Search for causality in ecological studies

Context1: estimating impact of the natural factors on the chemical and biological variables in Lake Peipsi

Context2: preparing papers on the topic

Natural variables influencing phosphorus variability in the growing period, in the large shallow Lake Peipsi (Estonia/Russia)

Olga Tammeorg[1], Tõnu Möls[1,2], and Külli Kangur[1]

[1]Centre for Limnology, Estonian University of Life Sciences, 61117 Rannu, Tartumaa, Estonia;
[2]Department of Bioinformatics, Institute of Cell and Molecular Biology, University of Tartu, Riia 23, 51010 Tartu, Estonia

7

Presented by T. Möls at the Bioinformatics Journal Club, 10th Dec 2012

History

Timm, H.; Käiro, K.; Möls, T.; Virro, T. (2011). An index to assess hydromorphological quality of Estonian surface waters based on macroinvertebrate taxonomic composition. *Limnologica*, 41(4), 398 – 410. (1.1)

Käiro, K.; Möls, T.; Timm, H.; Virro, T.; Järvekülg, R. (2011). The effect of damming on biological quality according to macroinvertebrates in some Estonian streams, Central – Baltic Europe: a pilot study. *River Research and Applications*, 27(7), 895 – 907 (1.1)

Buhvestova, O.; Kangur, K.; Haldna, M.; Möls, T. (2011). Nitrogen and phosphorus in Estonian rivers discharging into Lake Peipsi: estimation of loads and seasonal and spatial distribution of concentrations. *Estonian Journal of Ecology*, 60(1), 18 – 38 (1.2)

Haldna, M.; Möls, T.; Buhvestova, O.; Kangur, K. (2012). Predictive model for phosphorus in the large shallow Lake Peipsi: approach based on covariance structures. *Journal of Aquatic Ecosystem Health and Management* [**Accepted**] (1.1)

Timm, H.; Möls, T. (2012). Littoral macroinvertebrates in lowland lakes (Estonia, Baltic ecoregion of Europe): the effects of habitat, season, eutrophication and land use on some indices of biological quality. *Fundamental and Applied Limnology* [**Accepted**] (1.1)

Buhvestova, O.; Möls, T.; Kangur, K. Natural variables as factors behind phosphorus variability in the shallow eutrophic Lake Peipsi (Estonia/Russia) [**Not accepted** by *Hydrobiologia*]

Buhvestova, **O**., **Niemistö**, **J**., **Möls**, **T**., **Laugaste**, **R**., **Kangur**, **K**. Wind-induced sediment resuspension as a factor behind eutrophication of large shallow lake [**To be presented** to *Aquatic Sciences*]

Kangur, **K**., **Kangur**, **A**., **Kangur**, **P**., **Ginter**, **K**., **Orru K**., **Haldna**, **M**., **Möls**, **T**. Long-term effects of extreme weather events and eutrophication on the fish community of shallow Lake Peipsi (Estonia/Russia) [*Fisheries Management and Ecology* rejected. Rewritten will be presented to *Fundamental and Applied Limnology*]

History: actual state

Buhvestova, O.; Möls, T.; Kangur, K. Natural variables as factors behind phosphorus variability in the shallow eutrophic Lake Peipsi (Estonia/Russia) [**Not accepted** by *Hydrobiologia*]

↓ (Rewritten)

Tammeorg, O., Möls, T., and Kangur, K. Natural variables influencing phosphorus variability in the growing period, in the large shallow Lake Peipsi (Estonia/Russia) [**Presented to** ***Aquatic Sciences***]

Buhvestova, **O.**, **Niemistö**, **J.**, **Möls**, **T.**, **Laugaste**, **R.**, **Kangur**, **K.** Wind-induced sediment resuspension as a factor behind eutrophication of large shallow lake [~~To be~~ **presented** to *Aquatic Sciences*]

Kangur, **K.**, **Kangur**, **A.**, **Kangur**, **P.**, **Ginter**, **K.**, **Orru K.**, **Haldna**, **M.**, **Möls**, **T**. Long-term effects of extreme weather events and eutrophication on the fish community of shallow Lake Peipsi (Estonia/Russia) [rejected by *Fisheries Management and Ecology*]

↓ ↓ (Substantially rewritten) ↓

Authors and Title restated and the main problems differently focused

Factors, potentially influencing the total phosphorus (TP) in Lake Peipsi:

Constant

Year (*long-time trends: climate warming, pollution increase etc.)*
Day of the year *(regular seasonal changes during year, specific for the given site and ...)*
Coordinates of the sampling site *(morphometric and other condition in the site)*

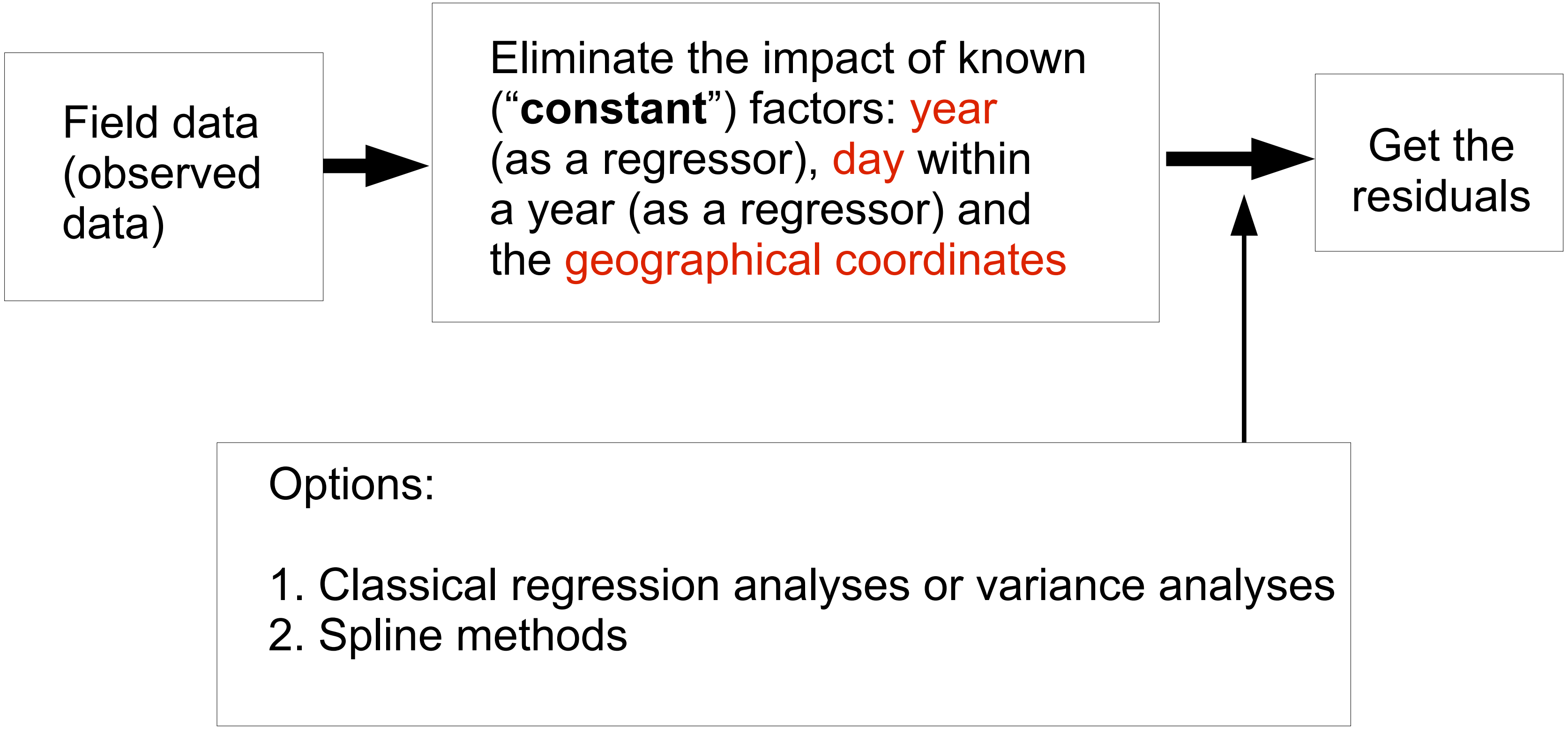
Natural

Water level (WL)
Water temperature (WT)
Photosynthetically active radiation (PAR)
Wind velocity (WV)

