

COST733 Working Group2

-

Implementation and development of weather types classification methods

Andreas Philipp

Judit Bartholy

Michel Erpicum

Sylvie Jourdain

Thomas Krennert

Spyros Lykoudis

Krystyna Pianko-Kluczynska

Piia Post

Domingo Rassilla Álvarez

Outline

I.) A retrospect of WG2

II.) Relevance of classification algorithms

III.) Randomcent as a benchmark method

IV.) Randomcent for classification tuning

I.) Retrospect of WG2

Development of cost733cat

1.) WG1 phase (09/2005 to 09/2006)

2.) WG2 collection phase (2007)

3.) version 1.0 (01.02.2008)

version 1.1 (17.02.2008)

version 1.2 (12.07.2008)

4.) version 2.0 (21.11.2009)

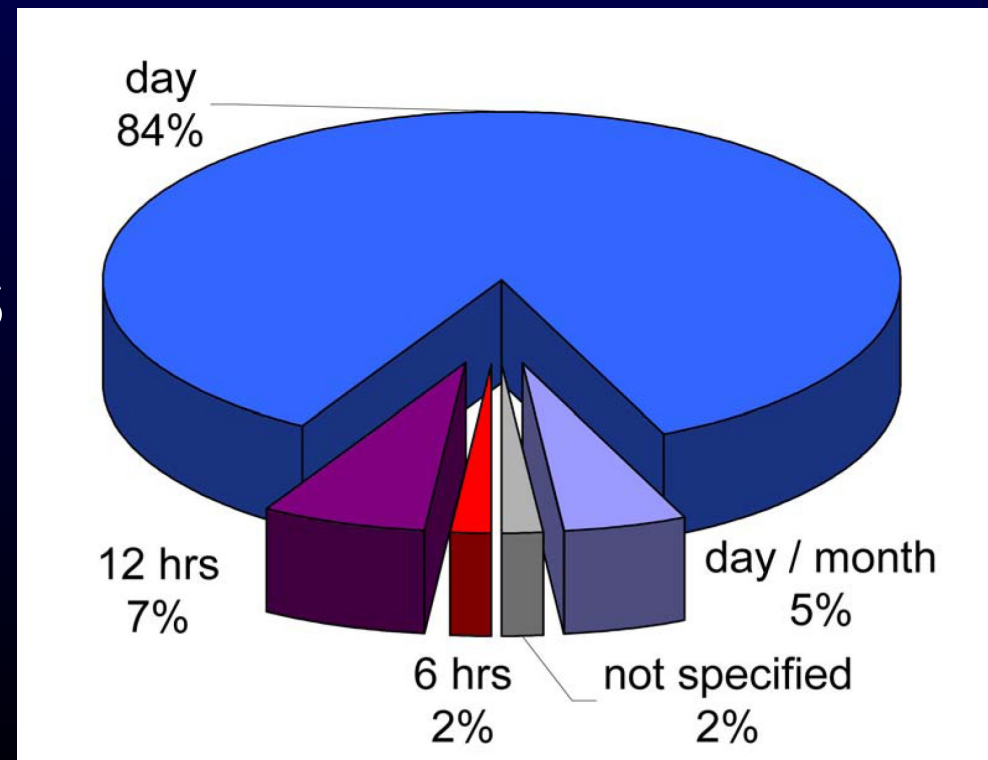
I.) Retrospect of WG2

Development of cost733cat

1.) WG1 phase (to 09/2006)

results of questionnaire:

- daily scale
- spatial scale
- circulation data
- selected methods



(Huth et al. 2007)

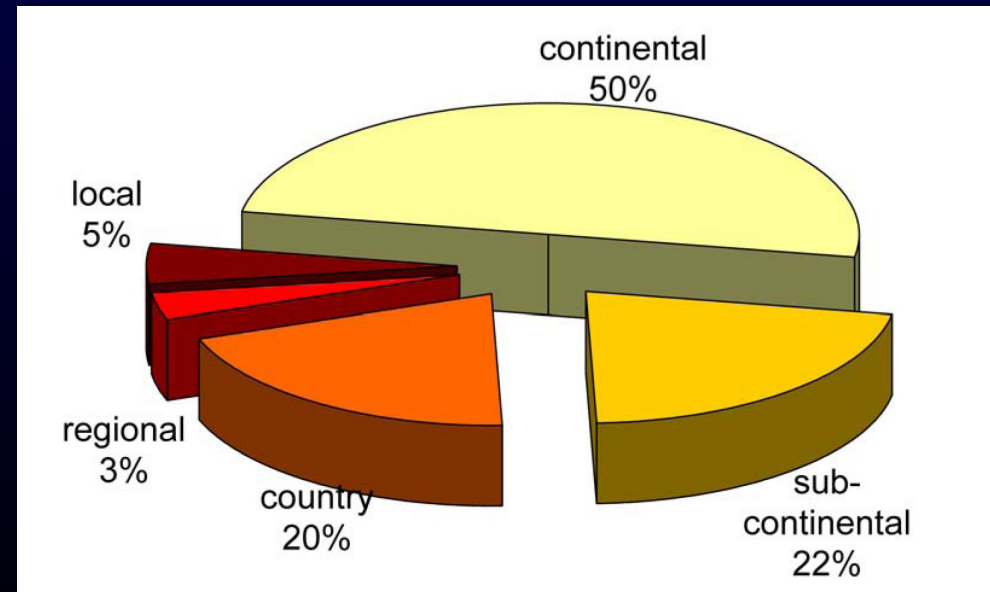
I.) Retrospect of WG2

Development of cost733cat

1.) WG1 phase (to 09/2006)

results of questionnaire:

- daily scale
- spatial scale
- circulation data
- selected methods



(Huth et al. 2007)

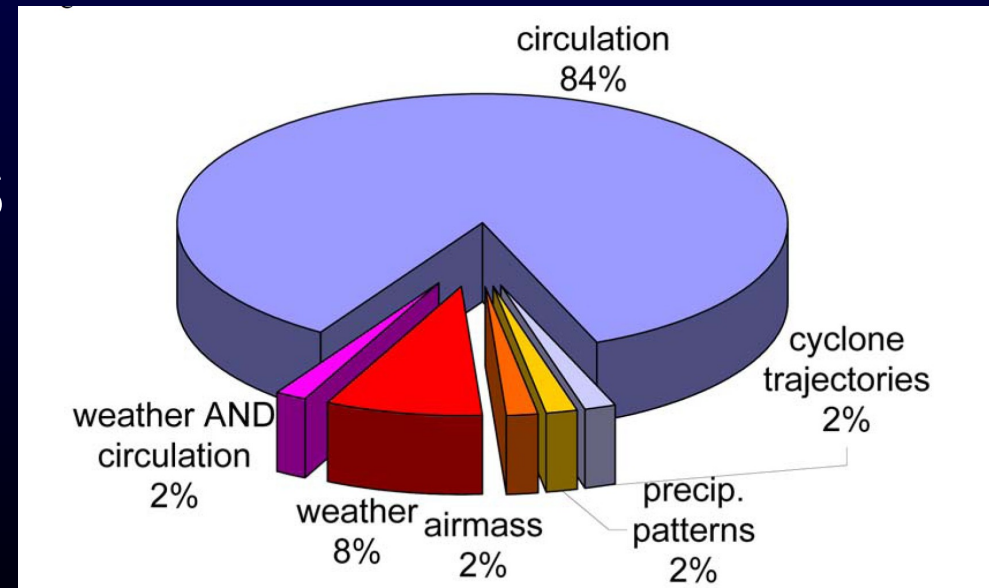
I.) Retrospect of WG2

Development of cost733cat

1.) WG1 phase (to 09/2006)

results of questionnaire:

- daily scale
- spatial scale
- circulation data
- selected methods



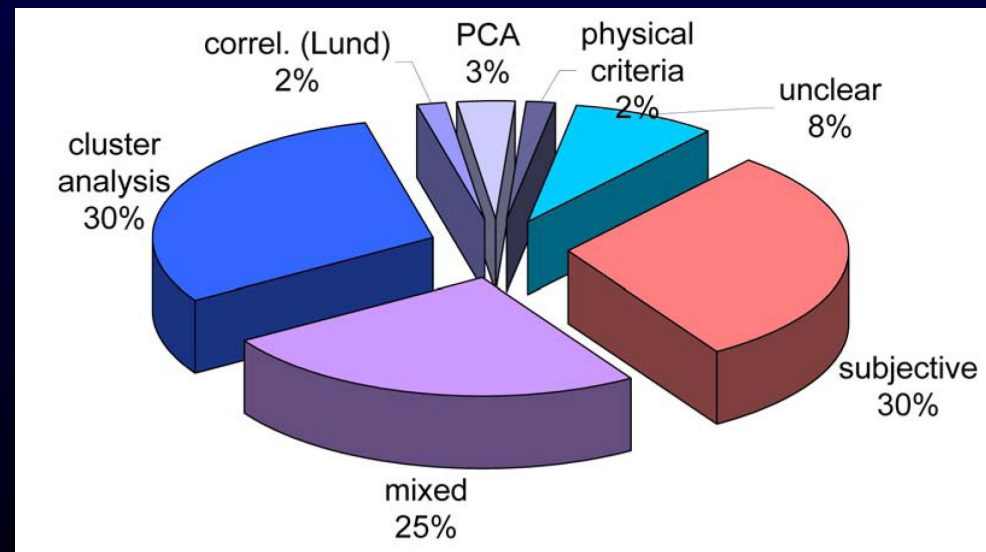
I.) Retrospect of WG2

Development of cost733cat

1.) WG1 phase (to 09/2006)

results of questionnaire:

- daily scale
- spatial scale
- circulation data
- selected methods



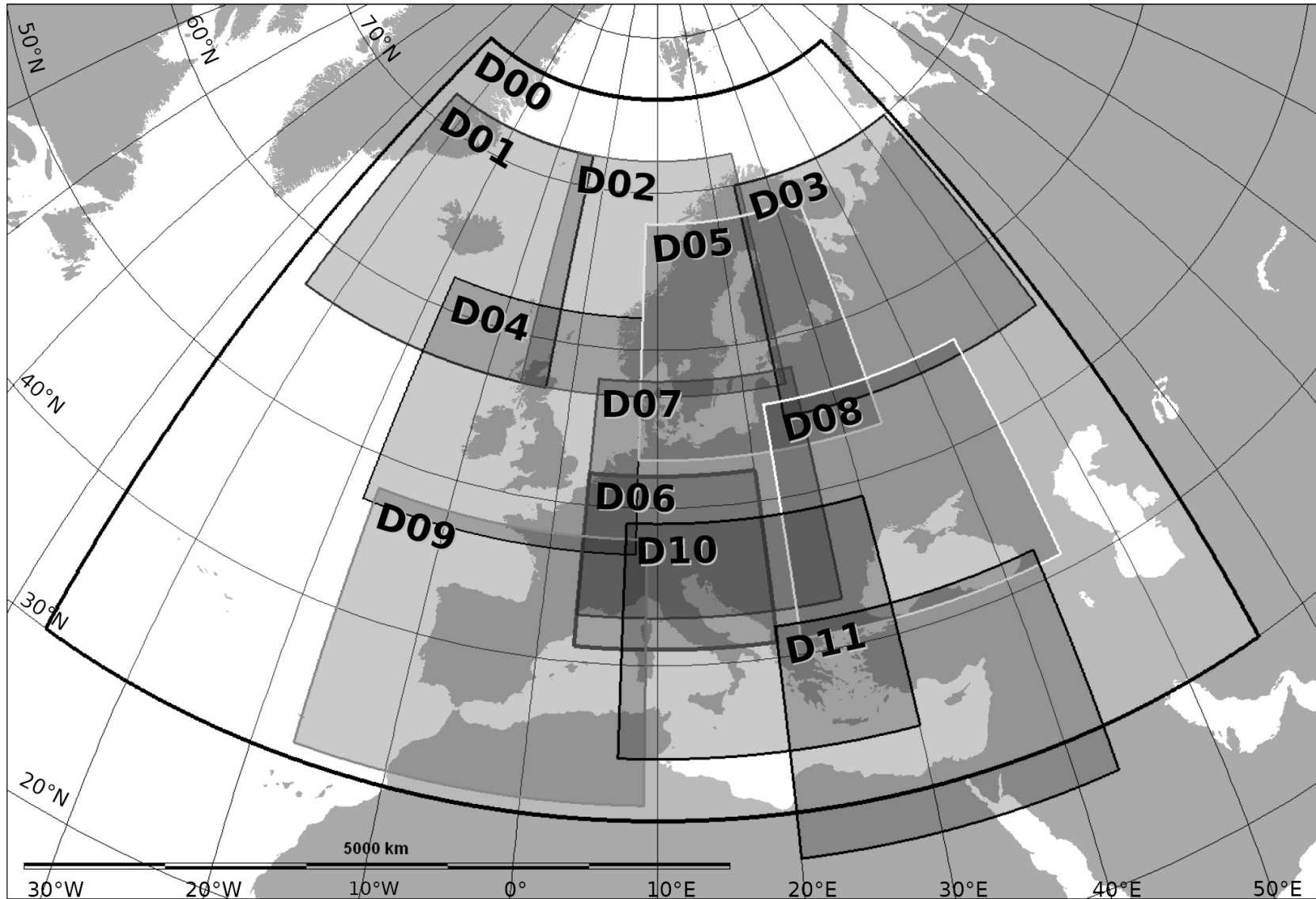
1.) Retrospect of WG2

Development of cost733cat

2.) WG2 collection phase (2006/2007)

- 25 methods selected
- daily ERA40 12:00 SLP data distributed
- definition of the 12 domains
- calculation by authors
- upload on directory cost733.org
- many authors - various formats
- need for versioning system

The COST733 spatial domains



1.) Retrospect of WG2

Development of cost733cat

2.) WG2 collection phase (2006/2007)

- 25 methods selected
- definition of the 12 domains
- daily ERA40 12:00 SLP data distributed
- calculation by authors
- upload on directory cost733.org
- many authors - various formats
- need for versioning system

