Applying SegRegA to the annual average temperature trend from 1900 to 2020 in the Netherlands resulting from global warming; analysis by segmented linear regression types and curved functions such as S-curve, Power curve, generalized quadratic and cubic regressions.

R.J. Oosterbaan 23-12-2020

Abstract
The data used were measured at the Royal Meteorological Institute (KNMI), De Bilt, Netherlands. Measurements were done since 1900 and continued up to now.
An initial trend-analysis was made by linear regression showing a rising trend. A second and third trend-analysis was done with segmented regressions Type 2 and 6 using the SegRegA software program. According to Type 2 the temperature increases at a faster rate after 1963. According to Type 6 there is a temperature jump around 1988 followed by a still faster rising trend. This jump is statistically significant and the ANOVA table shows a high statistical significance of the model. This significant trend could be the result of global climate change.

Contents

1. Introduction

In this article, different regression equations are used to characterize the trend of the yearly average of the daily maximum temperatures in °C over the years 1900 to 2019 in the meteorological station (KNMI) of de Bilt in the Netherlands. The prime objective is to describe the increasing trend resulting from global warming. The software used (SegRegA) is specified in section 9 (Reference) and in the Appendix (section 10).

2. Linear regression

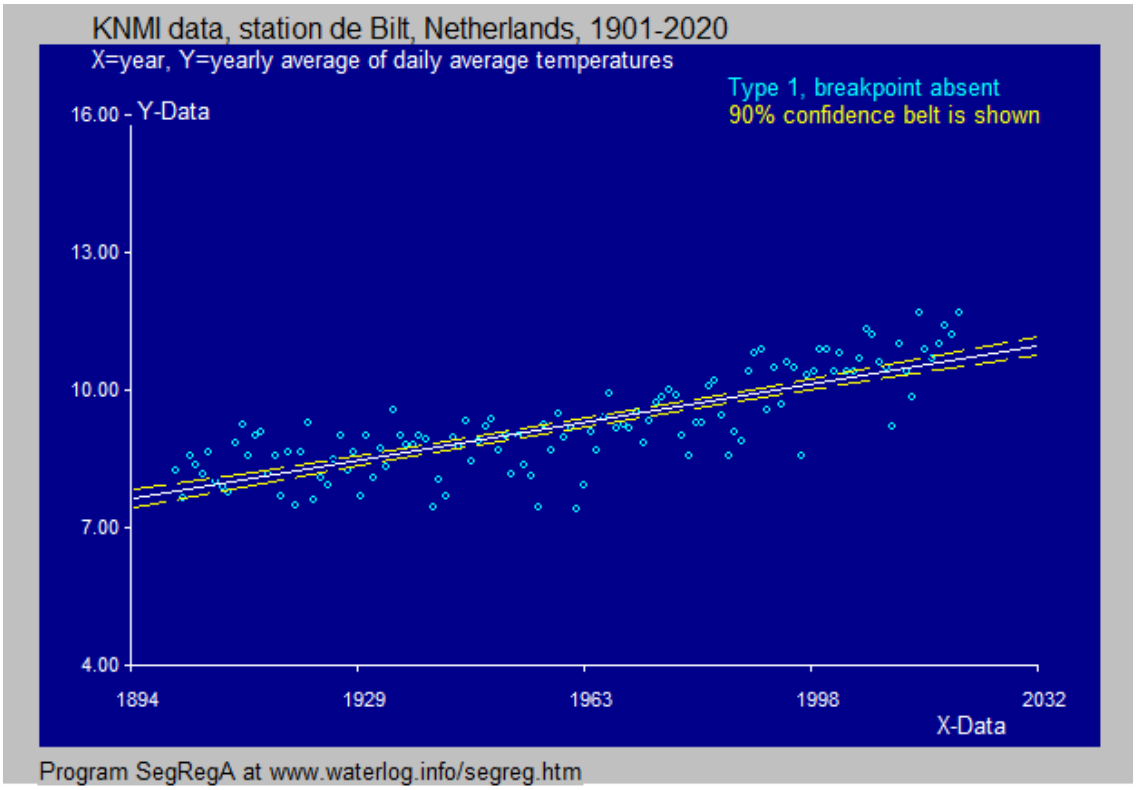


Figure 1. Linear regression of the yearly average of daily average temperatures on time in years. Data from the KNMI station, de Bilt, Netherlands. The trend is sloping upward, so the temperature rises steadily.

