VEÐURSTOFA
ÍSLANDS

Report
03009

Steen Henriksen

Report on the approximation of the annual cycle of temperature in Iceland

VÍ-ÚR09
Reykjavík
February 2003

1 The problem

We have twelve monthly mean temperatures, and want to make a function that approximates the annual temperature cycle. More formally the problem can be stated as:

Given twelve monthly mean temperatures $T_{Jan}, T_{Feb}, \ldots, T_{Dec}$ we want to find a smooth periodic curve $f:[1,365]\to \mathbb{R}$ that for each month fulfils

$$T_{I,month} := \frac{\int_{d_1}^{d_2} f(x)dx}{d_2 - d_1 + 1} = T_{month} \qquad (1)$$

where d_1 and d_2 are the first and last day of $month$ being Jan, Feb,..., Dec. $T_{I,month}$ is referred to as the monthly mean temperatures obtained from integrating f. In order to be acceptable the function f must furthermore resemble the annual temperature cycle obtained from averaging daily data of the manual stations in Iceland (with a running period from 21 to 30 years in the period 1961 to 1990). We remark that a year is assumed to have 365 days.

2 Tension splines

2.1 Introduction to tension splines

A tension spline is a spline curve that contains a parameter that determines the tension of the curve. That is the curve passing through the data constraint points can have different tension and thereby be more or less rigid. If the tension is small (near 0) the tension spline acts as a cubic spline. If the tension is large (going to infinity), then the tension spline approaches the piecewise linear function. A tension spline with tension larger than 0 is not like the cubic spline a polynomial.

2.2 The mathematics behind tension splines

In this section we will describe the theory of tension splines in outline. We are only going to work with 12 data constraint points (one pr. month) and we have periodic boundary conditions (December goes smoothly into January). For detailed and more general description, see **[**1**]**-**[**4**]**.

The mathematical model of the tension spline curve is described as follows. Given 13 knots

$$d_1 < d_2 < \cdots < d_{13}$$

and data y_i at each d_i. For the last knot d_{13} the corresponding data value $y_{13} = y_1$, this supplies periodicity. The tension spline that we seek is a function f having the following properties

(i) $f \in C^2_{per}[d_1, d_{13}]$

(ii) $f(d_i) = y_i \quad (1 \le i \le 13)$

(iii) On each open interval (d_i, d_{i+1}), f satisfies $f^{(4)} - \tau f'' = 0$

The properties (i) to (iii) mean that, f has two continuous derivatives globally, it interpolates the given data, and it satisfies a certain differential equation in each subinterval. It is clear that this prescription yields a cubic spline when $\tau = 0$ because the solutions of the equation $f^{(4)} = 0$ are cubic polynomials.

To determine f, we set $z_i := f''(d_i)$ and write down the conditions f must satisfy on the interval $[d_i, d_{i+1}]$

$$\begin{gathered} f^{(4)} - \tau f'' = 0 \\ f(d_i) = y_i \qquad f(d_{i+1}) = y_{i+1} \\ f''(d_i) = z_i \qquad f''(d_{i+1}) = z_{i+1} \end{gathered}$$

where $z_{13} = f''(d_{13}) = f''(d_1) = z_1$. One can verify that the solution of this two-point boundary-value problem is

$$\begin{aligned} f(x) = \{ & z_i \sinh[\tau(d_{i+1} - x)] + z_{i+1} \sinh[\tau(x - d_i)]\} / [\tau^2 \sinh(\tau h_i)] \\ & + (y_i - z_i / \tau^2)(d_{i+1} - x)/h_i + (y_{i+1} - z_{i+1}/\tau^2)(x - d_i)/h_i \end{aligned}$$

where $h_i = d_{i+1} - d_i$. After the coefficients z_i have been determined, this equation will be used to compute values of f on the interval $[d_i, d_{i+1}]$. The function f is what we will call the tension spline function.

In order that f have C^2 global smoothness, the conditions

$$\lim_{x \uparrow d_i} f'(x) = \lim_{x \downarrow d_i} f'(x) \qquad (1 \le i \le 13)$$

must be imposed at the knots. The tedious calculations involved in this are not given here. The result is a tridiagonal system of equations for the unknowns $z_1, z_2, \ldots, z_{13}$ that can be written in the form

$$\alpha_{i-1} z_{i-1} + (\beta_{i-1} + \beta_i) z_i + \alpha_i z_{i+1} = \gamma_i - \gamma_{i-1} \qquad (1 \le i \le 12) \tag{2}$$

with these abbreviations

$$\begin{aligned} \alpha_i &= 1/h_i - \tau \sinh(\tau h_i) \\ \beta_i &= \tau \cosh(\tau h_i)/\sinh(\tau h_i) - 1/h_i \\ \gamma_i &= \tau^2 (y_{i+1} - y_i)/h_i \end{aligned}$$

It is observed since (2) gives 12 equations with 13 unknowns that an additional condition is needed to determine the z-vector. This is supplied by the periodic boundary condition, defined by $y_1 = y_{13}$ and $z_1 = z_{13}$, (actually it is the case that $y_i = y_{i+12}$ and $z_i = z_{i+12}$). We now have the linear system of equations

$$\begin{bmatrix} \beta_{12}+\beta_1 & \alpha_1 & & & & \alpha_{12} \\ \alpha_1 & \beta_1+\beta_2 & \alpha_2 & & & \\ & \alpha_2 & \beta_2+\beta_3 & \alpha_3 & & \\ & & \ddots & \ddots & \ddots & \\ & & & \alpha_{10} & \beta_{10}+\beta_{11} & \alpha_{11} \\ \alpha_{12} & & & & \alpha_{11} & \beta_{11}+\beta_{12} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ \vdots \\ z_{11} \\ z_{12} \end{bmatrix} = \begin{bmatrix} \gamma_1 - \gamma_0 \\ \gamma_2 - \gamma_1 \\ \gamma_3 - \gamma_2 \\ \vdots \\ \gamma_{11} - \gamma_{10} \\ \gamma_{12} - \gamma_{11} \end{bmatrix} \tag{3}$$

where $\gamma_0 = \tau^2 (y_1 - y_{12})/h_{12}$. By solving the equations (invert the symmetric matrix in (3)) all z_i can be determined. That means the tension spline function f will be given. The constant τ is called the tension.

Note that for a given number of data points (e.g. twelve) the tension spline function is the solution to a differential equation. By a uniqueness theorem this tension spline curve is the unique tension spline curve going through the data points.

3 The solution to the problem

In order for a function f to be a solution to the problem defined in section one, it must fulfil the equation (1) and resemble the annual temperature cycles. The problem is solved (in Matlab) performing an iterative process applying tension spline functions. The tension is set to some fixed value and it remains the same for all functions during the whole process. We use a numerical method, so a small error is accepted, i.e. f is said to fulfil equation (1) if $|T_{month} - T_{I,month}|$ < 0.001°C. Meaning that the error for each month between the mean temperature calculated by integrating f and the exact measured value has to be below 0.001°C. We will now give a description of the iteration algorithm used.

