Statistical puzzles in ecological studies

(Context: writing new papers)

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)

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]

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 23b, 51010 Tartu, Estonia

6

Presented by T. Möls at the Bioinformatics Journal Club, 21th May 2012

History

Published

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)

Accepted

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*]

New mss

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

Consider first the rejected (by *Hydrobiologia*) paper: Natural variables as factors behind phosphorus variability in the shallow eutrophic Lake Peipsi (Estonia/Russia)

Olga Buhvestova[1], Tõnu Möls[1,2], Külli Kangur[1]*

The main problem raised in the manuscript:

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

Constant factors (values beforehand known):

The 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)*
(Sampling) depth *(depending on the lake depht in the sampling site)*

Natural factors (their values depend on the measurement time) :

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

Critical questions raised by the Editor and Referees:

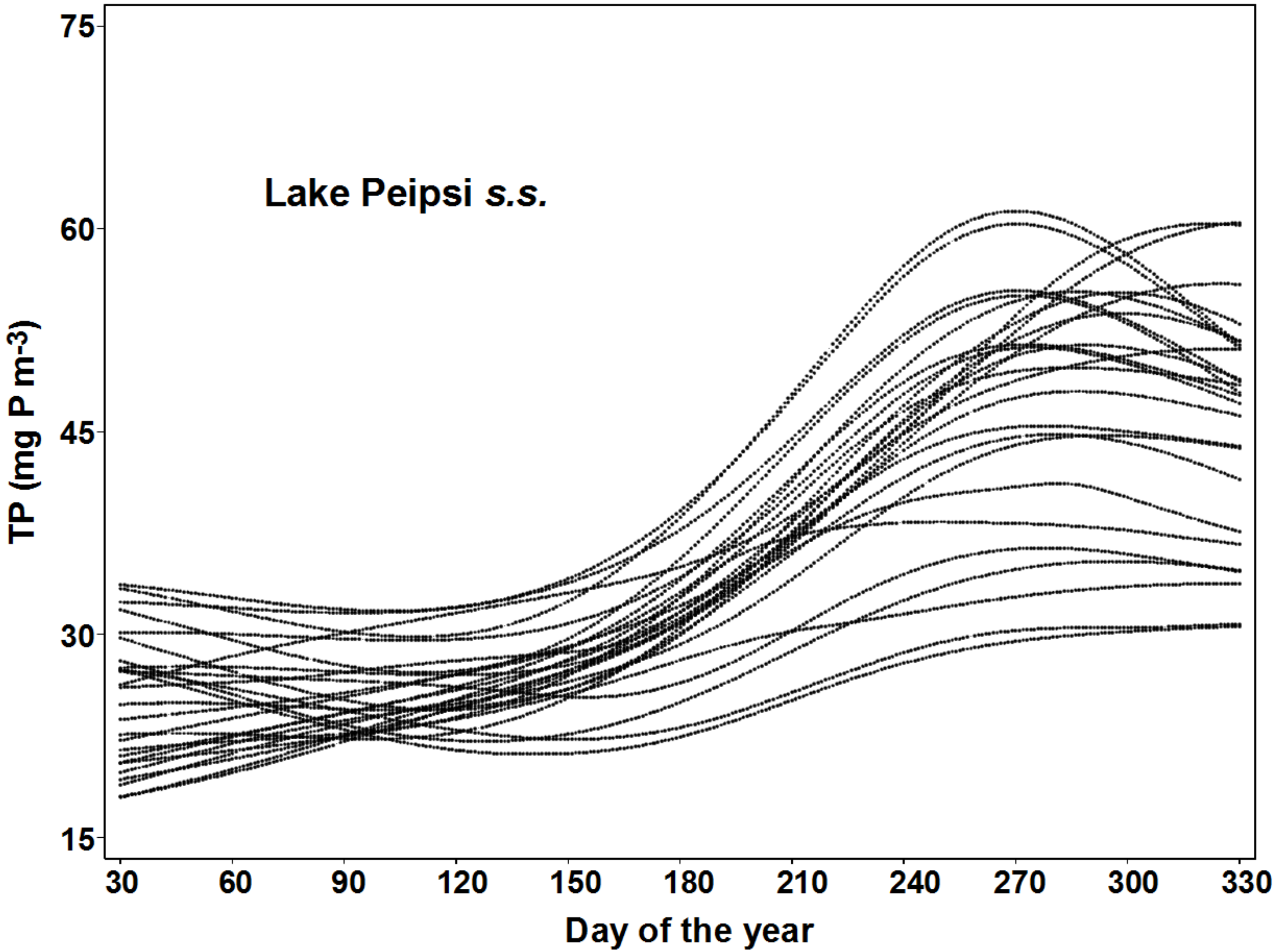
1) How would a lake manager change his or her actions based on knowledge about the role of temperature or PAR? These are not attributes that can be influenced by management actions.

This study does not address what those earlier studies have documented as a most important factor -- wind velocity -- why was this excluded. This seems a major problem.

2) The statistical methods section needs more detail. For example, the thin plate spline (TSP) method would involve choosing parameters controlling the rigidity of the fit, hence the magnitude of the residuals. This is important to this paper, but we have no idea how rigid the splines were. Reading p. 3, one might gather than all of the data were used to fit the TSP to the seasonal cycle, but it would seem from Fig. 4 that each year was fitted separately. No mention is made of the fitted spatial pattern. That 70% of the variation in TP is explained by the TSP is not interpretable without these issues being clearer.

The authors purport that the TSP fit removes the influence of space and time on TP, and look at the residuals with respect to factors that are themselves cyclical in nature (water level, light, and water temperature) using multiple regression. While significant regressions were obtained with the very large amount of data, the additional predictive power seems very weak (20% of the 30% residual from the TSP) and most of the significant predictors are lagged interaction terms. The approach seems inappropriate to me. The effects of WL, PAR, and WT have already been largely captured by fitting a seasonal cycle to the data. The reader cannot really judge the extent to which the residuals are variation that might be related to weather and other "natural factors" versus the fit of the TSP. Therefore the efforts to examine the residuals do not add much to the paper. That is, they do not tell us what causes the year to year variation.

A much more promising approach might be to look at indications of how the individual years depart from the general pattern, like height of the lake summer maximum, and relate these to the various independent variables. If this could be done, then paper much more informative and interesting.