The parameters of the linear regression line Y = A.X + B are A = 0.0236 and B= - 36.9. The coefficient of explanation (R^2) equals 0.634.

3. Segmented regression Type 2

Instead of using linear regression, one may try to detect a change of the relation in time. For this one can use the SegregA software (see reference), which admits different regression type options.

The segmented regression type 2 searches for a breakpoint (or separation point) and optimizes it by minimizing the sum of squares of deviations with the condition that the linear regression lines left and right of the breakpoint join at the breakpoint . SegRegA calculates the confidence intervals of the two regression coefficients and statistically verifies that they significantly different.

In addition, SegRegA, by means of variance analysis (ANOVA), checks statistically that the Type 2 model is a significant improvement of the overall linear regression (Type1, as in figure 1).

An example is shown in figure 2.

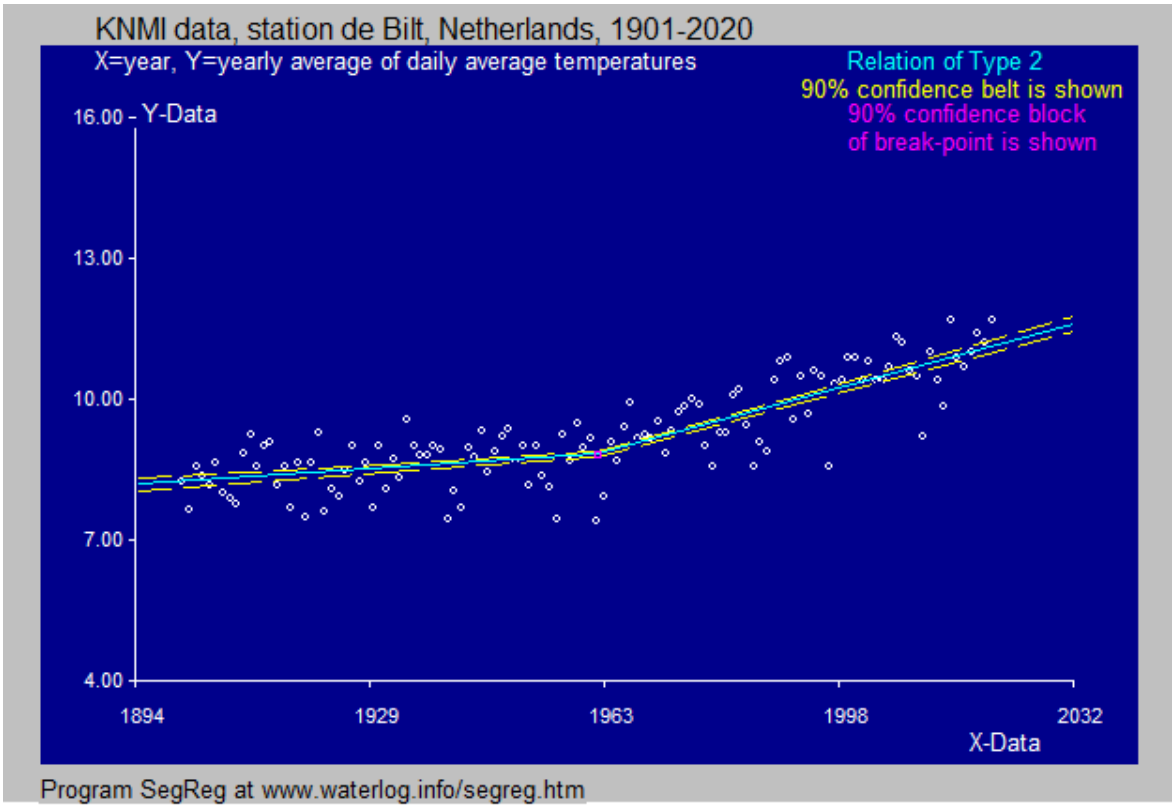


Figure 2. Segmented regression Type 2. From the year 1963 and onwards the regression line is steeper than the line before 1963. This indicates that the annual rise in temperature has increased after 1963, possibly due to global climatic change.

The coefficient of determination R^2 equals 0.704, which is higher than that for the linear regression, which signifies an improvement of the fit to the observed data. The confidence block of the breakpoint is very small and hardly visible.

.

In addition, SegRegA, by means of variance analysis (ANOVA), checks statistically that the Type 2 model is a significant improvement of the overall linear regression (Type1, as in figure 1). The ANOVA table is shown hereunder.

