oxyaquic Haploxerolls, Typic Argixerolls

slice-wise comparison via: clay, VCS, CEC, pH

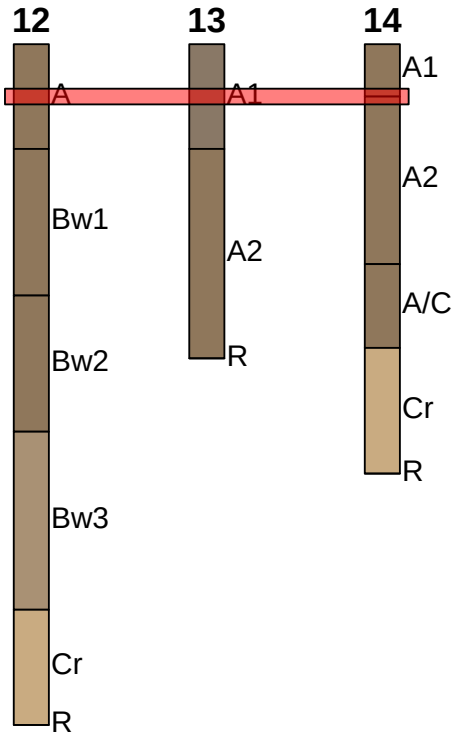
Slice 1

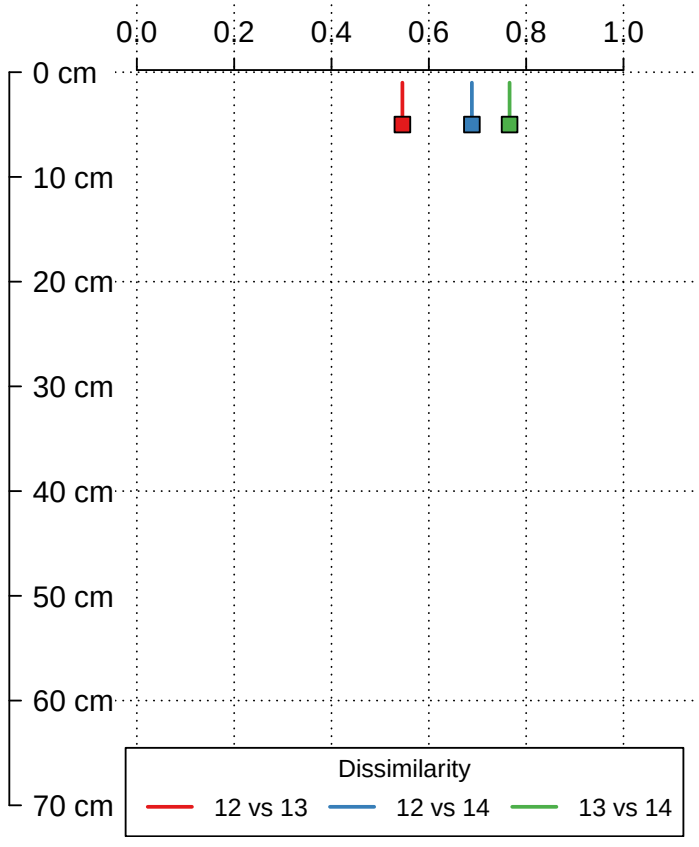

12
13
14
A
Bw1
Bw2
Bw3
Cr
R
A1
A2
R
A1
A2
A/C
Cr
R

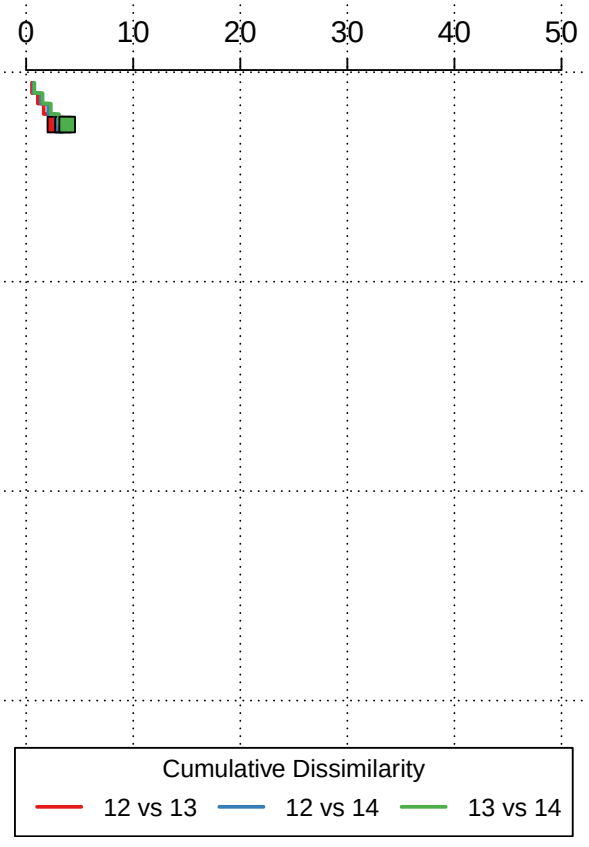

0.0
0.2
0.4
0.6
0.8
1.0
0 cm
10 cm
20 cm
30 cm
40 cm
50 cm
60 cm
70 cm
Dissimilarity
12 vs 13
12 vs 14
13 vs 14