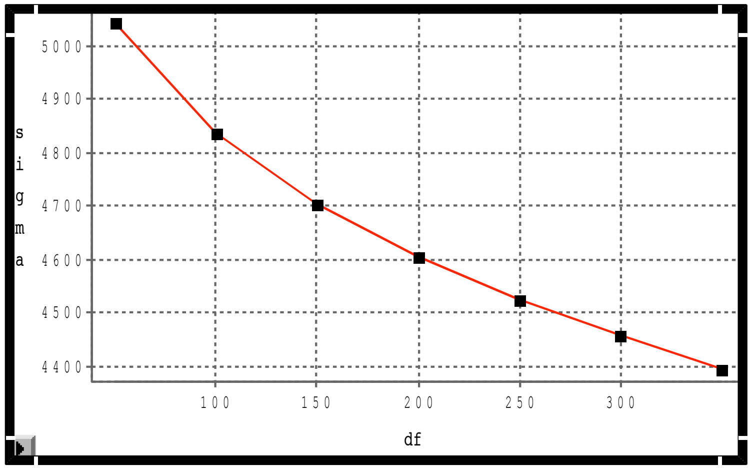
The main problem:

How do the natural factors – water temperature (WT), water level (WL), photosynthetically active radiation (PAR) and the wind velocity (WV) influence the concentration of the total phosphorus (TP) in the Lake Peipsi?

Flowchart of the first step of the analysis



Two basic methods for calculating residuals from constant factors:



(1)
Classical regression and variance analysis

(2)
Thin-Plate Splines

The TPSPLINE Procedure
Dependent Variable: LPTOT

Summary of Input Data Set

Non-Missing Observations	1862
Number of Missing Observations	0
Unique Smoothing Design Points	1858

Summary of Final Model

Number of Regression Variables	0
Number of Smoothing Variables	4
Order of Derivative in the Penalty	3
Dimension of Polynomial Space	15

Summary Statistics of Final Estimation

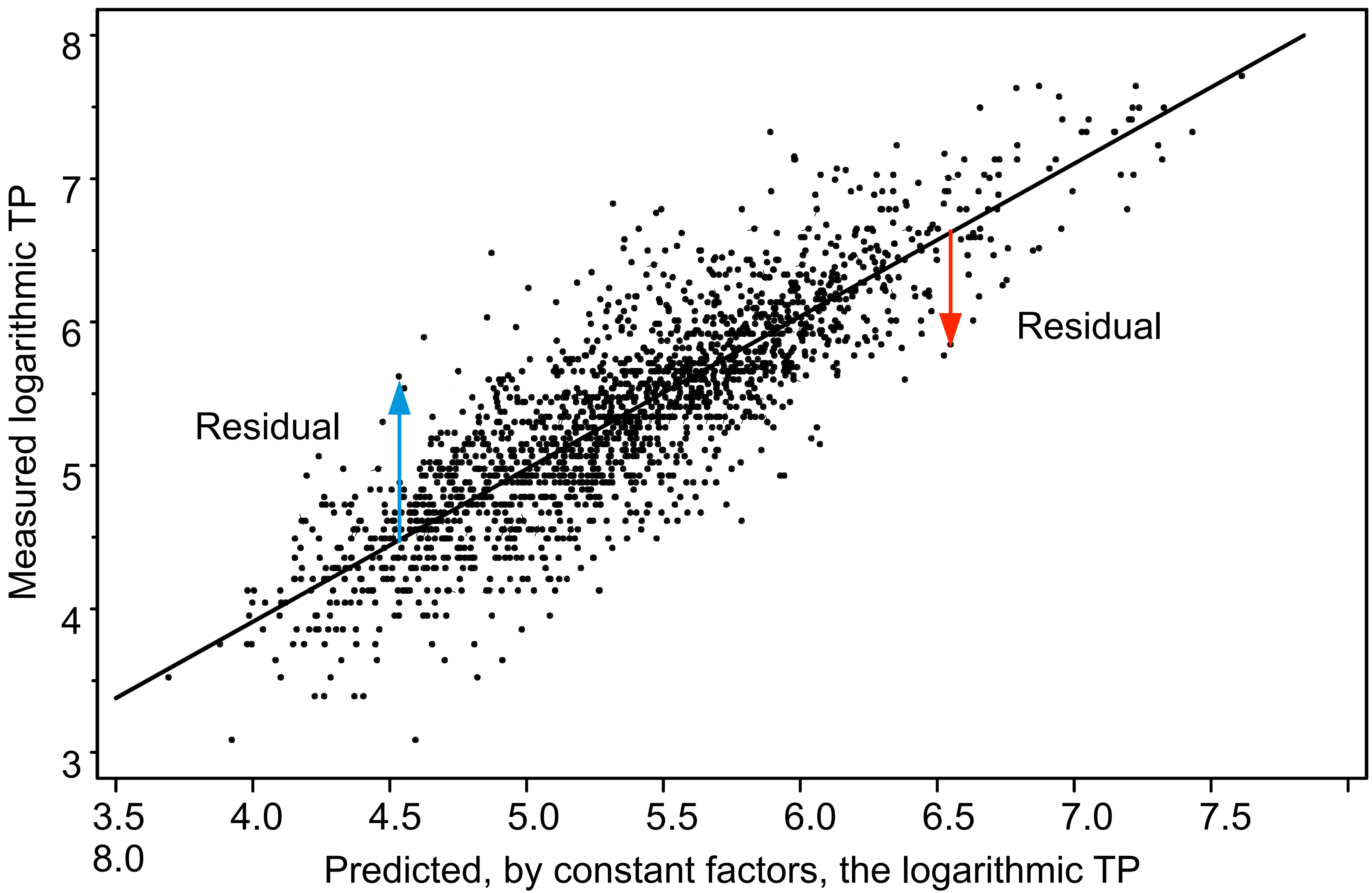
log10(n*Lambda)	-3.5198
Smoothing Penalty	157309.9076
Residual SS	324.0839
Tr(I-A)	1596.1030
Model DF	265.8970
Standard Deviation	0.4506

log$_2$(Ptot) =

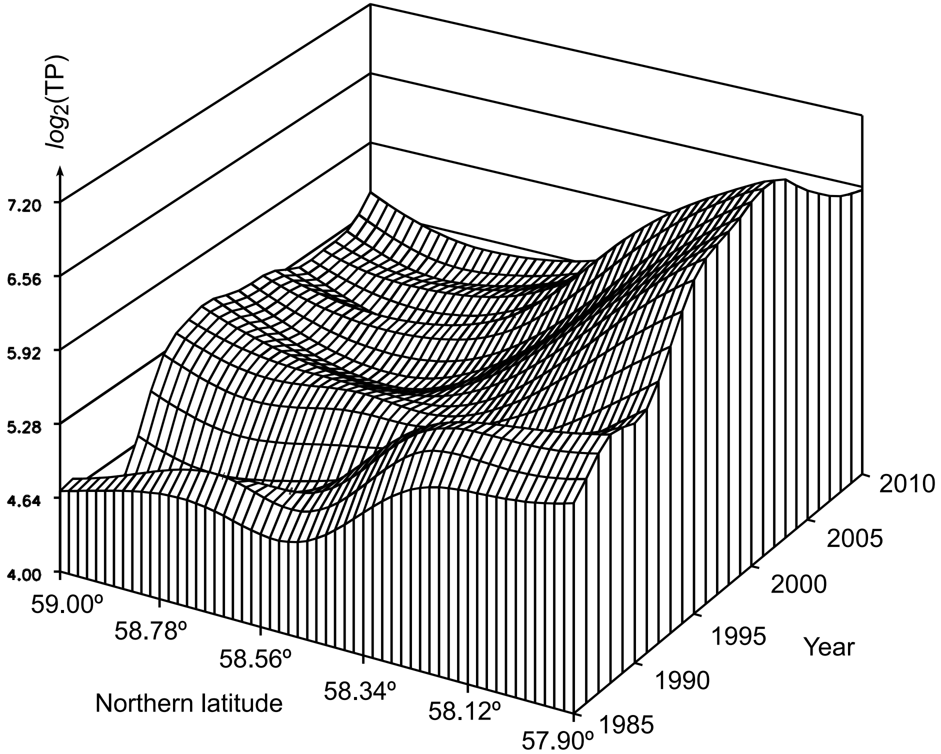
1+ /* vabaliige (intercept) */
a1+a2+a3+a4+a5+ /* aastaarvu (year) teisendid */
pl+pl*pl+pl*pl*pl+ip+ /* põhjalaiuse (latitude) ja idapikkuse (longitude) teisendid */
t35+t44+t53+ /* päeva number aastas (Julian day) teisendid */
sg+sg*sg+sg*sg*sg+sg*sg*sg*sg+ /* proovivõtu sügavus (sampling depth) */