Table 1.

Variance Analysis, ANOVA table, Regression Type: 2 (figure 2).
Sum[(Y-Av.Y)sq.] = 77.800 (total sum of squares of deviations)
Total nr. of data = 118
Degrees of freedom = 117

Sum of squares of deviations	Degrees of freedom	Variance	F-Test	Probability/ Significance
explained by linear. regression			F(1,116)=	
75.600	1	29.400	70.465	99.9 %
remaining unexplained				
47.400	116	0.417		
extra explanation by the break-point			F(3,113)=	
8.377	3	2.825	8.088	99.9 %

The extra explanation is highly significant.

4. Segmented regression Type 6

Selecting in SegRegA the option Type 6, one tries to obtain two regression lines that do not intersect each other at the breakpoint, provided that the jump between the two lines is statistically significant. The same data as used before produce a result a shown in figure 3. The type of regression is there indicated by 8, indicating that the two regression line are sloping, but with a different angle.

The breakpoint at year 1988 is optimized minimizing the sum of squares of errors and the jump is tested on statistical significance. Here it is highly significant.

The analysis of variance, to test the significance of the segmented regression of Type 6/8 compared to the linear regression is shown in the following table 2 as calculated by SegRegA.

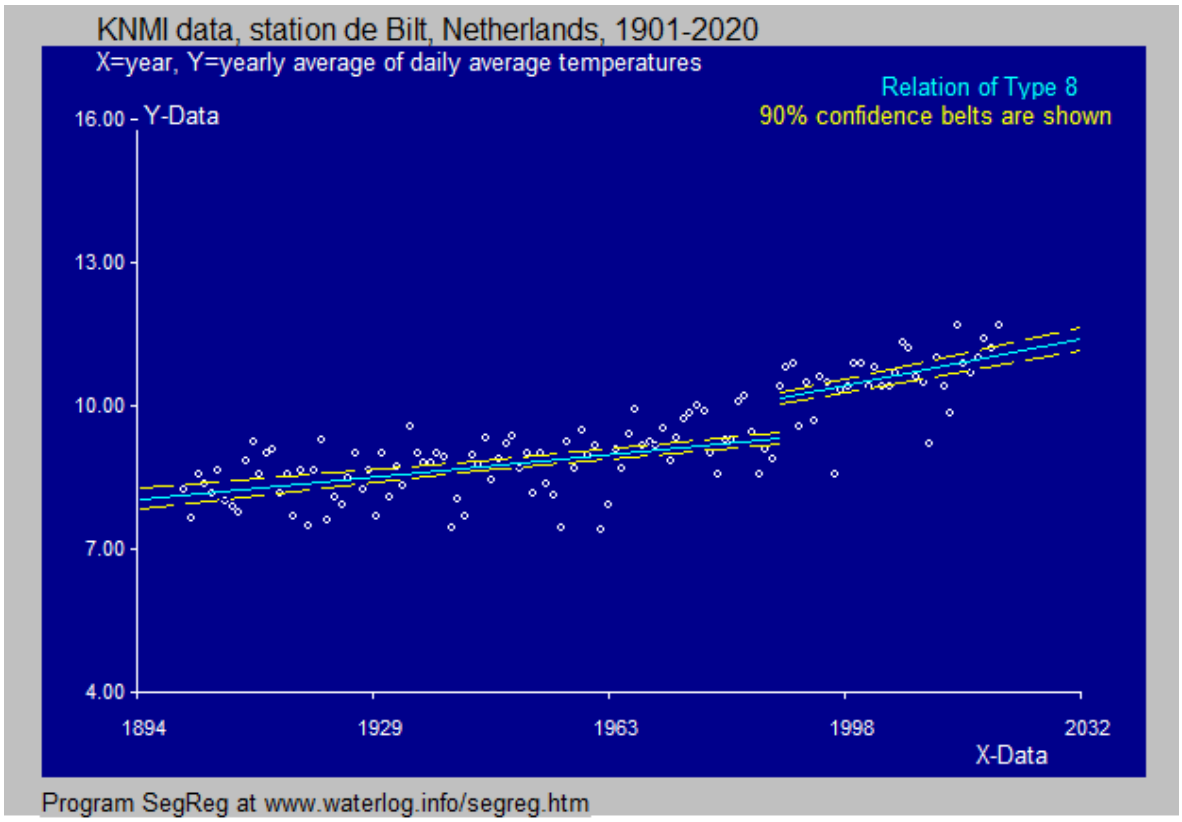


Figure 3. Segmented regression of Type 6/8. At the year 1988 there is a jump in temperature. The first segment is quite flat while thereafter there is a steeper segment. This could be the result of a climate change, which makes itself felt more strongly after 1988.