First step of iteration: Given twelve monthly mean temperatures T_{month}, one chooses a day d_{month} in each month and uses the twelve data points (d_{month}, T_{month}) to make a tension spline curve f. This tension spline curve f does not for each month fulfil equation (1). There is a difference between $T_{I,month}$ the mean temperatures obtained by integrating f and T_{month} the exact measured values. We can make these differences smaller by moving each of the points (d_{month}, T_{month}) either up or down a little in the T direction and thereby getting another tension spline curve f_1. More precisely we move T_{month} the amount $T_{month} - T_{I,month}$. Call these twelve new values T^1_{month}. From the new data points (d_{month}, T^1_{month}) we then make the new tension spline function, f_1, and calculate $T^1_{I,month}$ the monthly mean temperatures obtained from integrating f_1. This finishes the first step of the iteration.

By repetition n times the iteration process is explicitly given by

$$T^n_{month} = T^{n-1}_{month} + (T_{month} - T^{n-1}_{I,month})$$

where $n \geq 1$ and $T^0_{month} = T_{month}$, $T^0_{I,month} = T_{I,month}$ and where f_n is defined as the tension spline function obtained from the twelve values T^n_{month}.

Now the algorithm above is available, we can continue our search for the solution. I am due to my own tests almost confident, that for all choices of d_{month} the iteration above converges in common mathematical sense (and then it of course also converges to an accuracy better than 0.001°C). I have not got an explicit proof, but it is the case for all the tests I have been doing. Anyway we know it converges for many values, that is a fact, my assumption is just, it does it for all. It means that regarding the convergence, we then have (enormously) many choices of where to put (d_{month}, T_{month}). The problem is now to figure out, which leads to tension spline curves that resemble the temperature cycles for the manual stations.

By “guessing” I found the best resemblance when $d_{Jan} = 27$, $d_{Feb} = 27$, $d_{Mar} = 30$, $d_{Apr} = 16,\ldots$, $d_{Dec} = 16$. So application of the iteration algorithm with these choices of d_{month} and the measured T_{month} values, gives us a tension spline function, f_n, with good resemblance. And after just 8 to 10 iterations $|T_{month} - T^n_{I,month}| < 0.001$°C holds good.

Note that it never occurs that the data points (d_{month}, T^n_{month}) are moved up or down in such a fashion, that the curve of the final function for each month does not go through a corresponding point (d, T_{month}), d being just a day in the month.

Concerning the choice of the tension, I found the results satisfactory when it is between 0.16 and 0.24; and in this interval by my opinion 0.2 gives the best results.

4 The work process

This section is a brief description of the work process leading to the final solution and the Matlab program **tspline**.

I started the project using cubic splines and by putting all the d_{month}'s in the middle of each month. I then applied the technique from the iteration above of moving the points (d_{month}, T_{month}) up or down an amount equal to the size of the monthly error $|T_{month} - T_{I,month}|$. The method minimized the errors between T_{month} and $T_{I,month}$ to 0.001°C, but the cubic splines did not resemble the annual temperature cycle from the manual stations.

The method had shown to give very good result in Norway, but a problem occurred because the behaviour of the monthly mean temperatures in Iceland is different from Norway. In Iceland the mean temperature for March often drops below or equals the mean temperature of February. So the approach (d_{month}'s in middle of months) to the problem I had chosen failed because of this.

The idea was then to develop an algorithm for moving the data points, either horizontally, vertically or a combination of these, in order to find a cubic spline curve that would both minimize the errors between T_{month} and $T_{I,month}$, and resemble the annual temperature cycles from the manual stations. I did not succeed in finding this algorithm and eventually I gave up on the cubic splines. I started to look for other alternatives. My choice fell on tension splines, because I believed their "edge" behaviour at the data points, might do better than the cubic splines which are more "round".

Finally I wrote the tension spline program **tspline** that used the technique of moving the data points up or down. It minimized the errors between T_{month} and $T_{I,month}$ and I discovered immediately, that in order to get a resemblance to the manual stations, one has to put the first three data points in the end of the months and the rest in the middle of the month. I found that the results are optimal (by my opinion) when the tension constant is set to 0.2 and $d_{Jan} = 27,\ d_{Feb} = 27,\ d_{Mar} = 30$, $d_{Apr} = 16, \dots,\ d_{Dec} = 16$. In the end it actually turned out that for these choices of d_{month} the cubic splines (tension approx. $1 \cdot 10^{-6}$) also gave reasonable results. Turning to tension splines just made it clear for me right away which choices were advantageous.