*I*Interaktsioonitermid (Interaction terms) */*

a1*pl+a1*pl*pl+a2*pl+a2*pl*pl+a3*pl+a3*pl*pl+a4*pl+ a4*pl*pl+ a5*pl+a5*pl*pl+a1*ip+ a2*ip+a3*ip+a4*ip+
a5*ip+a1*t35+a1*t44+a1*t53+a2*t35+a2*t44+a2*t53+a3*t35+a3*t44+a3*t53+a4*t35+a4*t44+a4*t53+
a5*t35+a5*t44+a5*t53+pl*ip+pl*pl*ip+pl*t35+pl*t44+pl*t53+pl*pl*t35+pl*pl*t44+pl*pl*t53+ip*t35+ip*t44+
ip*t53+a1*pl*ip+a1*pl*t35+a1*pl*t44+a1*pl*t53+a2*pl*ip+a2*pl*t35+a2*pl*t44+a2*pl*t53+a3*pl*ip+
a3*pl*t35+a3*pl*t44+a3*pl*t53+a4*pl*ip+a4*pl*t35+a4*pl*t44+a4*pl*t53+a5*pl*ip+a5*pl*t35+a5*pl*t44+
a5*pl*t53+a1*pl*pl*t35+a1*pl*pl*t44+a1*pl*pl*t53+a2*pl*pl*t35+a2*pl*pl*t44+a2*pl*pl*t53+a3*pl*pl*t35+
a3*pl*pl*t44+a3*pl*pl*t53+a4*pl*pl*t35+a4*pl*pl*t44+a4*pl*pl*t53+a5*pl*pl*t35+a5*pl*pl*t44+a5*pl*pl*t53+
a1*ip*t35+a1*ip*t44+a1*ip*t53+a2*ip*t35+a2*ip*t44+a2*ip*t53+a3*ip*t35+a3*ip*t44+a3*ip*t53+a4*ip*t35+
a4*ip*t44+a4*ip*t53+a5*ip*t35+a5*ip*t44+a5*ip*t53+pl*ip*t35+pl*ip*t44+pl*ip*t53+a1*pl*ip*t35+a1*pl*ip*t44+
a1*pl*ip*t53+a2*pl*ip*t35+a2*pl*ip*t44+a2*pl*ip*t53+a3*pl*ip*t35+a3*pl*ip*t44+a3*pl*ip*t53+a4*pl*ip*t35+
a4*pl*ip*t44+a4*pl*ip*t53+a5*pl*ip*t35+a5*pl*ip*t44+a5*pl*ip*t53+sg*pl+sg*pl*pl+sg*t35+sg*t44+sg*t53+



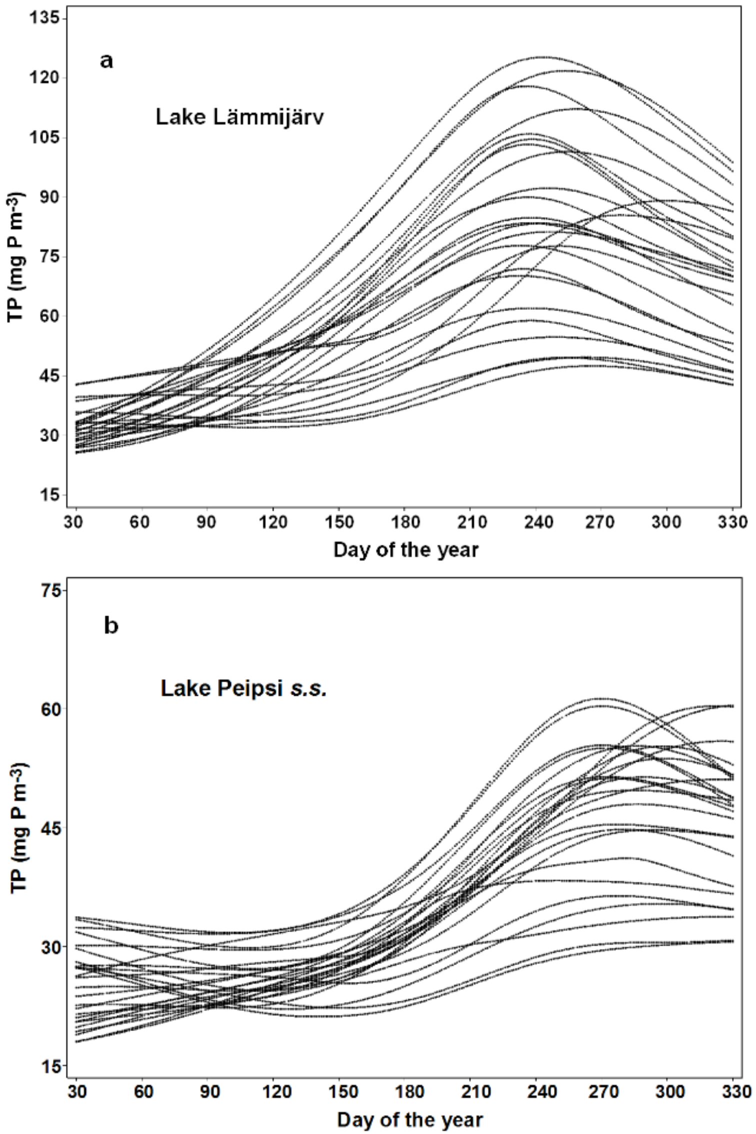
Measured logarithmic TP
8
7
6
5
4
3
Residual
Residual
3.5
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0
Predicted, by constant factors, the logarithmic TP



Section of the 4-dimensional Thin-Plate-Spline surface fitted to the 1862 LTP measurements made during years1985 to 2010 in Lake Peipsi.