I.) Retrospect of WG2

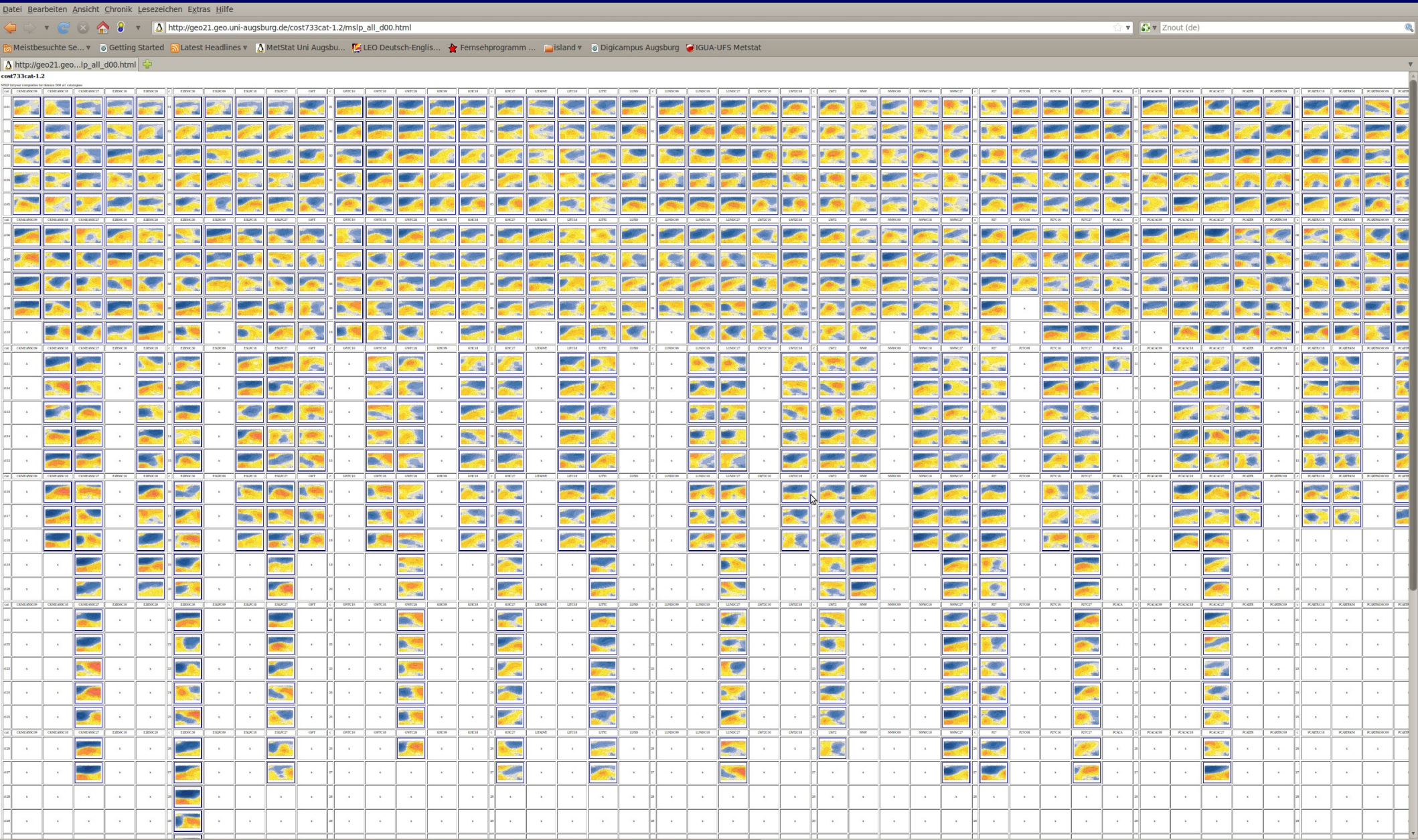
Development of cost733cat

3.) Version 1.0 to 1.2 (02-07/2008)

- numbers of types → 9, 18, 27
- SLP and full year (not seasonal)
- method revisions (e.g. E)
- 27 methods/catalogs (+ KH)
- single file version
- also available through wiki

#	abbrev & link	parameter	# classes	domains	author	contact	available	in 1.0	in 1.1	in 1.2
1	SEI	Z10007,Z900	40	000-D11	Erke et al.	Arne Spekat				
1	SEDC10	SLP	10	000-D07	Erke et al.	Arne Spekat				
1	SEDC20	SLP	20	000-D07	Erke et al.	Arne Spekat				
1	QMEANDC09	SLP	09	000-D11	Erke et al.	Arne Spekat				
1	QMEANDC10	SLP	10	000-D11	Erke et al.	Arne Spekat				
1	QMEANDC27	SLP	27	000-D11	Erke et al.	Arne Spekat				
2	BELPC10	SLP	10	000-D11	Ercpoum et al.	Michel Ercpoum				
2	BELPC30	SLP	30	000-D11	Ercpoum et al.	Michel Ercpoum				
2	EZ900C10	Z900	10	000-D11	Ercpoum et al.	Michel Ercpoum				
2	EZ900C30	Z900	30	000-D11	Ercpoum et al.	Michel Ercpoum				
2	EZ900C10	Z900	10	000-D11	Ercpoum et al.	Michel Ercpoum				
2	EZ900C20	Z900	20	000-D11	Ercpoum et al.	Michel Ercpoum				
2	EZ900C30	Z900	30	000-D11	Ercpoum et al.	Michel Ercpoum				
2	ESLPC09	SLP	09	000-D11	Ercpoum et al.	Michel Ercpoum				
2	ESLPC10	SLP	10	000-D11	Ercpoum et al.	Michel Ercpoum				
2	ESLPC27	SLP	27	000-D11	Ercpoum et al.	Michel Ercpoum				
3	GWT	SLP	15	000-D11	Beck	Christoph Beck				
3	GWTG10	SLP	10	000-D11	Beck	Christoph Beck				
3	GWTG10	SLP	10	000-D11	Beck	Christoph Beck				
3	GWTG20	SLP	20	000-D11	Beck	Christoph Beck				
3	GWTG30	SLP	30	000-D11	Beck	Christoph Beck				
4	KHC10	SLP	10	000-D11	Kirchhofer	Andreas Philipp				
4	KHC20	SLP	20	000-D11	Kirchhofer	Andreas Philipp				
4	KHC30	SLP	30	000-D11	Kirchhofer	Andreas Philipp				
4	KHC09	SLP	09	000-D11	Kirchhofer	Spyros Lykoudis				
4	KHC10	SLP	10	000-D11	Kirchhofer	Spyros Lykoudis				
4	KHC27	SLP	27	000-D11	Kirchhofer	Spyros Lykoudis				
5	LITAD06	SLP	9	000-D11	Litynski	Krystyna Plarico				
5	LIT10	SLP	10	000-D11	Litynski	Krystyna Plarico				
5	LIT10	SLP	27	000-D11	Litynski	Krystyna Plarico				
6	LUND	SLP	10	000-D11	Lund	Andreas Philipp				
6	LUNDG10	SLP	10	000-D11	Lund	Andreas Philipp				
6	LUNDG20	SLP	20	000-D11	Lund	Andreas Philipp				
6	LUNDG30	SLP	30	000-D11	Lund	Andreas Philipp				
6	LUNDG09	SLP	09	000-D11	Lund	Spyros Lykoudis				
6	LUNDG10	SLP	10	000-D11	Lund	Spyros Lykoudis				
6	LUNDG27	SLP	27	000-D11	Lund	Spyros Lykoudis				
7	LWT3010	SLP	10	000-D11	James	Paul James				
7	LWT3010	SLP	10	000-D11	James	Paul James				
7	LWT10	SLP	20	000-D11	James	Paul James				
8	NNW	Z900	9-30	000-D11	Michaelides et al.	Silas Michaelides				
8	NNWC09	SLP	09	000-D11	Michaelides et al.	Silas Michaelides				
8	NNWC10	SLP	10	000-D11	Michaelides et al.	Silas Michaelides				
8	NNWC27	SLP	27	000-D11	Michaelides et al.	Silas Michaelides				
9	PST	Z900	27	000-D11	Kruzinga	Pia Post				
9	PSTC09	SLP	09	000-D11	Kruzinga	Pia Post				
9	PSTC10	SLP	10	000-D11	Kruzinga	Pia Post				
9	PSTC27	SLP	27	000-D11	Kruzinga	Pia Post				
10	PCACA	SLP	4-9	000-D11	Rassilla	Domingo Rassilla				
10	PCACAC09	SLP	09	000-D11	Rassilla	Domingo Rassilla				
10	PCACAC10	SLP	10	000-D11	Rassilla	Domingo Rassilla				
10	PCACAC27	SLP	27	000-D11	Rassilla	Domingo Rassilla				
11	PEAXTR10	SLP	11-17	000-D11	Esteban	Pere Esteban				
11	PEAXTRC09	SLP	09-10	000-D11	Esteban	Pere Esteban				
11	PEAXTRC10	SLP	10-15	000-D11	Esteban	Pere Esteban				
11	PEAXTRK04	SLP	11-17	000-D11	Esteban	Pere Esteban				
11	PEAXTRKAC09	SLP	09-10	000-D11	Esteban	Pere Esteban				
11	PEAXTRKAC10	SLP	10-15	000-D11	Esteban	Pere Esteban				
12	PETISCO	SLP, Z900	25-30	000-D11	Pellico	Domingo Rassilla				
12	PETISCO09	SLP	09	000-D11	Pellico	Domingo Rassilla				
12	PETISCO10	SLP	10	000-D11	Pellico	Domingo Rassilla				
12	PETISCO27	SLP	27	000-D11	Pellico	Domingo Rassilla				
13	SANDRA	SLP	10-23	000-D11	Philipp	Andreas Philipp				
13	SANDRAC10	SLP	10	000-D11	Philipp	Andreas Philipp				
13	SANDRAC20	SLP	20	000-D11	Philipp	Andreas Philipp				
13	SANDRAC30	SLP	30	000-D11	Philipp	Andreas Philipp				
13	SANDRAC09	SLP	09	000-D11	Philipp	Andreas Philipp				
13	SANDRAC10	SLP	10	000-D11	Philipp	Andreas Philipp				
13	SANDRAC27	SLP	27	000-D11	Philipp	Andreas Philipp				
14	SANDRAS	Z920,Z900	30	000-D11	Philipp	Andreas Philipp				
14	SANDRASC09	SLP	09	000-D11	Philipp	Andreas Philipp				
14	SANDRASC10	SLP	10	000-D11	Philipp	Andreas Philipp				
14	SANDRASC27	SLP	27	000-D11	Philipp	Andreas Philipp				
15	TRCAV	SLP	6-12	000-D11	Huth	Radan Huth				
15	TRCAC07	SLP	7	000-D11	Huth	Radan Huth				
15	TRCAC09	SLP	09	000-D11	Huth	Radan Huth				
15	TRCAC10	SLP	10	000-D11	Huth	Radan Huth				
15	TRCAC27	SLP	27	000-D11	Huth	Radan Huth				
16	WLKC73	U700,V700,Z920,Z900,PW	40	000-D11	Oltmann et al.	Thomas Krennert				
16	WLKC09	U700,V700	09	000-D11	Oltmann et al.	Thomas Krennert				
16	WLKC10	U700,V700,Z920	10	000-D11	Oltmann et al.	Thomas Krennert				
16	WLKC20	U700,V700,Z920,Z900	20	000-D11	Oltmann et al.	Thomas Krennert				
not scalable: classifications restricted to one domain (for comparison)										
#	abbrev & link	parameter	# classes	domains	author	contact	available	in 1.0	in 1.1	in 1.2
17	HBOWL		29	Central Europe	Hess/Brezowski	Peter Werner				
17	HBOWL		10	Central Europe	Hess/Brezowski	Peter Werner				
18	ODWL		29	Europe	James	Paul James				
18	ODWLSLP	SLP	29	Europe	James	Paul James				
19	PECZELV		13	Hungary	Peczely	Judith Bartholy				
20	PERISBL		31	Switzerland	Ferret	Oliver Duding				
21	SCHUEPP		40	Switzerland	Schuepp	Reinhard Schiemann				
22	ZAMG		43	Austria	ZAMG, Krennert	Thomas Krennert				

I.) Retrospect of WG2 V1 composite plots



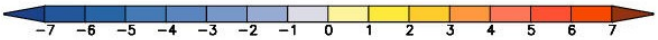
LUND Domain 11

Description:

Plot titles show 1.) method name, 2.) domain, 3.) class number, 4.) number of days, 5.) days in percent of season, 6.) months of season, 7.) dataset, 8.) variables.
 Spatial resolution used for all maps is 1x1".
 Red box shows spatial domain used for classification.

Mean Sea Level Pressure [hPa]: ERA40 12Z mean isobars shown as contour lines (fixed interval 2hPa).
 Isobars below/above the spatial mean shown as solid/dashed lines.

Temperature scale [K]: ERA40 mean 12Z anomaly referring to long term monthly mean 1957-2002.
 Insignificant values (alpha=0.05) blanked out.



Precipitation scale [mm]: ERA40 (CP+LSP) mean sum from 06Z to 12Z (to be changed to daily sum in the future).
 (note: The ENSEMBLES WP5.1 gridded RR dataset should replace/complement the ERA40 data in the future.)
 Values less than 0.5 and insignificant values (alpha=0.05) blanked out.



