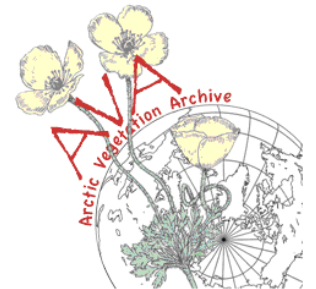


Approaches to storing and analyzing geobotanical data

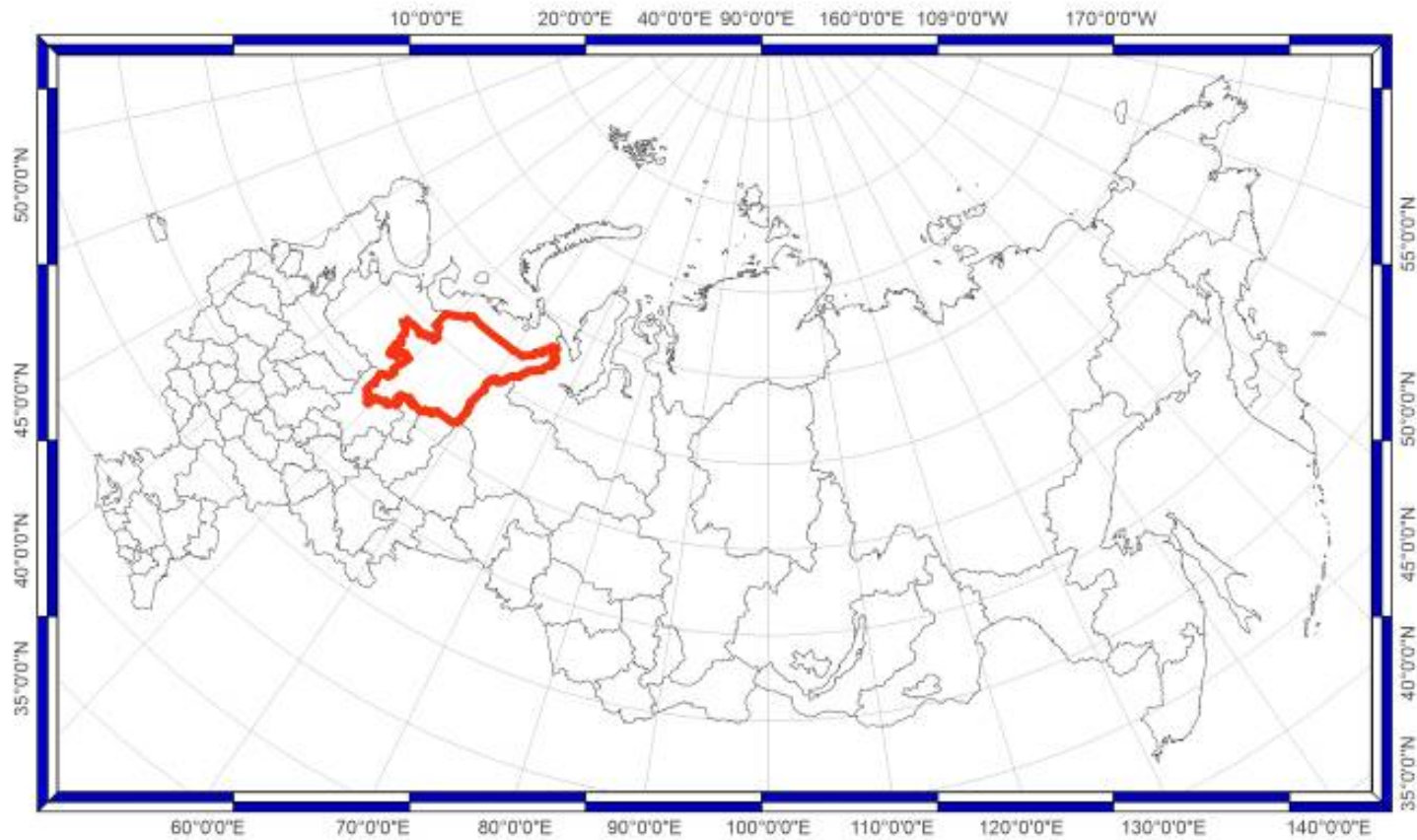
Alexander Novakovskiy

Institute of Biology, Komi Scientific
Centre, Russian Academy of Science

Russia, Syktyvkar



Area of the data collecting



Research is being conducted from 1950 to the present day

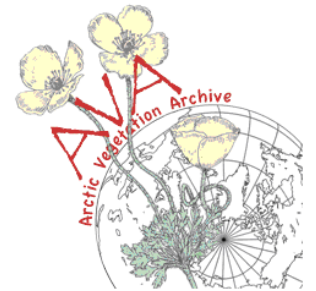
Main investigation regions: Komi republic (flatlands in the west, Northern and Subpolar Ural mountains in the east); more northern area - Nenets region (Bolshezemelskaya Tundra, the Pechora River delta, Polar Ural mountains)