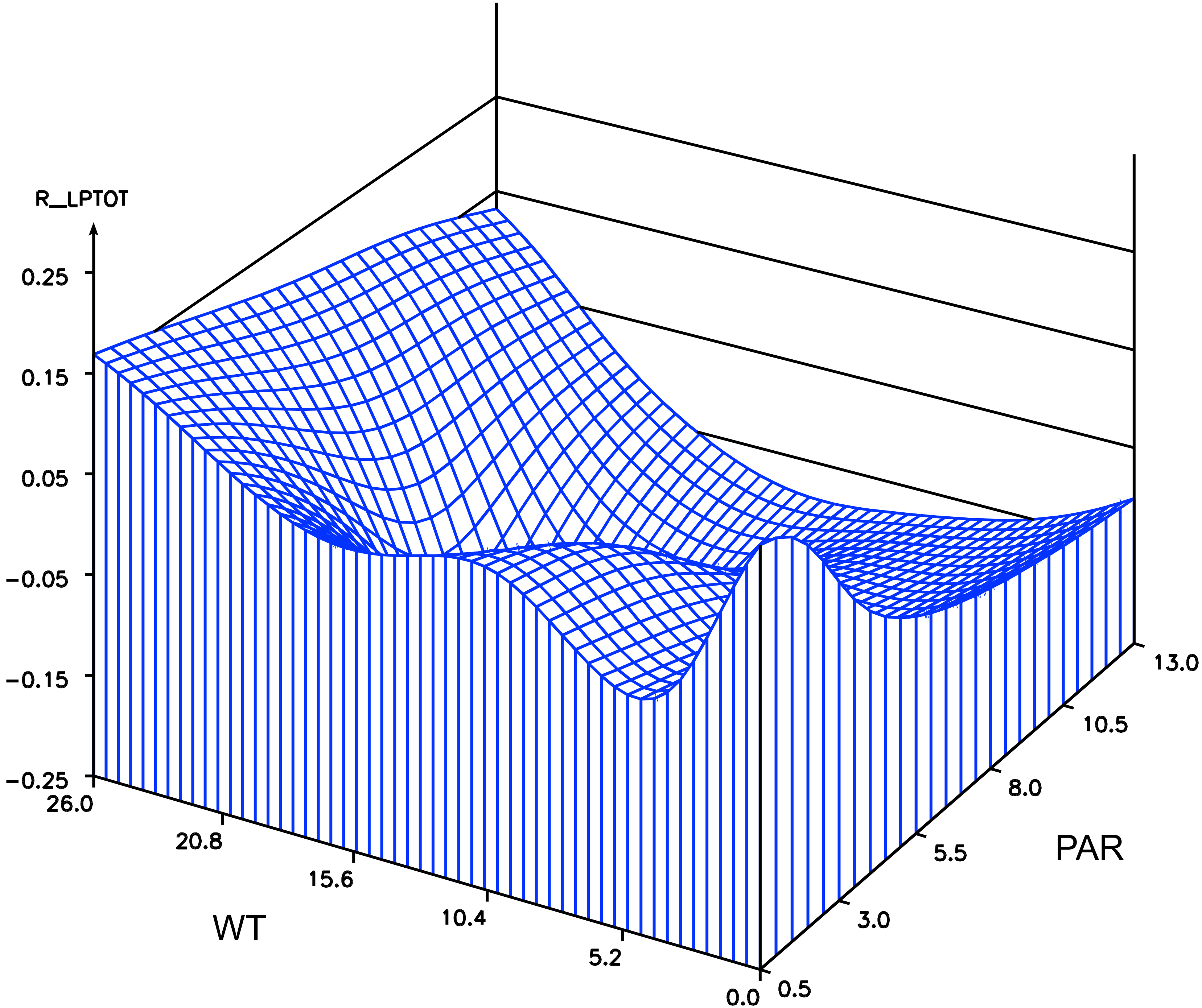
Section intersects the point corresponding to 183[th] day of the year and the geographical point at 27.5º eastern longitude

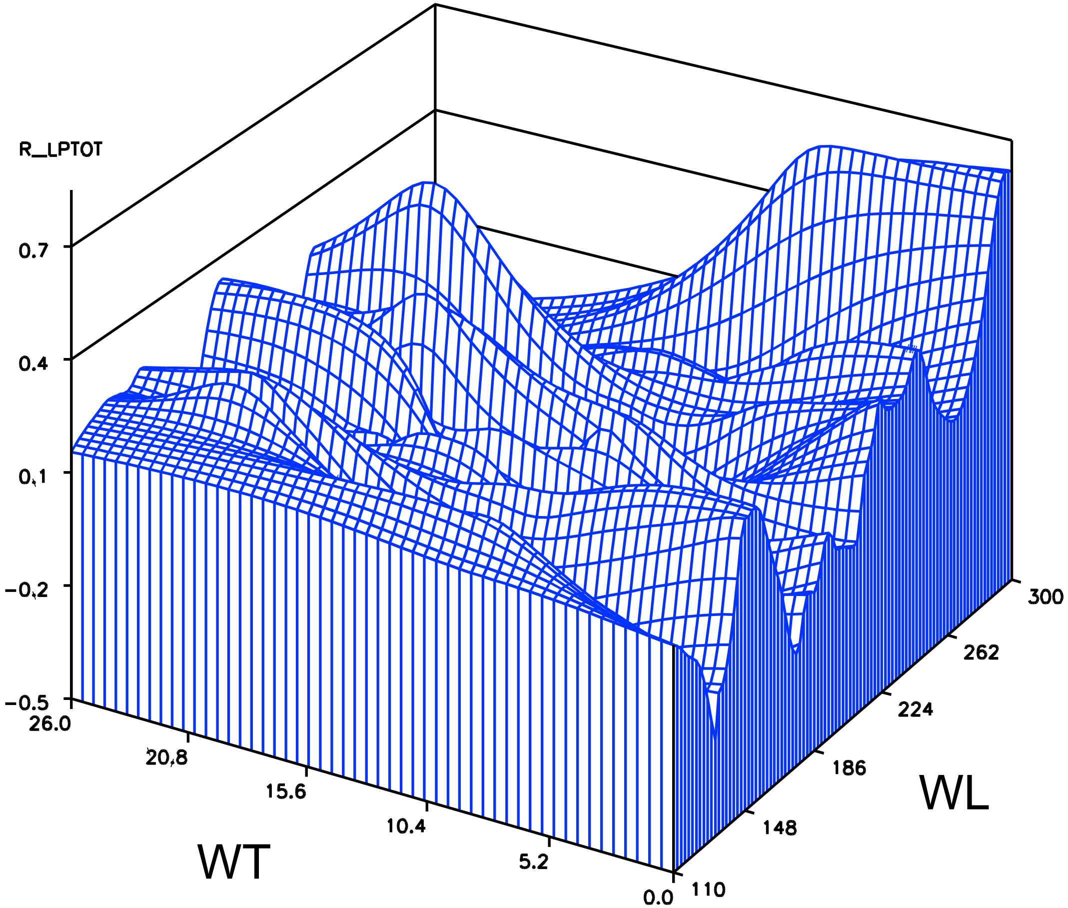

```
proc TPspline data = Peipsitihe_Ptotst;
    model LPtot =  (a t pl ip) / d=0.1 ;
    score data=to_pred out=vaip;
output  out = predy_Realdat pred resid;
run;
```



Sections of the 4-dimensional Thin-Plate-Spline surface fitted to the 1862 LTP measurements of years 1985 – 2010 in Lake Peipsi.

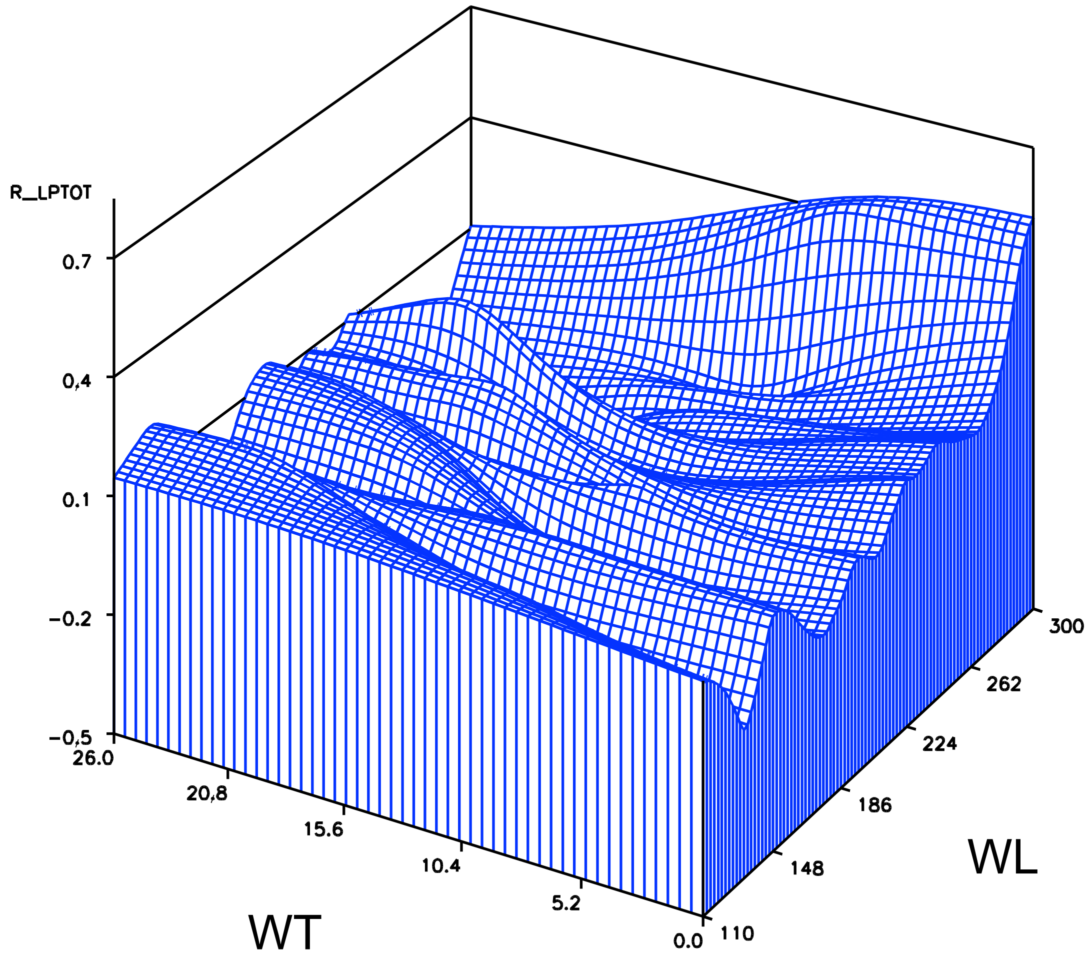
Each line corresponds to a specific year from 1985–2010. Sections on figure (a) intersect the point corresponding to the site <58.23N, 27.42E> in Lake Lämmijärv, lines on figure (b) run through site <58.80N, 27.36E > (Lake Peipsi *s.s.*)