Map table:

Type #	Winter (Dec-Jan-Feb)	Spring (Mar-Apr-May)	Summer (Jun-Jul-Aug)	Autumn (Sep-Oct-Nov)
01	LUND D11 C01 634d (15.6%) DJF ERA40 2mT/MSLP 	LUND D11 C01 1596d (38.6%) MAM ERA40 2mT/MSLP 	LUND D11 C01 3719d (89.8%) JJA ERA40 2mT/MSLP 	LUND D11 C01 2122d (51.8%) SON ERA40 2mT/MSLP
	LUND D11 C01 634d (15.6%) DJF ERA40 PREC/MSLP 	LUND D11 C01 1596d (38.6%) MAM ERA40 PREC/MSLP 	LUND D11 C01 3719d (89.8%) JJA ERA40 PREC/MSLP 	LUND D11 C01 2122d (51.8%) SON ERA40 PREC/MSLP

I.) Retrospect of WG2

Development of cost733cat

4.) Version 2.0 (11/2009)

- more meteorological parameters
- seasonal variants / sequences of days
- new nomenclature
- recalculations with cost733class
- 23 methods/catalogs (original and reproduced)
- 5 method groups established
- more than 5000 single catalogs

I.) Retrospect of WG2

Development of cost733cat

YR_S04 (whole year, 4-day-sequences)

dat	~k	GWLo	OGWo	PECo	PERo	SUEo	ZMGo	GWT	JCT	LWTo	LIT	LITo	WLK	WLKo	KRZ	KRZo	PXE	PXEo	PCT	PCTo	PTT	LND	KIR	ERP	ERPo	PTSo	CKM	CKMo	NNWo	SOM	CAP	CAPo	PXK	PXKo	SAN	RAC
SP	09														X		X		X		X	X		X			X				X		X		X	X
SP	18														X		X		X		X	X		X			X				X		X		X	X
SP	27														X		X		X		X	X		X			X				X		X		X	X
SP-Z5	09														X		X		X		X	X		X			X				X		X		X	X
SP-Z5	18														X		X		X		X	X		X			X				X		X		X	X
SP-Z5	27														X		X		X		X	X		X			X				X		X		X	X
SP-K5	09														X		X		X		X	X		X			X				X		X		X	X
SP-K5	18														X		X		X		X	X		X			X				X		X		X	X
SP-K5	27														X		X		X		X	X		X			X				X		X		X	X
SP-Y5	09														X		X		X		X	X		X			X				X		X		X	X
SP-Y5	18														X		X		X		X	X		X			X				X		X		X	X
SP-Y5	27														X		X		X		X	X		X			X				X		X		X	X
SP-Z5-Y5-K5	09														X		X		X		X	X		X			X				X		X		X	X
SP-Z5-Y5-K5	18														X		X		X		X	X		X			X				X		X		X	X
SP-Z5-Y5-K5	27														X		X		X		X	X		X			X				X		X		X	X

SE_S01 (seasonal, single days)

dat	~k	GWLo	OGWo	PECo	PERo	SUEo	ZMGo	GWT	JCT	LWTo	LIT	LITo	WLK	WLKo	KRZ	KRZo	PXE	PXEo	PCT	PCTo	PTT	LND	KIR	ERP	ERPo	PTSo	CKM	CKMo	NNWo	SOM	CAP	CAPo	PXK	PXKo	SAN	RAC	
SP	27																							X			X				X				X	X	
Z1-Z5	27																											X									
SP-Z5	27																							X			X				X					X	X
SP-K5	27																							X			X				X					X	X
SP-Y5	27																							X			X				X					X	X
SP-Z5-Y5-K5	27																							X			X				X					X	X

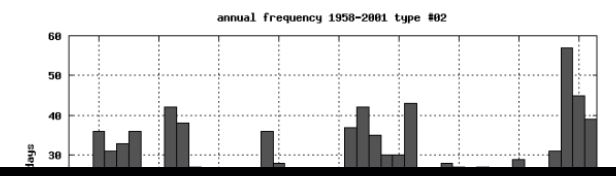
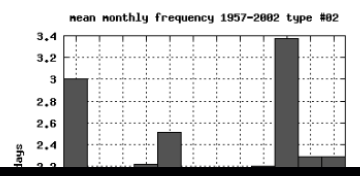
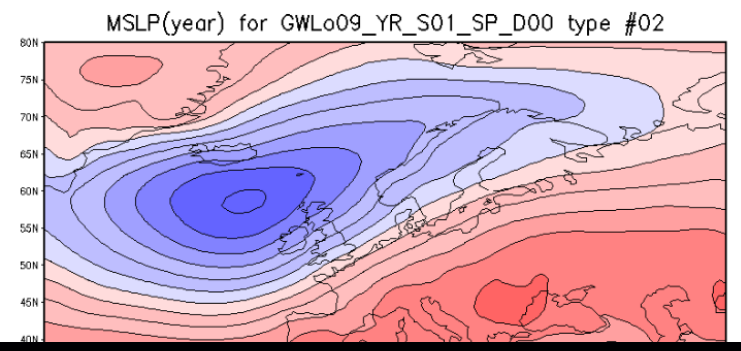
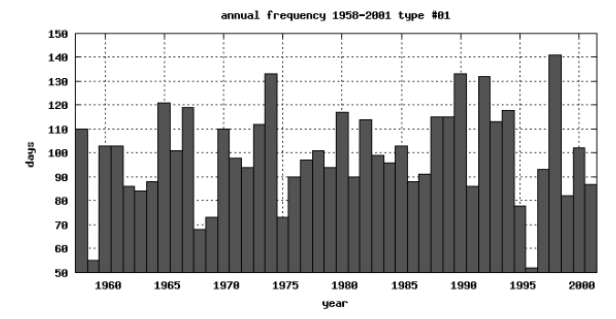
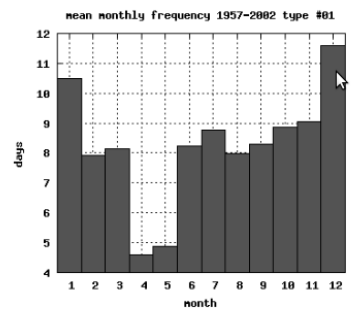
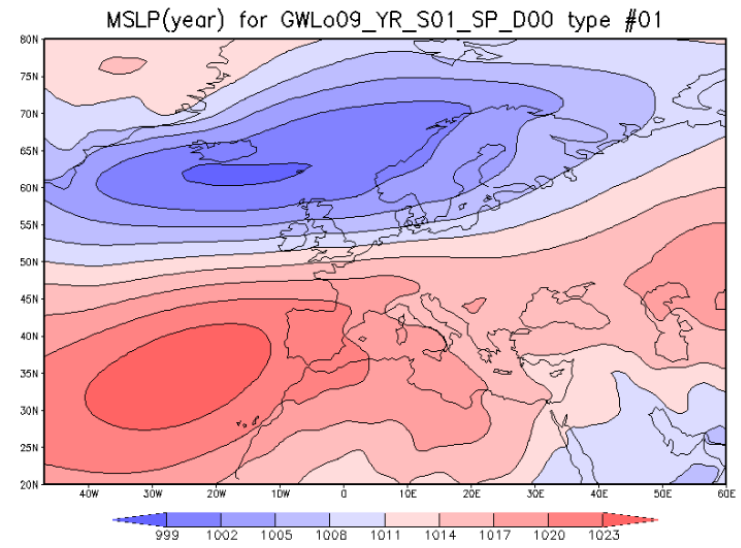
cost733cat-2.0 centroid plots

See [catalog table](#) for existing configurations and explanation of abbreviations, visit [cost733wiki](#) or [cost733 main page](#) for information about the COST Action 733.

classification method	number of types	preprocessing mode	classification input data	classification domain	centroid parameter	centroid level (for Z,R,T,VORT only)	centroid season	
<ul style="list-style-type: none"> GWLo OGWo PECo PERo SUEo ZMGo GWT JCT LWTo LIT LITo WLK WLKo KRZ 	<ul style="list-style-type: none"> 00 09 18 27 	<ul style="list-style-type: none"> YR_S01 YR_S04 SE_S01 	<ul style="list-style-type: none"> SP Z8 Z5 Z1-Z5 SP-Z5 SP-K5 SP-Y5 SP-Z5-Y5-K5 U7-V7 U7-V7-Z9 U7-V7-Z9-Z5 U7-V7-Z9-Z5-TW SP-RR SP-TT 	<ul style="list-style-type: none"> 00 01 02 03 04 05 06 07 08 09 10 11 	<ul style="list-style-type: none"> Z MSLP PRC CP LSP R TWC 2mT T Z500-Z850 10mU 10mV U700 V700 	<ul style="list-style-type: none"> 925 850 700 500 300 	<ul style="list-style-type: none"> year win spr sum aut 	<input type="checkbox"/> anomalies <input type="checkbox"/> normalized <input type="checkbox"/> map size: lat: +20: +80 lon: -47: +60 800x600 <input type="checkbox"/> draw box <input type="checkbox"/> draw labels <input type="checkbox"/> centroid data <input checked="" type="checkbox"/> seasonal cycle <input checked="" type="checkbox"/> interann. freq.

searching GWLo09 YR_S01 SP/GWLo09 YR_S01 SP_D00.txt...
[GWLo09 YR_S01 SP_D00: GWLo09 YR_S01 SP/GWLo11 YR_S01 SP_D00.txt](#) (actually 11 types)
 See statistics at [GWLo09 YR_S01 SP/GWLo11 YR_S01 SP_D00.stat](#)

- Download [GWLo09_YR_S01_SP.tbz](#) (bzip2-compressed tar-archive including data, grads script and eps-plots).



cost733cat-2.0 centroid plots

See [catalog table](#) for existing configurations and explanation of abbreviations, visit [cost733wiki](#) or [cost733 main page](#) for information about the COST Action 733.

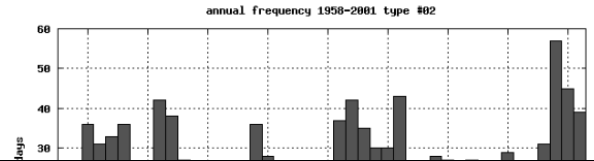
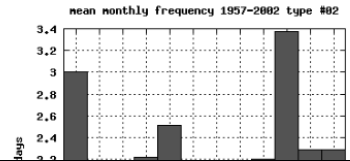
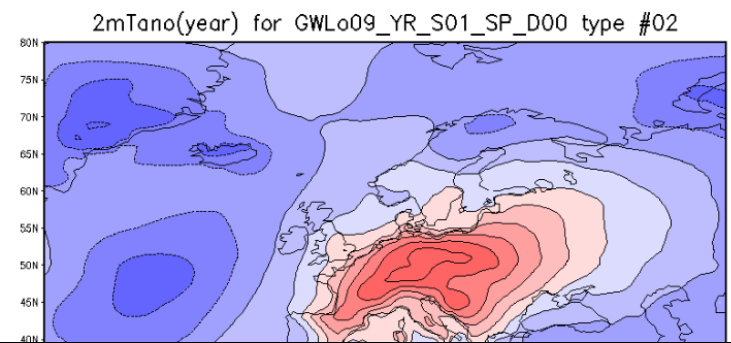
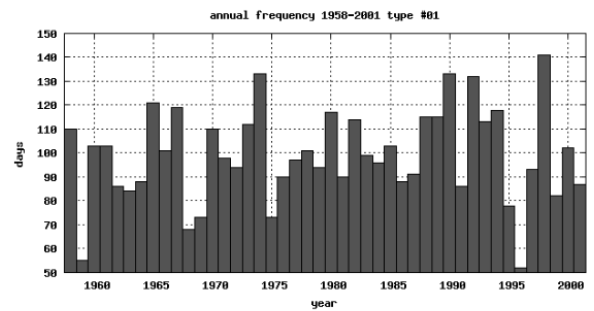
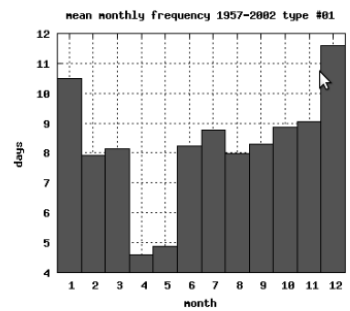
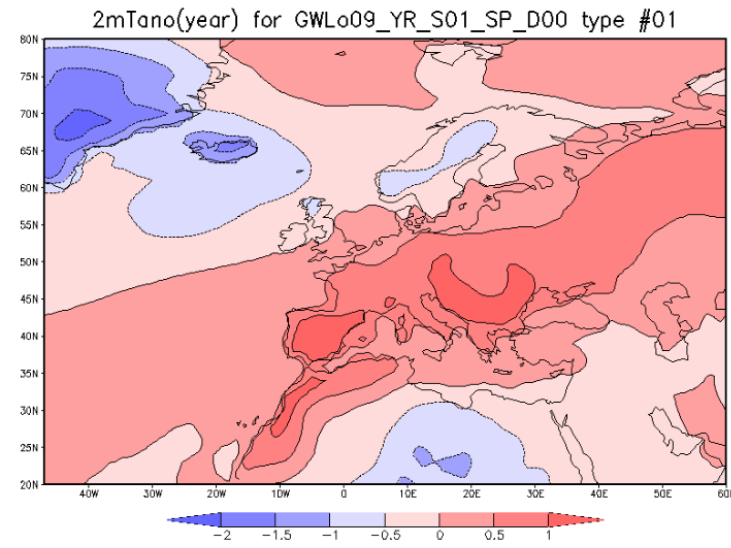
classification method	number of types	preprocessing mode	classification input data	classification domain	centroid parameter	centroid level (for Z,R,T,VORT only)	centroid season	run
<ul style="list-style-type: none"> GWLo OGWo PECo PERo SUEo ZMGo GWT JCT LWTo LIT LITo WLK WLKo KRZ 	<ul style="list-style-type: none"> 00 09 18 27 	<ul style="list-style-type: none"> YR_S01 YR_S04 SE_S01 	<ul style="list-style-type: none"> SP Z8 Z5 Z1-Z5 SP-Z5 SP-K5 SP-Y5 SP-Z5-Y5-K5 U7-V7 U7-V7-Z9 U7-V7-Z9-Z5 U7-V7-Z9-Z5-TW SP-RR SP-TT 	<ul style="list-style-type: none"> 00 01 02 03 04 05 06 07 08 09 10 11 	<ul style="list-style-type: none"> Z MSLP PRC CP LSP R TWC 2mT T Z500-Z850 10mU 10mV U700 V700 	<ul style="list-style-type: none"> 925 850 700 500 300 	<ul style="list-style-type: none"> year win spr sum aut 	<input checked="" type="checkbox"/> anomalies <input type="checkbox"/> normalized <input type="checkbox"/> map size: lat: +20 +80 lon: -47 +60 <input type="checkbox"/> draw box <input type="checkbox"/> draw labels <input type="checkbox"/> centroid data <input type="checkbox"/> seasonal cycle <input checked="" type="checkbox"/> interann. freq.

searching GWLo09 YR_S01 SP/GWLo09 YR_S01 SP_D00.txt...
 GWLo09 YR_S01 SP_D00: GWLo09 YR_S01 SP/GWLo11 YR_S01 SP_D00.txt (actually 11 types)
 See statistics at GWLo09 YR_S01 SP/GWLo11 YR_S01 SP_D00.stat

- Download [GWLo09_YR_S01_SP.tbz](#) (bzip2-compressed tar-archive including data, grads script and eps plots).