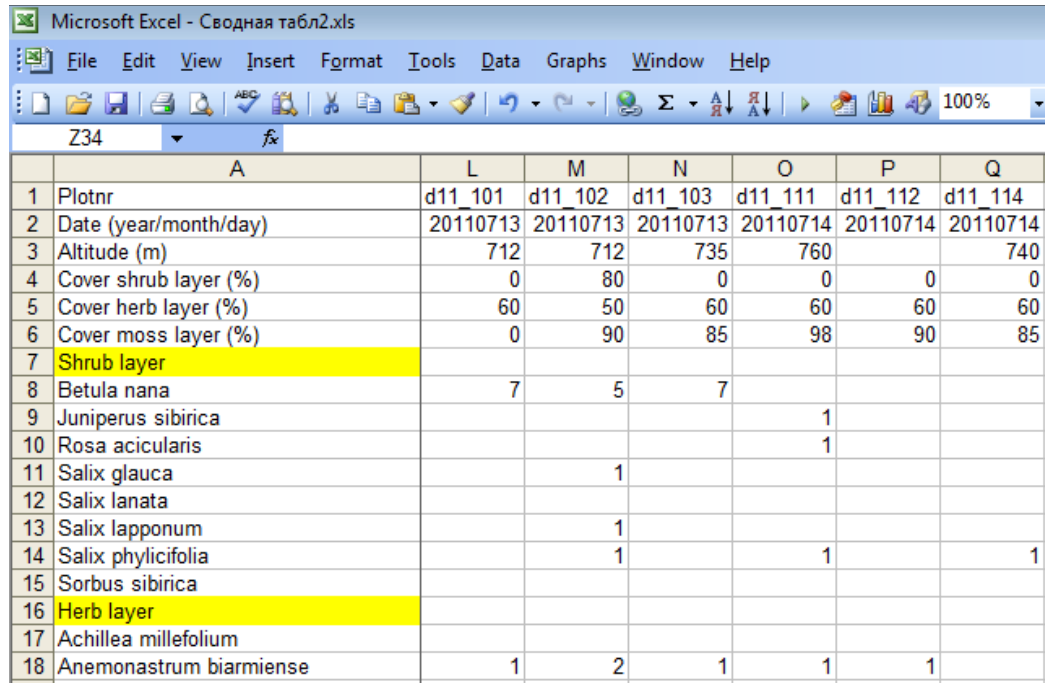
Input (basic) data



There are about 5000-6000 original relevés , collected in 1950-1990s in the vegetation plots collection of the Institute. These relevés still don't convert into a digital format.



Using Excel



Microsoft Excel - Сводная табл2.xls

File Edit View Insert Format Tools Data Graphs Window Help

Z34

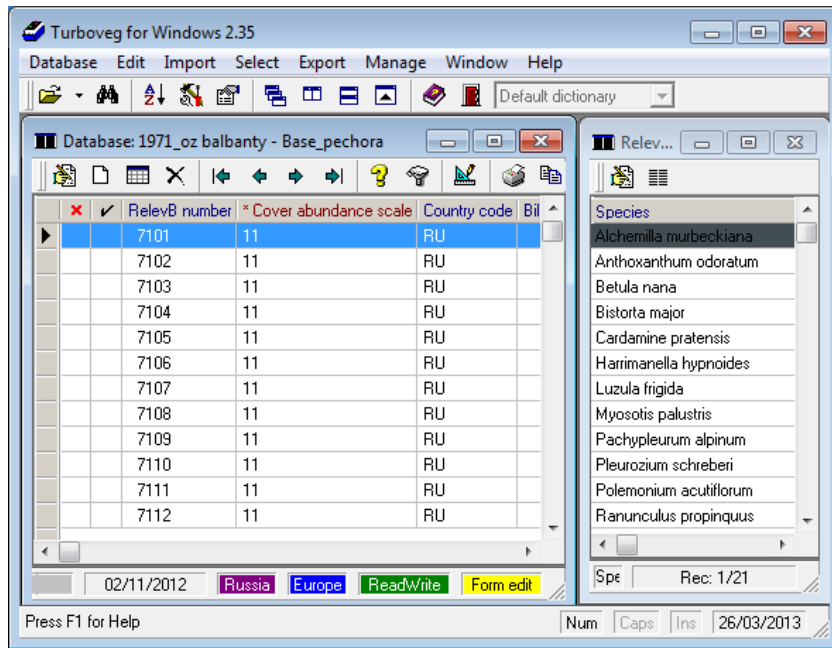
	A	L	M	N	O	P	Q
1	Plotnr	d11_101	d11_102	d11_103	d11_111	d11_112	d11_114
2	Date (year/month/day)	20110713	20110713	20110713	20110714	20110714	20110714
3	Altitude (m)	712	712	735	760		740
4	Cover shrub layer (%)	0	80	0	0	0	0
5	Cover herb layer (%)	60	50	60	60	60	60
6	Cover moss layer (%)	0	90	85	98	90	85
7	Shrub layer						
8	Betula nana	7	5	7			
9	Juniperus sibirica				1		
10	Rosa acicularis				1		
11	Salix glauca		1				
12	Salix lanata						
13	Salix lapponum		1				
14	Salix phylicifolia		1		1		1
15	Sorbus sibirica						
16	Herb layer						
17	Achillea millefolium						
18	Anemonastrum biarmense	1	2	1	1	1	