0
10
20
30
40
50
Cumulative Dissimilarity
12 vs 13
12 vs 14
13 vs 14

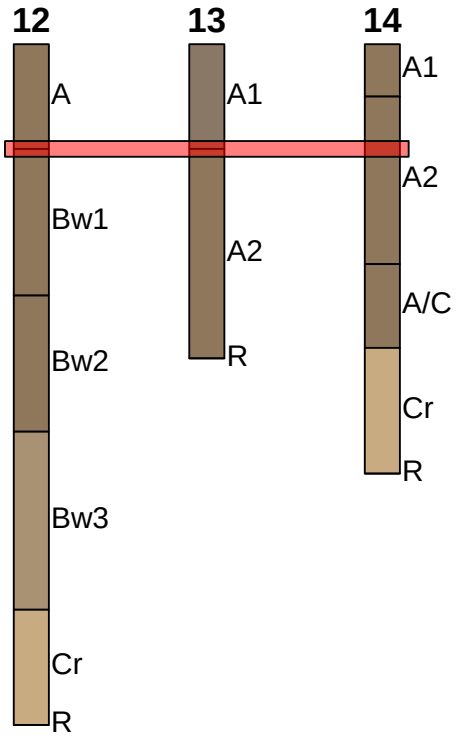
Slice 5

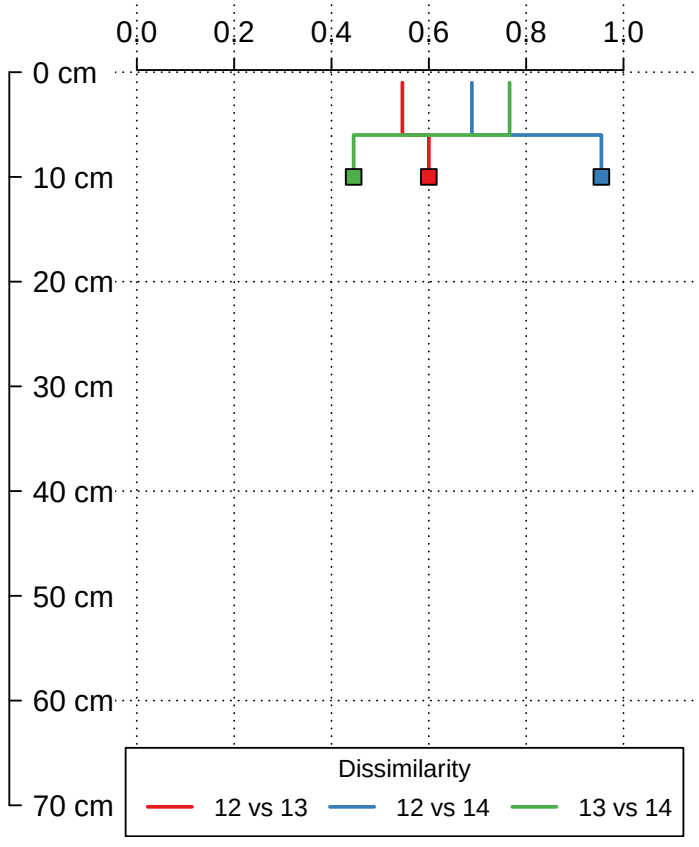

12
13
14
A
A1
A1
Bw1
Bw2
Bw3
Cr
R
A2
R
A2
A/C
Cr
R

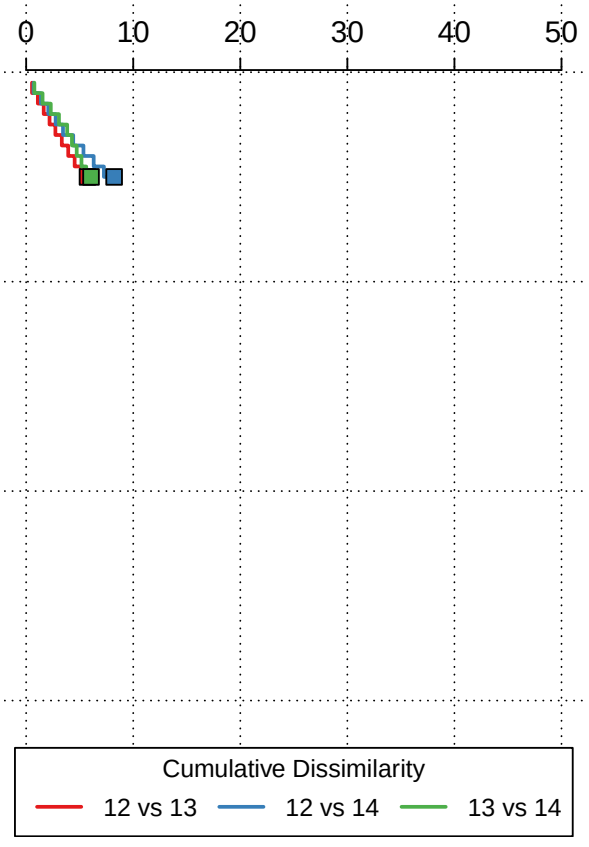

0.0
0.2
0.4
0.6
0.8
1.0
0 cm
10 cm
20 cm
30 cm
40 cm
50 cm
60 cm
70 cm
Dissimilarity
12 vs 13
12 vs 14
13 vs 14


0
10
20
30
40
50
Cumulative Dissimilarity
12 vs 13
12 vs 14
13 vs 14

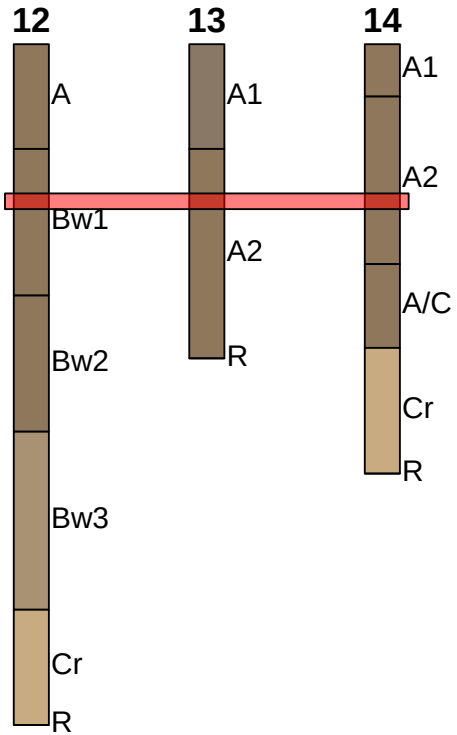
Slice 10

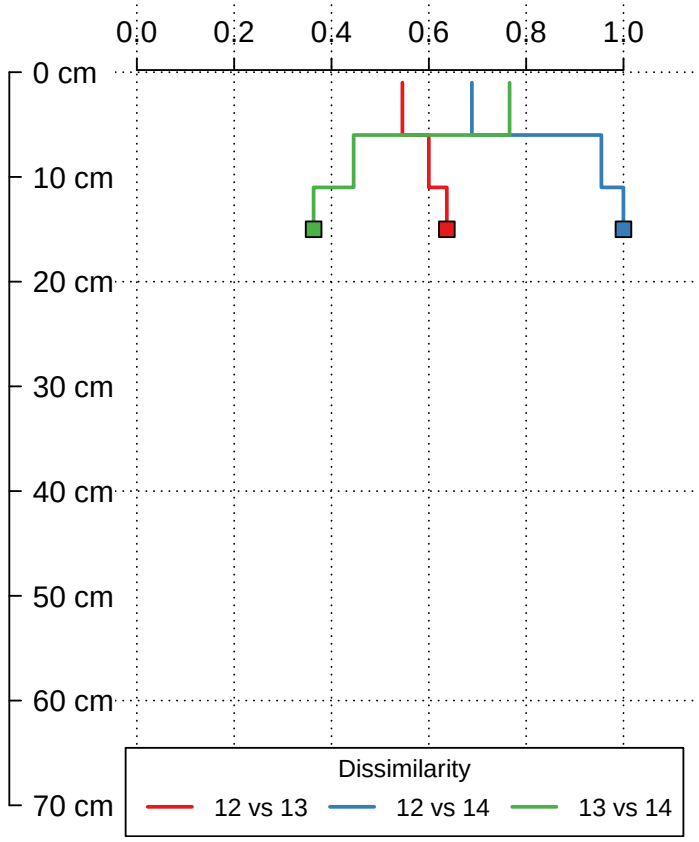

12
13
14
A
Bw1
Bw2
Bw3
Cr
R
A1
A2
R
A1
A2
A/C
Cr
R

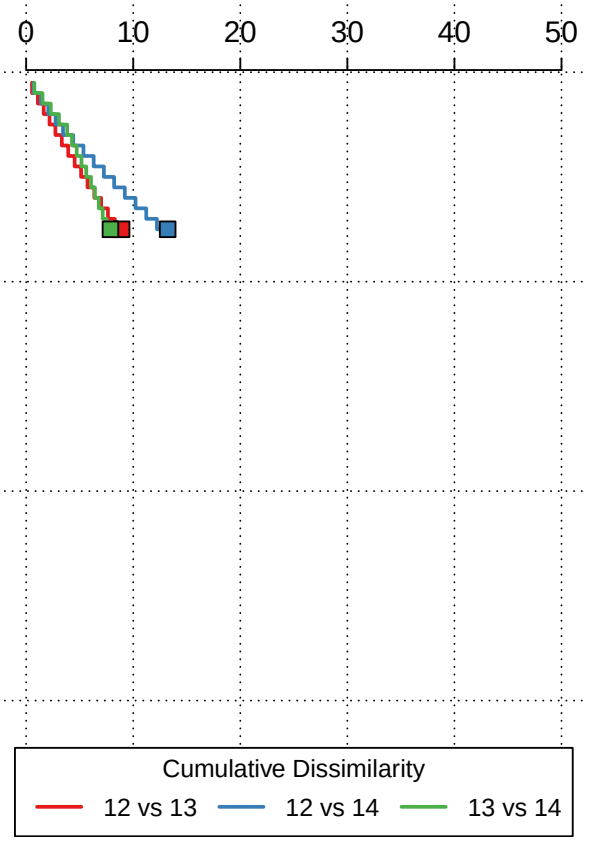

0.0
0.2
0.4
0.6
0.8
1.0
0 cm
10 cm
20 cm
30 cm
40 cm
50 cm
60 cm
70 cm
Dissimilarity
12 vs 13
12 vs 14
13 vs 14