Table 2.

Variance Analysis, ANOVA table, Regression Type: 6 (with jump)
Sum[(Y-Av.Y)sq.] = 77.80 (total sum of squares of deviations)
Total nr. of data = 118
Degrees of freedom = 117

Sum of squares of deviations	Degrees of freedom	Variance	F-Test	Probability/ Significance
explained by lin. regr. 75.600	1	29.400	F(1,116)= 70.465	99.9 %
remaining unexplained 47.400	116	0.417		
extra expl. by break-point 8.986	3	2.995	F(3,113)= 8.811	99.9 %

The above table tells us that the variation explained by the break point (99.9%) is highly significant. Hence, the segmented regression is significantly better than the linear regression. Also the coefficient of determination is slightly higher (R^2=0.707) is slightly higher than the previous ones. However, it would need an additional study to test whether it is also significantly better than the Type 2 regression.

5. Quadratic regression

Both the type 2 and type 6/8 suggest that a quadratic regression may be relevant. It has the advantage to produce a fluid parabola instead of segmentation with breakpoints.

SegRegA gives the option to select a quadratic function, with or without transformation of the data. The transformation may serve to obtain a still better fit by generalization. Here, however, the second option is used.

Figure 4 depicts the results of this treatment. The analysis of variance is just as positive as discussed in the previous cases. Therefore the ANOVA table needs not be shown.