Since 2000-s the most of releves were converted into digital format. Initially, Microsoft Excel was used for it.

Excel tables are not a real database, and there were a lot of difficulties to use these tables. But this form of data storage allow us to estimate some characteristics (average abundance, species fidelity, calculate numbers of species etc.) and also makes it easy to work with vegetation plots tables for classification purposes.



Using TurboVeg



Last 5-7 years releves from hard (paper) copies and Excel format were being converted into Turboveg format. Today there are about 3500 releves

The screenshot shows the 'Edit relev 6508' form. The fields and their values are:

- E_coord: 5837377
- N_coord: 6303247
- Site: (empty)
- Plotnr: y0917.1
- Diam_htcm: 0
- Diam_lcm: 0
- Land_unit: FSs
- EII_f: 6.5
- EII_n: 3.1
- EII_r: 4.2
- EII_l: 6.6
- Relief: (empty)
- Frst_type: (empty)
- Land_un_id: 0
- Ram_mois: 74
- Ram_rs: 6
- Cg_acid: 5
- Cg_illum: 4

Buttons on the right side include: Confirm (checked), Next, Previous, Save, Exit, and Help.

Extra information in Turboveg database

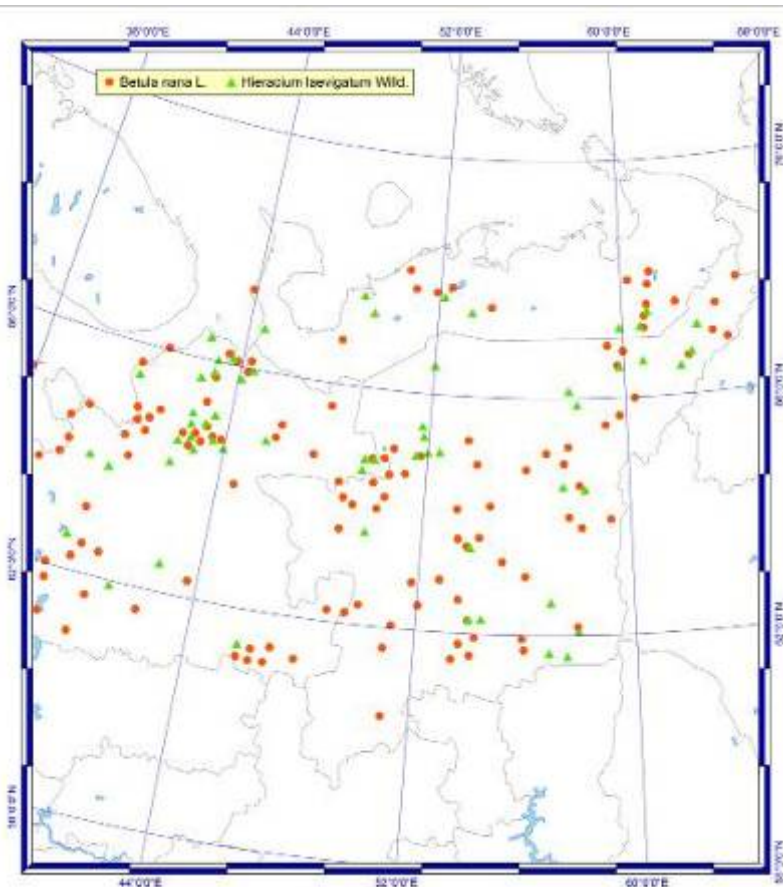
E_coord, N_coord – latitude and longitude of the revele

Plotnr – unique identification code

Land_unit – index of vegetation type

EII_f, EII_n, EII_r, EII_l, Ram_mois, Ram_rs, Cg_acid, Cg_illum – Values of Ellenberg, Ramensky and Ciganov ecological scales, calculated for certain releves according to their species composition

Digital version of four-volume “Flora of the USSR European Northeast”



Distribution map of *Betula Nana* L. and *Hieracium Laevigatum* Wild.