0
10
20
30
40
50
Cumulative Dissimilarity
12 vs 13
12 vs 14
13 vs 14

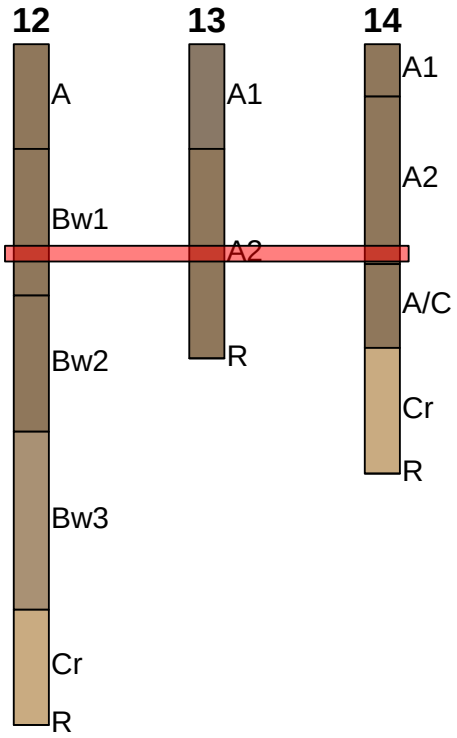
Slice 15

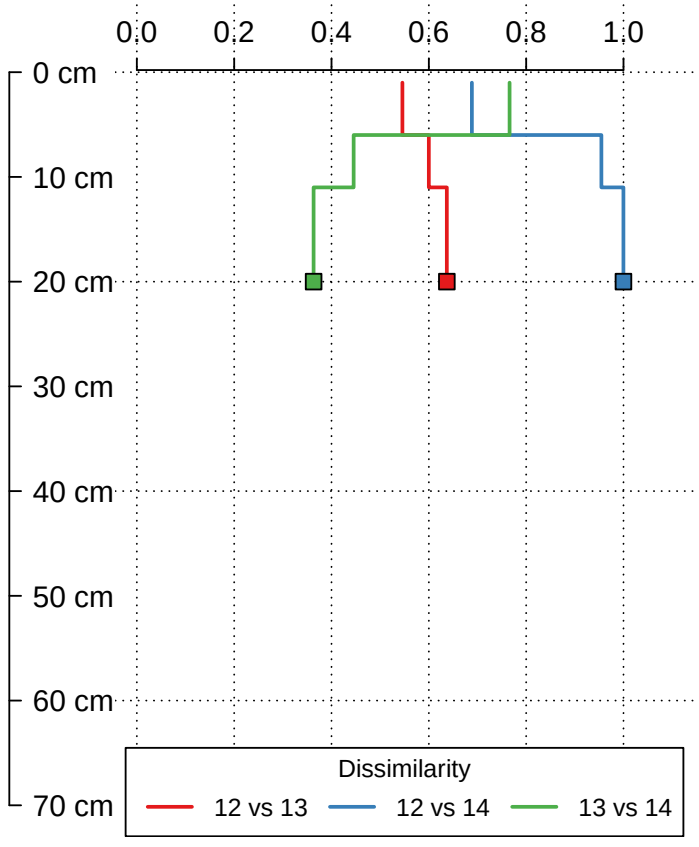

12
13
14
A
Bw1
Bw2
Bw3
Cr
R
A1
A2
R
A1
A2
A/C
Cr
R

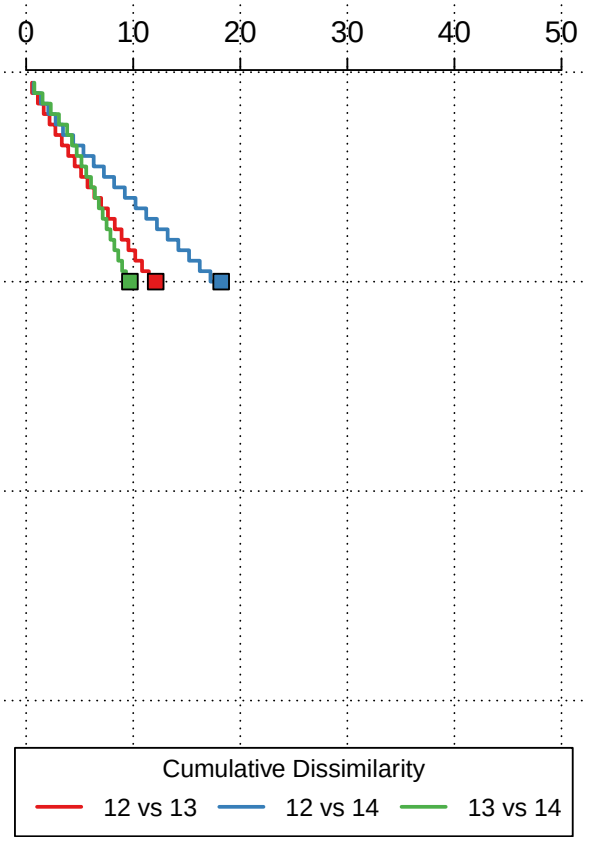

0.0
0.2
0.4
0.6
0.8
1.0
0 cm
10 cm
20 cm
30 cm
40 cm
50 cm
60 cm
70 cm
Dissimilarity
12 vs 13
12 vs 14
13 vs 14


