519. 2

ON SMOOTHING OF TIME SERIES

by

S. Uusitalo

Institute of Marine Research, Helsinki

Abstract

Filtering with a specific method is presented. In using suitable chosen symmetric weighting functions, the curve near the maximum of the Fourier transform will be broadened preserving better the information of the slow variations of the original time series.

1. *General*

Data for scientific and practical purposes are often obtained at equal time intervals. As an example meteorological observations may be mentioned. Data are sometimes rather irregular. To minimize the effect of occasional variations, averaged values of different kind are used, see Milne [2]. Also other filtering methods to eliminate unwanted irregularities are used, *cf.* Barlett [1], Scheid [3], Wiener [4]. Here another approach to the problem has been made.

2. *The method*

The mathematical approach is made by using continuous functions. Let us assume that the function to be smoothed is $f(t)$ and the corresponding weighting function $p(t)$. Then the smoothed function will be

$$f_1(t) = \int_{-\infty}^{\infty} p(\tau) f(t + \tau) \, d\tau \, . \tag{1}$$

The smoothing has lessened the original function by an amount

$$f(t) - f_1(t) \, . \tag{2}$$

To alleviate the general tendency of this effect in $f_1(t)$, the amount (2) is distributed among values around t using the same weight $p(t)$.

$$f_2(t) = f_1(t) + \int_{-\infty}^{\infty} p(\tau) \, [f(t + \tau) - f_1(t + \tau)] d\tau \, . \tag{3}$$

This operation can be repeated at will. After the nth step we have

$$f_n(t) = f_{n-1}(t) + \int_{-\infty}^{\infty} p(\tau) \, [f(t + \tau) - f_{n-1}(t + \tau)] \, d\tau \, , \; n = 1, 2, 3, \ldots \tag{4}$$

3. *The Fourier transforms*

To get a better insight into the properties of Eq (4), Fourier transforms are performed. The function and its Fourier transform are denoted by small and capital letters respectively. Thus

$$P(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega t} p(t) dt \, . \tag{5}$$

$$F(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega t} f(t) dt \, . \tag{6}$$

Then the Fourier transform of $f_1(t)$ is found from Eq (1).

$$F_1(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega t} \, dt \int_{-\infty}^{\infty} p(\tau) f(t + \tau) d\tau$$

or

$$F_1(\omega) = 2\pi \bar{P}(\omega) \, F(\omega) \, . \tag{7}$$

Our corrected function (3) has the transform

$$F_2(\omega) = \frac{1}{2\pi}\int\limits_{-\infty}^{\infty} e^{-i\omega t}dt\left\{f_1(t) + \int\limits_{-\infty}^{\infty} p(\tau)[f(t+\tau) - f_1(t+\tau)]\,d\tau\right\} \tag{8}$$

or after simplifications and omitting the argument ω

$$F_2 = (4\pi\bar{P} - 4\pi^2\bar{P}^2)\,F\,. \tag{9}$$

This can also be written

$$F_2 = [1 - (1 - 2\pi\bar{P})^2]\,F\,. \tag{9$'$}$$

Repeated calculations suggest that the following general formula may be valid

$$F_n = [1 - (1 - 2\pi\bar{P})^n]\,F\,. \tag{10}$$

By induction it is readily shown to be true. When the denotation

$$\bar{Q}_n = 1 - (1 - 2\pi\bar{P})^n \tag{11}$$

is introduced, Eq (10) assumes the simple form

$$F_n = \bar{Q}_n F\,. \tag{10$'$}$$

From Eq (11) it is decuced that

$$\bar{Q}_n - \bar{Q}_1 = (1 - \bar{Q}_1)\,\bar{Q}_{n-1}\,. \tag{11$'$}$$

4. *Examples of weighting functions and their Fourier transforms*

Symmetrical weighting functions are of importance. Therefore our first choice is

$$\begin{aligned} p^{(1)}(t) &= \frac{t_0 - |t|}{t_0^2}\,, \quad \text{when } t_0 \geq t\,, \\ &= 0 \qquad\qquad , \quad \text{when } t_0 < t \end{aligned} \tag{12}$$