Contain information about distribution of:

- 1500 vascular plants
- 2000 lichens and mosses



Typical way of geobotanical data analysis



Hard copies

Microsoft Excel - Сводная таблица.xls

	L	M	N	O	P	Q
1 Plotnr	d11_101	d11_102	d11_103	d11_111	d11_112	d11_114
2 Date (year/month/day)	20110713	20110713	20110713	20110714	20110714	20110714
3 Altitude (m)	712	712	735	760		740
4 Cover shrub layer (%)	0	80	0	0	0	0
5 Cover herb layer (%)	60	50	60	60	60	60
6 Cover moss layer (%)	0	90	85	98	90	85
7 Shrub layer						
8 Betula nana	7	5	7			
9 Juniperus sibirica				1		
10 Rosa acicularis					1	
11 Salix glauca		1				
12 Salix lanata						
13 Salix lapponum			1			
14 Salix phylicifolia			1			1
15 Sorbus sibirica					1	
16 Herb layer						
17 Achillea millefolium						
18 Anemonastrum biarmense	1	2	1	1	1	

Turboveg for Windows 2.35

Reliab number	Cover abundance scale	Country code	BI
7101	11	RU	
7102	11	RU	
7103	11	RU	
7104	11	RU	
7105	11	RU	
7106	11	RU	
7107	11	RU	
7108	11	RU	
7109	11	RU	
7110	11	RU	
7111	11	RU	
7112	11	RU	

TurboVeg

PC-ORD

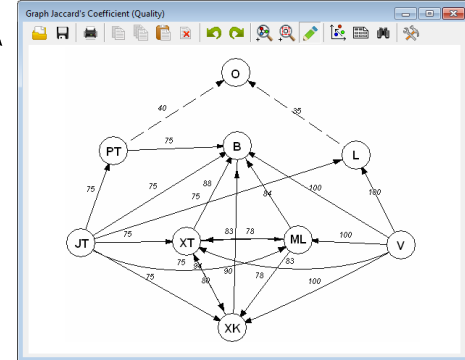
	C	C	C
22			
81			
	Leptolyn	Leptolyn	Pseudana
s17	0	0	0
s18	1	0	0
s19	0	1	0
s20	0	0	0

SPSS

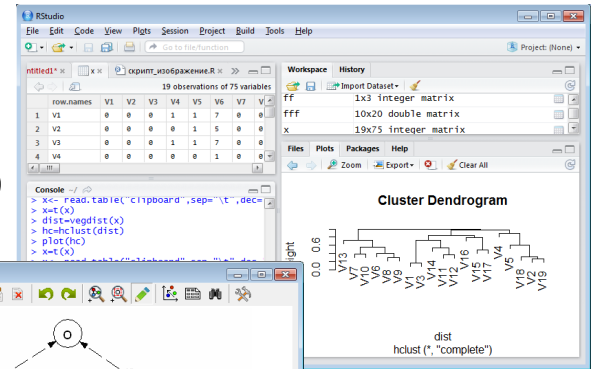
*короткий.sav (Набор данных)3 - Редактор данных IBM SPSS Statistics

PlotnNR	Num...	Mi	Ei	Land_u	Eil_f	Eil_n
1	d11_165	ЦП9	.00	679.00	6.9	.0
2		ЦП19	.00	646.00		.0
3	d11_184a	ЦП13	.00	724.00	6.9	7.0
4	d11_216b	ЦП20	.00	634.00	7.9	.0
5		ЦП8	.00	717.00		.0
6		ЦП16	.00	.00		.0
7	d11_155	ЦП7	.00	556.00	6.8	.0
8		ЦП1	.00	692.00		.0
9	d11_220a	ЦП22	.00	621.00	6.5	3.0
10	d11_137a	ЦП5	.00	690.00	6.5	.0