R_LPTOT
0.25
0.15
0.05
−0.05
−0.15
−0.25
26.0
20.8
15.6
10.4
5.2
0.0
WT
0.5
3.0
5.5
8.0
10.5
13.0
PAR



Summary Statistics of Final Estimation

log10(n*Lambda)	-0.6959
Smoothing Penalty	70.5186
Residual SS	190.2115
Tr(I-A)	1663.3565
Model DF	185.6435
Standard Deviation	0.3382



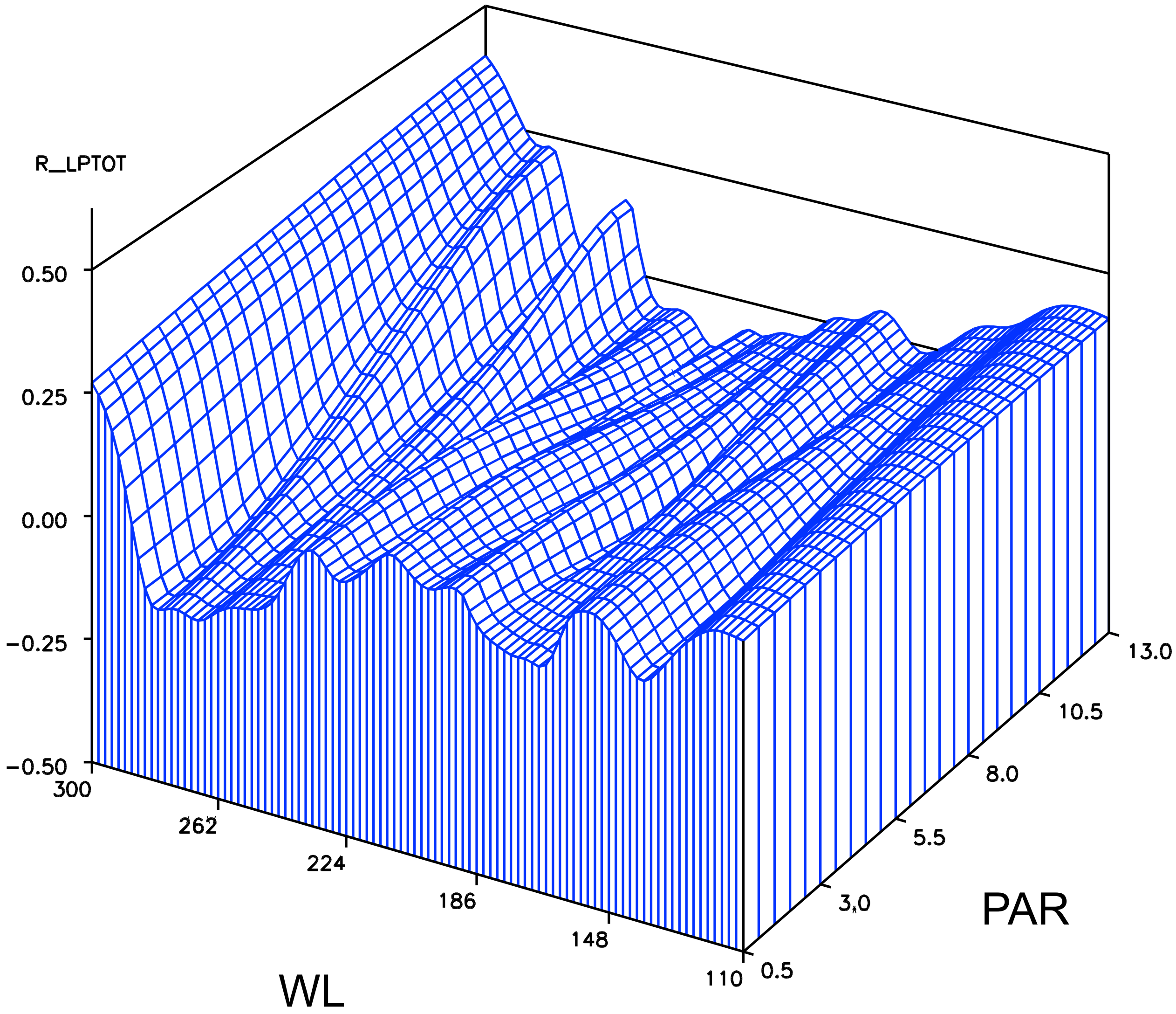
Non-Missing Observations 1849
Missing Observations 13
Unique Smoothing Design Points 255

Summary of Final Model

Number of Regression Variables	0
Number of Smoothing Variables	3
Order of Derivative in the Penalty	2
Dimension of Polynomial Space	4

Summary Statistics of Final Estimation

log10(n*Lambda)	0.8149
Smoothing Penalty	3.2342
Residual SS	252.3960
Tr(I-A)	1799.0000
Model DF	50.0000
Standard Deviation	0.3746



Non-Missing Observations 1849
Missing Observations 13
Unique Smoothing Design Points 255

Summary of Final Model

Number of Regression Variables	0
Number of Smoothing Variables	3
Order of Derivative in the Penalty	2
Dimension of Polynomial Space	4

Summary Statistics of Final Estimation

log10(n*Lambda)	0.8149
Smoothing Penalty	3.2342
Residual SS	252.3960
Tr(I-A)	1799.0000
Model DF	50.0000
Standard Deviation	0.3746

Using the time shifts in estimating of influence of natural factors makes the results more trustworthy

(Generation of new model terms with shifts and interactions)

WT: water temperature. When measured?
In the time of sampling?
Three days ago?
Maximal WT last year?

???