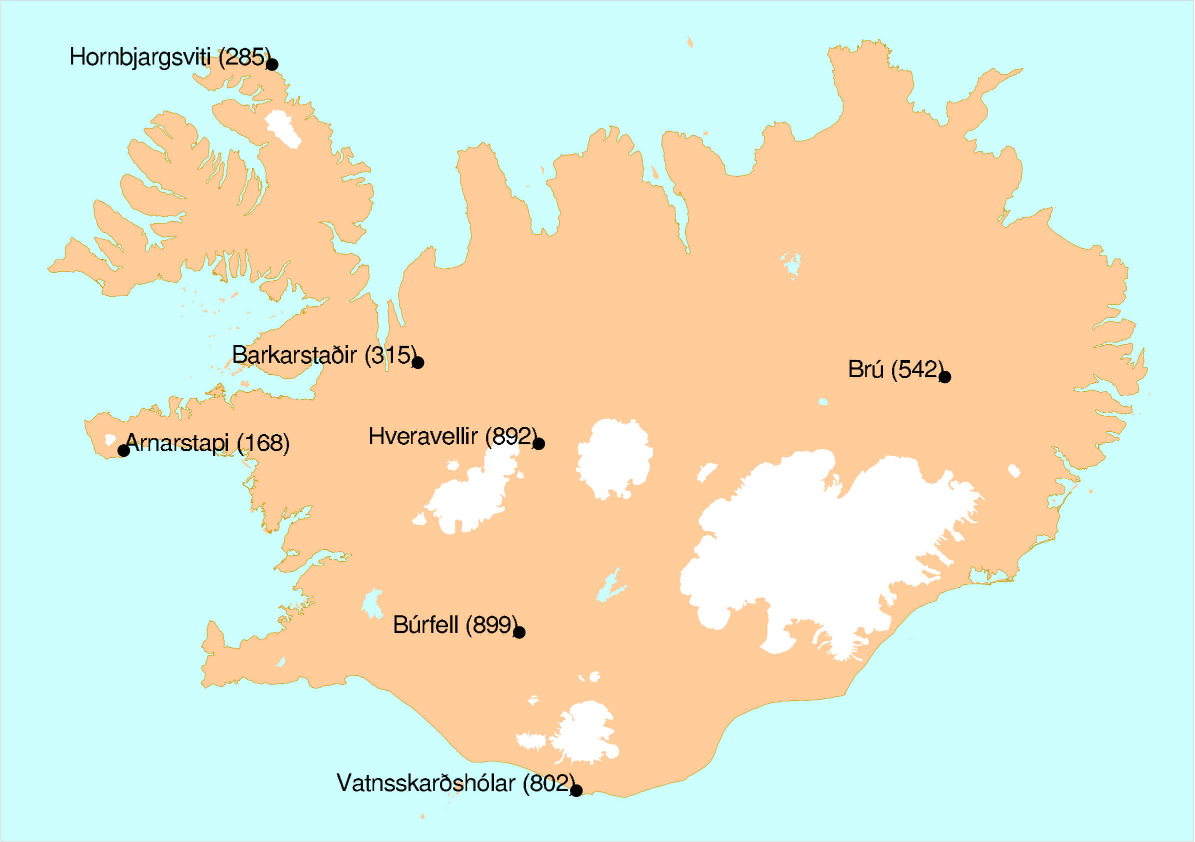
5 Examples

Though the cubic splines turned out to be reasonable after all, I still find the results coming from the tension spline model better. Let us first look at some examples to see the difference. Plots are added in the end of the report. The stations used were chosen to highlight various difficulties encountered with the method. The locations of the stations used in the examples are shown on the map below. The tension spline curve is always at the upper figure and the cubic spline at the lower. Keep in mind that we are only given twelve values, it is therefore remarkable how good the tension spline approximations (red curve) actually follows the exact annual temperature cycles (blue curve). In the examples below the tension is set to 0.2 and $1 \cdot 10^{-6}$ respectively. It takes a long time, but one has the possibility of running **tsplinegraphic** in order to see plots of all the manual stations; for more info see section 6.

Example 1: For (almost) all stations, there is a little drop in temperature in March, often it is quite steep. The tension splines tend to follow this steep drop better than the cubic. This is the main argument for using the tension spline method.

Example 2: It is a fact that during the winter the temperature often rises in February and drops again in March. So if we want to determine the freezing season, it is sometimes divided in two periods, because the temperature can rise above the freezing point and continue again to drop down below zero for some time. The tension spline follows this last drop below zero better than the cubic.

Example 3: In some cases the daily temperatures are a little "flat" at the top (July, August), tension splines often approximates this better. This is also the case in example 2.



Let us now look at some results gotten from tension splines. The dots are the measured monthly mean temperatures. Notice that a measured monthly mean temperature can be reached more than once in one month.

Example 4: A typical station.

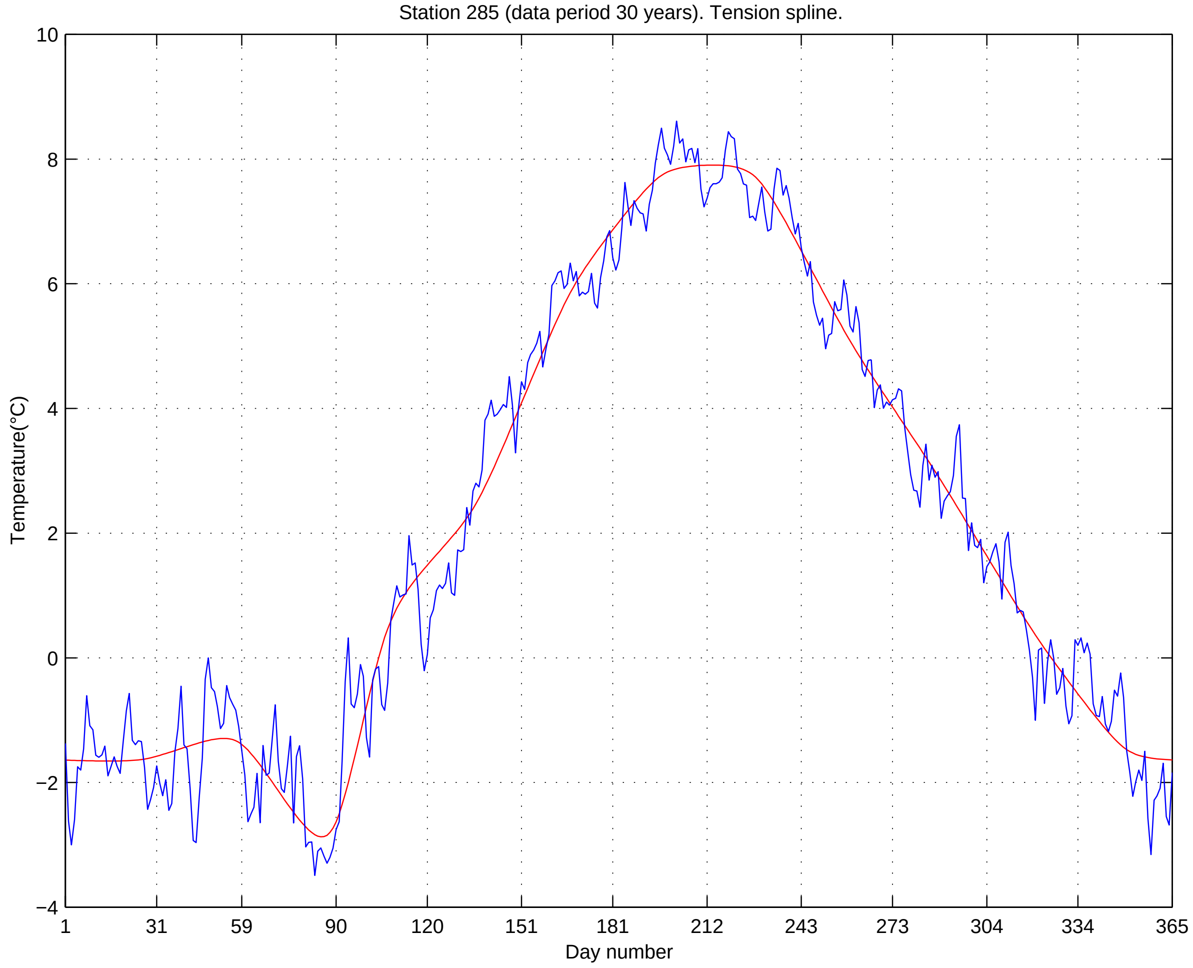
Example 5: One of the very few stations that do not have the drop in temperature in the end of March.