0
10
20
30
40
50
Cumulative Dissimilarity
12 vs 13
12 vs 14
13 vs 14

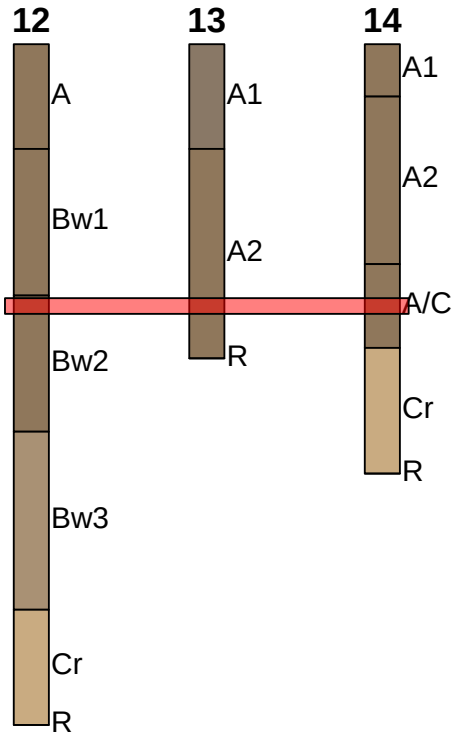
Slice 20

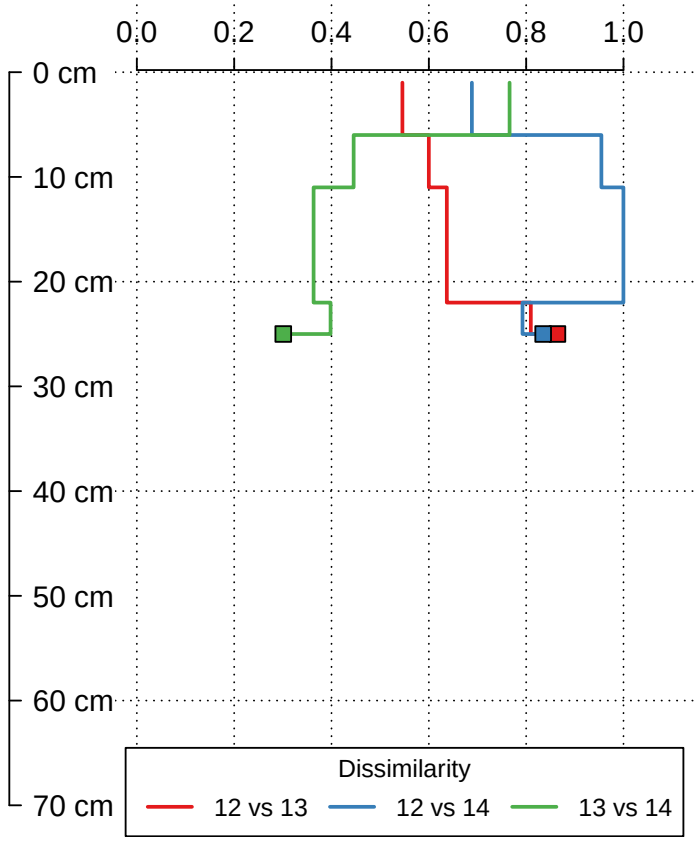

12
13
14
A
Bw1
Bw2
Bw3
Cr
R
A1
A2
R
A1
A2
A/C
Cr
R

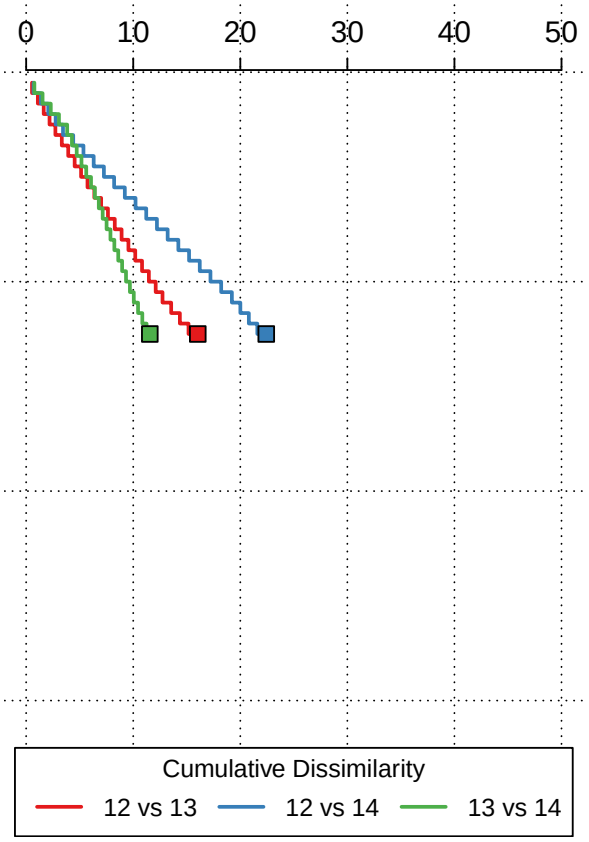

0.0
0.2
0.4
0.6
0.8
1.0
0 cm
10 cm
20 cm
30 cm
40 cm
50 cm
60 cm
70 cm
Dissimilarity
12 vs 13
12 vs 14
13 vs 14


0
10
20
30
40
50
Cumulative Dissimilarity
12 vs 13
12 vs 14
13 vs 14

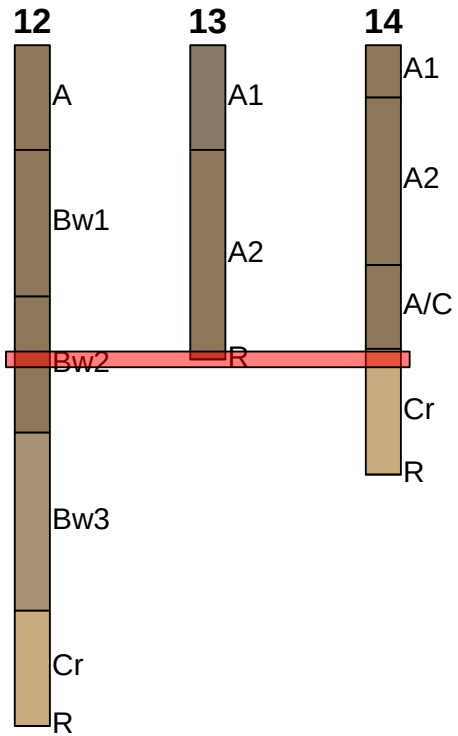
Slice 25

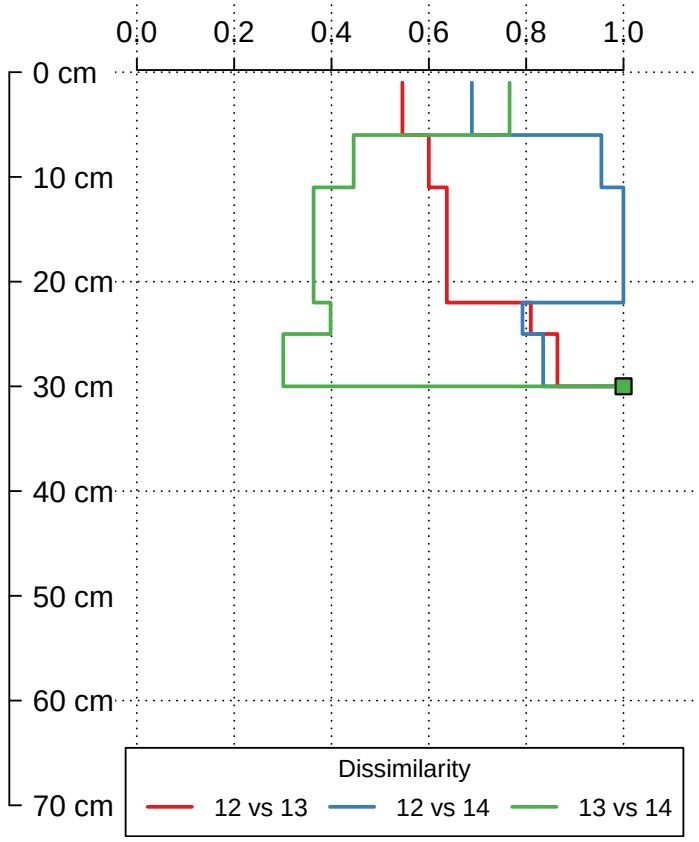

12
13
14
A
Bw1
Bw2
Bw3
Cr
R
A1
A2
R
A1
A2
A/C
Cr
R

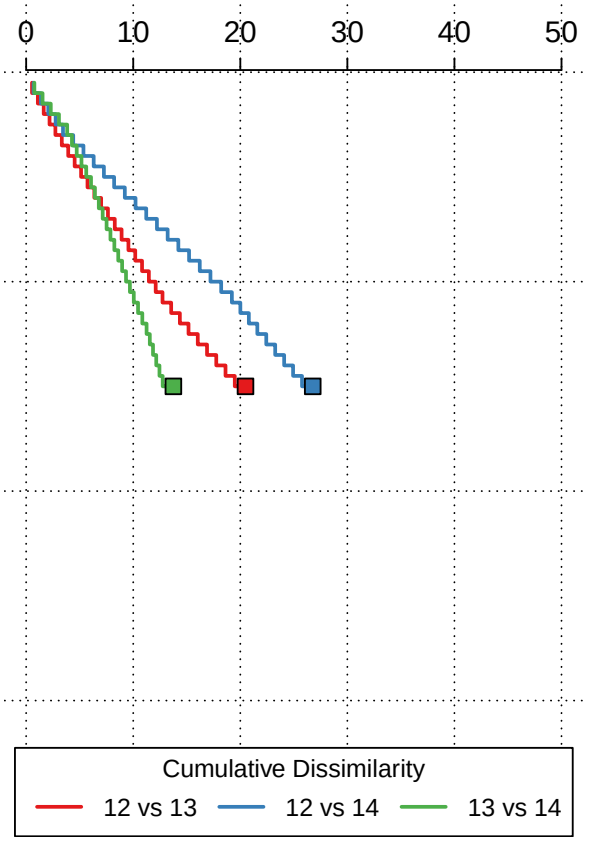

0.0
0.2
0.4
0.6
0.8
1.0
0 cm
10 cm
20 cm
30 cm
40 cm
50 cm
60 cm
70 cm
Dissimilarity
12 vs 13
12 vs 14
13 vs 14