Mean Seasonal Cycle of type frequencies

Long term time series of type frequencies



II.) Relevance of classification algorithms for properties of the resulting frequencies

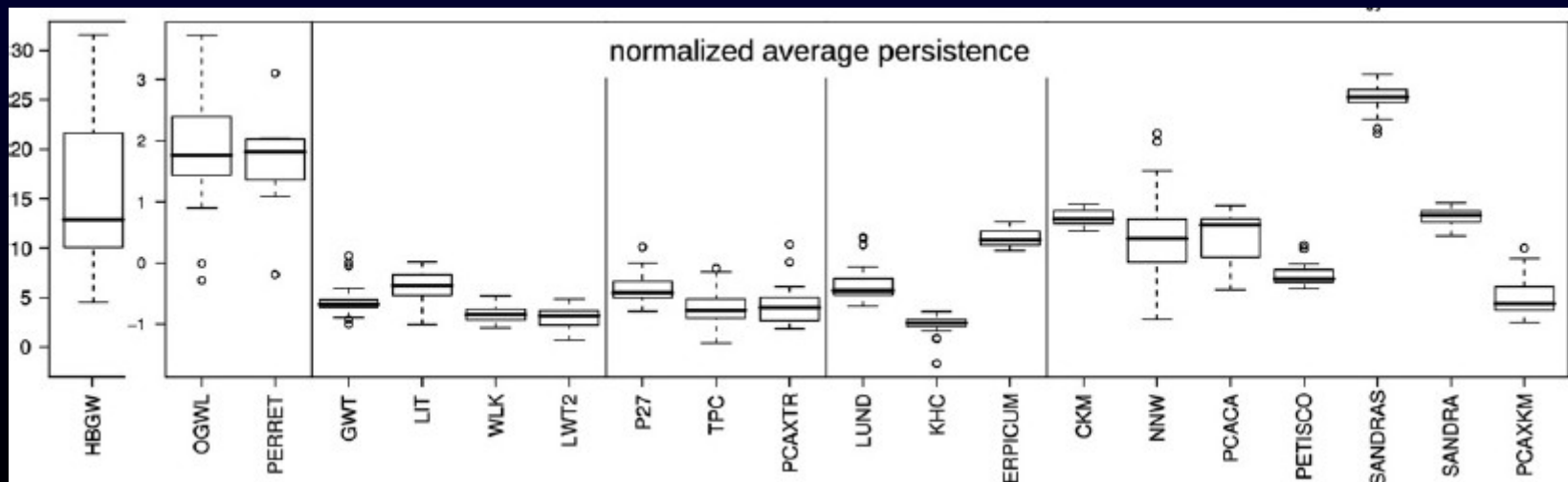
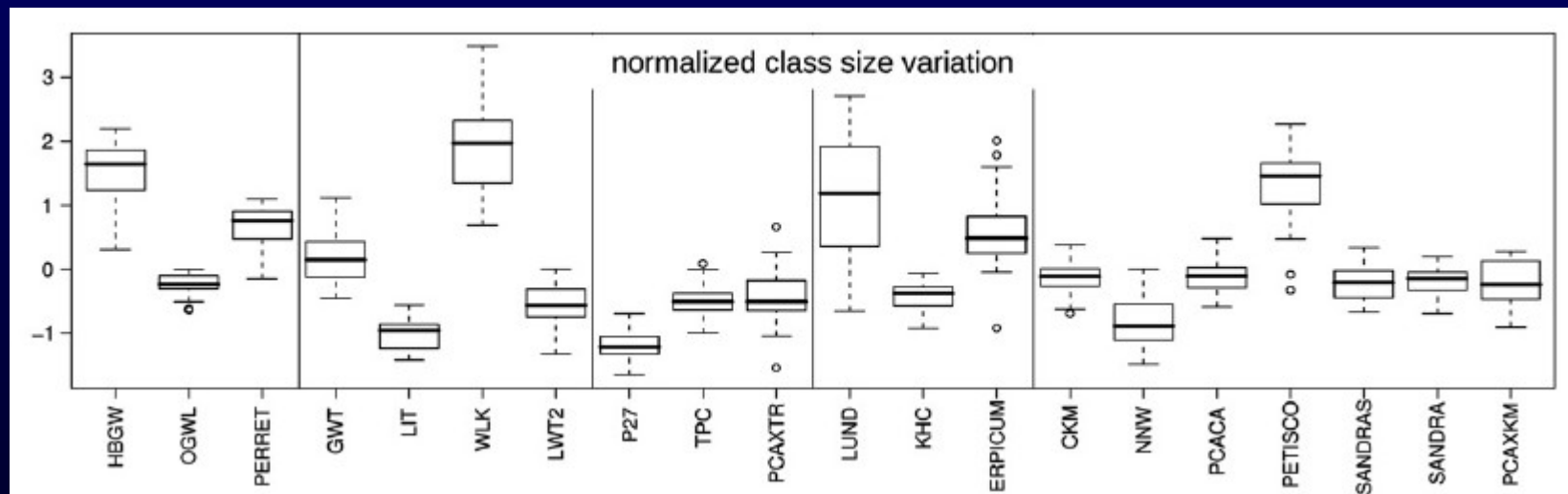
SUB

THR

PCA

LDR

OPT



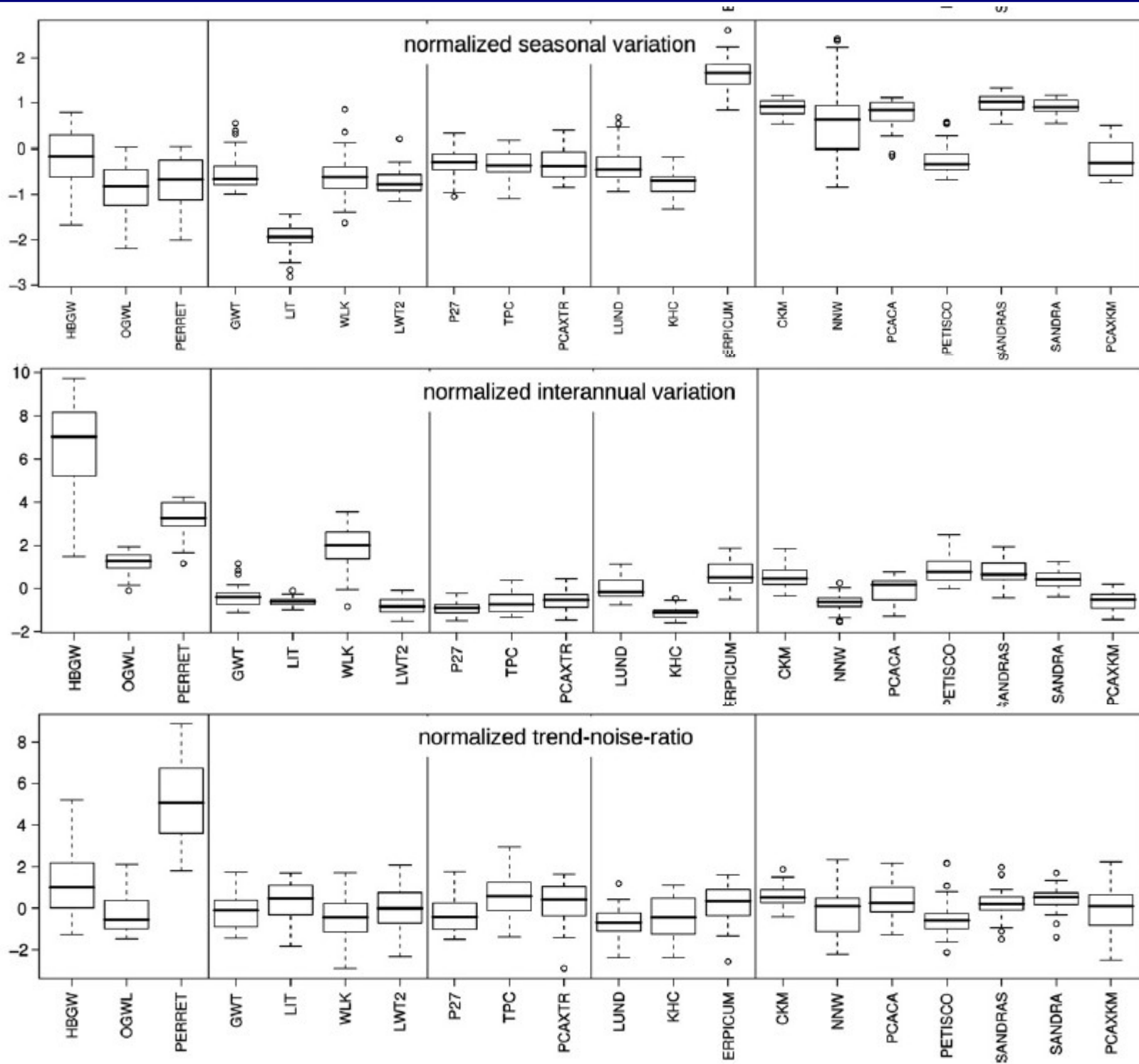
SUB

THR

PCA

LDR

OPT



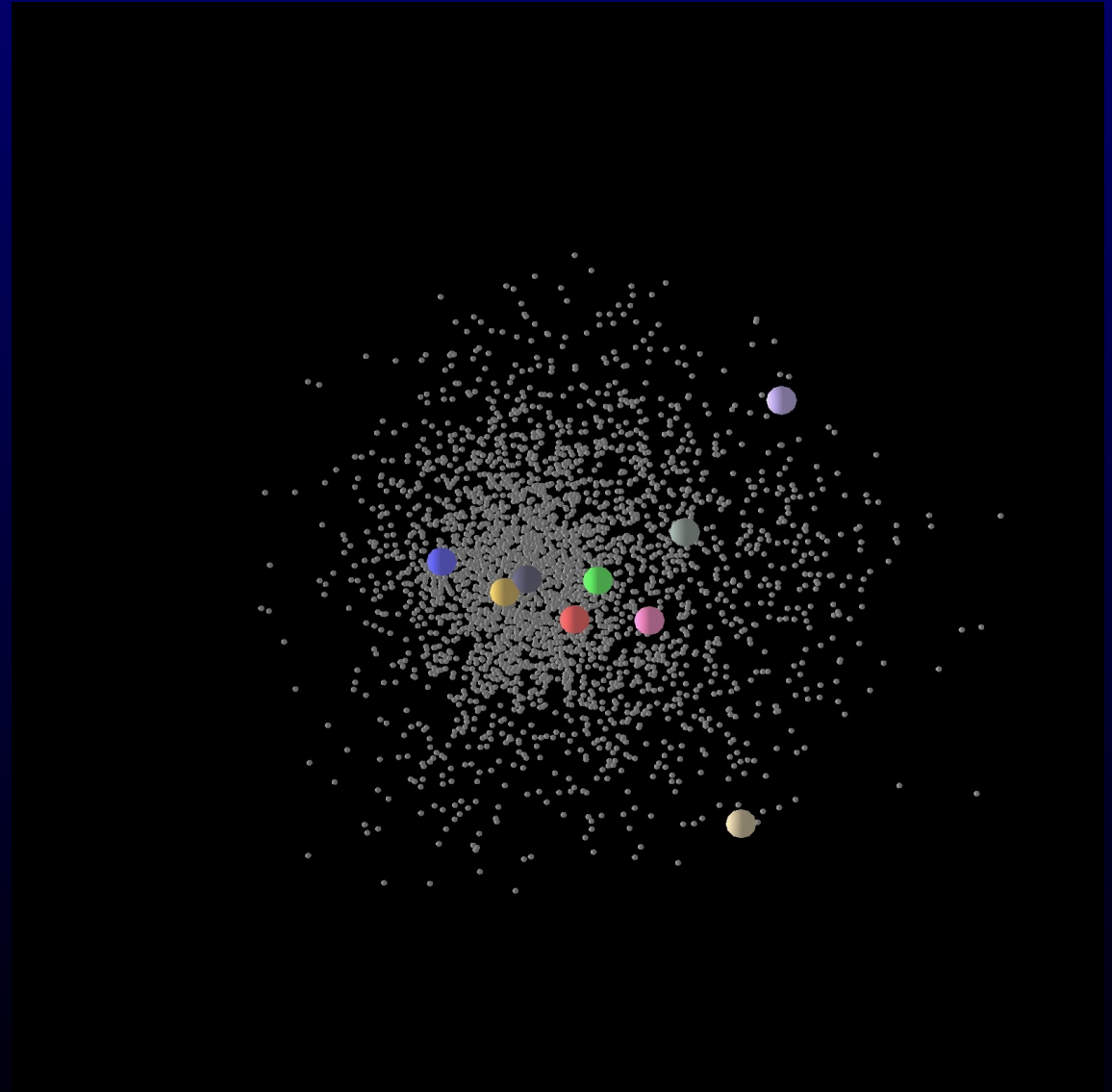
II.) Relevance of the classification Algorithms for applications

Feedback from WG3 and WG4 concerning:
temperature, precipitation and other applications:

- no general significant differences!
- does the method matter?
 - try (pseudo)randomly generated classifications