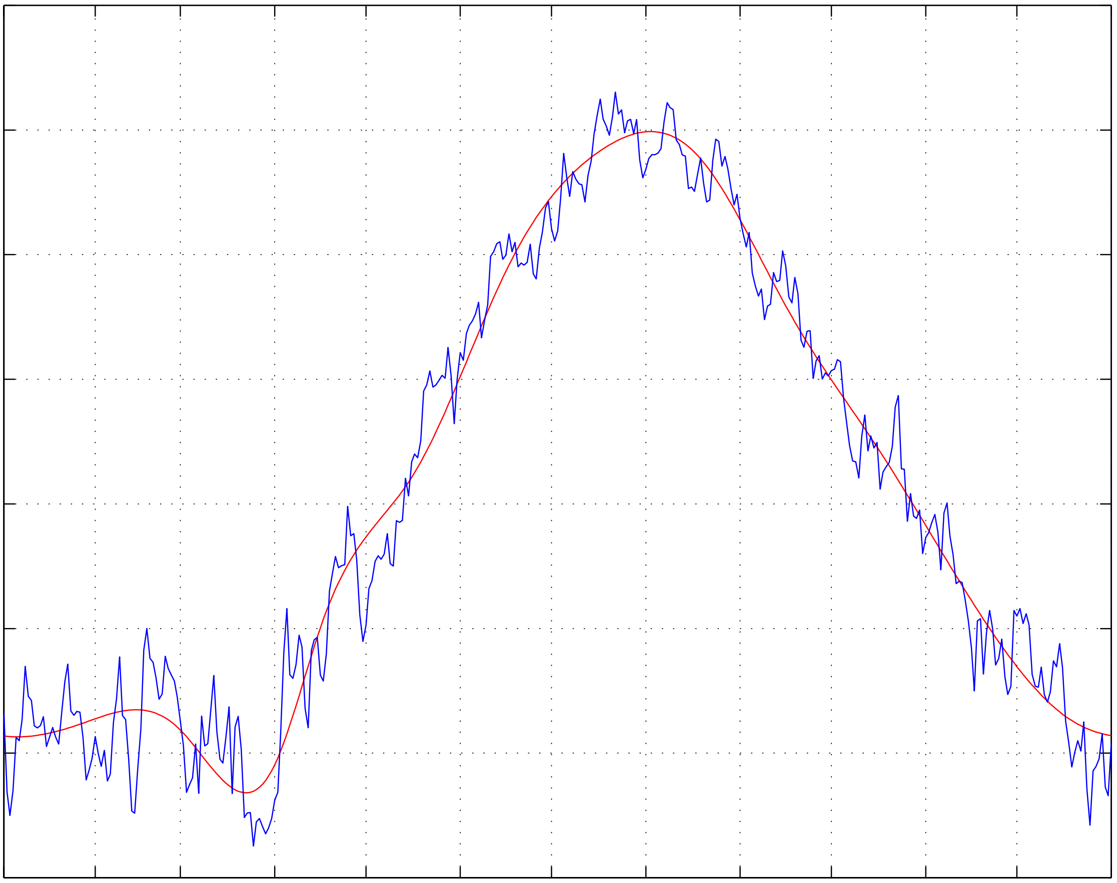
Example 6: A station where the temperature does not go below zero.

Example 7: A station with low temperatures.

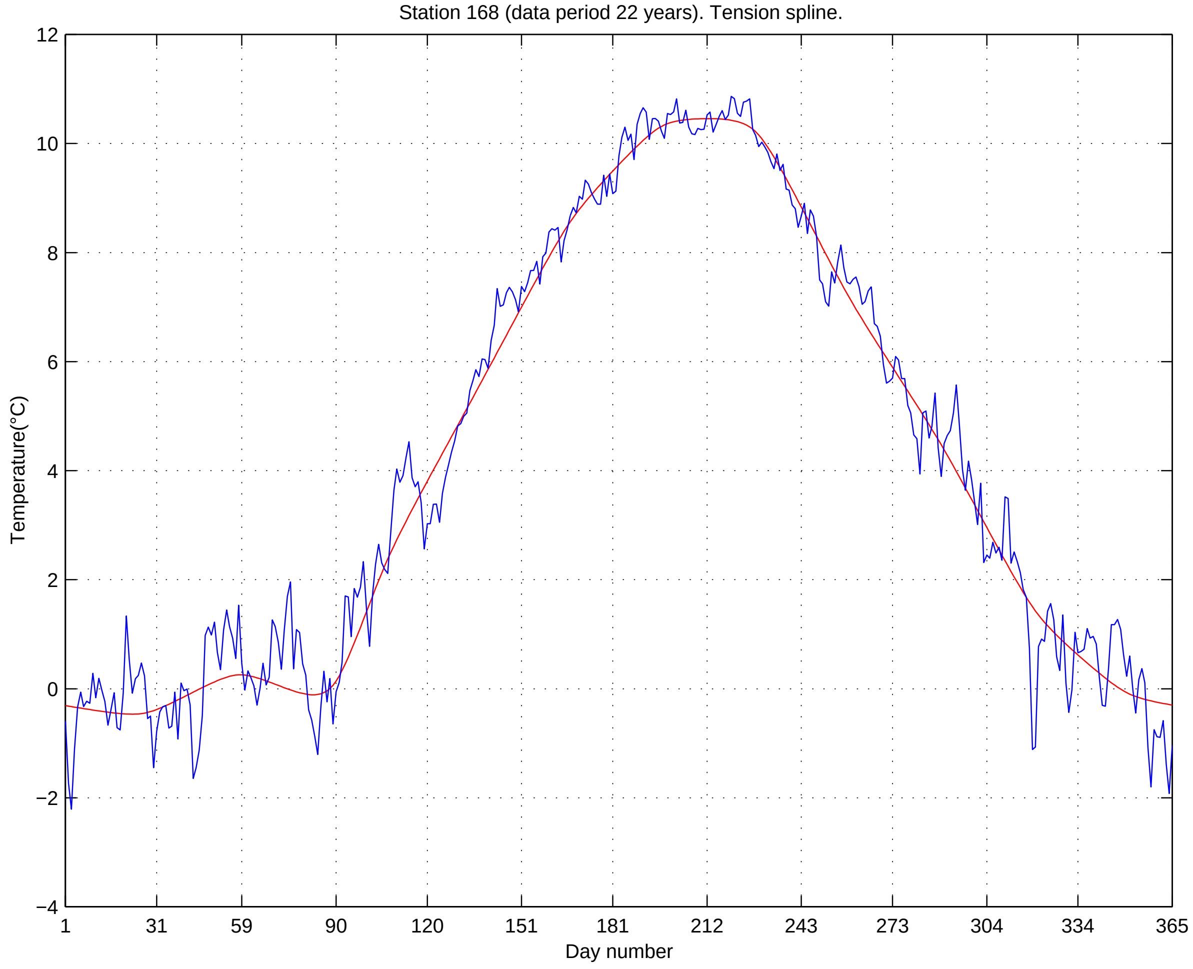
More examples can be found in the report “Applications of the tension spline method”.