PC-ORD



GRAPHS

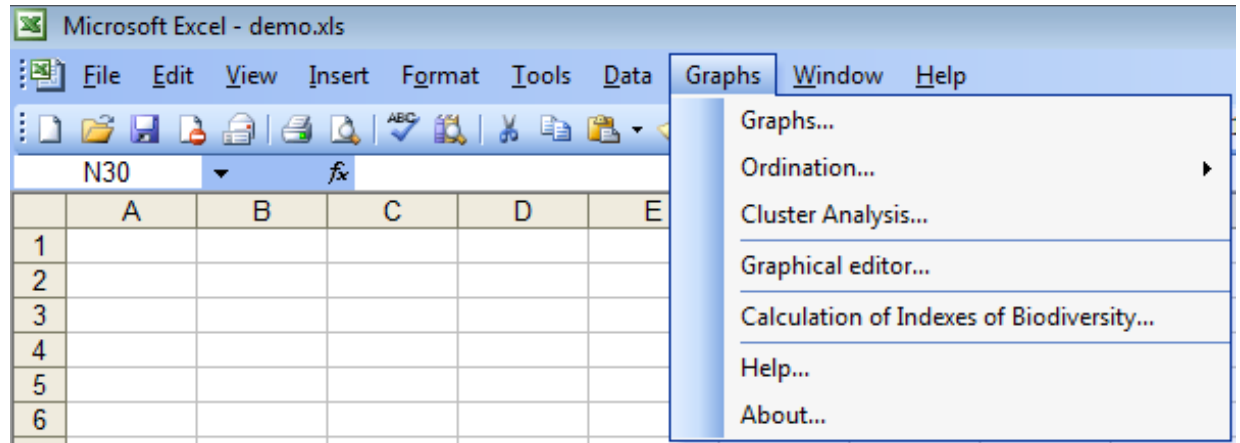


R



Module GRAPHS

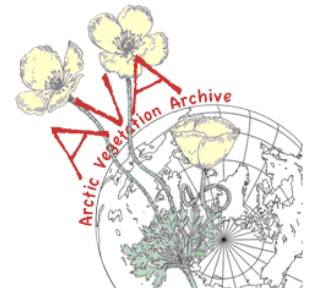
The module GRAPHS is an Excel add-on. After installation the additional sub-menu appears. Researchers can use any data formats compatible with Microsoft Excel and all Excel abilities to prepare their data



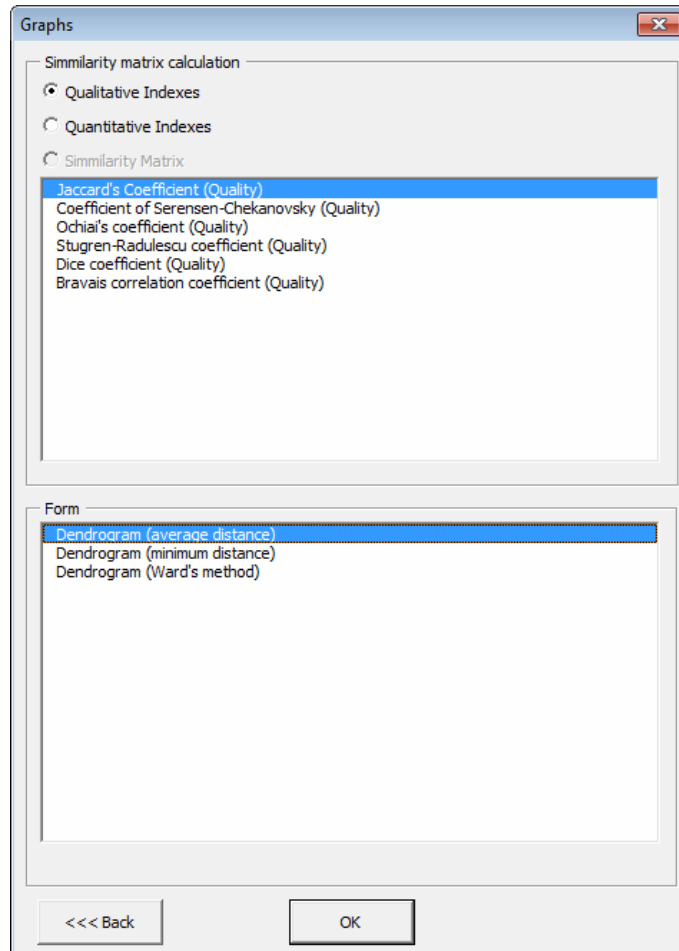
A screenshot of the Microsoft Excel interface showing a data table. The table has columns labeled A through I and rows numbered 1 through 13. The first row is a header for 'Species\Communities' and sites. The following rows list species names and their corresponding values across the sites. Annotations include a circle around the species names and a line pointing to the site headers.