III.) Randomcent

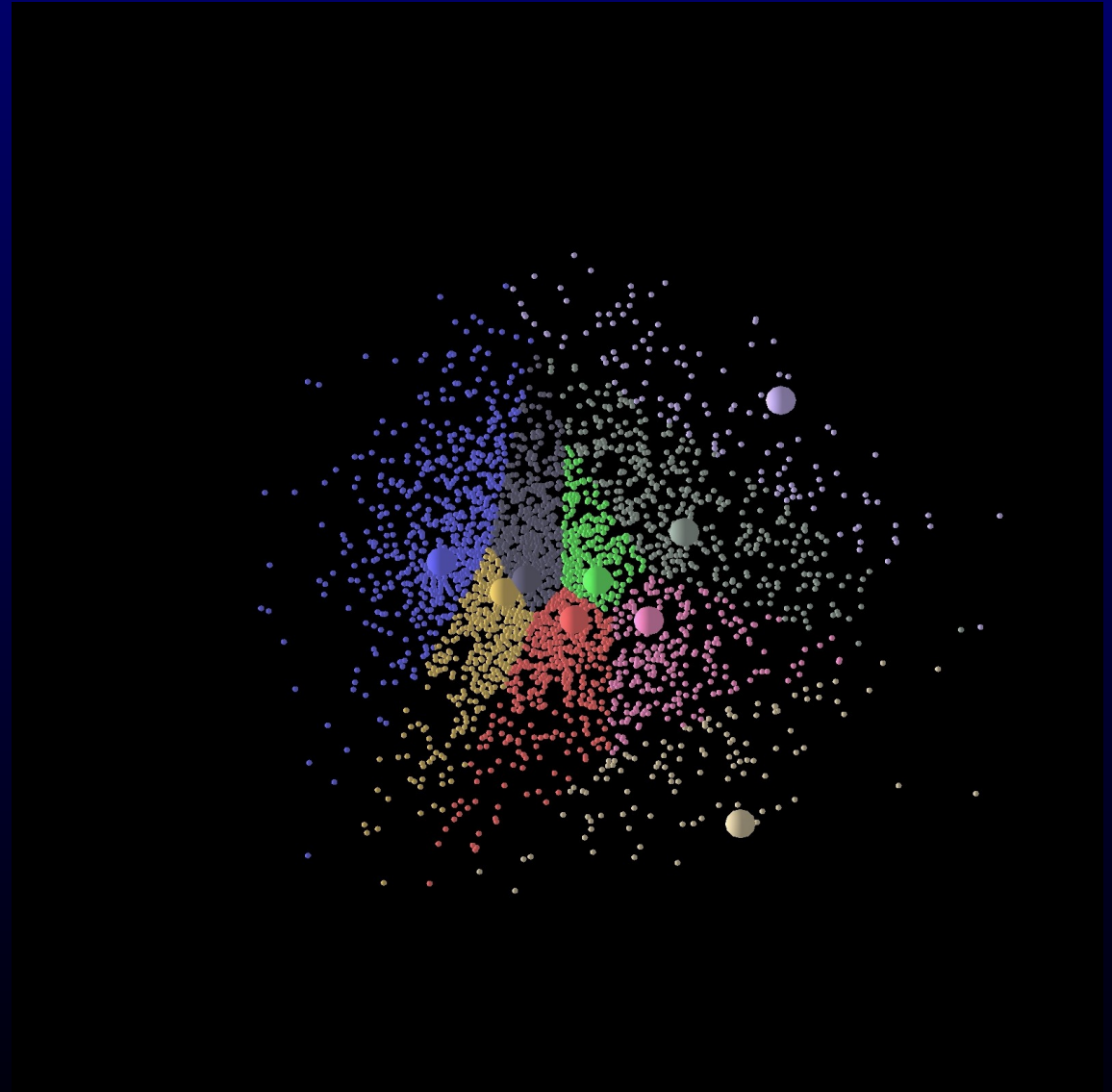
1.) select some daily patterns by random as key day



III.) Randomcent

1.) select some daily patterns by random as key day

2.) assign each day to the most similar key day

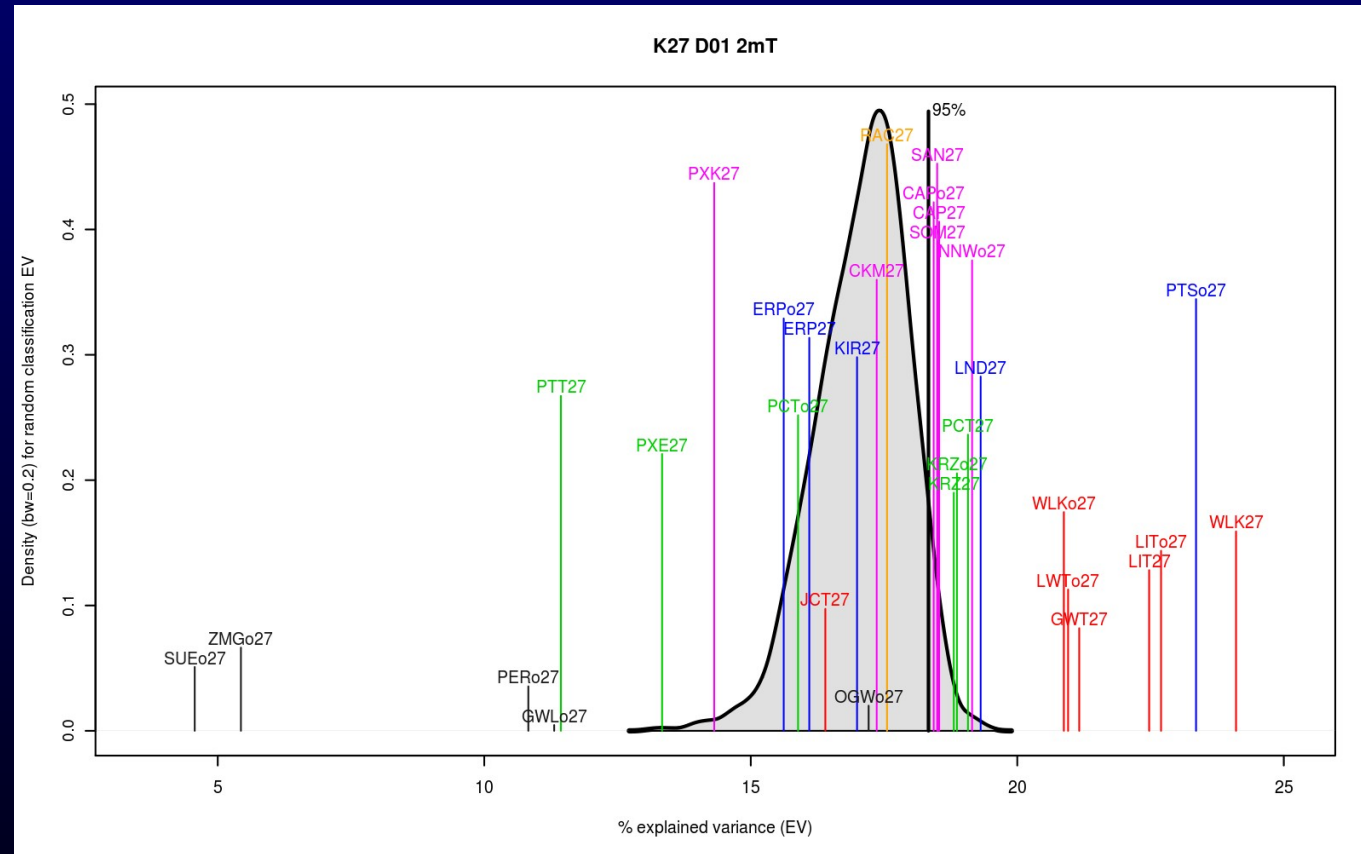


III.) Randomcent

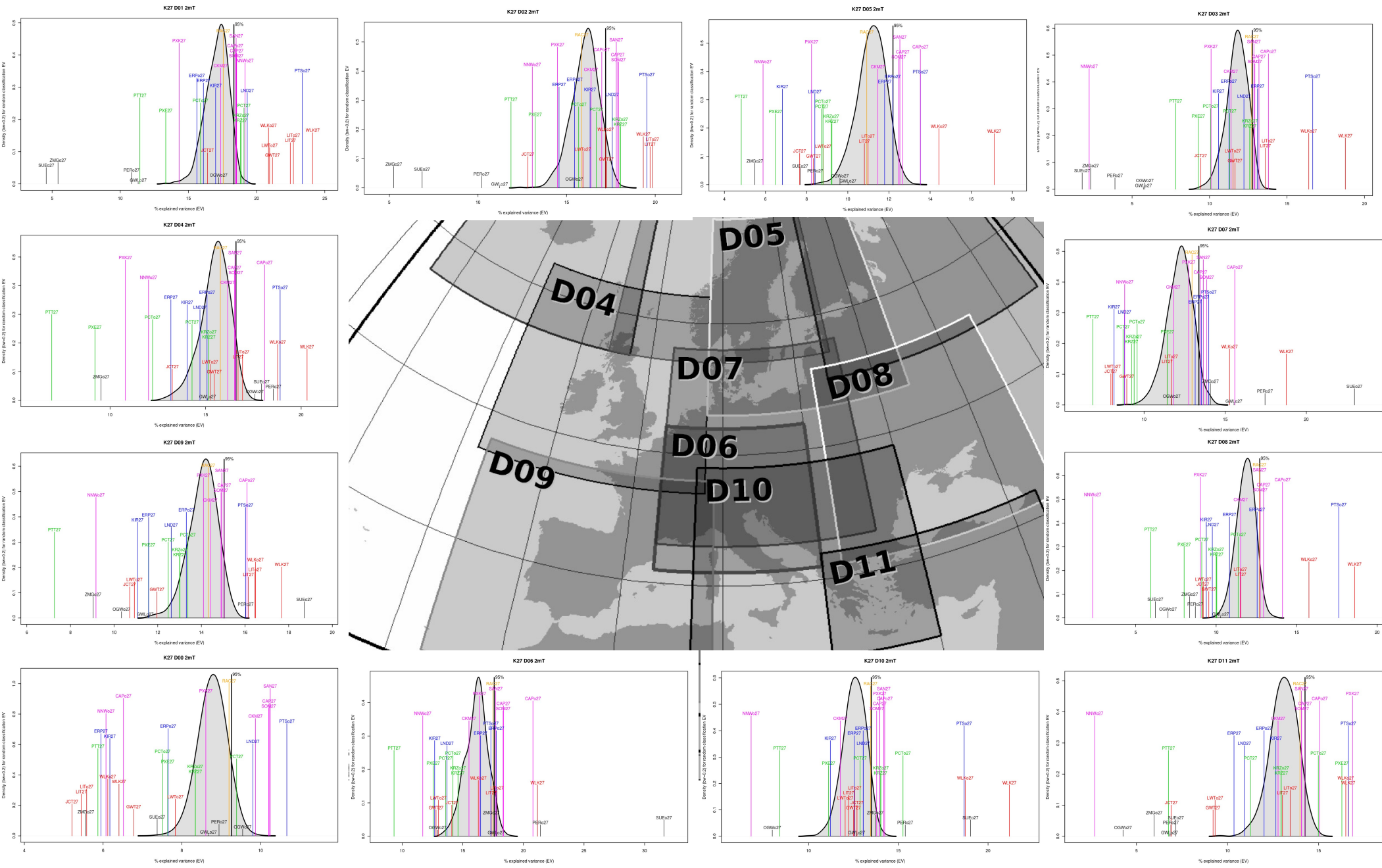
1.) select some daily patterns by random as key day as key day

2.) assign each day to the most similar key day

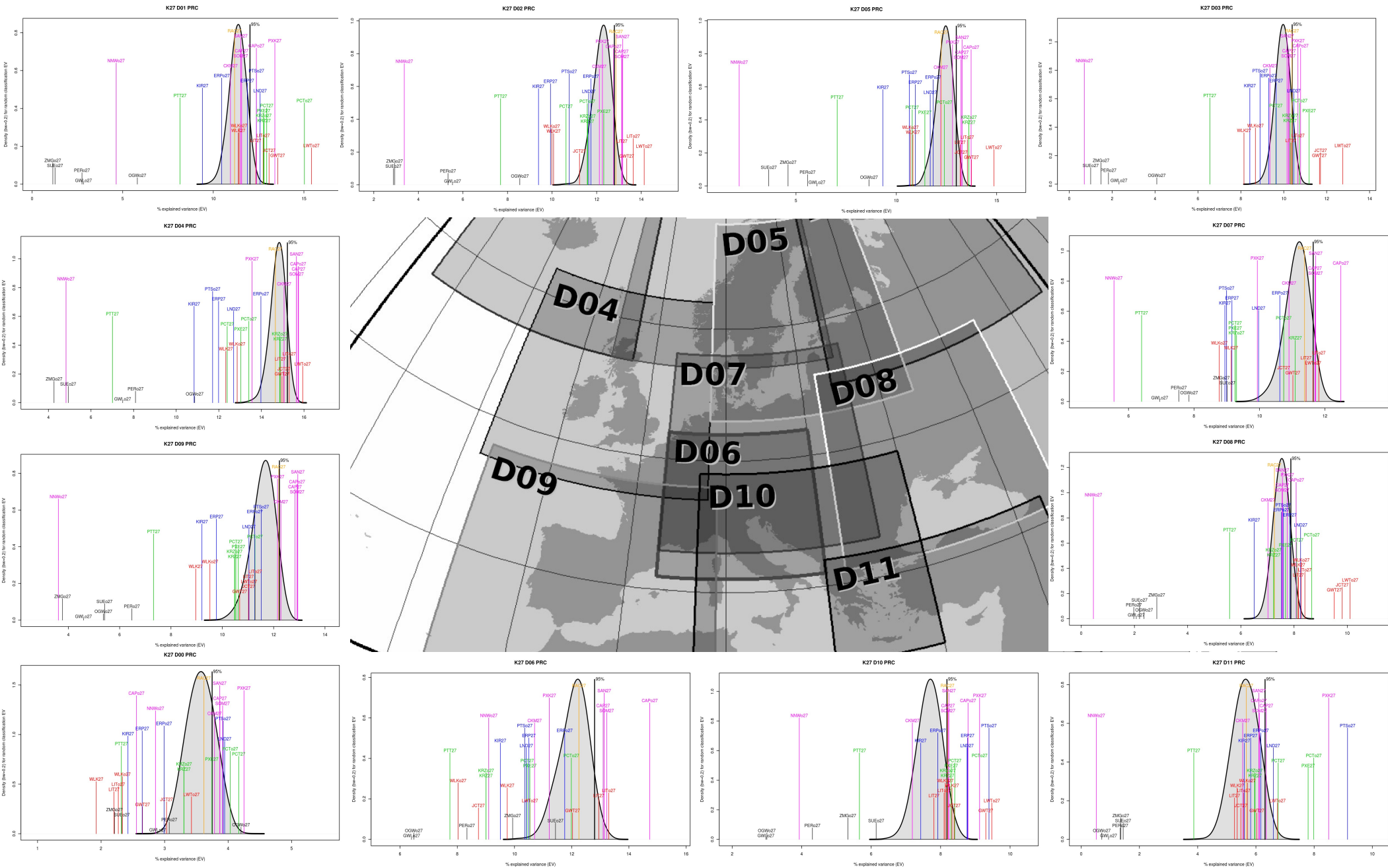
3.) do that 1000 times and compare those to the deliberate methods