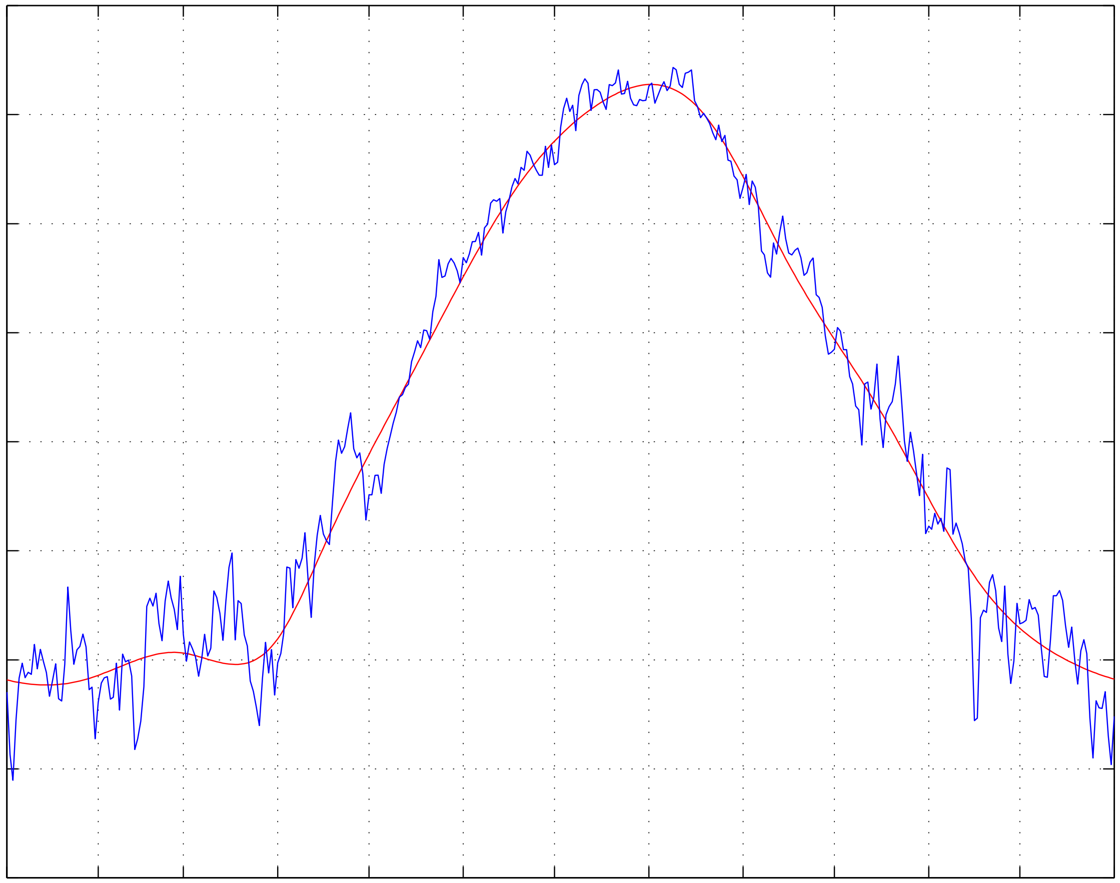
Example 1


Station 285 (data period 30 years). Tension spline.
Temperature(°C)
10
8
6
4
2
0
−2
−4
1
31
59
90
120
151
181
212
243
273
304
334
365
Day number