Lake Peipsi *s.s.*
TP (mg P m-3)
75
60
45
30
15
30
60
90
120
150
180
210
240
270
300
330
Day of the year

```
proc TPspline data = Peipsitihe_Ptotst;
    /* Teeme jääkide faili 'PREDY_REALDAT', kus jäägid R_LPtot
        on arvutatud konstantseid faktoreid kasutades*/
    model LPtot =  (a t pl ip) / M=2 DF=150 ;
              /*  a, t, pl, ip  lähendava TP-splaini suhtes */
    output  out=predy_Realdat   pred residual;
run;
```

```
%macro SDmaking;
    proc TPspline data = Ptot;
        model LPtot = (a t pl ip) / d=0.1;
        score data = fikt out=SD;
        output  out=predicted_Realdata  pred residual;
    run;
%mend SDmaking;
```

Parametrization and shifting-averaging of natural factors

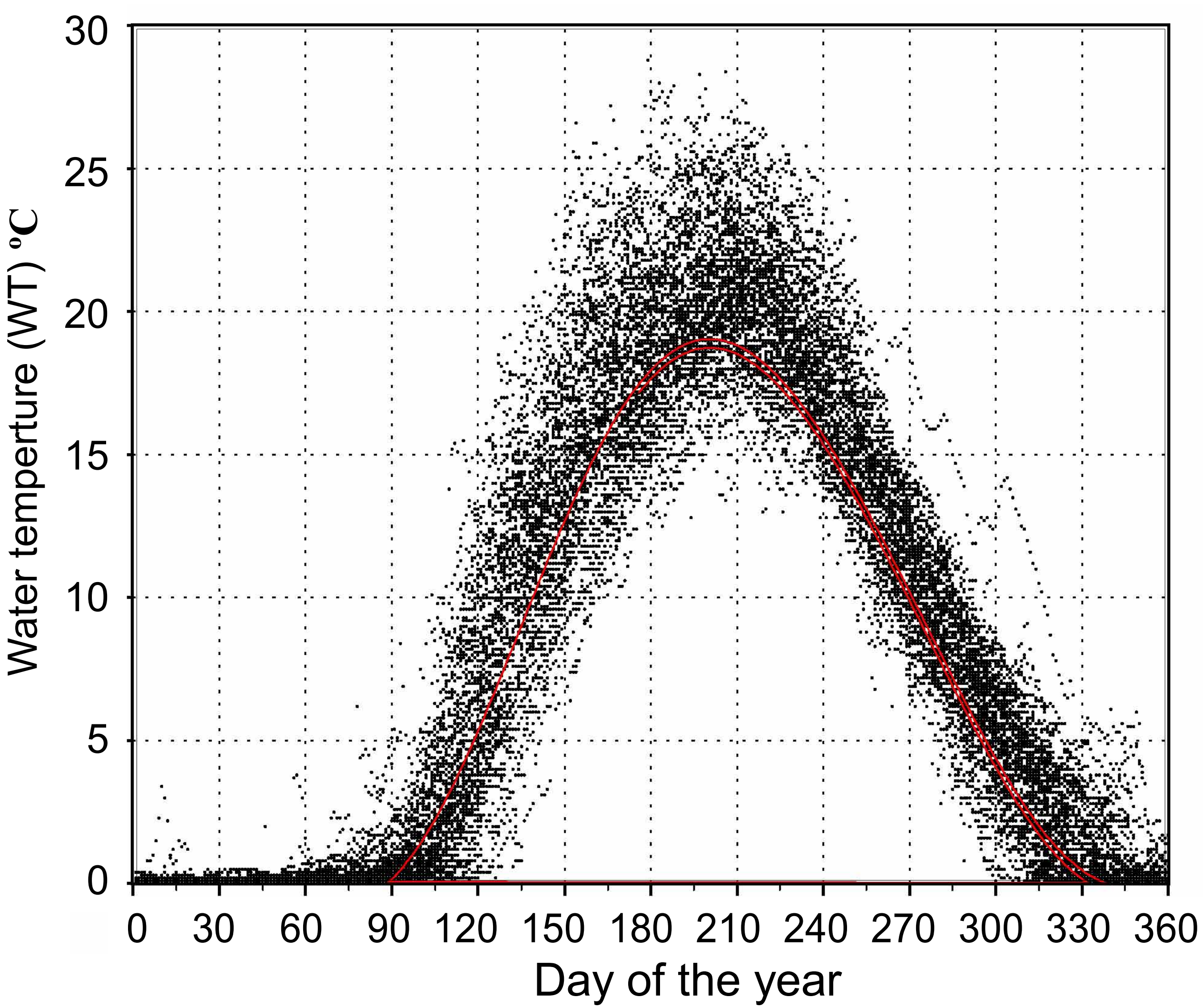
Consider, for example, the WT. And the question: how depends TP on WT? To carry out the analysis, we have to parametrise WT curves with numeric Parameters. For different years the parameters have different values.

In addition, it is possible, that the value of the dependent variable in a year depends on WT parameters in previous years. Therefore, we have introduced specific notations. For example, a parameter, say U, averaged over five consecutive years 7 days earlier, is denoted by U_{5_7}.

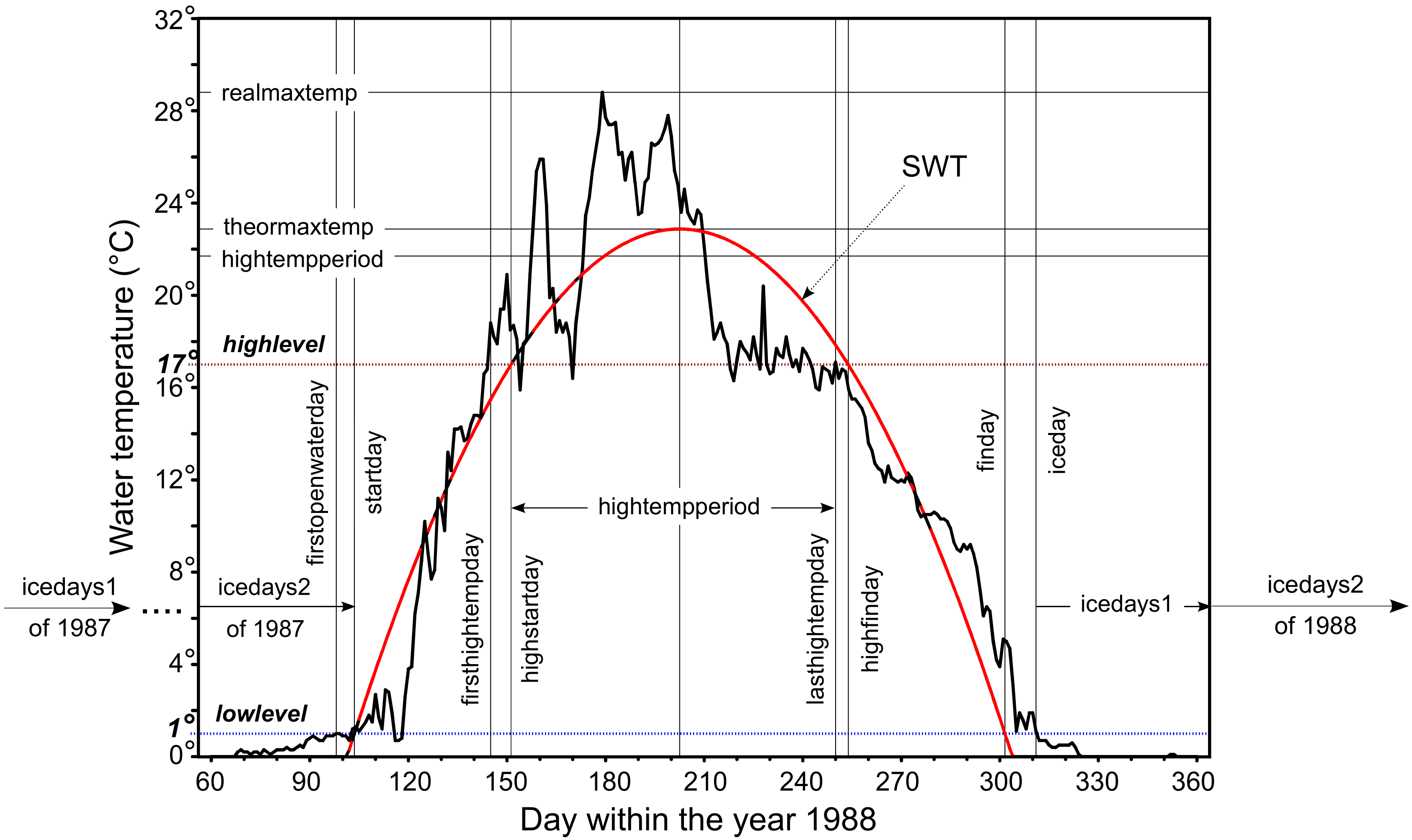
In that way we might have for a parameter 40 or even more different modifications. The number of WT parameter modifications over 30, so we may have for WT 1200 modified parameters.