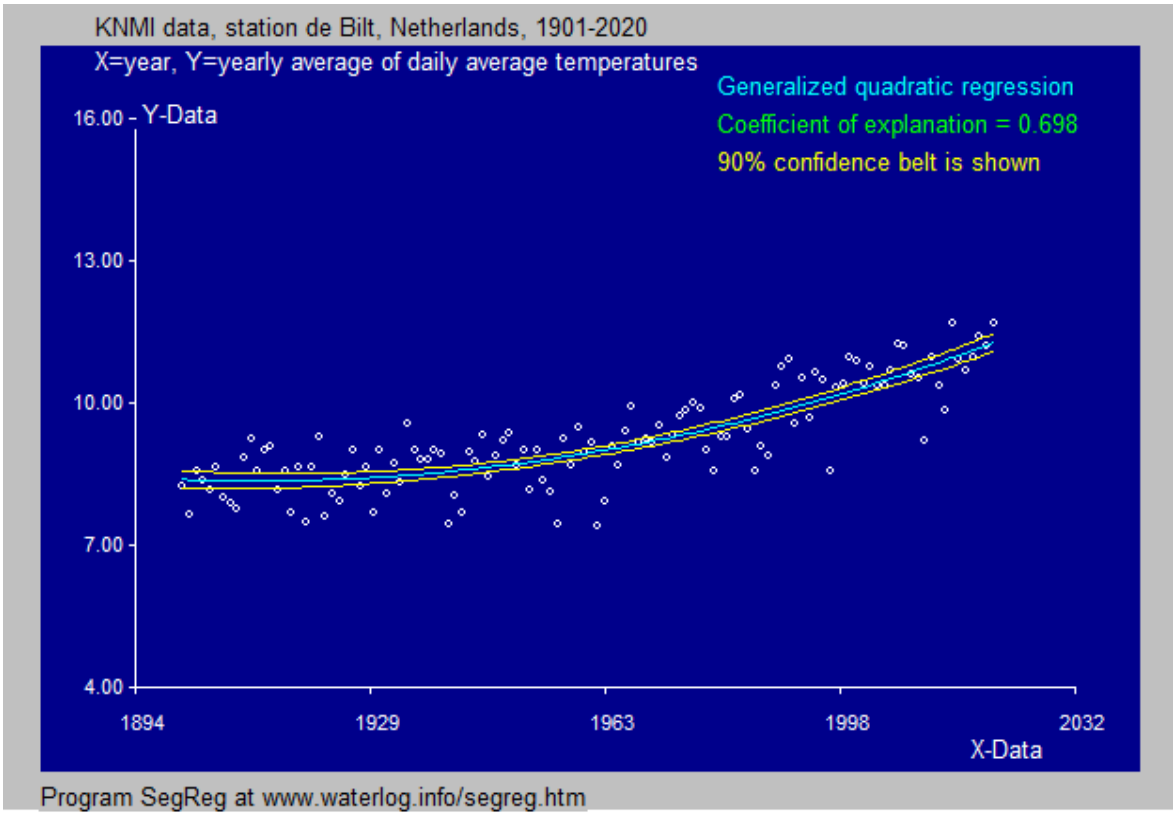


Figure 4. Quadratic regression resulting in a parabolic function. In the last decades the picture show an increasing curvature, suggesting that the global warming up proceeds at an increasing rate. There is yet no theoretical explanation available for this phenomenon. However when this trend continues for some more years, the evidence becomes overwhelming and it will be required to undertake theoretical research to confirm the empirical facts.

6. Power curve

The SegRegA software employed, also provides the option to us a power curve instead of quadratic. The results are shown in figure 5 below.

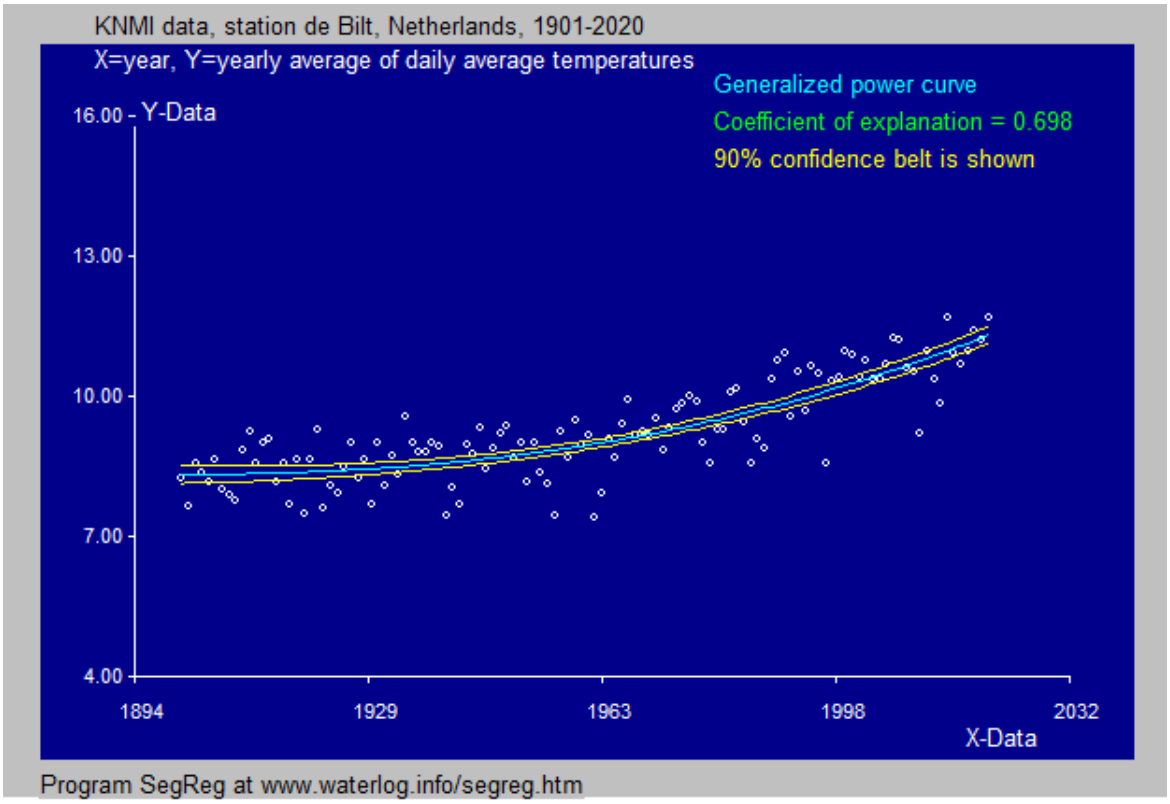


Figure 5. The shape of the power curve, in this case is: $Y = 3.72 * X_t^{2.43} + 8.32$, where X_t is the transformed X value (year nr.), being $X_t=(X-X_{minim})/(X_{maxim}-X_{minim})$, does not differ much from the quadratic curve. In both cases, the coefficient of explanation is the same (69.8 %)

7. Logistic S-curve

Instead of applying the Type 6 segmented regression (figure 3), one could also try the logistic S-curve (figure 6).