Solution by using indexes:

1) WT5_3 = mean WT of five days, measured three days earlier (green asterisks)

2) Realmaxtemp2 = maximal WT two years earlier

Oluliste termide leidmine suure hulga termide seast

Faktorite ja nendest moodustatud termide hulk võib olla väga suur – nt. ≈30 000

Miks nii palju terme?
5 looduslikku faktorit (Wt, Wl, PAR, Wvmean, Wvmax)
6 erineva aknaga keskmistamine (1 … 6 päeva)
8 erinevat nihet ajas varasemaks (0 … 7 päeva)
Kokku 5x6x8 = 240 termi
Kuna mudelis on ka interaktsioonid (termide korrutised), siis
veel 240x241 / 2 = 28920 interaktsioonitermi
Kokku 240 + 28920 = 29160 termi
Järjekorras oleksid nad näiteks nii:
Wt1_0, Wt2_0,..., Wt6_0, …, Wvmax6_7*Wvmax6_7

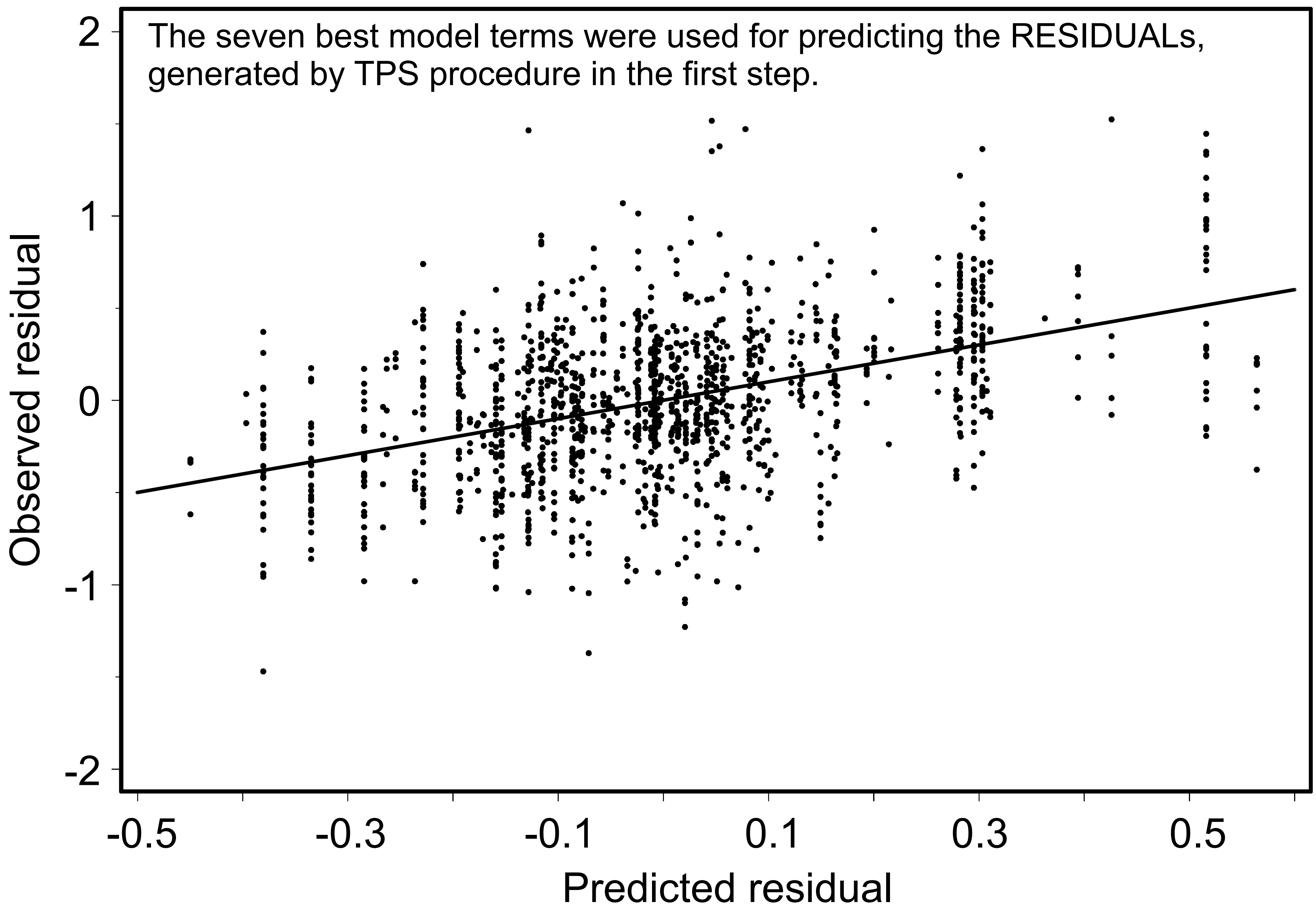
Nende 29160 termide seast tuleb leida kõige olulisemad TP mõjutajad

Ilus otsimise meetod on: Samm-sammuline (stepwise) regressioon
Teine, meil vähe proovitud on GLMSELECT

Results of the statistical analysis are difficult to interpret and present

For example, the stepwise regression analysis has found amongst 29160 model terms seven significant terms:

Term	**Effect**	***P***
WT1_1	+	<0.0001
WL1_4*WL1_3	-	<0.0001
WT1_3*WL2_7	+	<0.0001
PAR1_0*WL1_0	+	<0.0001
PAR4_0*WT1_0	-	<0.0001
PAR4_0*WL1_6	-	0.0002
PAR3_1*PAR3_1	+	<0.0001


The seven best model terms were used for predicting the RESIDUALs, generated by TPS procedure in the first step.
Observed residual
Predicted residual
2
1
0
-1
-2
-0.5
-0.3
-0.1
0.1
0.3
0.5

Analysis of residuals

```
proc glm data= set0;
    model R_LPtot = wt3_4 | wl3_4 | qf3_4 | wv_av3_4 | wv_mx3_4 |
    wt3_4 | wl3_4 | qf3_4 | wv_av3_4 | wv_mx3_4 @2 / ss1 solution;

    estimate "Wt3_4"
        wt3_4                1
        wt3_4*wl3_4          mean_wl
        wt3_4*qf3_4          mean_qf
        wt3_4*wv_av3_4       mean_wv_av
        wt3_4*wv_mx3_4       mean_wv_mx
        wt3_4*wt3_4          mean_wt2;

    estimate "Wl3_4"
        wl3_4                1
        wl3_4*wt3_4          mean_wt
        wl3_4*qf3_4          mean_qf
        wl3_4*wv_av3_4       mean_wv_av
        wl3_4*wv_mx3_4       mean_wv_mx
        wl3_4*wl3_4          mean_wl2;
    ……..................................
run; quit;
```

Analysis of residuals (kernel of the SAS macro)

```
proc glm data= set0;
    model R_LPtot = wt&i._&j|wl&i._&j|qf&i._&j|wv_av&i._&j|wv_mx&i._&j|
    wt&i._&j|wl&i._&j|qf&i._&j|wv_av&i._&j|wv_mx&i._&j @2 / ss1 solution;

    estimate "Wt&i._&j"
        wt&i._&j                1
        wt&i._&j*wl&i._&j       &a0wl
        wt&i._&j*qf&i._&j       &a0qf
        wt&i._&j*wv_av&i._&j    &a0wv_av
        wt&i._&j*wv_mx&i._&j    &a0wv_mx
        wt&i._&j*wt&i._&j       &a0wt2;

    estimate "Wl&i._&j"
        wl&i._&j                1
        wl&i._&j*wt&i._&j       &a0wt
        wl&i._&j*qf&i._&j       &a0qf
        wl&i._&j*wv_av&i._&j    &a0wv_av
        wl&i._&j*wv_mx&i._&j    &a0wv_mx
        wl&i._&j*wl&i._&j       &a0wl2;
    …....................................
run; quit;
```

Table 1 Significant effects of natural factors on the residuals of logarithmic TP (LTP)