with t_0 being constant. The Fourier transform is

$$P^{(1)}(\omega) = \frac{1}{2\pi}\int\limits_{-t_0}^{t_0} e^{-i\omega t}\,\frac{t_0 - |t|}{t_0^2}\,dt\,,$$

which after simplifications and using Eq (11) with $n = 1$ gives

$$Q_1^{(1)}(\omega) = \frac{4\sin^2\frac{\omega t_0}{2}}{\omega^2 t_0^2} . \tag{13}$$

This function is plotted in Fig. 1 a.

As an other weighting function we choose

$$p^{(2)}(\tau) = \frac{1}{\vartheta\sqrt{2\pi}}\, e^{-\frac{\tau^2}{2\vartheta^2}} \tag{12$'$}$$

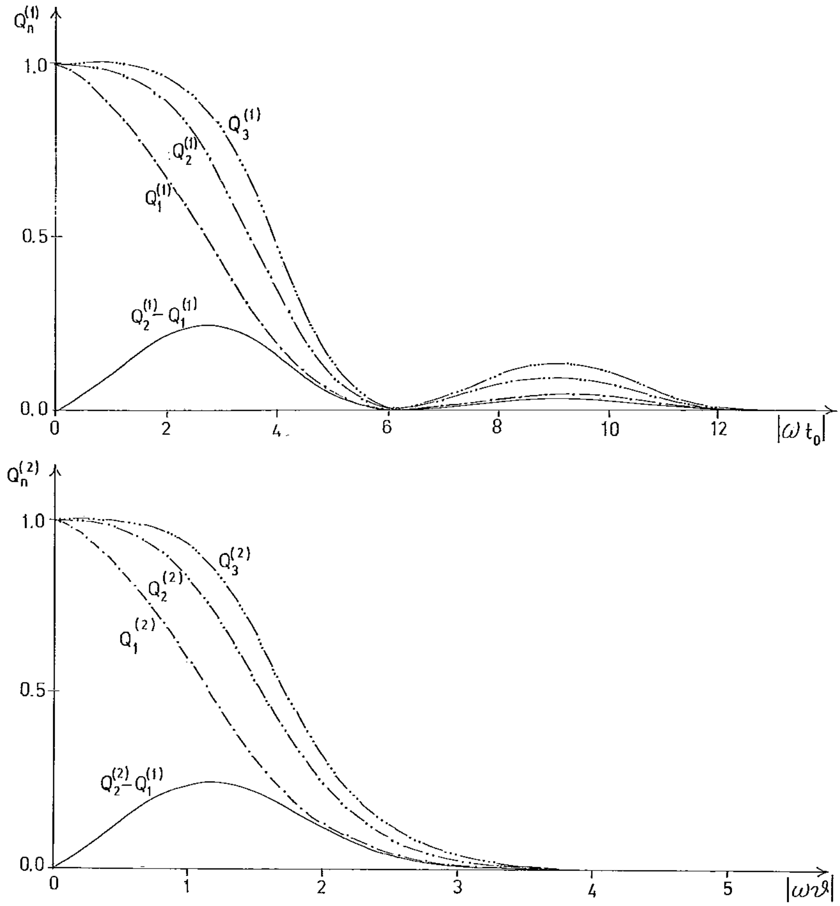

Fig. 1. Functions $Q_1 = Q$, $Q_2 = 1 - (1 - Q)^2$, $Q_3 = 1 - (1 - Q)^3$ are plotted as functions of $|\omega t_0|$ and $|\omega\vartheta|$ in Figs 1 a and 1 b respectively. The quantity Q represents correspondingly either the value $4 \sin^2(\omega t_0/2) / (\omega t_0)^2$ or the value $\exp(-1/2\ \omega^2\vartheta^2)$.

with ϑ being constant. This function has the Fourier transform

$$P^{(2)}(\omega) = \frac{1}{2\,\pi} \int\limits_{-\infty}^{\infty} e^{i\omega\tau} \frac{1}{\vartheta\sqrt{2\pi}} e^{-\frac{\tau^2}{2\vartheta^2}} d\tau \,,$$

which after simplifications becomes

$$Q_1^{(2)}(\omega) = e^{-\frac{\omega^2\vartheta^2}{2}} \,. \tag{13$'$}$$

This function is depicted in Fig. 1 b.