III.) Randomcent: Explained variance of 2mT

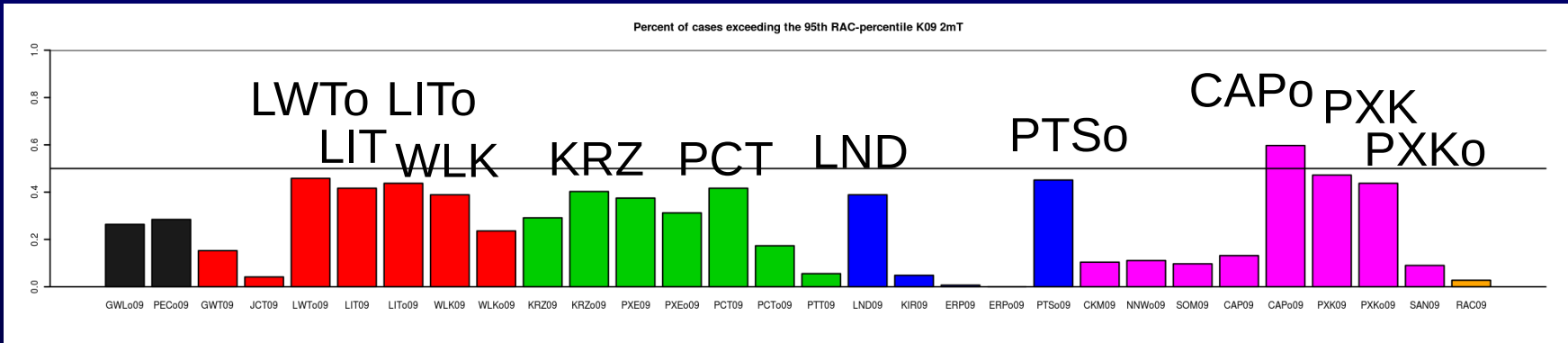


III.) Randomcent: Explained variance of PRC

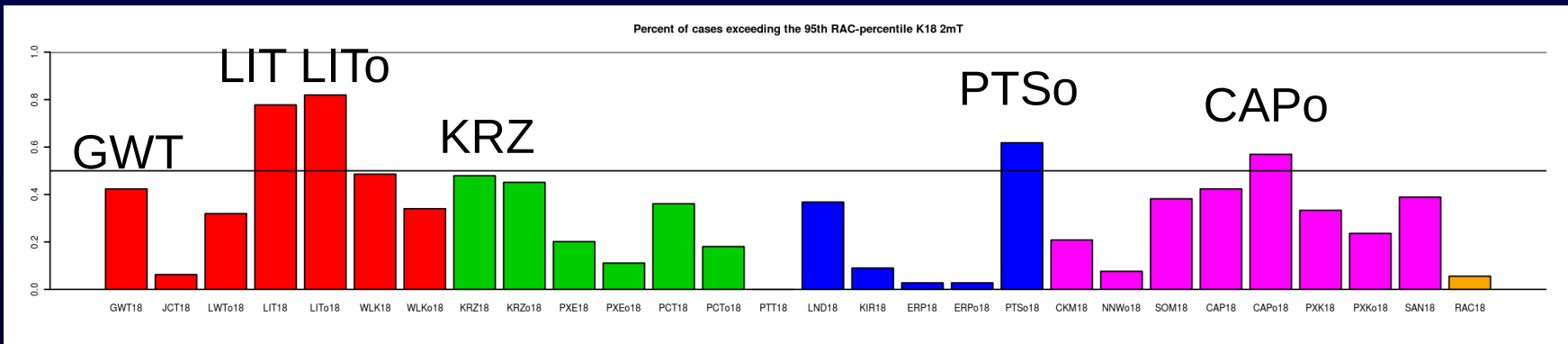


Counting 95th percentile exceedence for 2mT in all seasons and all domains

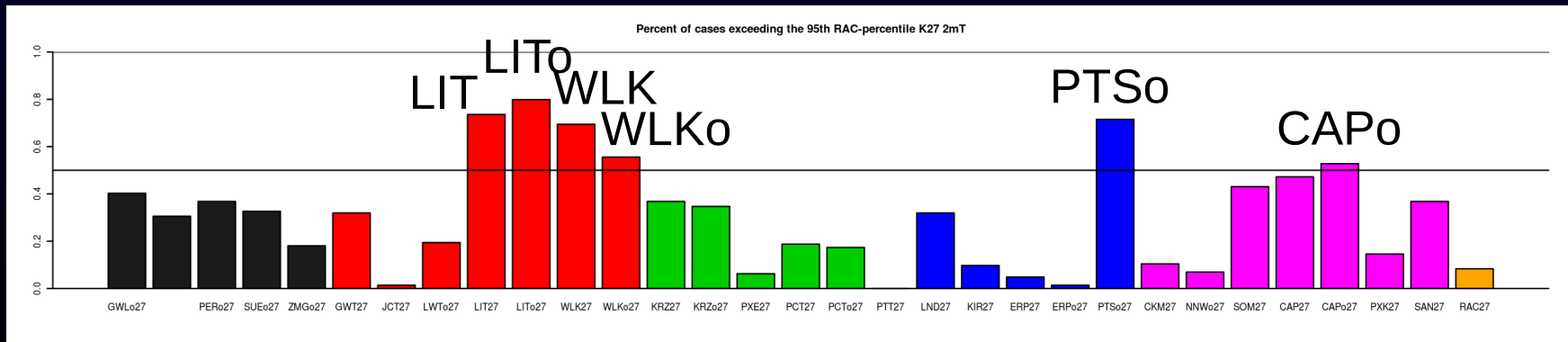
~99 types



~18 types

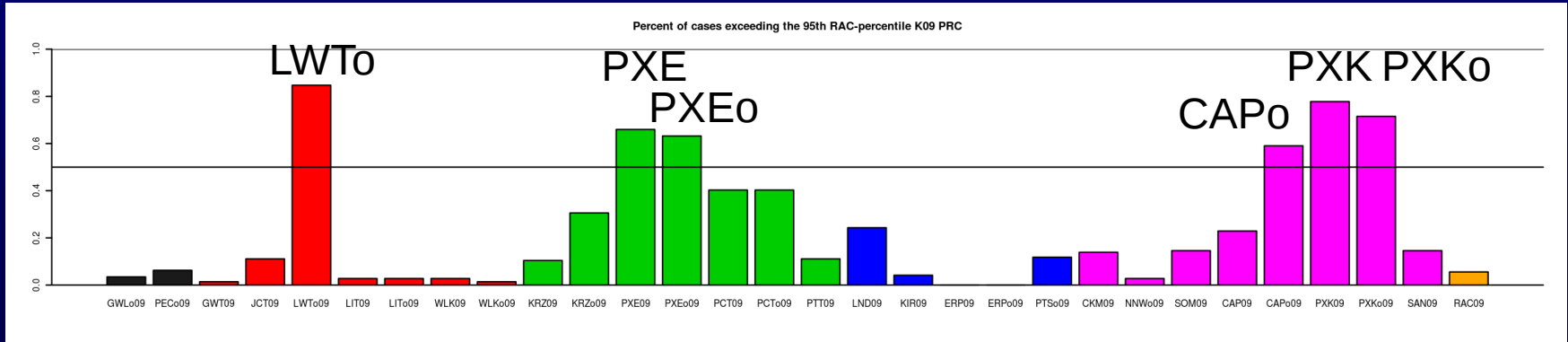


~27 types

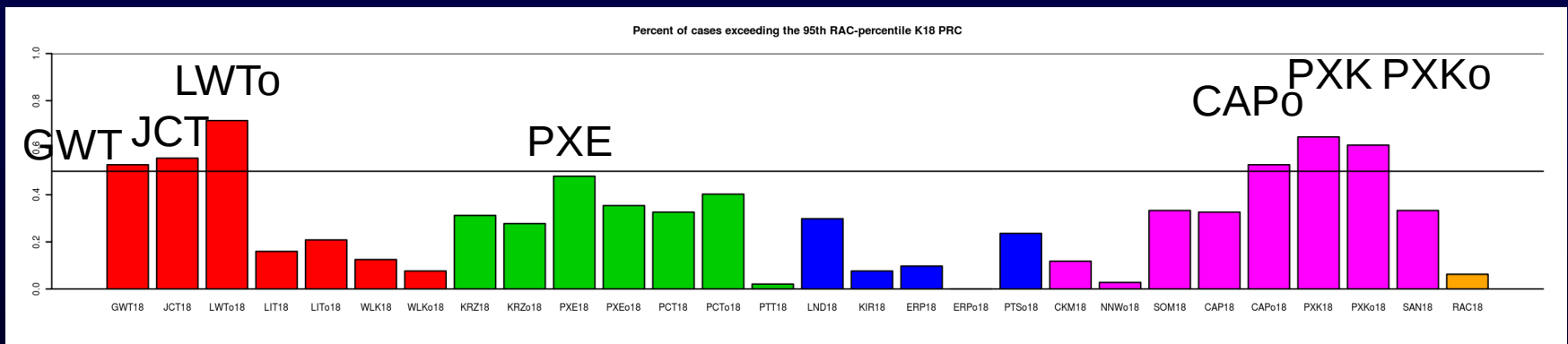


Counting 95th percentile exceedence for PRC in all seasons and all domains

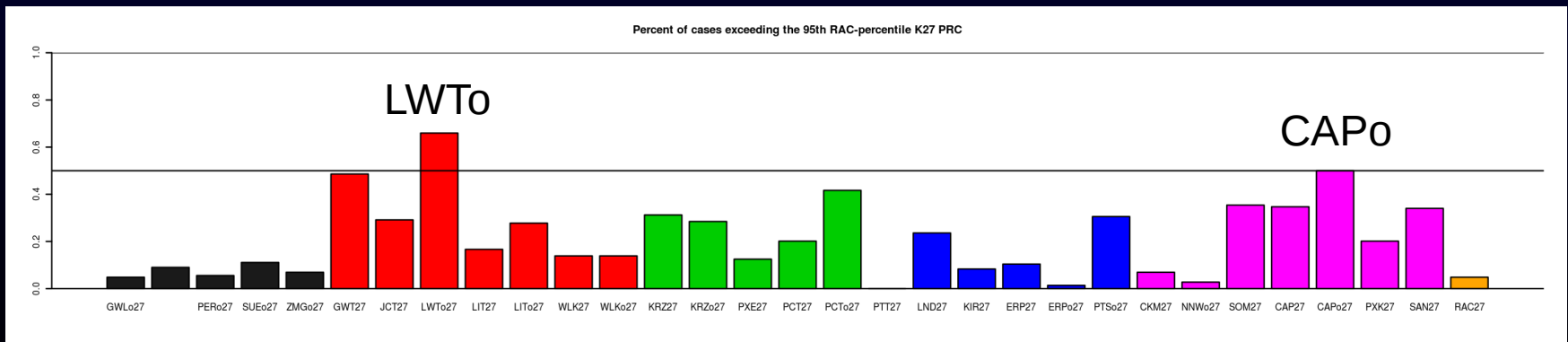
~09 types



~18 types



~27 types



III.) Randomcent

- A benchmark method for classifications
- A method for testing classification tuning parameters excluding the (possibly random) influence of the classification algorithm:

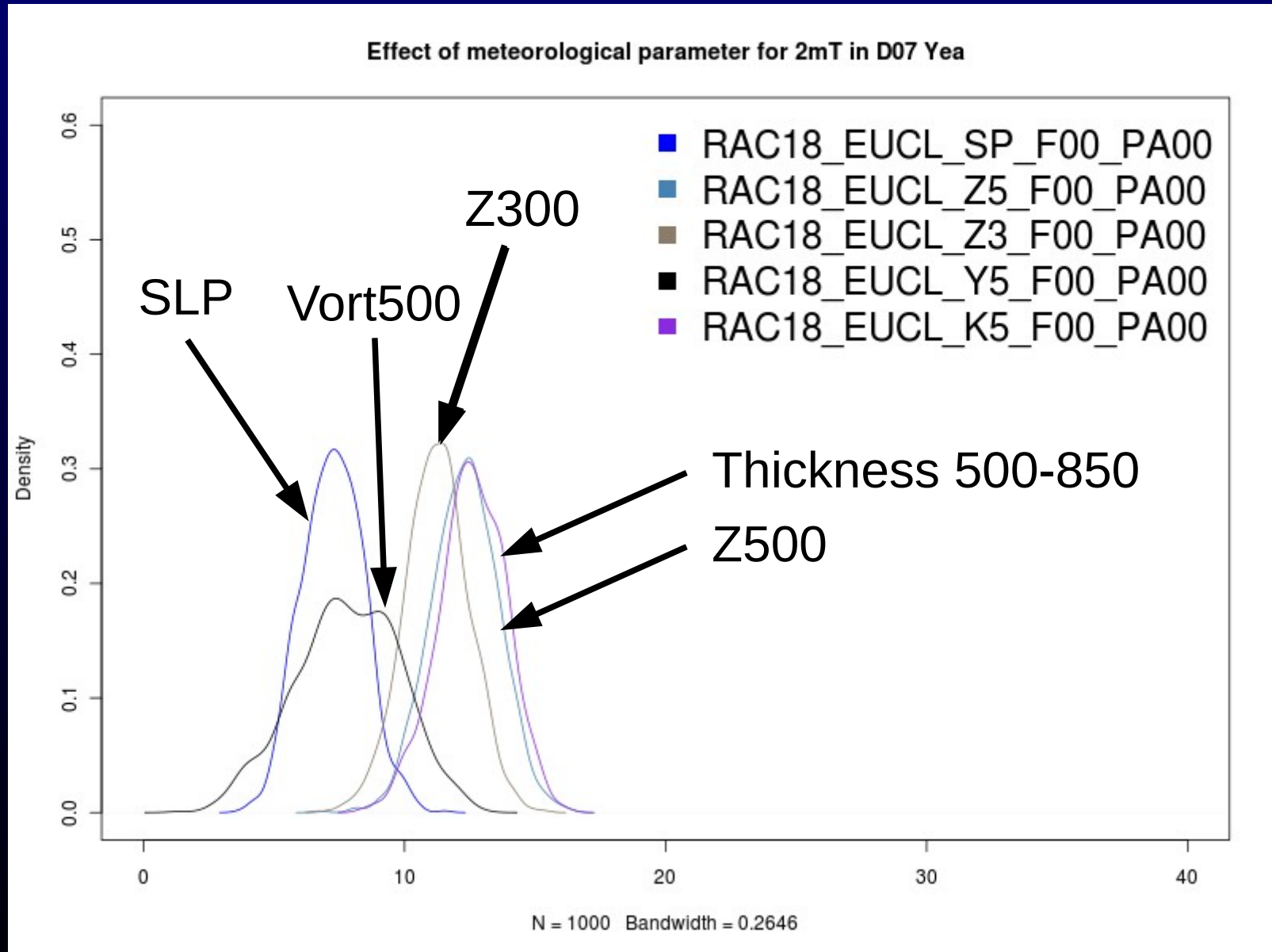
Many (1000) realizations of RAC allow a robust estimate whether a certain tuning parameter is good in general or only by chance!