	A	B	C	D	E	F	G	H	I
1	Species\Communities	Site1	Site2	Site3	Site4	Site5	Site6	Site7	Site8
2	Aconitum excelsum		0.1			10			
3	Adoxa moschatellina			2			2		8
4	Alopecurus pratense	0.1		5					
5	Angelica archangelica	1							5
6	Anthoxantum odoratum				10				
7	Arctous alpina								
8	Astragalus frigidus								
9	Bartsia alpina	3							
10	Betula nana	50	0.1	20	0.1	5		5	5
11	Bistorta major	3							
12	Bistorta viviparum		0.1						
13	Cardamina pratense		0.1						

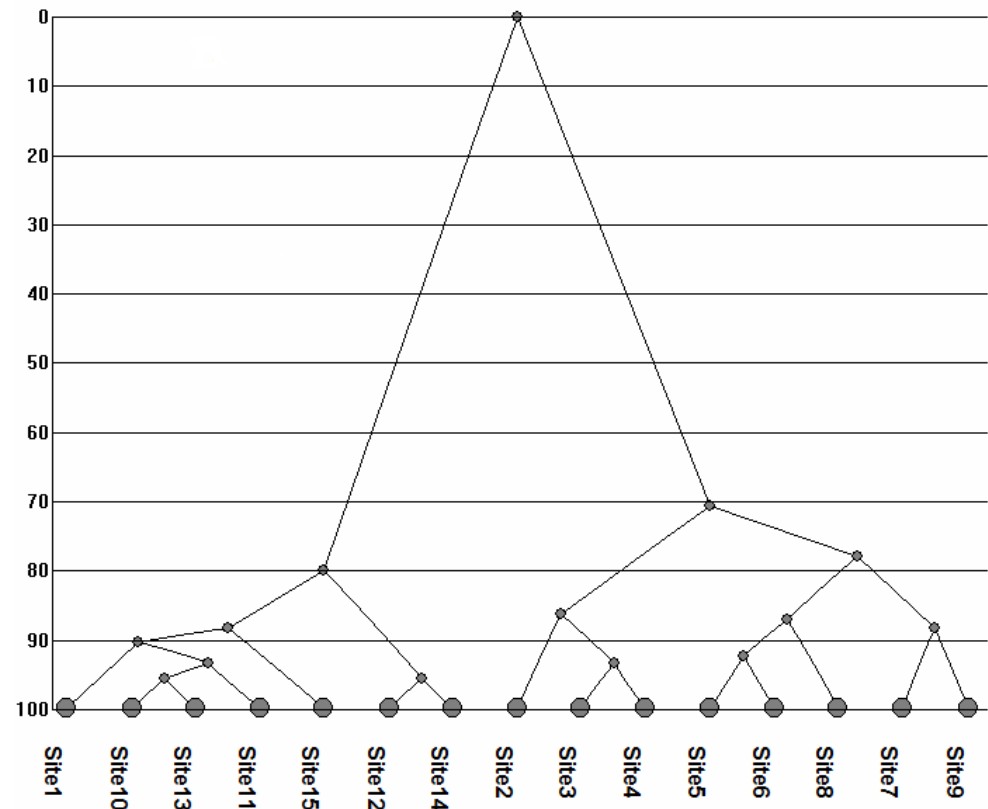
Data source – classic geo-botanical table where columns are plots (objects), – rows are species (properties).



Cluster analysis



Cluster analysis settings



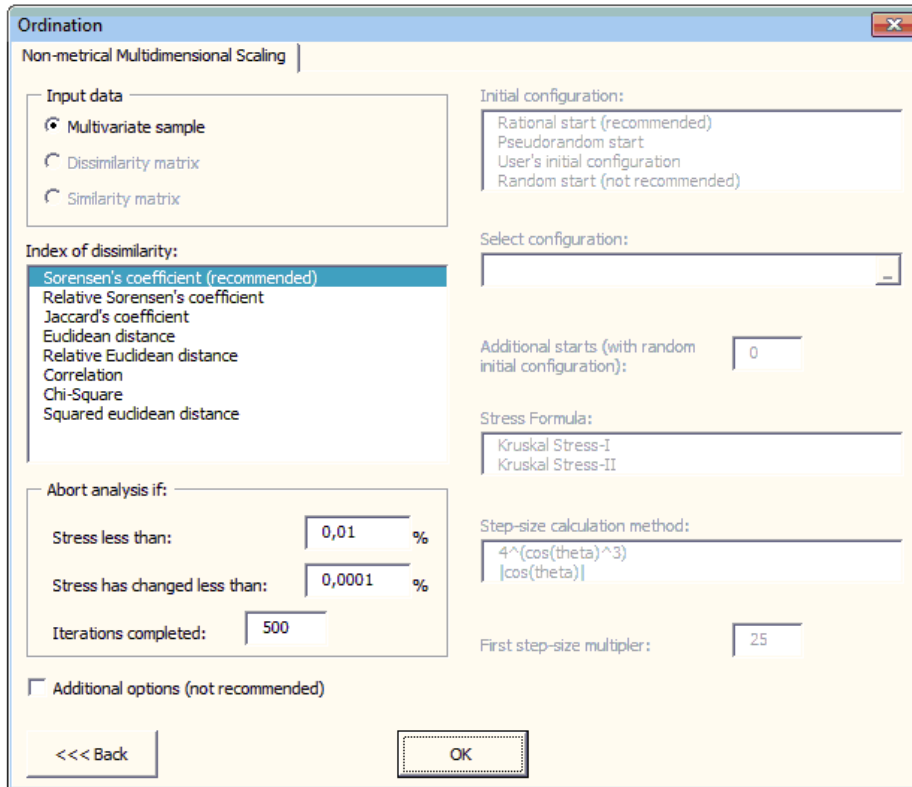
Result of cluster analysis (Ward clustering)

Most common similarity indexes: Jaccard, Sorensen, correlations and conjugation between species.