0
10
20
30
40
50
Cumulative Dissimilarity
12 vs 13
12 vs 14
13 vs 14

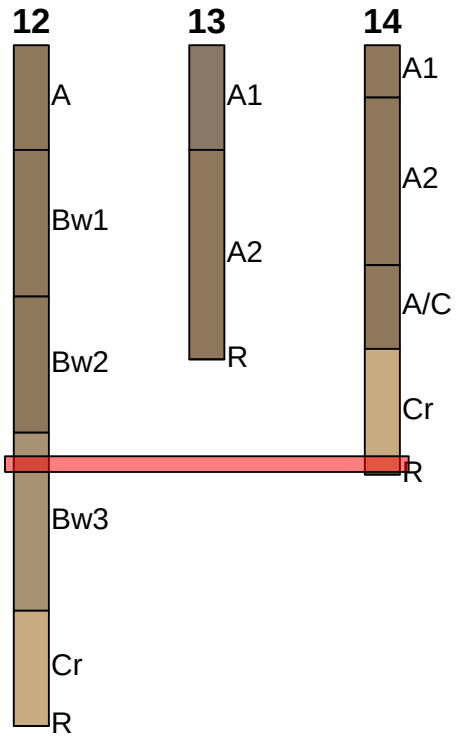
Slice 30

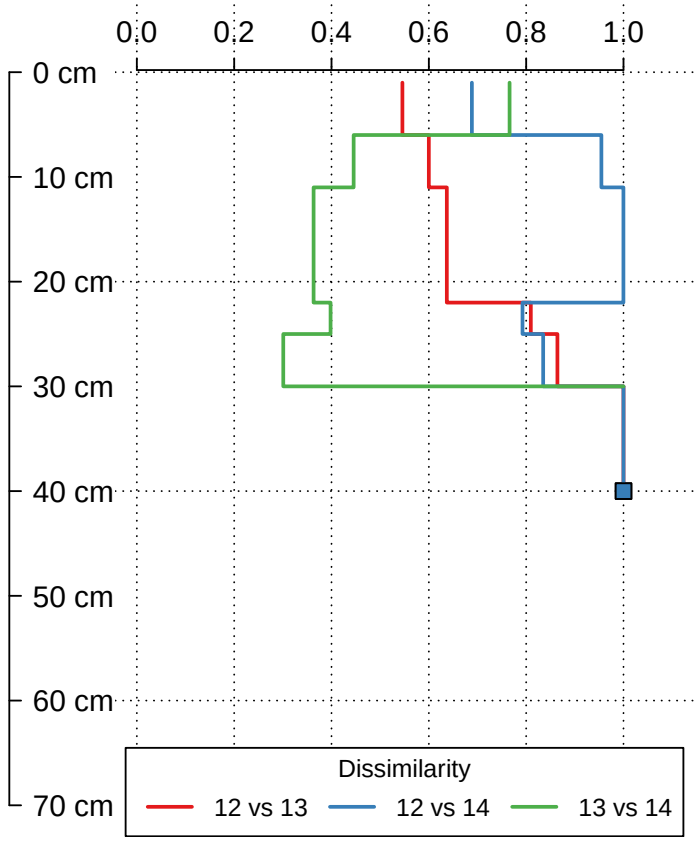

12
13
14
A
Bw1
Bw2
Bw3
Cr
R
A1
A2
R
A1
A2
A/C
Cr
R

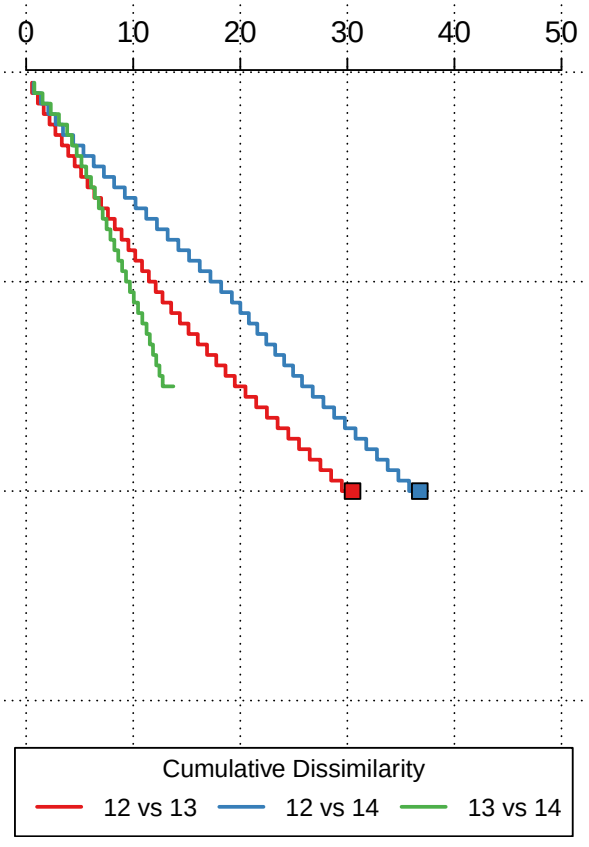

0.0
0.2
0.4
0.6
0.8
1.0
0 cm
10 cm
20 cm
30 cm
40 cm
50 cm
60 cm
70 cm
Dissimilarity
12 vs 13
12 vs 14
13 vs 14