Search for parameters modifications, which are influential, is carried out with the stepwise regression analysis.

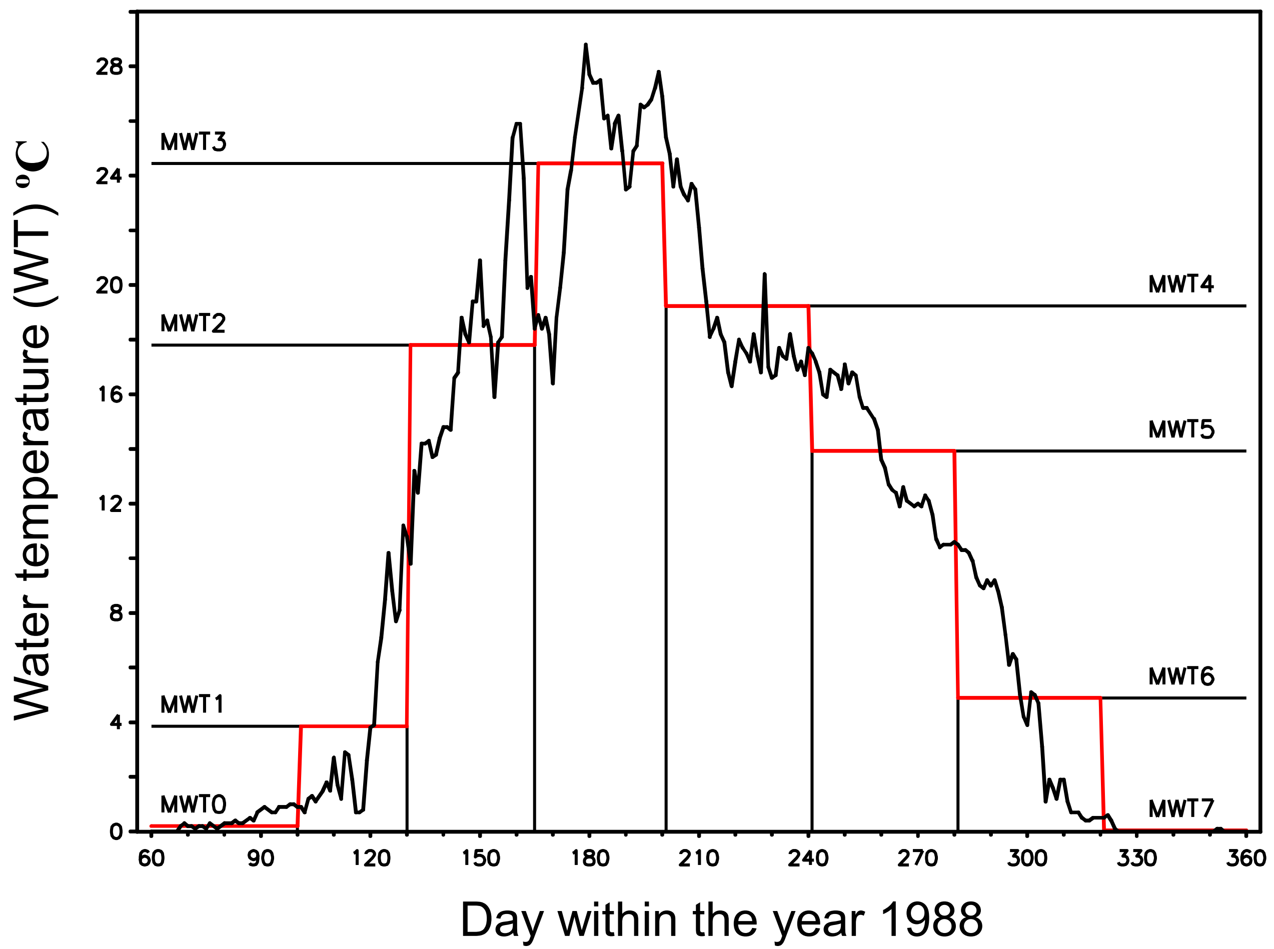
Water temperature (WT) in Lake Peipsi



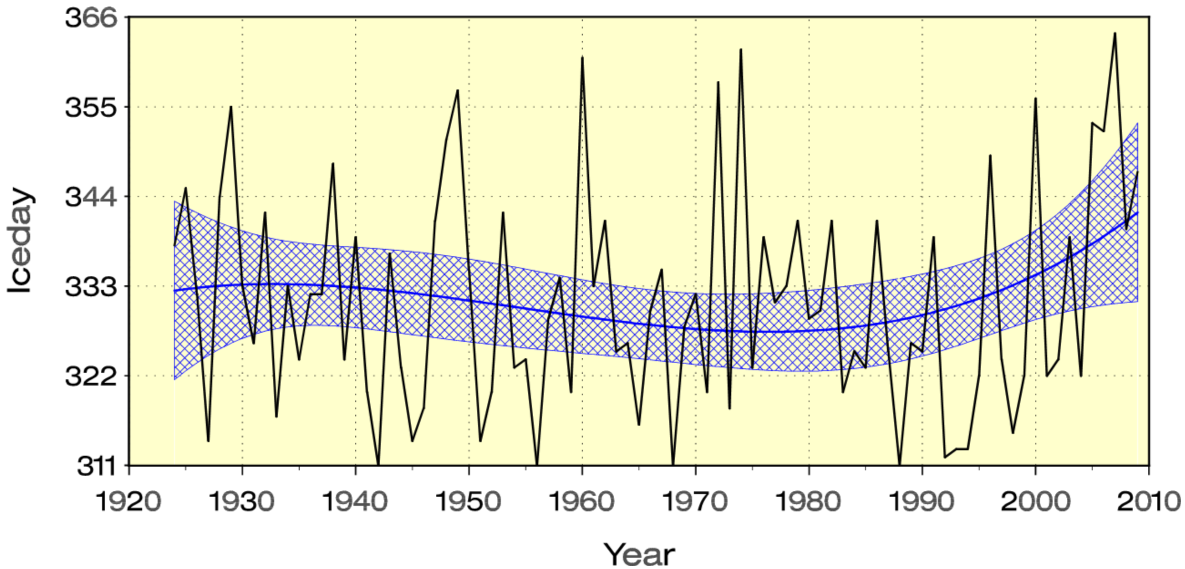
WT patterns in 1924 – 2009 (Mustvee Hydro Station, Lake Peipsi) Red lines are predicted for the war years (1041 and 1944).