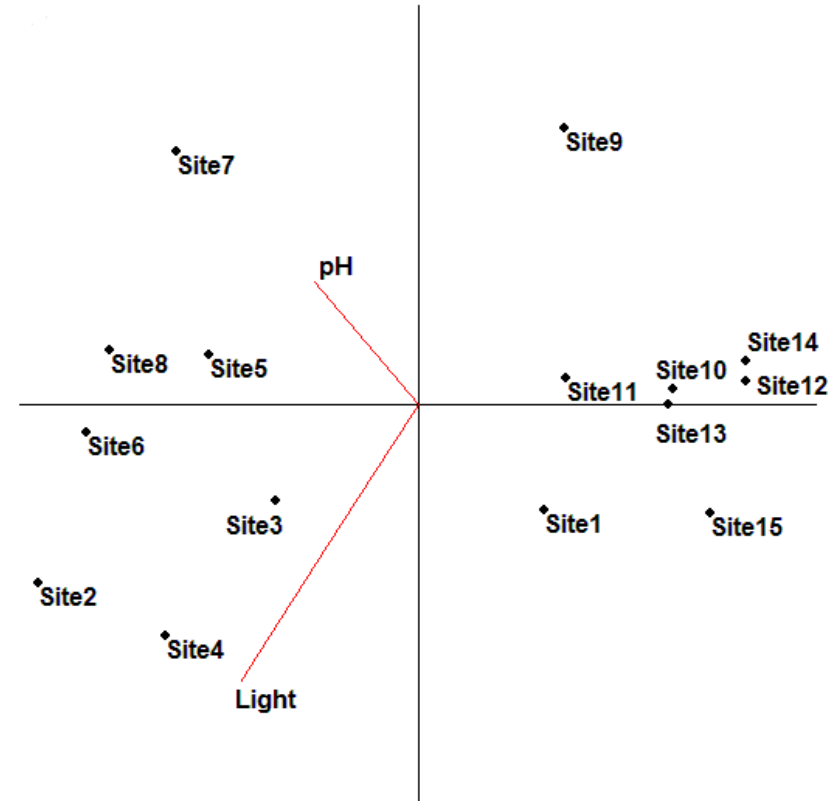
Methods of grouping: nearest neighbor, UPGMA, Ward's method