IV.) RAC for testing classification tuning parameters

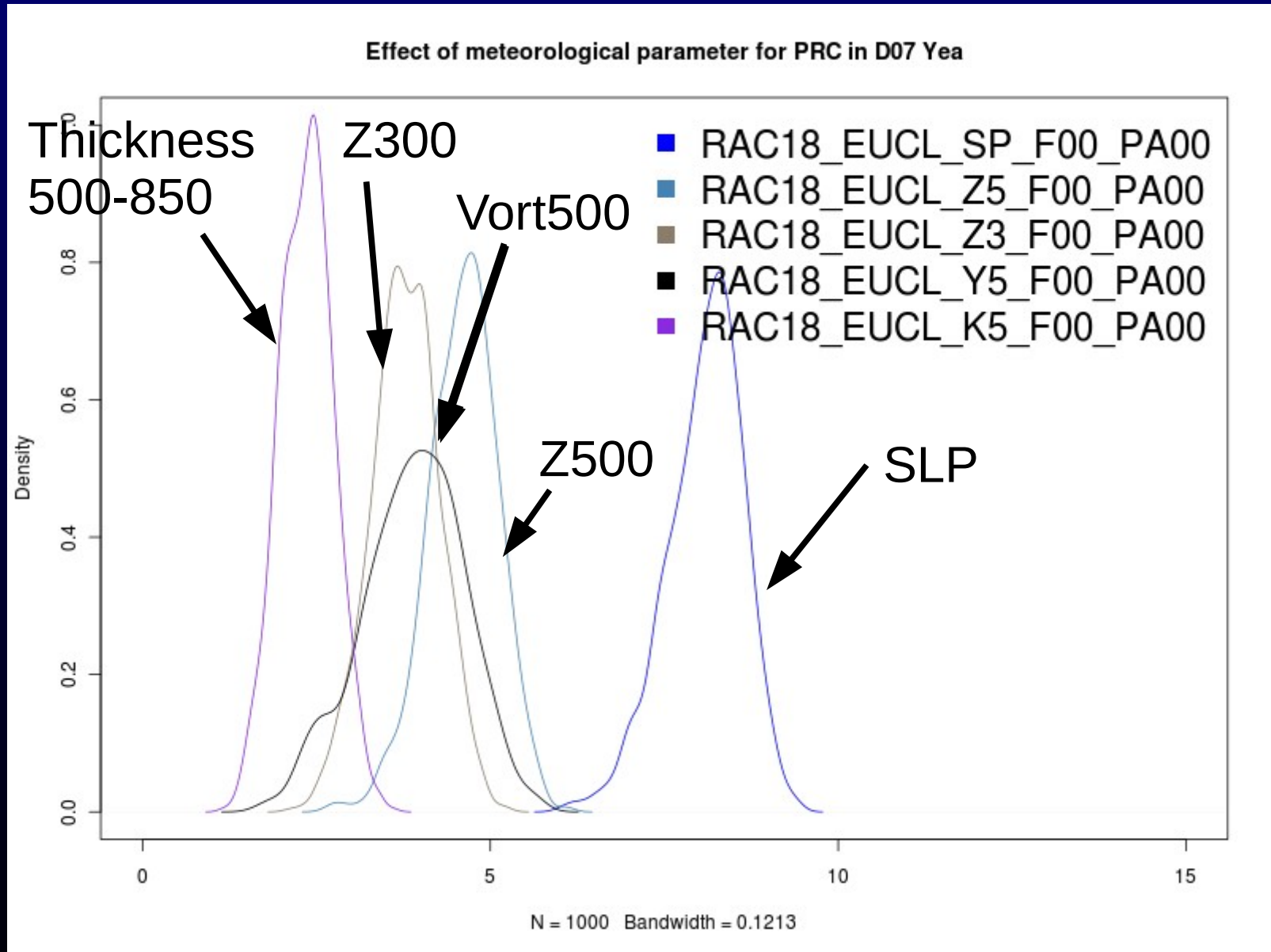
RAC-variants done with cost733class
software:

- meteorological parameter (input variables)
- PCA preprocessing (scores)
- time filtering of input data (high-/low-pass)
- sequence length (1-10-day sequences)
- distance metrics (Euclid, Pearson, Minkowski)

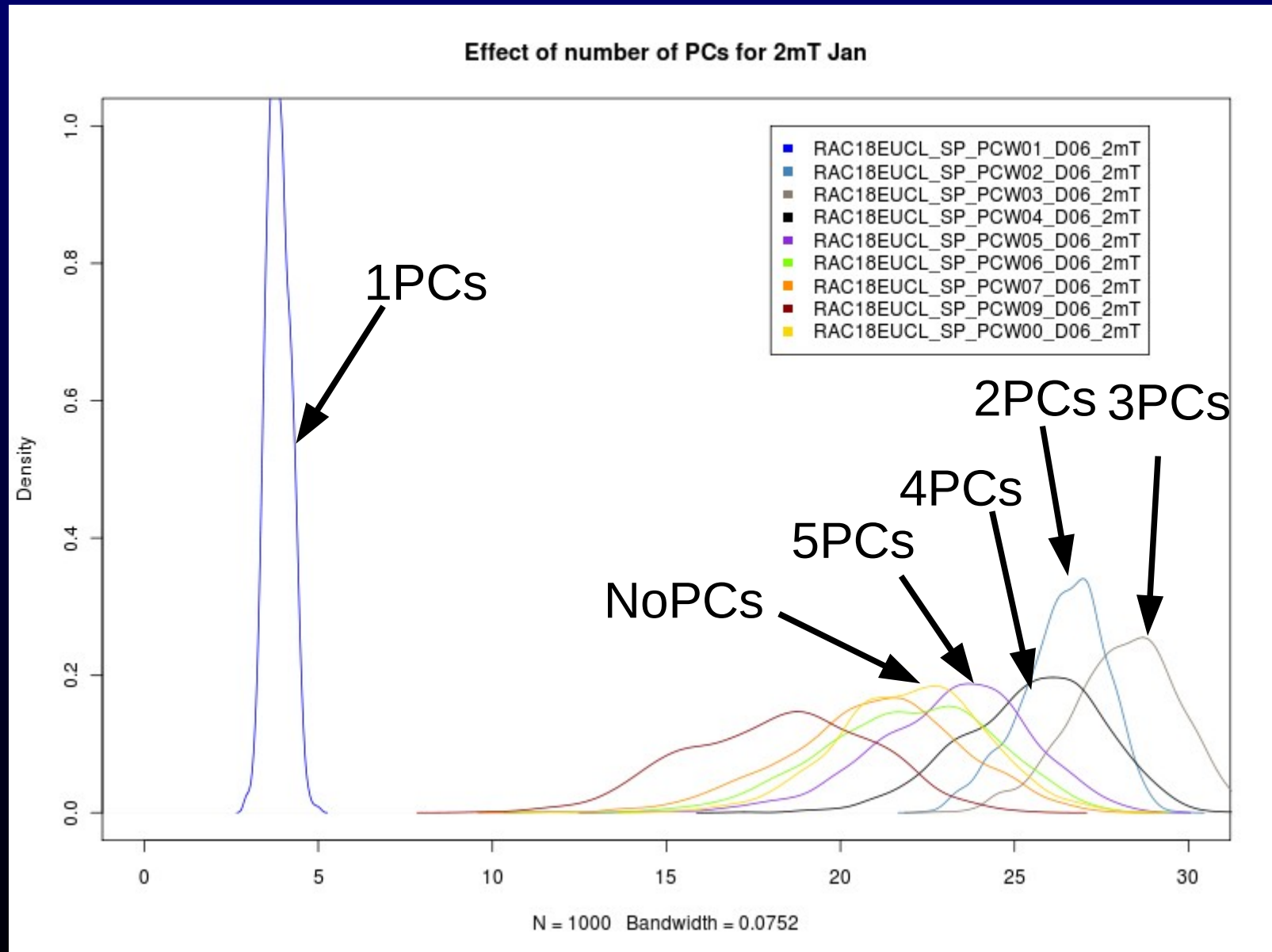
RAC testing meteorological parameter for 2mT D07



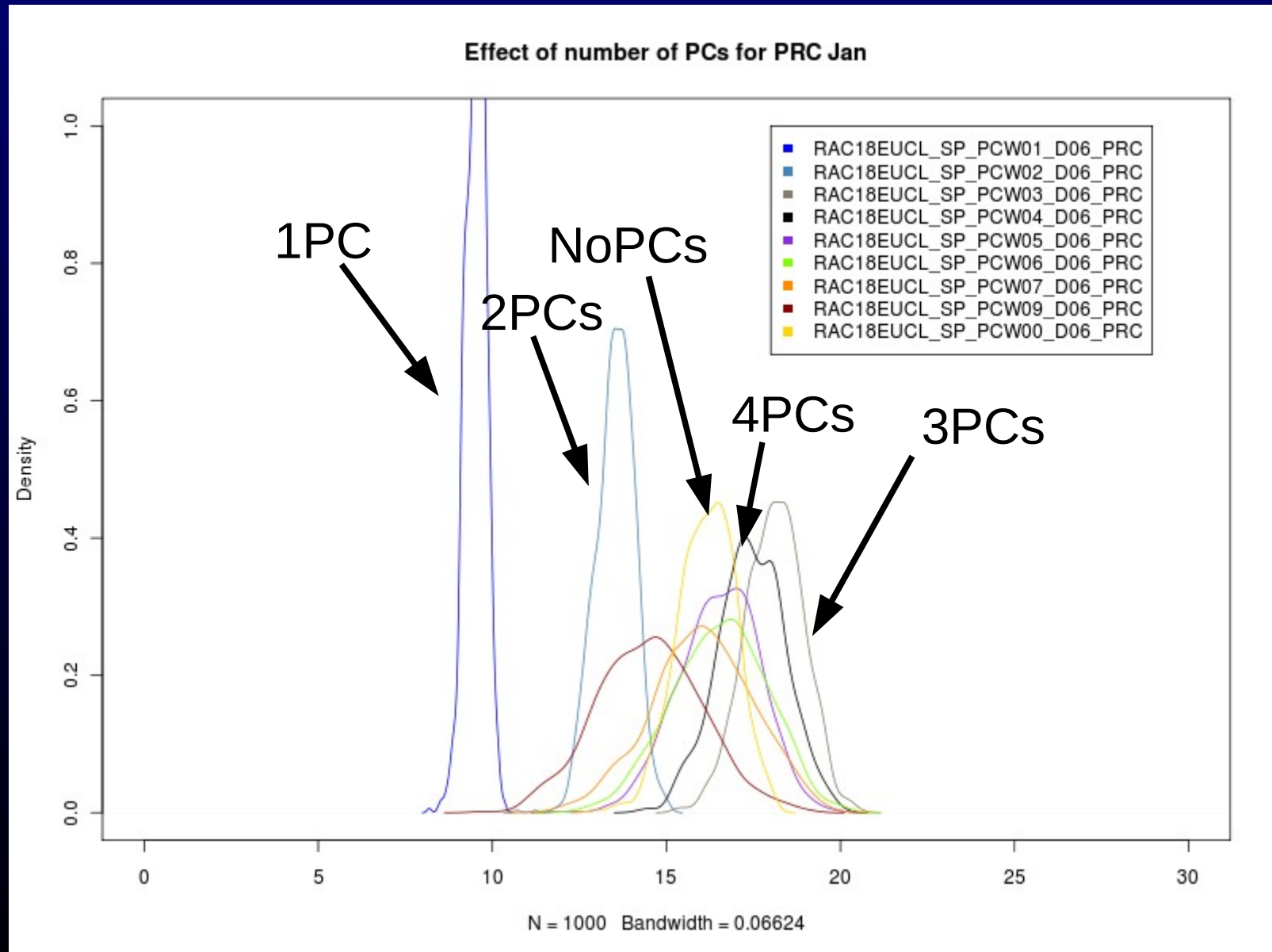
RAC testing meteorological parameter for PRC D07