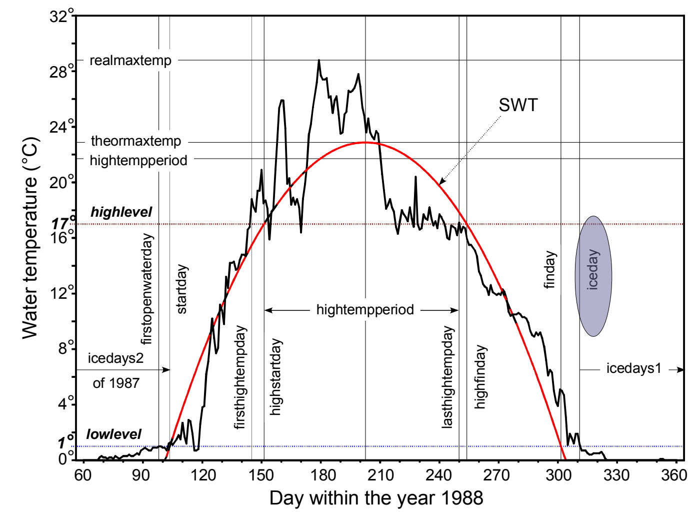
Dynamic parameterisation of WT curve for a year. For each year, 21 non-trivial dynamic parameters were defined.



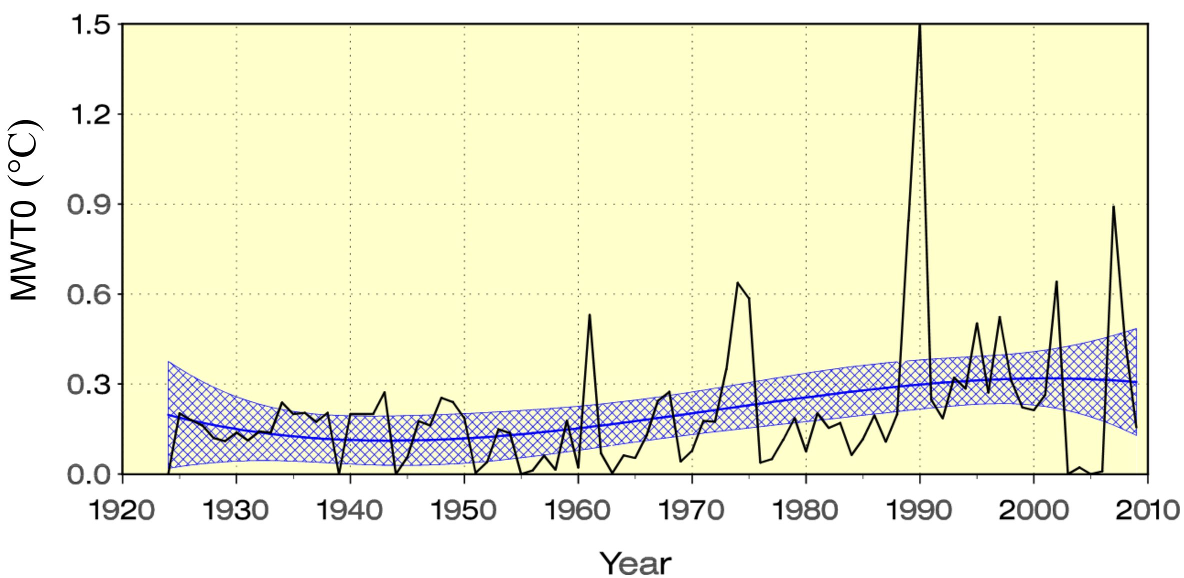
Static parameterisation of WT curve for a year. For each year, 13 non-trivial static parameters were defined. For example, MWT5 means the average WT for the fifth time period in the year.



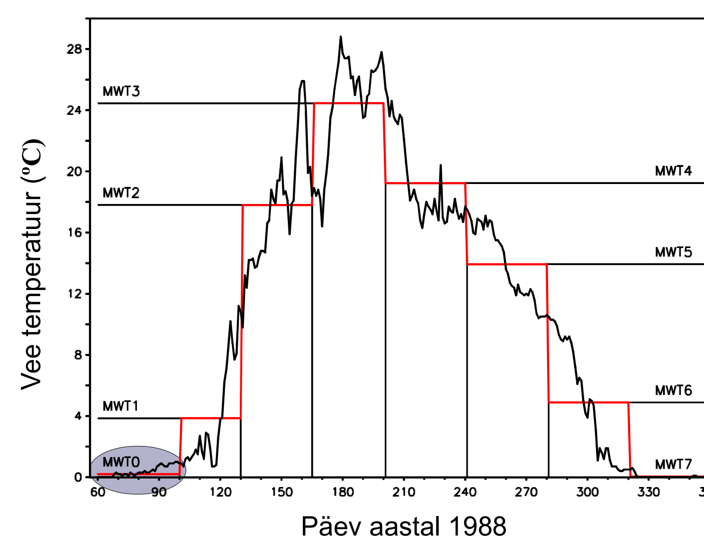
Iceday
366
355
344
333
322
311
1920
1930
1940
1950
1960
1970
1980
1990
2000
2010
Year