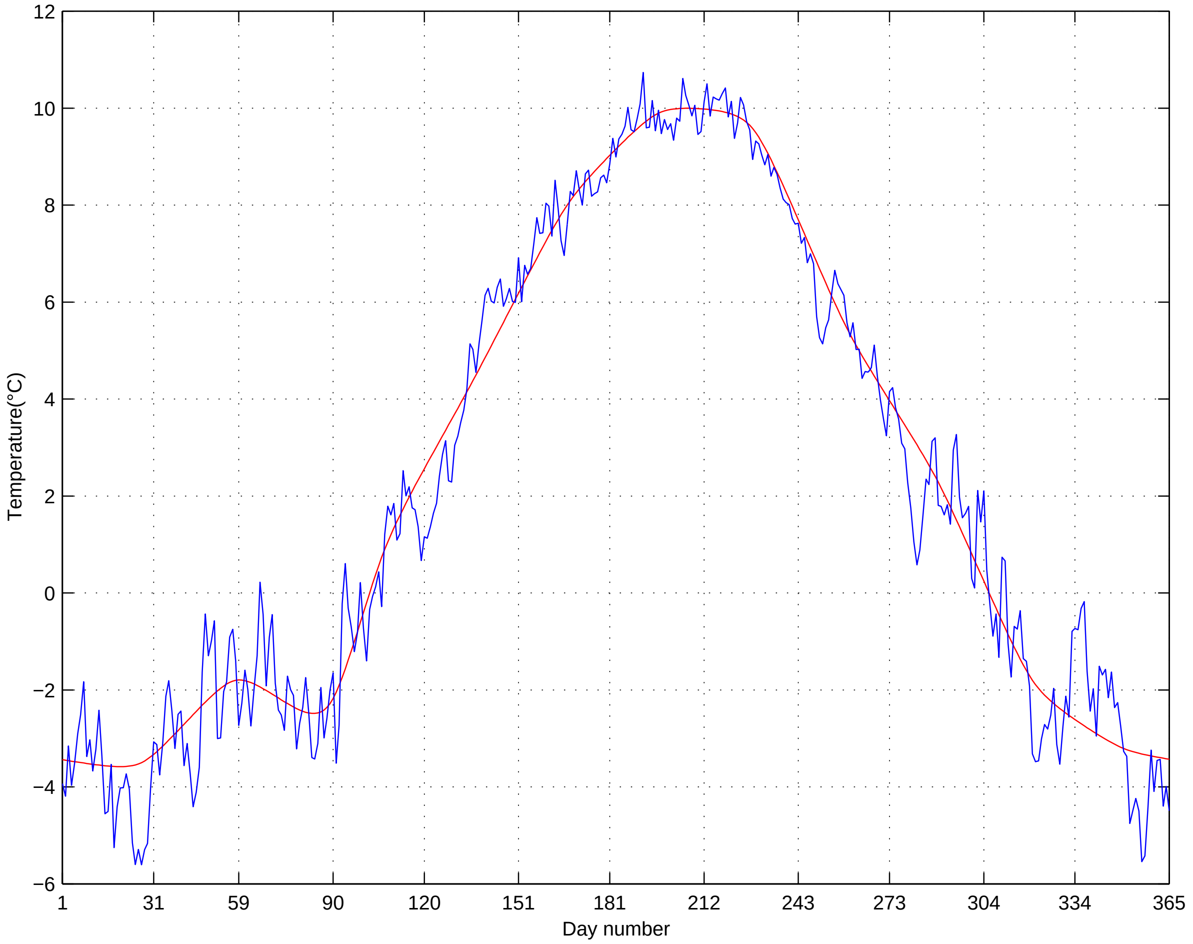


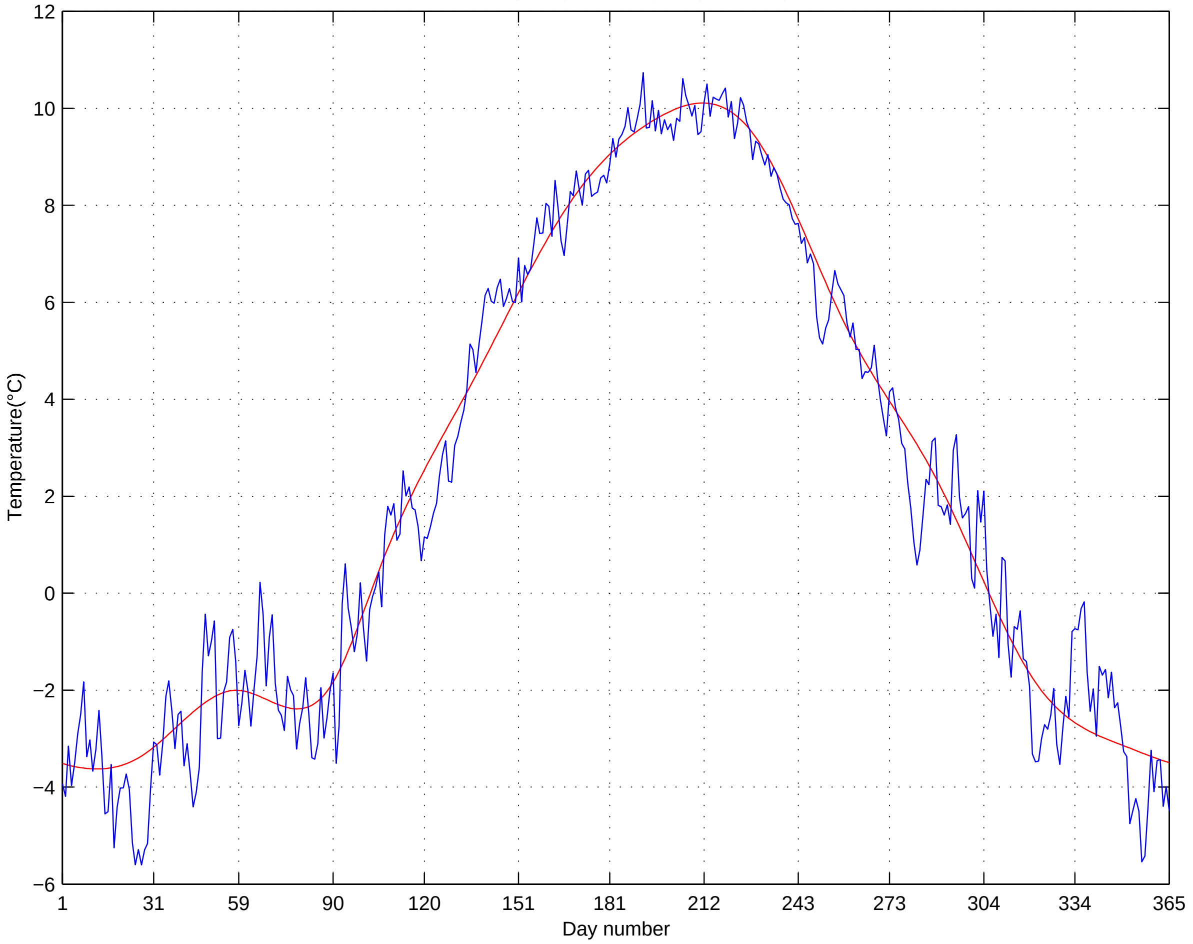
Example 2


Station 168 (data period 22 years). Tension spline.
Temperature(°C)
12
10
8
6
4
2
0
−2
−4
1
31
59
90
120
151
181
212
243
273
304
334
365
Day number