0
10
20
30
40
50
Cumulative Dissimilarity
12 vs 13
12 vs 14
13 vs 14

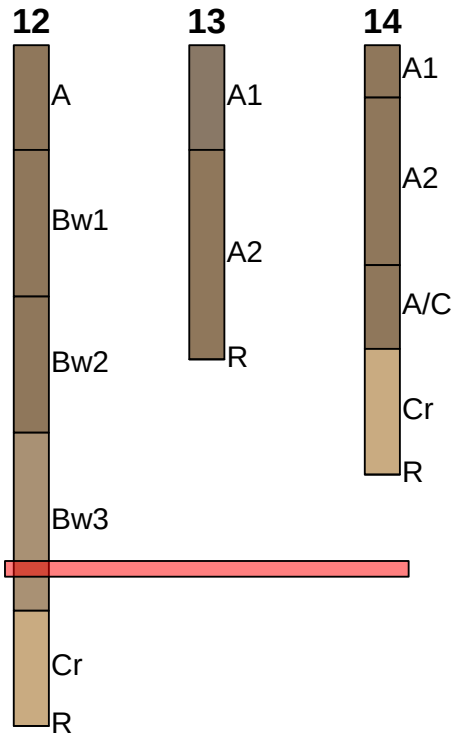
Slice 40

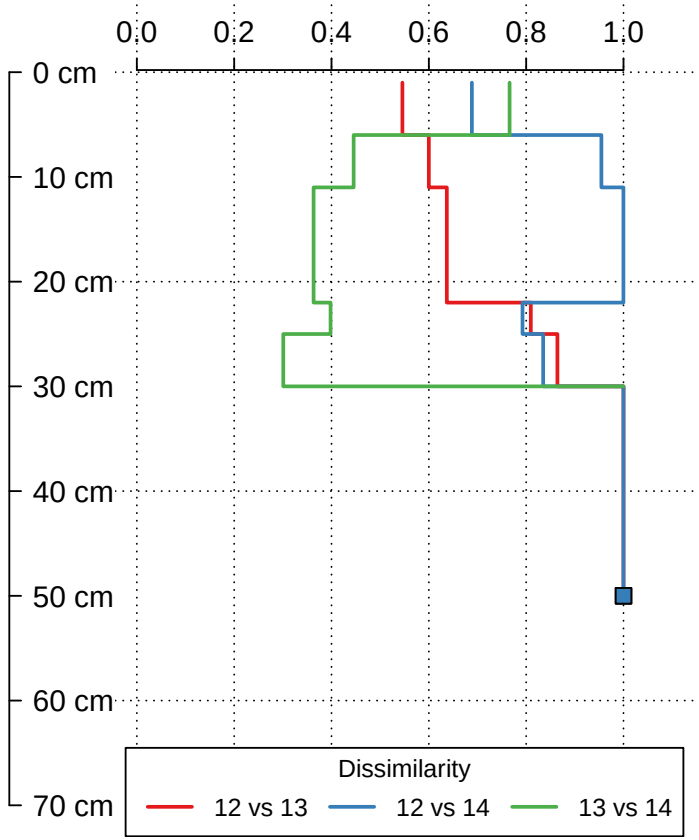

12
13
14
A
Bw1
Bw2
Bw3
Cr
R
A1
A2
R
A1
A2
A/C
Cr
R

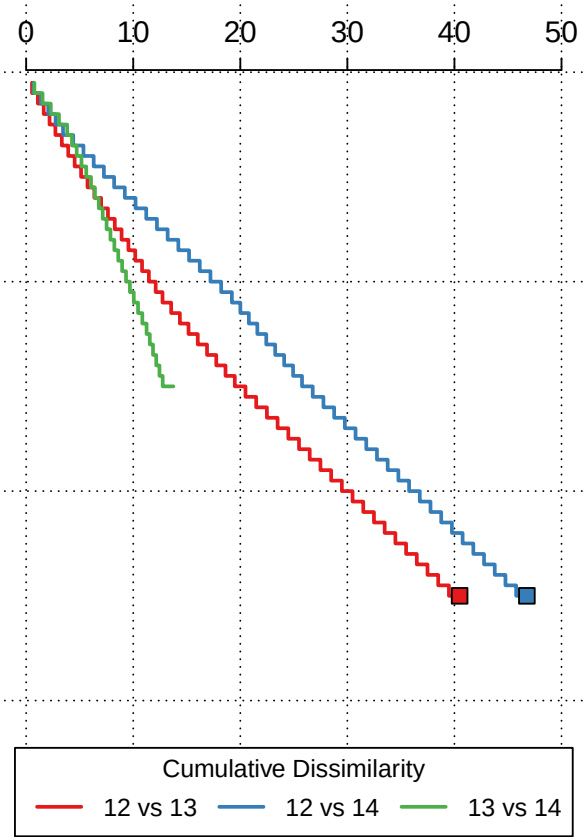

0.0
0.2
0.4
0.6
0.8
1.0
0 cm
10 cm
20 cm
30 cm
40 cm
50 cm
60 cm
70 cm
Dissimilarity
12 vs 13
12 vs 14
13 vs 14


0
10
20
30
40
50
Cumulative Dissimilarity
12 vs 13
12 vs 14
13 vs 14

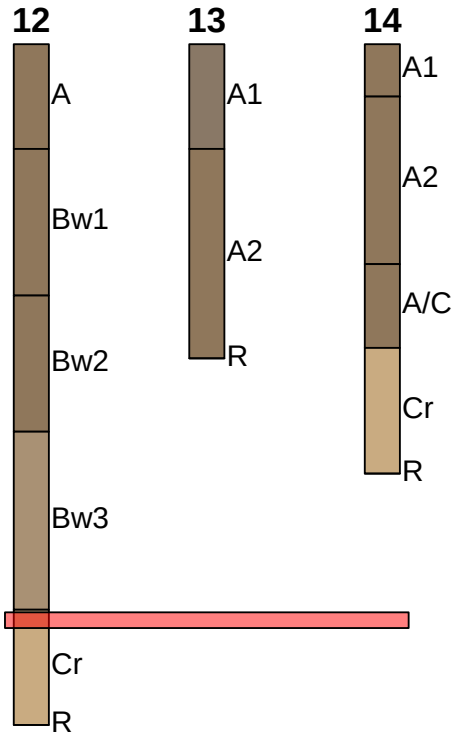
Slice 50

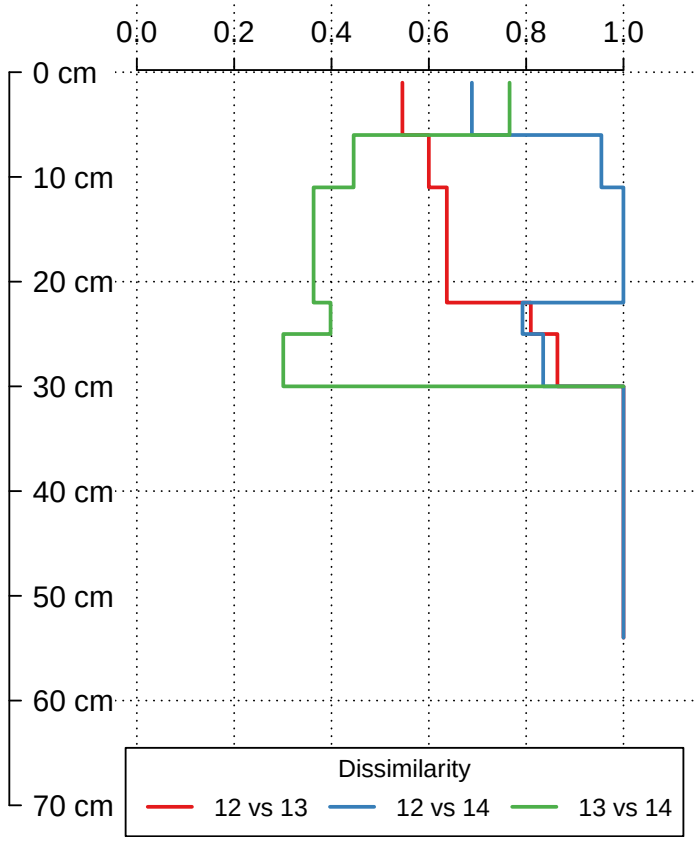

12
A
Bw1
Bw2
Bw3
Cr
R
13
A1
A2
R
14
A1
A2
A/C
Cr
R

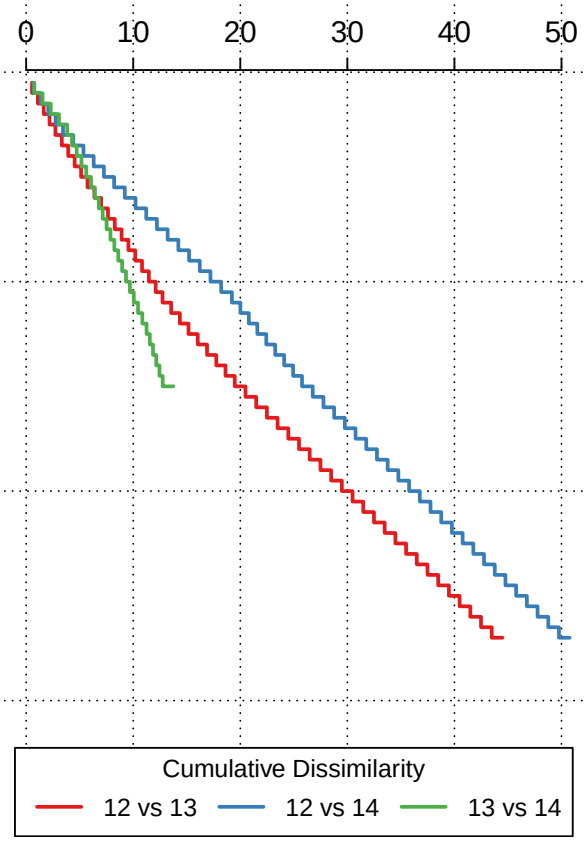

0.0
0.2
0.4
0.6
0.8
1.0
0 cm
10 cm
20 cm
30 cm
40 cm
50 cm
60 cm
70 cm
Dissimilarity
12 vs 13
12 vs 14
13 vs 14