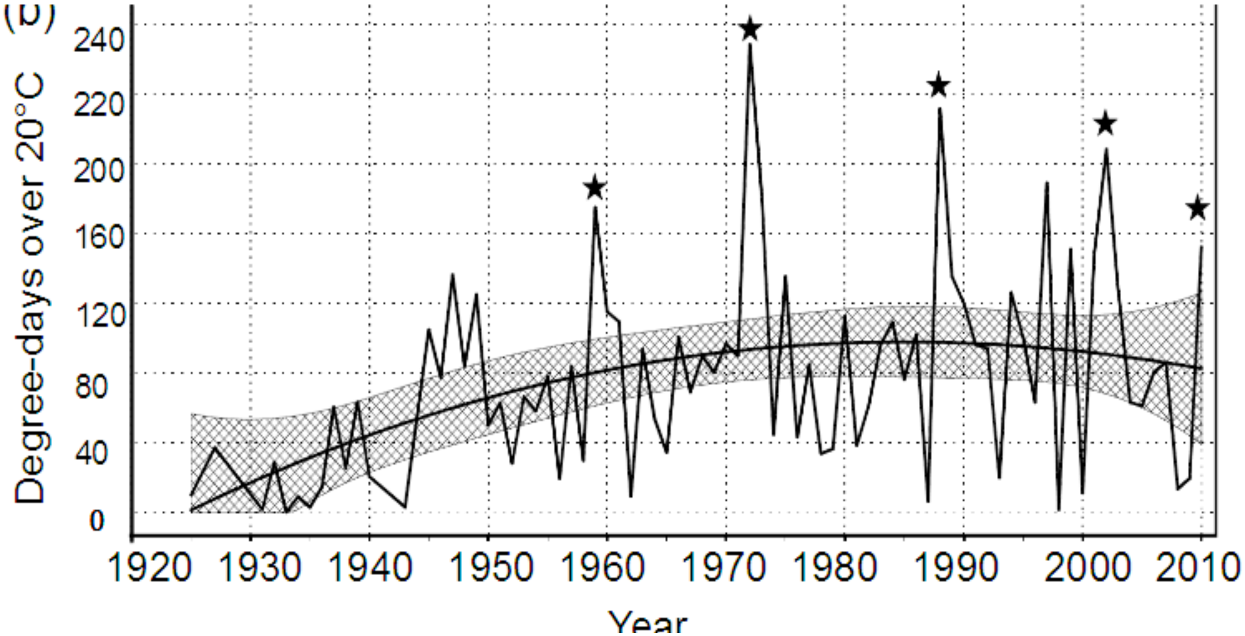
Natural factor	Unit	Effect days	Effects count	Mean effect	Min effect	Max effect
All the days (1175 observations)						
Water temperature	C	4 - 8	4	1.3%	1.0%	1.6%
Water level	dm	0 - 12	46	-1.4%	-1.4%	-2.1%
Photosynthetically active radiation	MJ m^{-2}	0 - 9	31	-2%	-1.3%	-2.8%
Wind velocity (mean)	m s^{-1}	1 - 12	8	-24%	-87%	32%
Wind velocity (max)	m s^{-1}	3 - 6	4	-9%	-23%	7%
Days before the 180th Julian day (479 observations)						
Water temperature	C	0 - 8	19	-2.8%	-2.5%	-5.5%
Water level	dm	0 - 10	16	-2.9%	-2.1%	-3.5%
Photosynthetically active radiation	MJ m^{-2}	0 - 12	6	0.5%	-7.3%	4.8%
Wind velocity (mean)	m s^{-1}	2 - 12	10	-47%	-646%	114.4 %
Wind velocity (max)	m s^{-1}	-	-	-	-	-
Days after the 180th Julian day (703 observations)						
Water temperature	C	0 - 8	8	2.5%	2%	3%
Water level	dm	2 - 5	4	-1.4%	-1.4%	-1.4%
Photosynthetically active radiation	MJ m^{-2}	0 - 9	25	-3.6%	-2.2%	-3.6%
Wind velocity (mean)	m s^{-1}	7 - 7	1	31%	31%	
Wind velocity (max)	m s^{-1}	1 - 5	3	-22%	-9%	30.9%

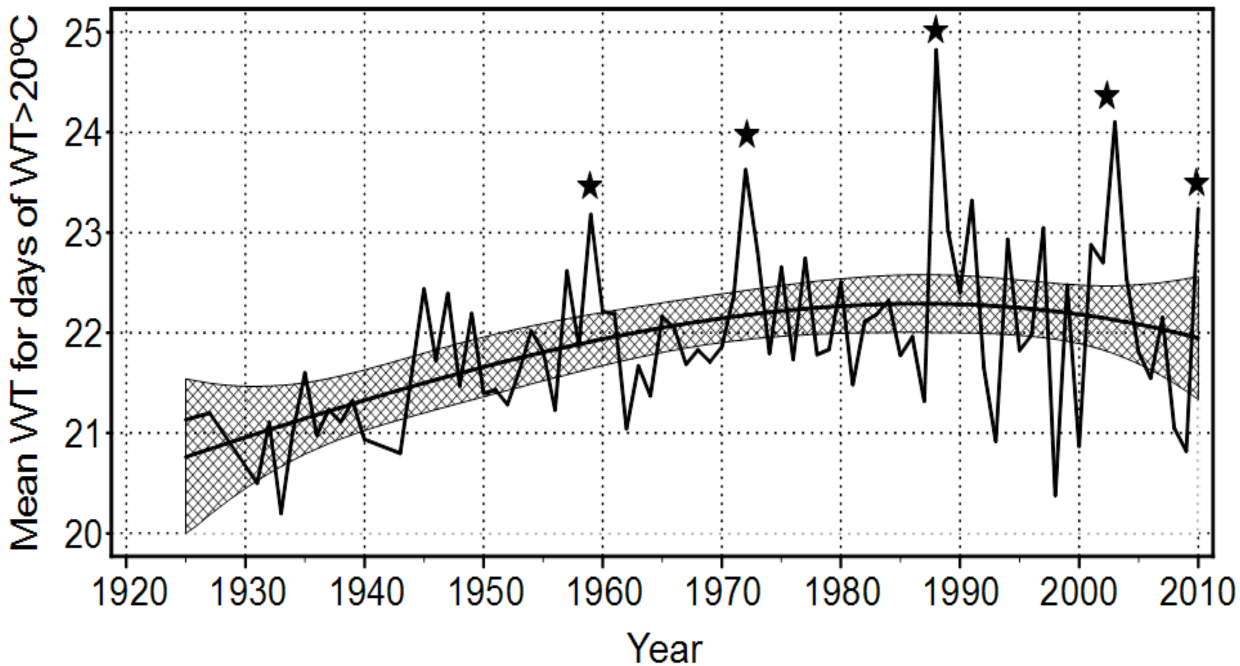
Kangur, **K**., **Kangur**, **A**., **Kangur**, **P**., **Ginter**, **K**., **Orru K**., **Haldna**, **M**., **Möls**, **T**. Long-term effects of extreme weather events and eutrophication on the fish community of shallow Lake Peipsi (Estonia/Russia) [rejected by *Fisheries Management and Ecology*]

The original version was split into **three separate papers**

1) One presented, but not related with bioinformatics chair

2) Two papers are in course of preparation


(b)
Degree-days over 20°C
0
40
80
120
160
200
220
240
1920
1930
1940
1950
1960
1970
1980
1990
2000
2010
Year


Mean WT for days of WT>20°C
20
21
22
23
24
25
1920
1930
1940
1950
1960
1970
1980
1990
2000
2010
Year

Moodustame uue uuritava tunnuse Dec3

Arvutame iga kombinatsiooni <aasta× traalimine×kalaliik> kohta suhte

Liigi osakaal antud traalimisel saadud kalade summaarses kaalus

ja analoogilise suhte, kuid arvutatud eelmise aasta kõigi traalimiste andmetest:

Liigi osakaal kõigil eelmise aasta traalimistel saadud kalade summaarses kaalus

Dec3 on nende kahe suhte jagatise logaritm:

$$\textbf{Dec3} = log_2 \frac{\textit{Liigi osakaal antud traalimisel saadud kalade summaarses kaalus}}{\textit{Liigi osakaal eelmise aasta traalimistel saadud kalade kaalus}}$$

Dec3 mõõdab liigi osakaalu muutust kahel järjestikkusel aastal (kaalu absoluutseid muute ei saa võrrelda, sest traalimise intensiivsus eri aastatel võib erineda).