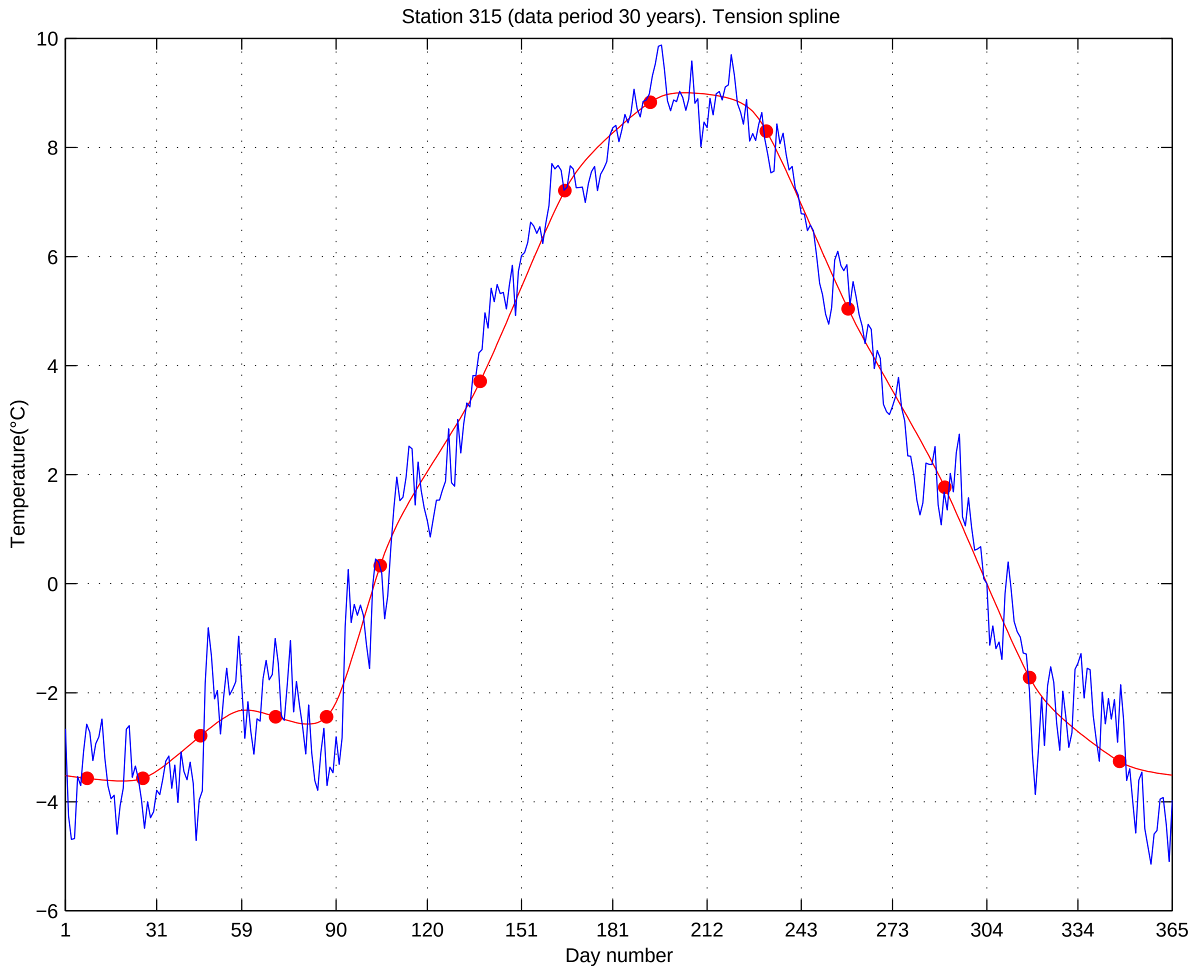


Example 3

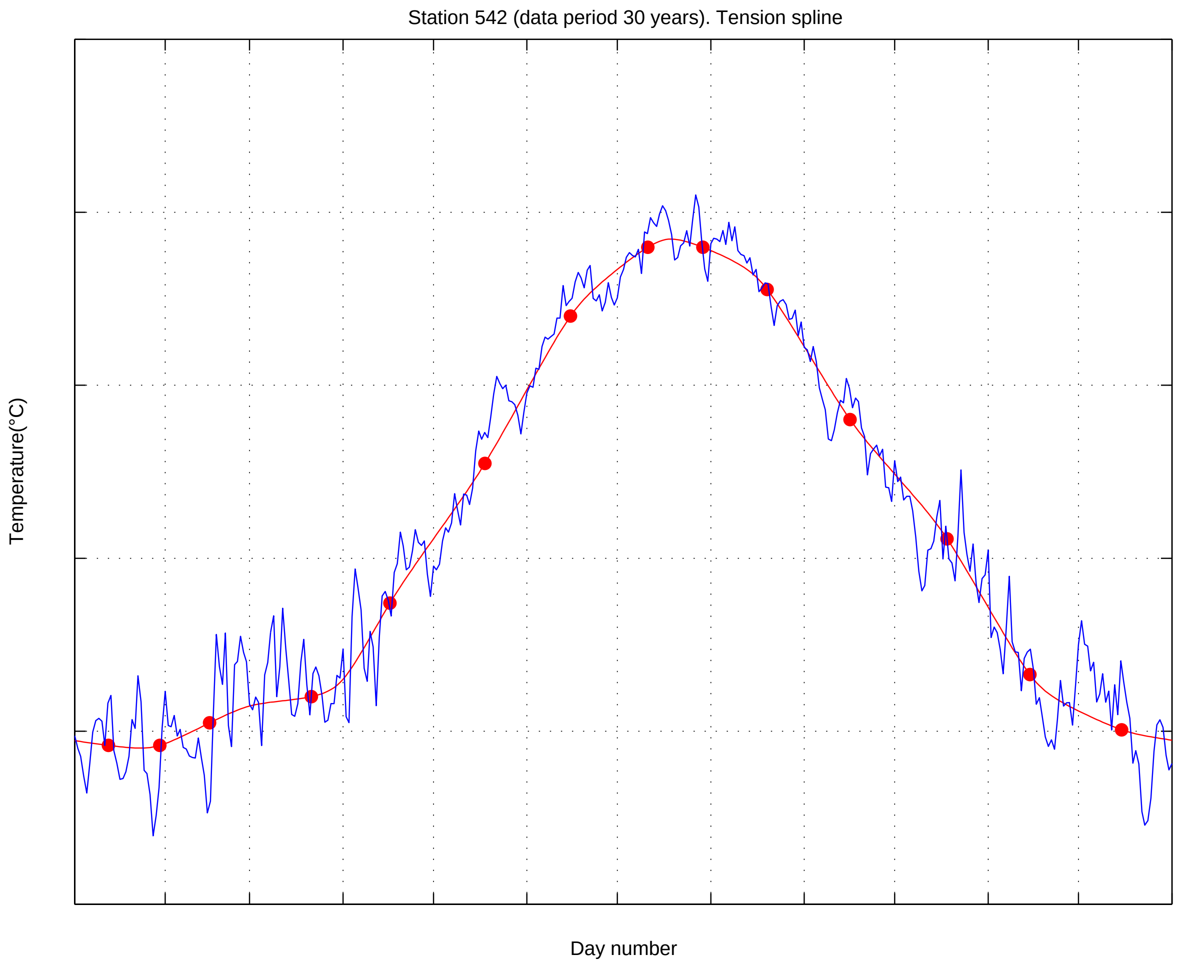

Station 899 (data period 21 years). Tension spline.


Station 899 (data period 21 years). Cubic spline.