5. *Smoothing using a symmetrical weighting function*

The Fourier transform for a symmetrical function p(t) is

$$\frac{Q_1(\omega)}{2\,\pi} = P(\omega) = \frac{1}{2\,\pi} \int\limits_{-\infty}^{\infty} \cos\,(\omega t) p(t) dt \,. \tag{14}$$

When $\omega = 0$ then the integral on the right hand side $= 1$ because of the properties of the weighting function. On the other hand,

$$1 - Q_1(\omega) = \int\limits_{-\infty}^{\infty} p(t)[1 - \cos\,(\omega t)]\, dt$$

$$= 2 \int\limits_{-\infty}^{\infty} p(t) \sin^2 \frac{\omega t}{2}\, dt \,.$$

Thus for $p > 0$, $Q_1 \leq 1$. It can also be shown that for $p > 0$, $Q_1 \geq -1$.

In Fig. 1 a few of the functions Q_n are plotted. Also a curve for $Q_2 - Q_1$ is included for both cases treated. This curve shows a maximum, when $Q_1 = 0.5$. In general, the maximum is to be found, when

$$\frac{d}{dP} (Q_n - Q_1) = 0 \,,$$

i.e., when

$$Q_1 = 1 - \frac{1}{\sqrt[n-1]{n}} \,.$$

This value tends to zero, when n tends to infinity, *i.e.* the maximum effect of the method takes place for the smaller P the larger n is. This means that the maximum moves toward larger values of $|\omega|$ with increasing n.

6. *Numerical experiments*

In Fig. 2 we have the daily mean temperatures of July 1970 at Sodankylä*) as a fulldrawn line and the corresponding smoothed values as a dashed line. The weighting function (12) was replaced by the set

$$p_i^{(1)} = \frac{3 - |i|}{9}, \quad i = -2, -1, 0, 1, 2. \qquad (12)''$$

The basic property of weights $\Sigma p_i = 1$ is thus fulfilled. Periods corresponding to three data are practically damped out, but the main period of about nine data is maintained almost unchanged.

As another example, we performed the smoothing of daily mean temperatures for the years 1931 to 1960 computed as means of calendar days at Ilmala Observatory*), Helsinki, see Fig. 3.

This time the weighting function (12)′ was replaced by the set

$$p_i^{(2)} = ce^{-ai^2}, \quad i = 0, \pm 1, \pm 2, \pm 3, \ldots, \qquad (12)'''$$

with $a = 0.002$, $c = 0.0252$. This set satisfies the condition $\Sigma p_i^{(2)} = 1$. The choice corresponds to $\vartheta = 15.8$ according to Eq (12)′.

7. *Discussion*

The amplifications $Q_2, Q_3, \ldots$ of our method show a much broader maximum as the corresponding simple amplification Q_1, see Fig. 1. The small positive values of the amplification increase about by the amount Q_1 every time the method (4) is used thus increasing the value of Q_n to about nQ. If Q_1, and thus Q_n are very small, then the shifting of the maximum of $Q_n - Q_1$ toward larger values with increasing n causes the slope of the Q_n-curve to become steeper between the

*) All meteorological data published in this paper are callected from archives of the Finnish Meteorological Institute, Helsinki, Finland.