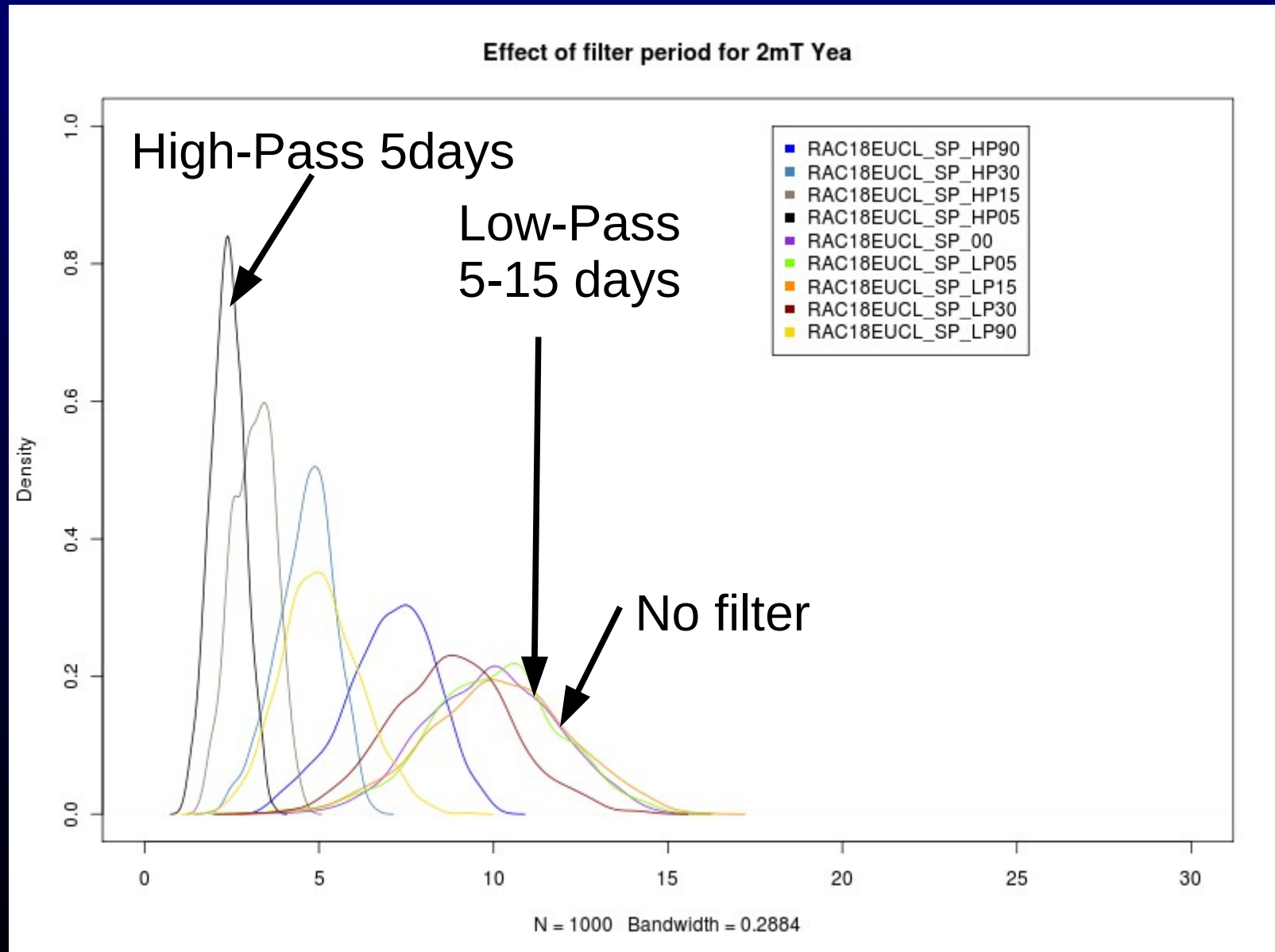
RAC testing number of PCs for 2mT D06



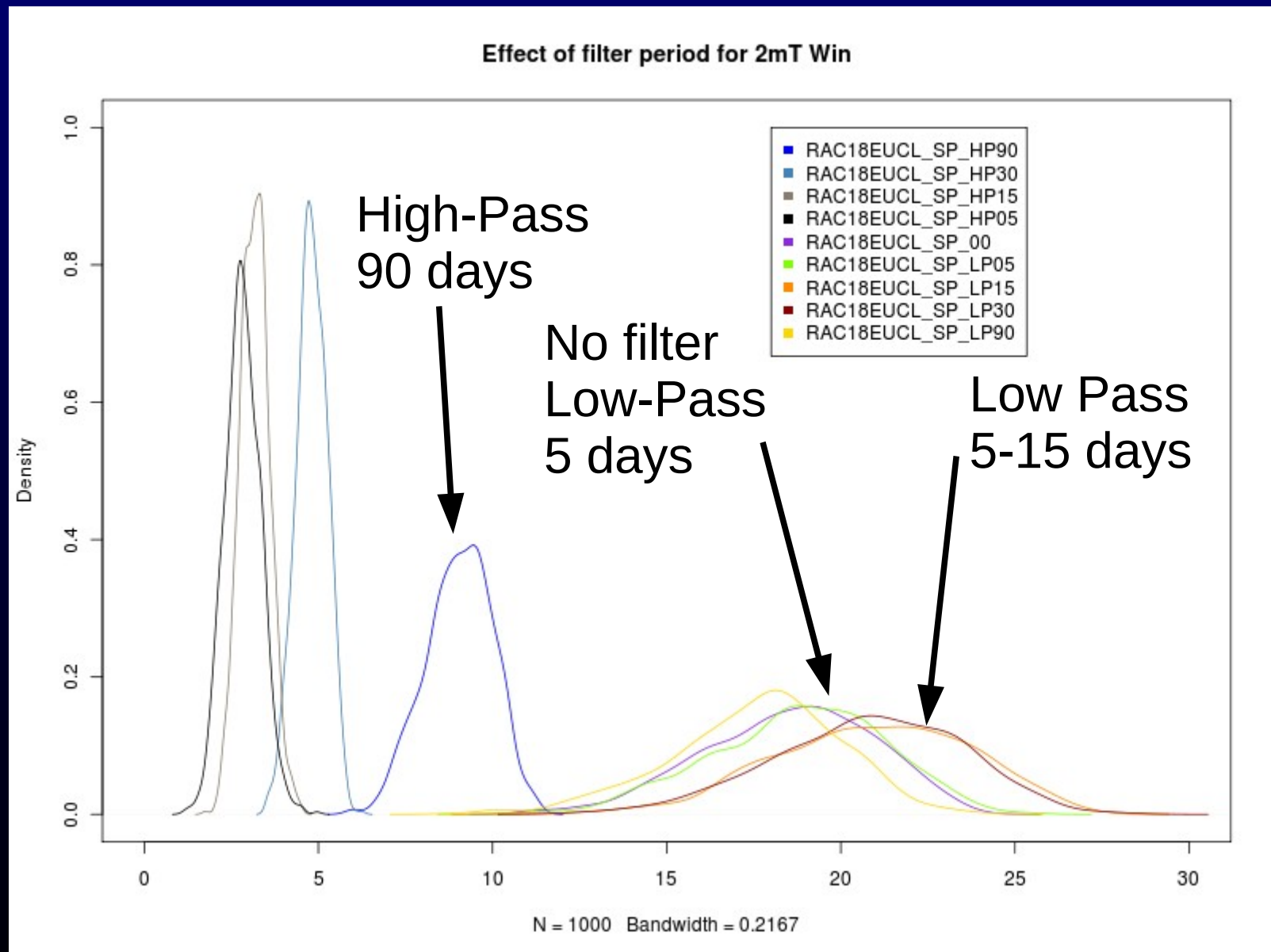
RAC testing number of PCs for PRC D06



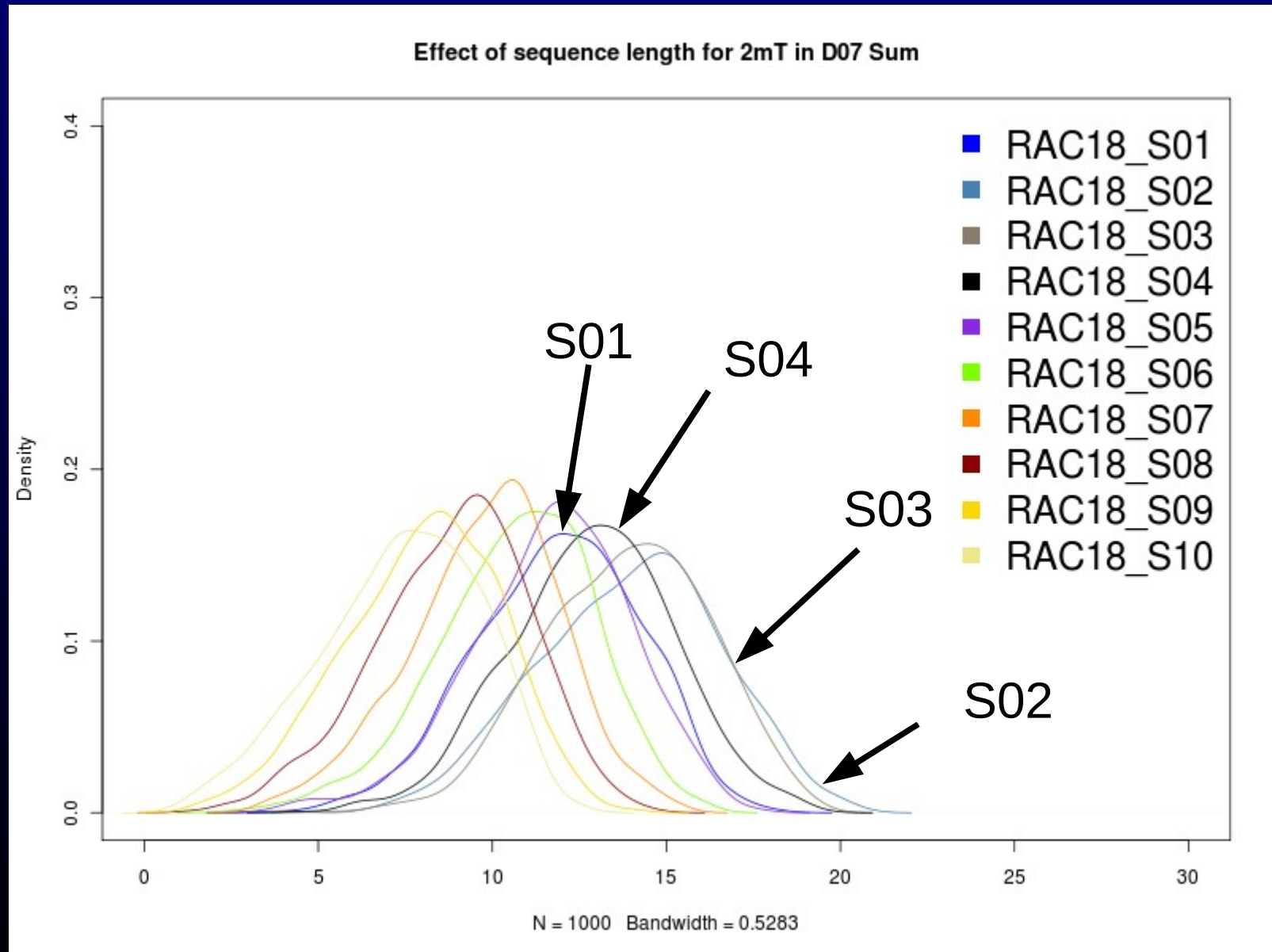
RAC testing time filter period for 2mT (Yea) D06



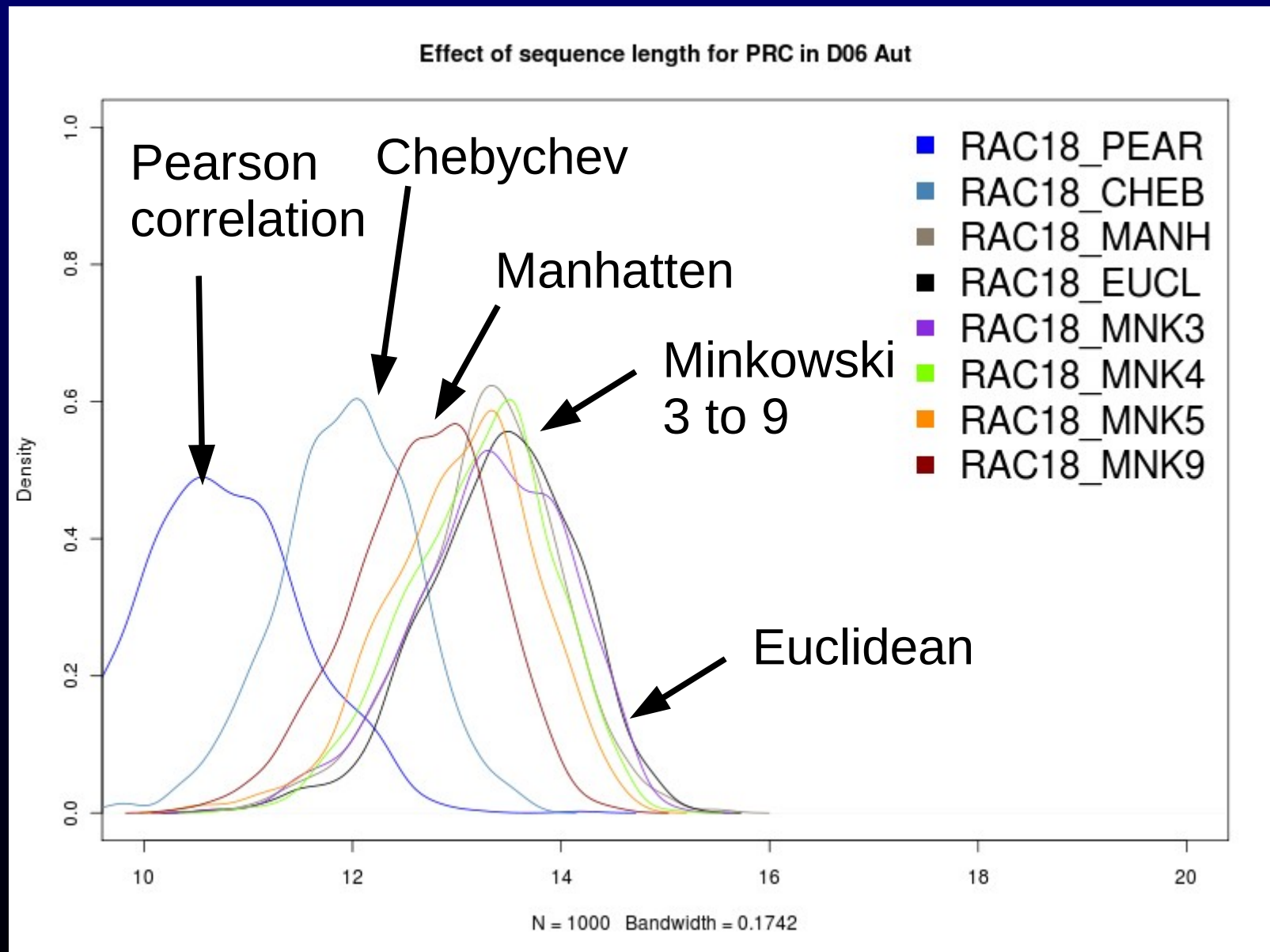
RAC testing time filter period for 2mT (DJF) D07



RAC testing sequence length for 2mT D07



RAC testing distance metrics for PRC D06



Conclusions

- we haven't found/developed the ultimate classification
- but from the methodological point of view we have now much more insight into classification procedures
- for applications it seems not possible to highlight single method
RAC can be used as benchmark
- data preprocessing/selection is more important, testing is simplified by cost733class software, RAC shows systematic rules
- optimum parameter search in conditioned classification still has a large potential to improve applications → Downscaling
- further developments should distinguish between pure circulation classifications for basic (regime) research and conditioned weather type classification for application purposes

Classification tuning for applications in synoptic climatology

OPS Optimal Parameter selection

- not only optimize assignment of days to classes for optimizing downscaling skill but
 - also optimize:
 - number of classes
 - weight of grid points
 - weight of parameters
 - sequences
 - time filtering

Bump-hunting for regimes

in the PDF-plane

→ Circulation
Dynamics