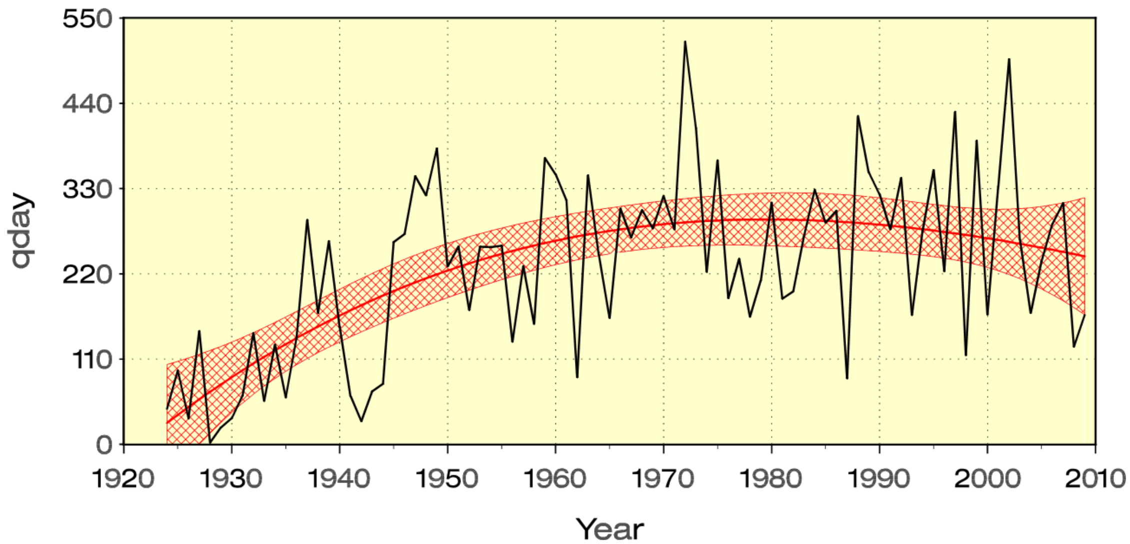
Water temperature (°C)
realmaxtemp
theormaxtemp
hightempperiod
highlevel
SWT
firstopenwaterday
startday
firsthightempday
highstartday
hightempperiod
lasthightempday
highfinday
finday
iceday
icedays2
of 1987
icedays1
lowlevel
Day within the year 1988

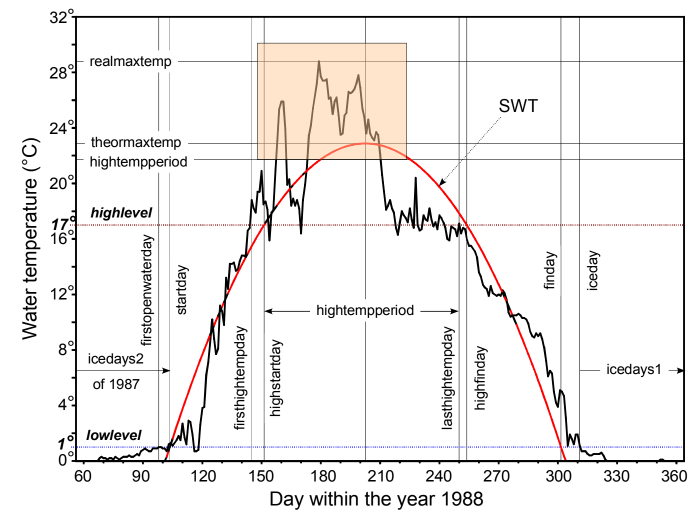


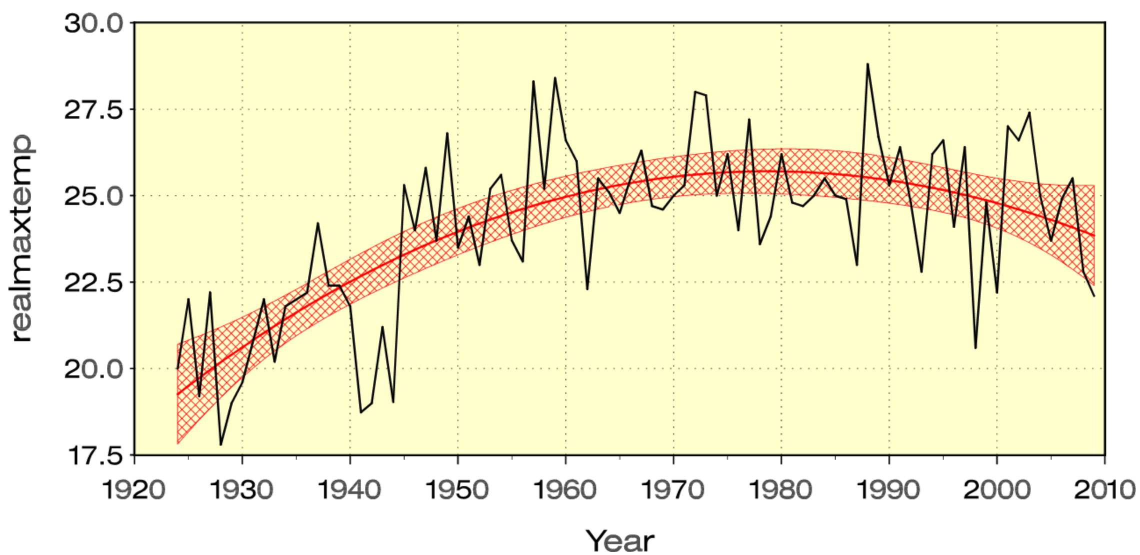
MWT0 (°C)
1.5
1.2
0.9
0.6
0.3
0.0
1920
1930
1940
1950
1960
1970
1980
1990
2000
2010
Year