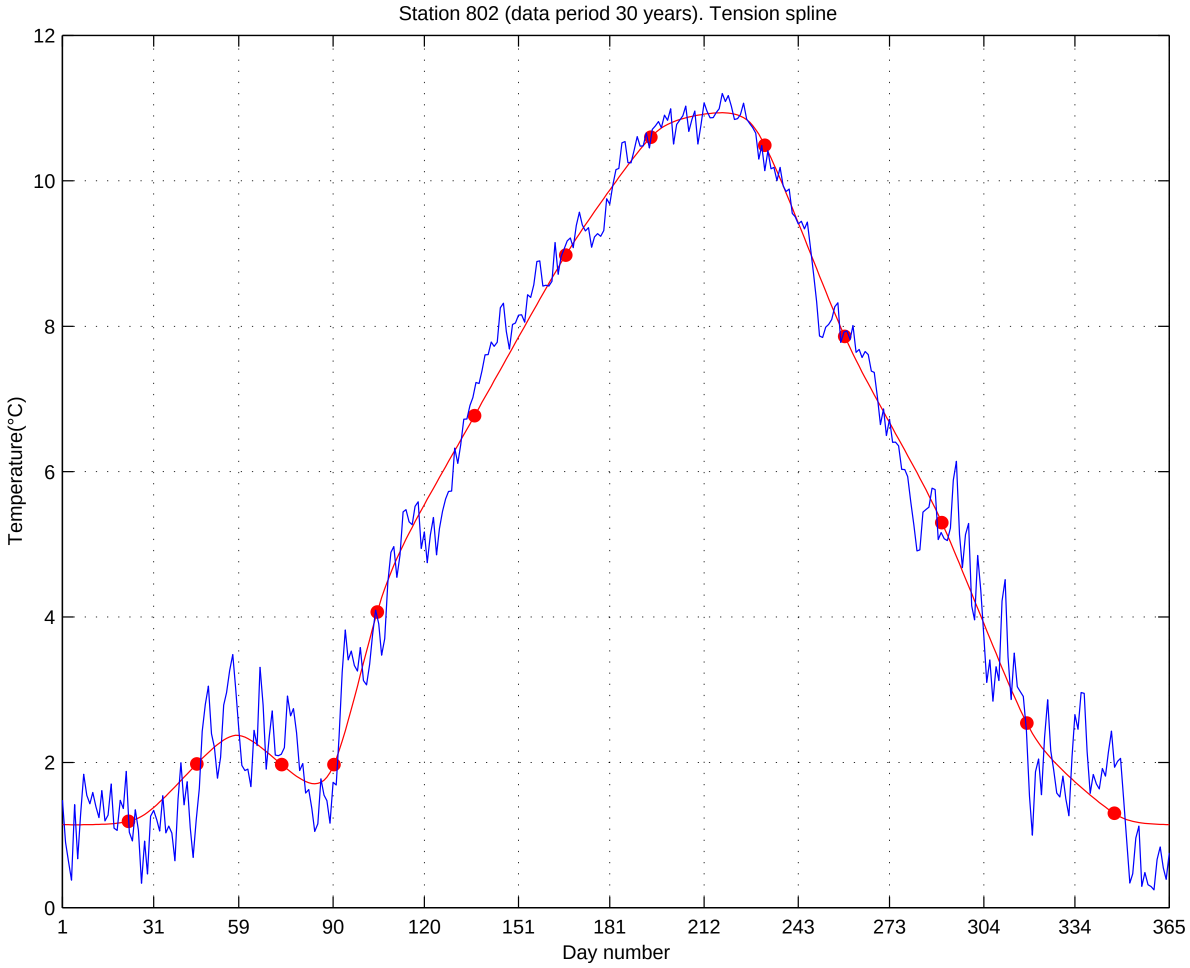
Example 4



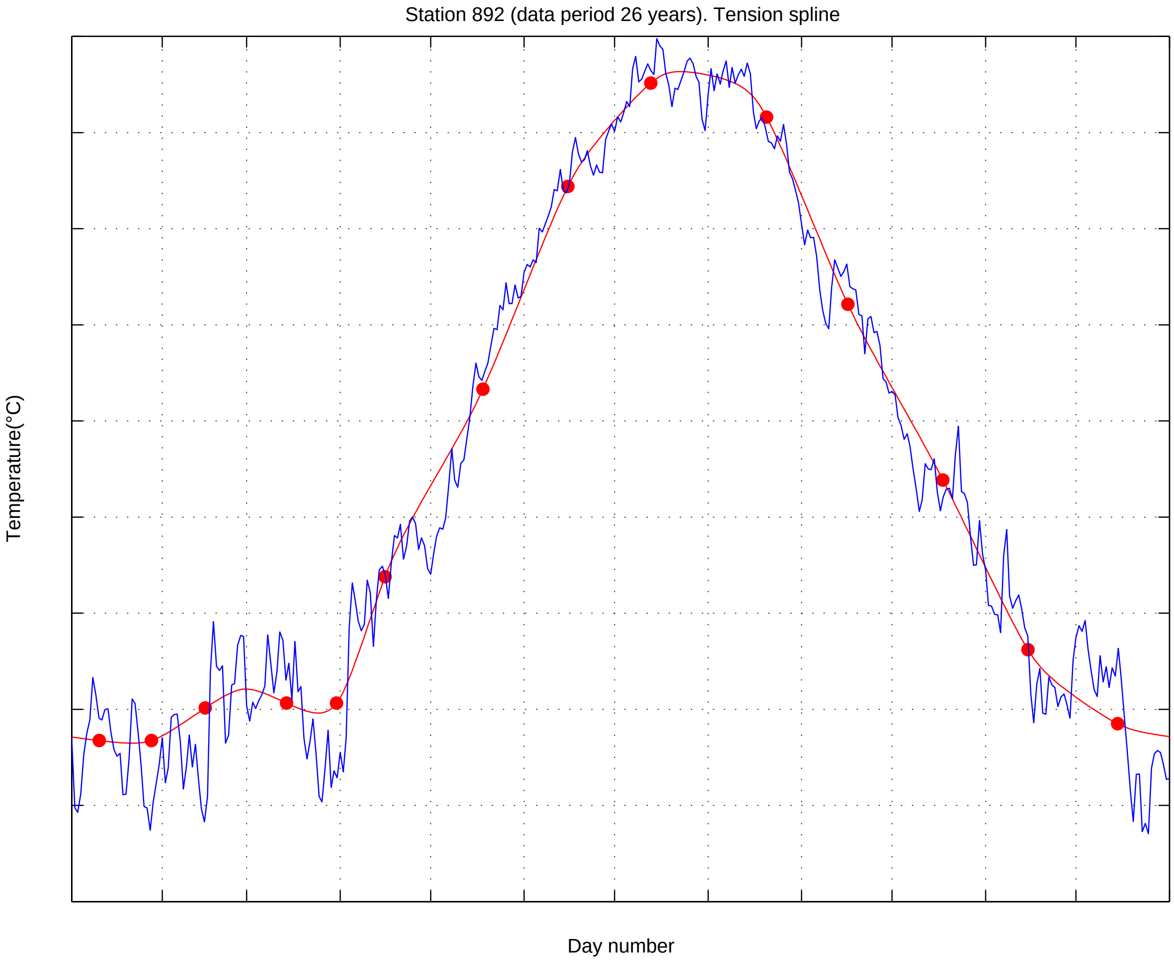
Example 5



Example 6



Example 7



6 Matlab scripts tspline.m and tsplinegraphic.m

The Matlab scripts **tspline** and **tsplinegraphic** are stored an IMO's server "**skuggi**", in the "**urs/steen**" area. This report can also be found as a word file in the directory.

tspline This Matlab program makes from twelve monthly mean temperatures the approximation of the annual temperature cycle; it outputs 365 temperature values. With **tspline** a finer accuracy than 0.001°C can easily be obtained it is only a matter of processor time; the iteration to the accuracy of 0.001°C takes approximately 1.1 sec. on a rather slow computer.

IMPORTANT! The tspline program is not a general tension spline approximation, it is designed to approximate temperature data from Iceland, i.e. following the drop in temperature in March. If one wants to make it a general tension spline function, please see the **tspline** script how it is done, though it is only possible in the case of 12 data values and periodic boundary conditions. Read remarks in the beginning of the program code to see how it works.

tsplinegraphic This Matlab program plots the measured annual temperature cycles of the manual stations in Iceland and the approximations found by **tspline**. It is optional whether to plot all stations or just some selected ones. See the script how it is done, two changes have to be made. Read remarks in the beginning of the program code to see how it works.

References

[1] Kincaid, D and Cheney, W. 1996. *Numerical Analysis*, second edition. Brooks/Cole.

[2] Schweikert , D. G 1966. “An interpolation curve using splines in tension.” Journal of Mathematics & Physics **45**, 312-317.

[3] Cline, A. K. 1974a. “Scalar and planar valued curve-fitting using splines under tension.” ACMCOM **17**, 218-220.

[4] Pruess, S. 1976. “Properties of splines in tension.” JAT **17**, 86-96.