/Dec3 on nn. kolmandat tüüpi dekrement/

Andmed traalimiste kaupa (trawl data)

M	Y	catch	Smelt	Vendace	Whitefish	Pike	Eel	Roach	jne	Tot
9	1994	1	0.5784	0.00000	0.00000	13.7137	0	2.471		150.228
9	1994	2	0.3039	0.00000	0.00000	16.0398	0	11.133		193.015
9	1994	3	0.4216	0.00000	0.00000	20.4347	0	7.420		154.672
9	1994	4	1.1373	0.00000	0.80392	21.0660	0	8.249		189.064
9	1994	5	0.5882	0.00000	1.01961	11.5097	0	9.312		140.850
9	1994	6	0.5294	0.00000	0.00000	19.2156	0	8.920		180.709
9	1995	1	1.8376	0.00000	0.74510	3.9300	0	38.382		405.388
9	1995	2	2.4235	0.02745	1.01961	1.6679	0	26.667		211.200
10	1996	1	15.6863	0.00000	0.00000	2.3897	0	0.000		270.031
11	1996	1	5.1765	0.00000	0.00000	8.6339	0	127.741		631.647
10	1996	2	15.6863	0.00000	0.00000	3.5525	0	84.451		501.170
11	1996	2	11.7647	0.00000	0.00000	25.3219	0	105.914		483.747
10	1996	3	16.8235	0.00000	0.00000	1.6471	0	39.529		282.516
11	1996	3	15.2941	0.00000	0.00000	26.0470	0	94.667		528.233

Andmed aastate kaupa (trawl data, summarised by years)

Y	Smelt	Vendace	Whitefish	Pike	Roach	Ruffe	Tott
1994	3.559	0.00000	1.82353	101.979	47.506	41.105	1008.54
1995	4.261	0.02745	1.76471	5.598	65.049	18.286	616.59
1996	80.431	0.00000	0.00000	67.592	452.302	45.270	2697.34

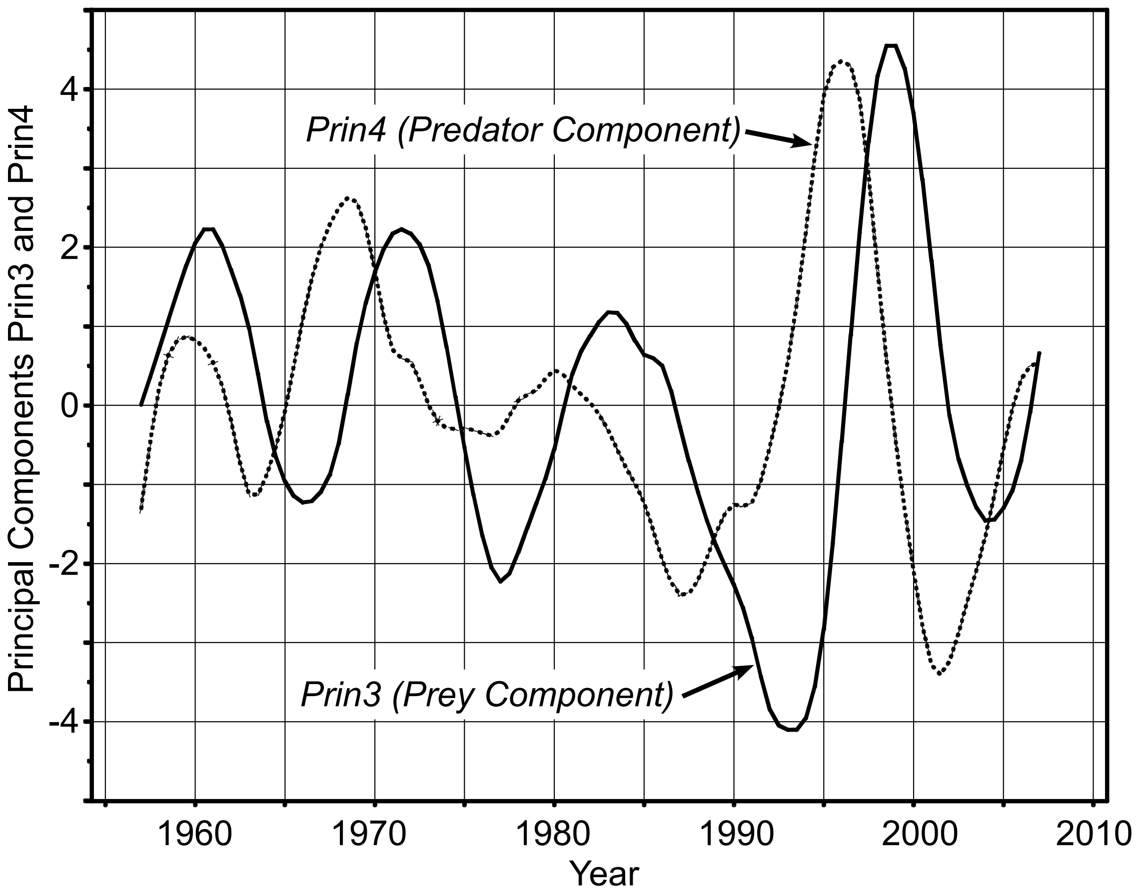
(Särg = Roach)

Lugeja (numerator) = 26.667/211.200 = 0.126264

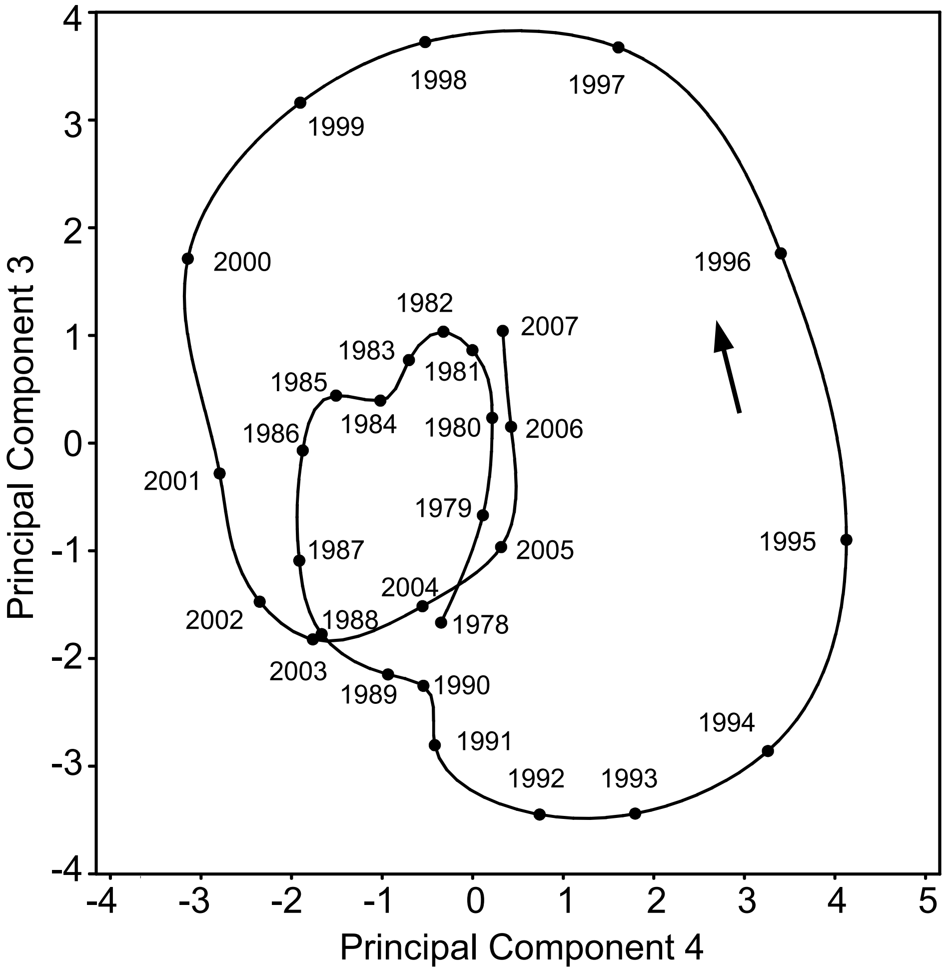
Nimetaja (denominator) = 47.506/1008.54 = 0.04710

Dec3 = $\log_2$(lugeja / nimetaja) = $\log_2(0.2815)$ = **1.422644**

Positiivne Dec3 näitab, et *roach* osakaal oli 1995. aasta 2. traalimisel suurem kui 1994. a. traalimiste andmetel keskmiselt.



It is not clear, which of the principal components is the “predator component”



Phase portrait of two principal components based on fish catches (different species)

Aitäh!