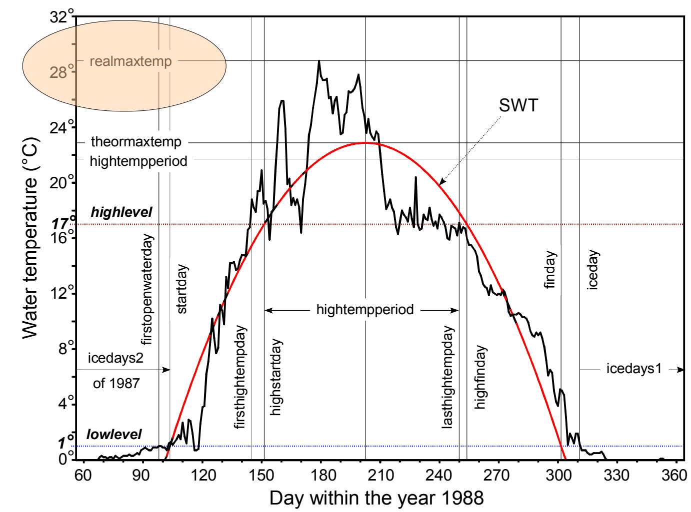


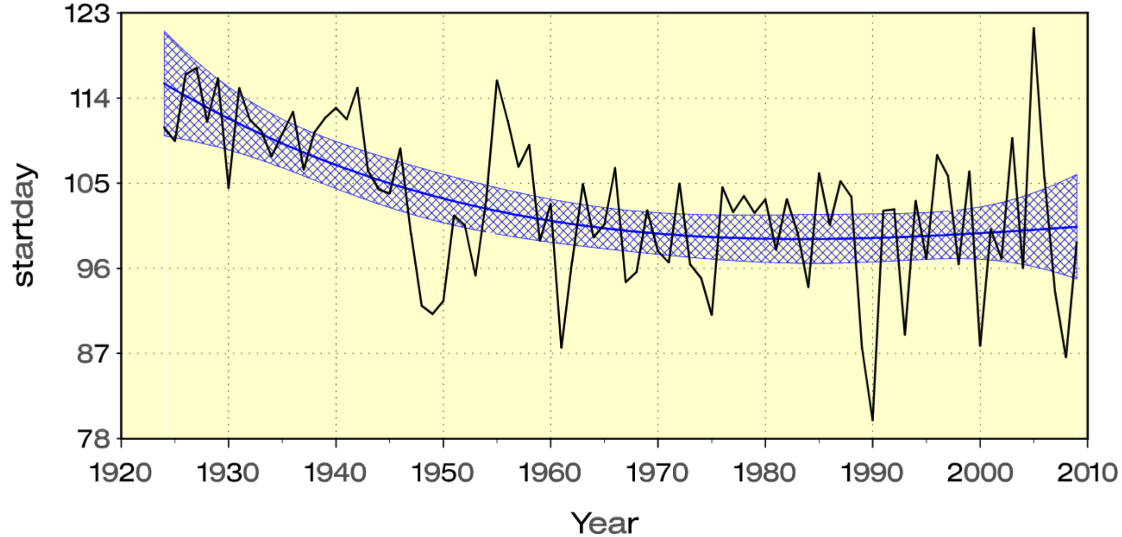
Vee temperatuur (°C)
MWT3
MWT2
MWT1
MWT0
MWT4
MWT5
MWT6
MWT7
Päev aastal 1988

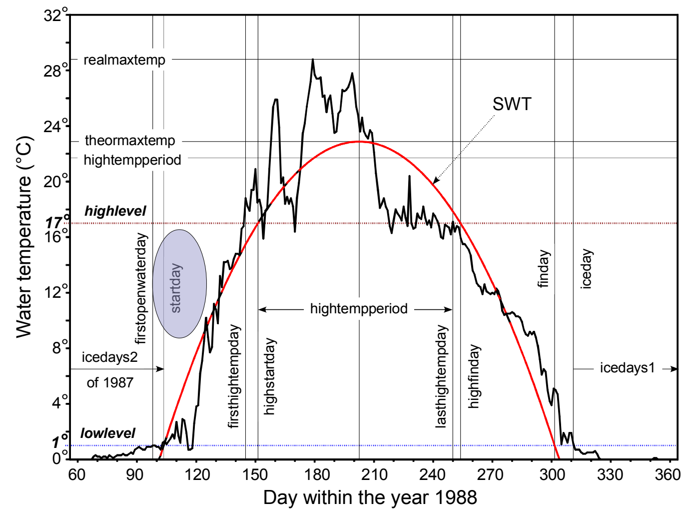

qday
550
440
330
220
110
0
1920 1930 1940 1950 1960 1970 1980 1990 2000 2010
Year

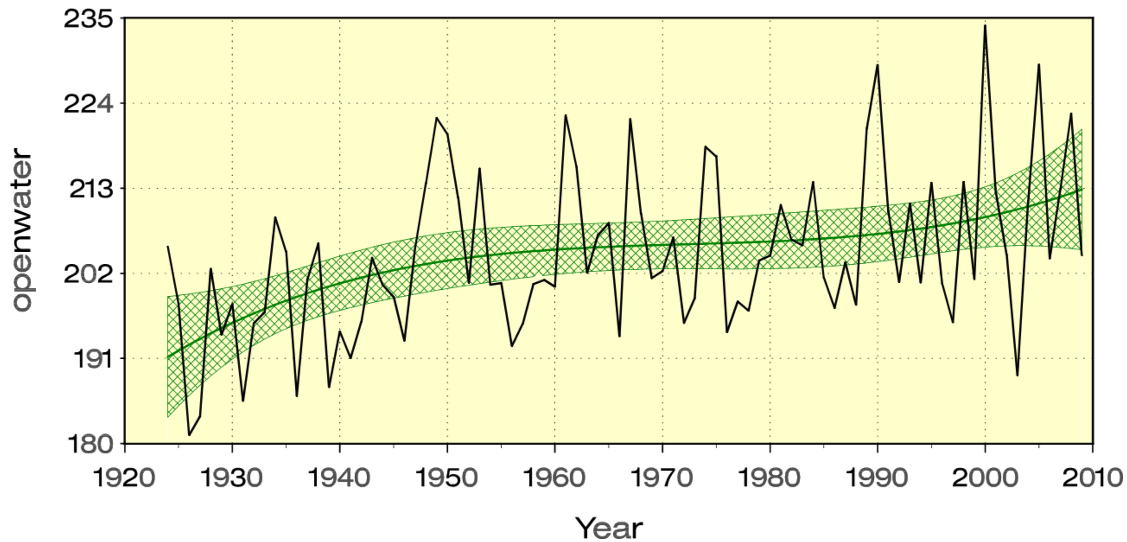

Water temperature (°C)
realmaxtemp
theormaxtemp
hightempperiod
highlevel
lowlevel
SWT
firstopenwaterday
startday
firsthightempday
highstartday
hightempperiod
lasthightempday
highfinday
finday
iceday
icedays2
of 1987
icedays1
Day within the year 1988

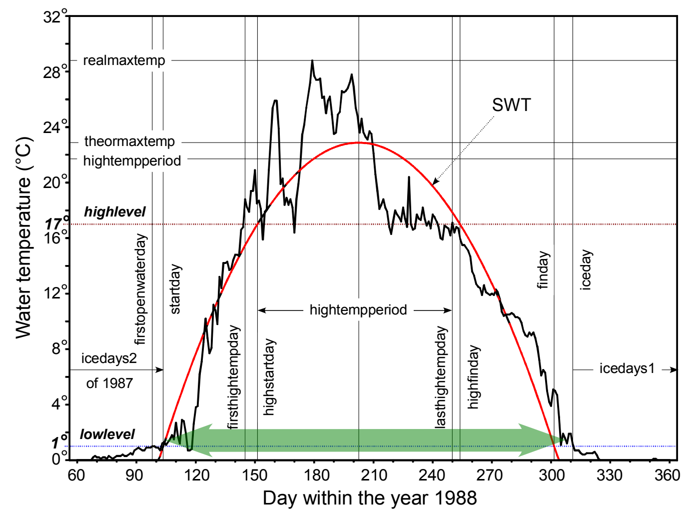

realmaxtemp
30.0
27.5
25.0
22.5
20.0
17.5
1920 1930 1940 1950 1960 1970 1980 1990 2000 2010
Year


Water temperature (°C)
realmaxtemp
theormaxtemp
hightempperiod
highlevel
lowlevel
SWT
firstopenwaterday
startday
firsthightempday
highstartday
hightempperiod
lasthightempday
highfinday
finday
iceday
icedays2
of 1987
icedays1
Day within the year 1988


startday
Year


Water temperature (°C)
Day within the year 1988
realmaxtemp
theormaxtemp
hightempperiod
highlevel
lowlevel
SWT
firstopenwaterday
startday
firsthightempday
highstartday
lasthightempday
highfinday
finday
iceday
icedays2
of 1987
icedays1


openwater
Year


Water temperature (°C)
Day within the year 1988
realmaxtemp
theormaxtemp
hightempperiod
highlevel
lowlevel
SWT
firstopenwaterday
startday
firsthightempday
highstartday
lasthightempday
highfinday
finday
iceday
icedays2
of 1987
icedays1