0
10
20
30
40
50
Cumulative Dissimilarity
12 vs 13
12 vs 14
13 vs 14

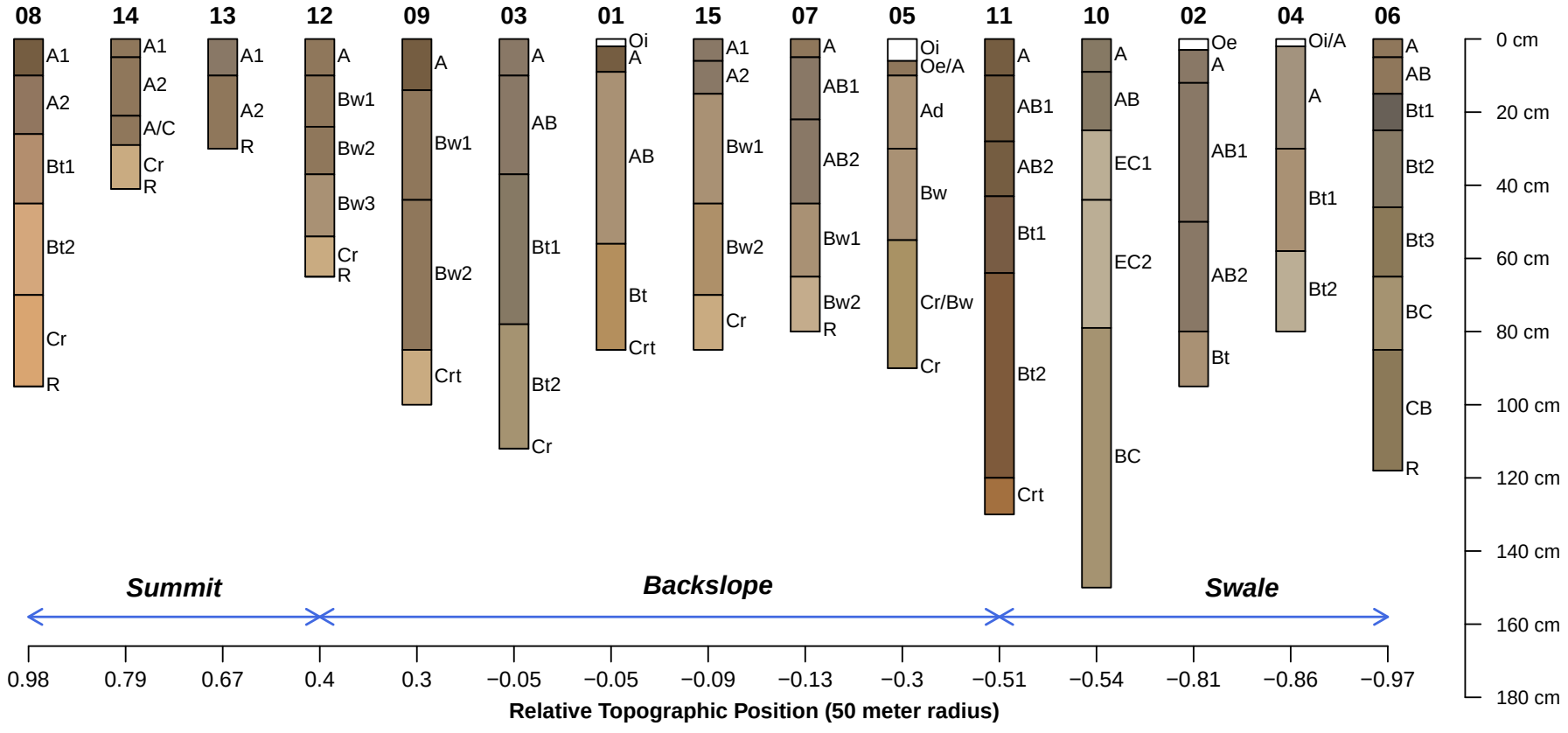
Slice 55


12
A
Bw1
Bw2
Bw3
Cr
R
13
A1
A2
R
14
A1
A2
A/C
Cr
R


0.0
0.2
0.4
0.6
0.8
1.0
0 cm
10 cm
20 cm
30 cm
40 cm
50 cm
60 cm
70 cm
Dissimilarity
12 vs 13
12 vs 14
13 vs 14


0
10
20
30
40
50
Cumulative Dissimilarity
12 vs 13
12 vs 14
13 vs 14

Example: Algorithm Applied to Entire Collection



- summit: Lithic Haploxerolls, Typic Haploxerolls, Mollic Haploxeralfs
- backslope: Typic Haploxerepts, Ultic Haploxerepts
- swale: Oxyaquic Haploxerolls, Typic Argixerolls