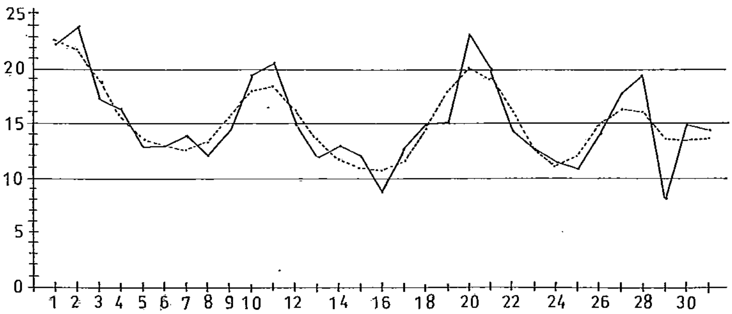

Fig. 2. Temperature values at Sodankylä in July 1970 are plotted as a fulldrawn line and the values are smoothed using the set (12)″.

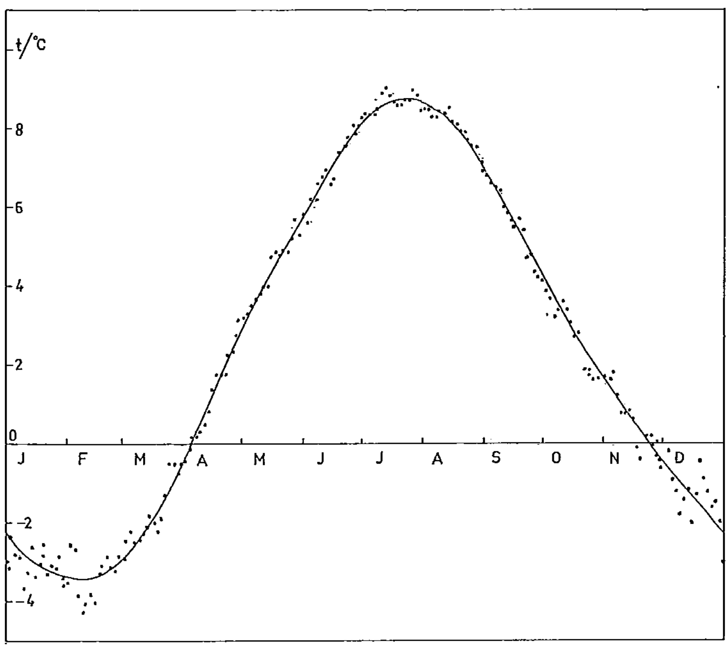

Fig. 3. Daily mean temperatures for the years 1931—61 computed as mean of calendar days, at Ilmala Observatory, Helsinki, are indicated by dots. Smoothed values are found by using the set (12)‴ and expressed by a fulldrawn line.

named two values, the maximum and the very small Q_n-value. This causes the slow variations of the original information to be preserved rather well. The separation of the slow variation part from the rapid one becomes better.

8. *Conclusions*

Using a suitable chosen weighting function, *e.g.* (12)′, it is possible to increase the filtering of rapid fluctuations from a time series by repeating the procedure (4), because this cause the main maximum of Q_n in (11) to broaden and the separation of its large and small values to be more distinct. While this effect is still more emphasized, when using (12) as weighting function, it may not indeed be as advantageous, because of the side maxima, which grow larger every time Eq (4) is used.

Acknowledgement: The author is greatly indepted to Dr. P. Mälkki and others for fruitful discussions.

REFERENCES

1. Bartlett, M. S., 1950: Periodogram analysis and continuous spectra. *Biometrika* 37, 1–2, 1–16.
2. Milne, W. E., 1949: *Numerical calculus,* pp. 275–280. Princeton University Press, Princeton, New Jersey.
3. Scheid, F., 1968: *Theory and problems of numerical analysis.* Schaum's outline series. McGraw-Hill Book Company. New York, St. Louis, San Francisco, Toronto, Sydney. Pp. 243–247.
4. Wiener, N., 1970: *Extrapolation, interpolation and smoothing of stationary time series with engineering applications. Time series,* MIT-press, Cambridge, Mass. 163 pp.