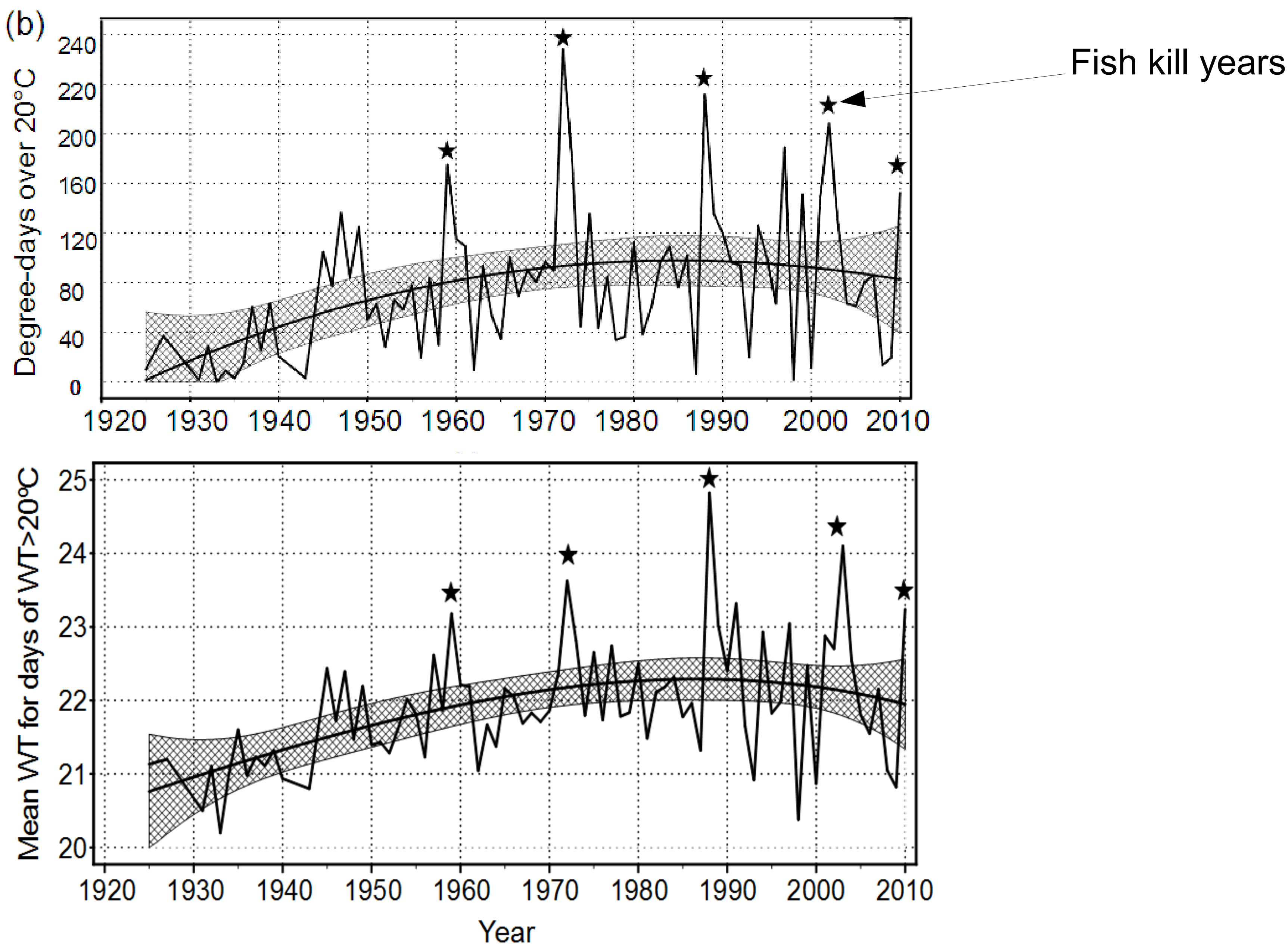
The manuscript

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)

Was initiated by comparison of WT parameters with the days of massive Fish kills. For authors, the expressive graphs seemed to be sufficient for the evidence of causal relation between warm waters and the fish kills.

But the Referees remained skeptical. Additional statistical analysis based on industrial catches did not confirm our vision. Limited experimental trawling data seems to verify the effect of extreme weather conditions.

Strong cyanobacterial bloom, high water temperature and progressive water level reduction led to fish-kill



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

. The mean abundance of fish species (CPUE – individuals per trawl hour ± standard error) and their proportion (%) of the total catch according to the autumn trawls (September and October) in Lake Peipsi in 2009 and 2010.

Fish species	2009		2010	
	CPUE	%	CPUE	%
Ruffe	3078.3±770.8	64.1	532.0±152.2	18.0
Eurasian perch	684.9±217.4	14.2	1201.4±147.1	40.7
Pikeperch	965.9±510.9	20.1	689.5±197.9	23.4
Roach	29.0±15.6	0.6	45.6±9.5	1.6
Common bream	39.1±12.7	0.8	468.7±85.0	15.9
Lake smelt	2.6±1.3	0.1	1.2±1.2	0.0
Northern pike*	2.2±0.7	0.0	6.2±1.3	0.2
Burbot	2.7±1.5	0.1	5.2±1.7	0.2
Peipsi whitefish	1.3±1.0	0.0	0.4±0.4	0.0
Total catch	4802.0±986.5	100	2950.7±304.7	100

* *Esox lucius* L.

The manuscript

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].

This paper was initiated by O. Buhvestova on the example of Finnish study group, who is studying sedimentation and resuspension problems. O. Buhvestova manages to get special traps for collectting sediment regularly during year. The collected experimental data enable to clarify, how the wind speed influences TP in the Lake Peipsi. The result help to get a much more clear picture of the factors, determining TP level in Lake Peipsi.