Ordination



Appearance of the settings window for NMS

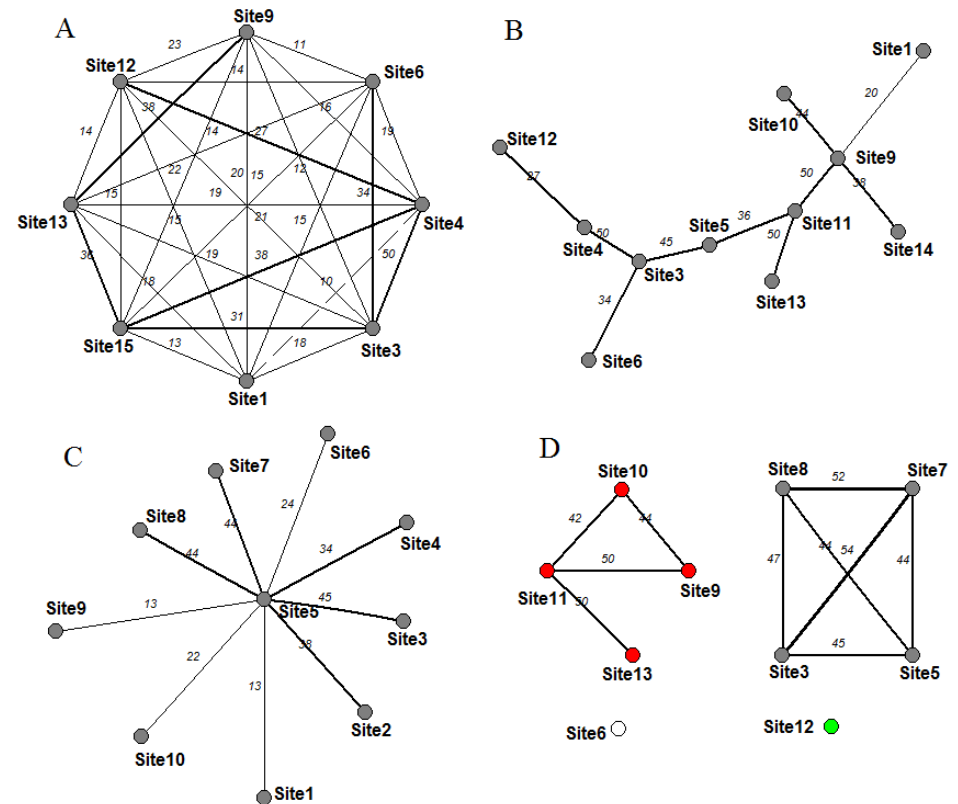
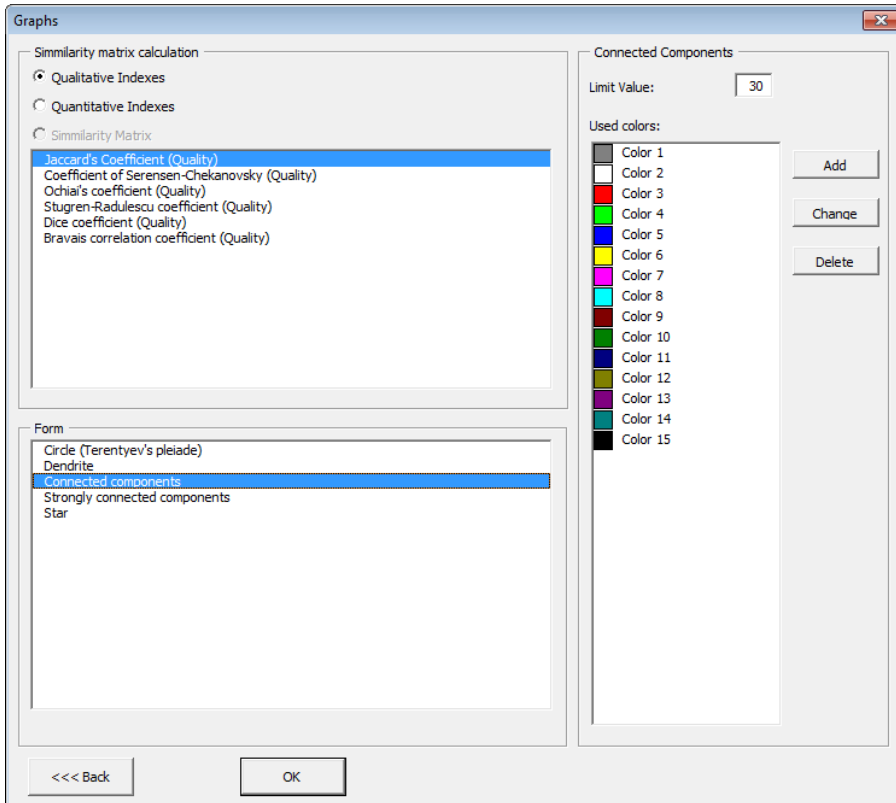


Result of ordination algorithm

PCA – Principal components analysis,
CA – Correspondence analysis
NMS – Non-metric Multidimensional scaling



Graph theory



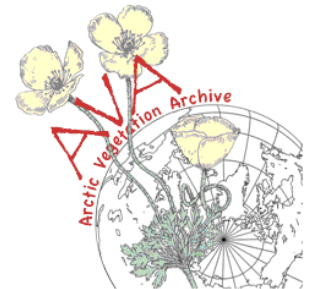
Appearance of the window for graph theory algorithms

<http://m-graphs.com/index.php/en>

Different forms of graphs presentation:
 A – circle graph; B – tree graphs (connected graph with no cycles); C – star graph; D – connected components

Summary

- 5000-6000 vegetation plots in paper format (hard copy)
- 1000-1500 Excel format
- 2000-3000 IBIS (mostly aquatic vegetation)
- About 3500 releves in Turboveg format
- Digital version of the “Flora of the USSR European Northeast”
- Herbarium, which contains more than 150,000 samples of vascular plants and about 60,000 samples of moss and lichens



ZooCode

Family-genus-species-subspecies – hierarchical system

Species name – Potentilla nivea L. (Family – Rosaceae)

AA (Rosaceae), BB (Potentilla), CC (nivea)

Unique key – AABBC

Potentilla gelida C.A.Mey ssp. borea-asiatica Jurtz. et R. Kam

AA (Rosaceae), BB (Potentilla), CD (Gelida), AF (borea-asiatica)

Unique key – AABBCDAF

Select “AA*” – all species from Rosaceae family

Select “AABB*” – all species from Potentilla genus

**Thank you very much for your
attention!**