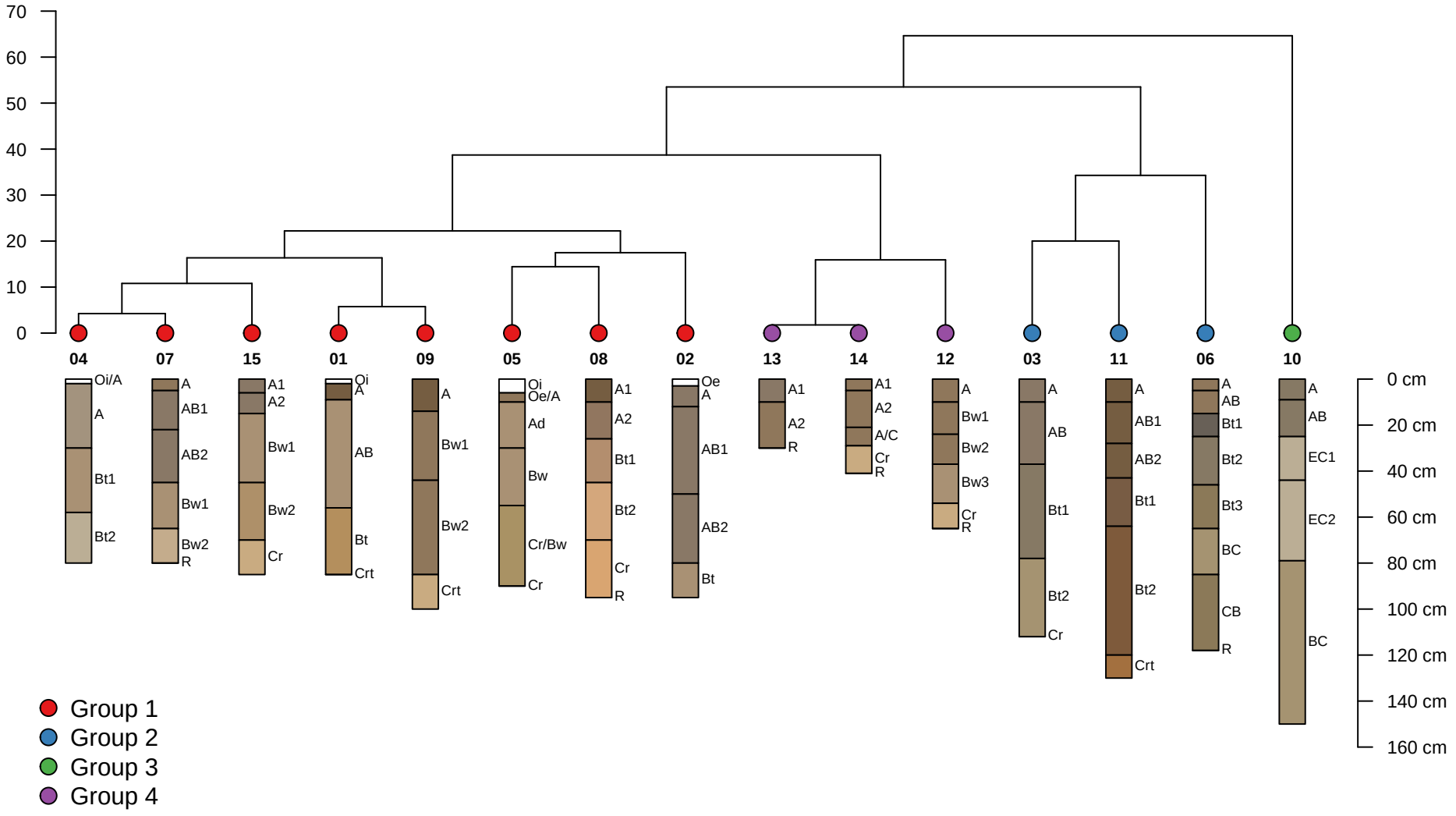
slice-wise comparison via: clay, VCS, CEC, pH

Results: Pair-Wise Dissimilarities

	01	02	03	04	05	06	07	08	09	10	11	12	13	14
02	25													
03	41	33												
04	22	25	35											
05	19	23	38	24										
06	68	49	42	60	64									
07	20	27	35	7	22	61								
08	30	27	39	27	22	63	25							
09	9	28	43	20	22	72	20	30						
10	71	68	50	67	68	70	68	67	69					
11	53	48	31	51	53	53	47	51	54	68				
12	32	46	57	32	39	77	29	44	29	84	61			
13	50	60	70	44	56	83	42	55	48	100	74	24		
14	51	60	71	44	56	83	41	55	49	100	75	25	3	
15	25	34	44	17	29	70	15	32	24	75	51	21	33	35

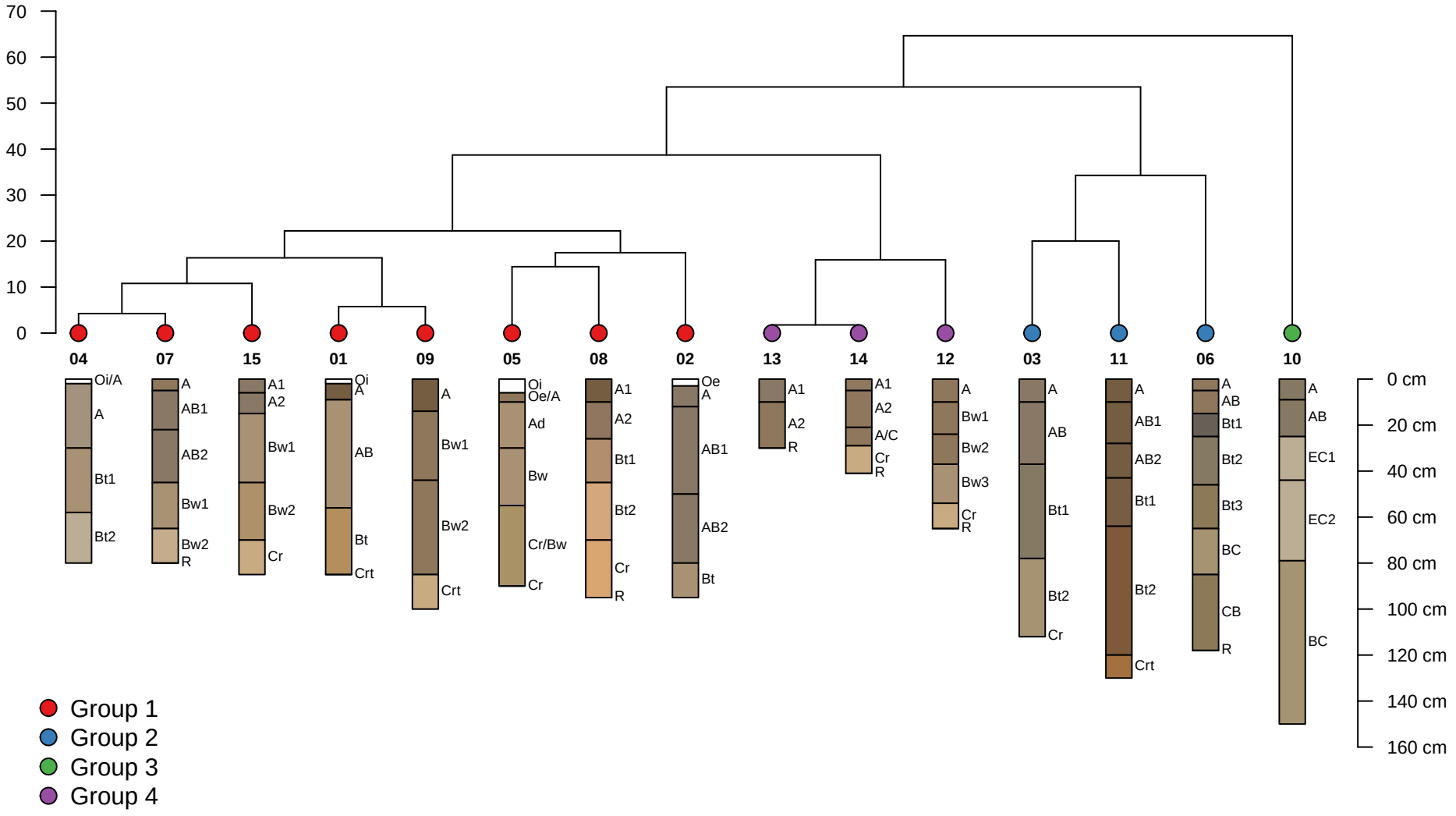
how is this useful?

Results: Dendrogram Representation



divisive, hierarchical clustering

Results: Dendrogram Representation



divisive, hierarchical clustering

what does it mean?

Conclusions, Ideas, Caveats

Innovation(?) Relative to (Moore et al., 1972) and Others

- direct parameterization of soil depth
- accomodation of binary/nominal/ordinal/ratio variables: Gower's Distance
- simple implementation, readily scaled to HPC
- integration of hz-scale and pedon-scale variables: $D = \frac{w_{hz} D_{hz} w_p D_p}{2}$

Possible Applications

- similar/dissimilar decisions (map unit composition / OSD house-cleaning)
- automated allocation / identification of outliers
- bridging classification systems
- distance between taxa (Inceptisol $<<<<$ Alfisol $<<$ Ultisol)
- evaluation of *functional* differences

Caveats

- fundamental assumption: comparison along depth-slices makes sense
- missing data heavily bias results $\rightarrow$ limit application


Thank You
Algorithms for Quantitative Pedology:
http://aqp.r-forge.r-project.org