Working with residuals

```
proc GLM data=I_II_III;      /* Elimination of station and day caused biases */
      class station;
      model dependent = station day|day|day /ss3 solution;
      output out=Resres r=res;        /* Resres consists all data and residuals */
run; quit;
```

Observation station
Cubic polynomial of day

Mean wind speed
Maximum wind speed
Water temperature
Water level

Residuals of:

Sedimentation
Resuspension
TP_w
FMB_w
Fs_w
SRP_w
CY_w
Chl_w
SS_w
Ft_w
Fr_sed
SRP_sed_pore_w

Dependent variable / sls	Fact.in	Fact.out	Fact	Part.R^2	Model R^2	F	P
dep: resuspension / sls: 0.001	TK1_5		1	0,6028	0,6028	227,6122	7,07E-032
dep: resuspension / sls: 0.001	WL1_0		2	0,0582	0,6610	25,5954	1,22E-006
dep: resuspension / sls: 0.001	TK1_0		3	0,0412	0,7022	20,5008	1,22E-005
dep: sedimentation / sls: 0.001	TK1_5		1	0,6553	0,6553	285,1465	1,63E-036
dep: sedimentation / sls: 0.001	WL1_0		2	0,0681	0,7234	36,7052	1,07E-008
dep: sedimentation / sls: 0.001	WL2_0		3	0,0350	0,7584	21,4428	7,92E-006
dep: fs / sls: 0.001	wt4_5		1	0,2122	0,2122	40,3973	2,36E-009
dep: fs / sls: 0.001	TM1_1		2	0,0658	0,2779	13,5695	3,21E-004
dep: TP_w / sls: 0.001	WL6_5		1	0,6145	0,6145	239,1178	7,38E-033
dep: TP_w / sls: 0.001	TM2_2		2	0,0773	0,6918	37,3847	8,12E-009
dep: TP_w / sls: 0.001	TK8_5		3	0,0398	0,7316	21,9426	6,31E-006
dep: SRP_w / sls: 0.001	TM1_2		1	0,3657	0,3657	86,4744	1,58E-016
dep: SRP_w / sls: 0.001	TK3_0		2	0,0792	0,4449	21,2527	8,59E-006
dep: SRP_w / sls: 0.001	TM4_5		3	0,0614	0,5062	18,3902	3,23E-005
dep: FBM_w / sls: 0.001	wt7_5		1	0,2897	0,2897	61,1853	8,56E-013
dep: FBM_w / sls: 0.001	WL1_3		2	0,2629	0,5526	87,5665	1,17E-016
dep: FBM_w / sls: 0.001	TK3_2		3	0,1192	0,6718	53,7426	1,38E-011
dep: FBM_w / sls: 0.001	wt7_4		4	0,0728	0,7446	41,9034	1,35E-009
dep: FBM_w / sls: 0.001	WL5_0		5	0,0280	0,7726	17,9967	3,91E-005
dep: CY_w / sls: 0.001	TK3_2		1	0,2343	0,2343	45,8936	2,66E-010
dep: CY_w / sls: 0.001	TM1_0		2	0,3719	0,6062	140,7048	2,85E-023
dep: CY_w / sls: 0.001	TM4_5		3	0,0512	0,6573	22,1004	5,87E-006
dep: Chl_w / sls: 0.001	WL1_1		1	0,3610	0,3610	84,7271	2,78E-016
dep: Chl_w / sls: 0.001	wt5_5		2	0,1334	0,4944	39,3199	3,69E-009
dep: Chl_w / sls: 0.001	TM1_5		3	0,0834	0,5778	29,2369	2,51E-007
dep: Chl_w / sls: 0.001	TK6_2		4	0,0814	0,6592	35,1131	2,12E-008
dep: Chl_w / sls: 0.001		wt5_5	3	0,0007	0,6585	0,2932	5,89E-001
dep: Chl_w / sls: 0.001	TM8_1		4	0,0256	0,6842	11,9362	7,20E-004

'sedimentation'	0,872894	TK4_4	0,997373	WL1_1	0,99984	WL4_3
'sedimentation'	0,870058	TK8_2	0,996756	WL7_1	0,99982	WL2_4
'sedimentation'	0,867917	TK7_3	0,996742	WL8_1	0,9998	WL7_1
'sedimentation'	0,865603	TK3_4	0,99672	WL2_1	0,99978	WL7_3
'sedimentation'	0,862447	TK6_4	0,996708	WL3_1	0,99973	WL2_5
'sedimentation'	0,857268	TK5_5	0,996587	WL6_1	0,99972	WL5_2
'sedimentation'	0,851393	TK8_3	0,996497	WL5_3	0,9997	WL8_3
'sedimentation'	0,849261	TK6_2	0,996397	WL6_3	0,99968	WL6_1
'sedimentation'	0,838706	TK5_3	0,996353	WL5_1	0,99967	WL5_4
'sedimentation'	0,834454	TK7_4	0,996314	WL4_1	0,99965	WL8_0
'sedimentation'	0,834337	TK8_4	0,996294	WL7_3	0,99962	WL3_5
'sedimentation'	0,80684	TK2_4	0,996259	WL4_3	0,99961	WL1_5
'sedimentation'	0,800288	TK6_5	0,996224	WL7_2	0,99958	WL8_4
'sedimentation'	0,786088	TK7_5	0,996198	WL8_2	0,99955	WL6_4
'sedimentation'	0,768238	TK4_3	0,996173	WL3_5	0,99955	WL7_0
'sedimentation'	0,757472	TK5_2	0,99617	WL6_2	0,99952	WL7_4
'sedimentation'	0,749091	TK8_5	0,99612	WL4_4	0,99948	WL8_5
'sedimentation'	0,699306	TK7_1	0,995933	WL2_5	0,99945	WL3_3
'sedimentation'	0,642043	TK3_3	0,995866	WL5_4	0,99936	WL5_1
'sedimentation'	0,570299	TK6_1	0,995845	WL3_3	0,99935	WL7_5
'sedimentation'	0,559354	TK4_2	0,995838	WL5_2	0,99932	WL4_2
'sedimentation'	0,547806	TK8_0	0,995824	WL3_4	0,99927	WL6_0
'sedimentation'	0,488097	TK1_4	0,995779	WL6_4	0,9992	WL6_5
'sedimentation'	0,394742	TK7_0	0,995641	WL4_5	0,99916	WL4_5
'sedimentation'	0,32	TK5_1	0,995621	WL1_3	0,99911	WL5_5
'sedimentation'	0,215734	TK2_3	0,995572	WL5_5	0,99905	WL4_1
'sedimentation'	0,136934	TK6_0	0,995505	WL8_3	0,99899	WL5_0
'sedimentation'	-0,10018	TK3_2	0,995493	WL1_5	0,99891	WL1_4

Averaged Days	Shift					
	0	1	2	3	4	5
1	TK1_0					
2	TK2_0	TK1_1				
3	TK3_0	TK2_1	TK1_2			
4	TK4_0	TK3_1	TK2_2	TK1_3		
5	TK5_0	TK4_1	TK3_2	TK2_3	TK1_4	
6	TK6_0	TK5_1	TK4_2	TK3_3	TK2_4=0.80684	TK1_5=1.0000
7	TK7_0	TK6_1	TK5_2=0.75747	TK4_3=0.76824	TK3_4=0.86560	TK2_5=0.92312
8	TK8_0	TK7_1	TK6_2=0.84926	TK5_3=0.83871	TK4_4=0.87289	TK3_5=0.93484
9		TK8_1=0.90344	TK7_2=0.88856	TK6_3=0.88563	TK5_4=0.89237	TK4_5=0.89882
10			TK8_2=0.87006	TK7_3=0.86792	TK6_4=0.86245	TK5_5=0.85727
11				TK8_3=0.85139	TK7_4=0.83445	TK6_5=0.80029
12					TK8_4=0.83434	TK7_5=0.78609
13						TK8_5

Tänan kuulamast !