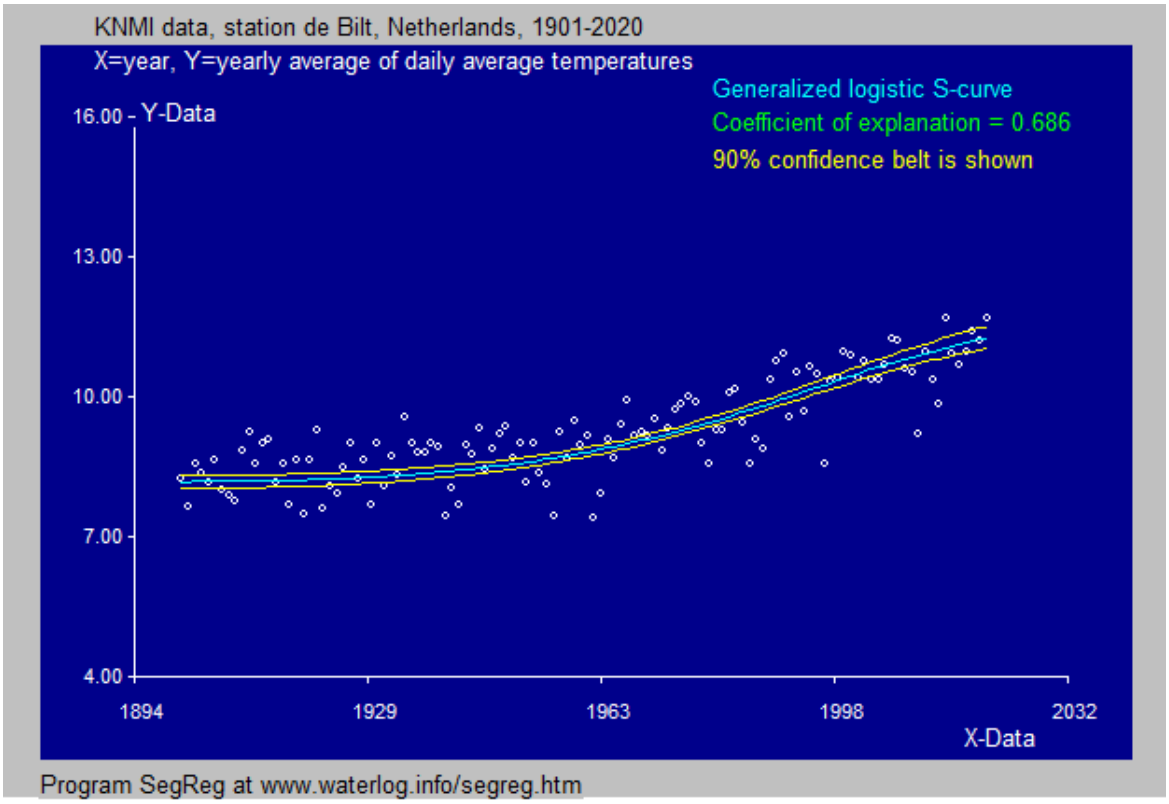


Figure 6. Logistic S-curve of the average temperature trend in de Bilt, The Netherlands. Towards the year 2020 the slope is steepest (like in figure 4 for quadratic regression), but it tends to flatten beyond that date. A longer time range of observations will be required to confirm or reject the S-curve tendency. Also the coefficient of explanation is slightly less than in the two previous cases.

The S-curve equation reads: $Y = Y_{minim} + (Y_{maxim}-Y_{minim}) / [1+ \exp \{ A * (X-C)^{E} + B \}]$
where $Y_{minim} = 7.35$, $Y_{maxim} = 11.80$, $A = -0.00008$, $C = 1901$, $E = 2.12$ *and* $B = 1.47$

8. Conclusion

Whatever regression type one employs, it leads to the conclusion that the temperature is rising over the years. The analysis according to Type 1 (purely linear) is well know and accepted. The analysis with Type 2, leading to the conclusion that the temperature has risen with increasing speed, is difficult to contradict but has so far not been generally accepted. The exceptionally high temperatures in the 21st century need to be explained before the trend shown in figure 3 (segmented Type 8 with a jump) can be theoretically accepted. However, this last phenomenon is essentially an empirical fact and cannot be denied. As an alternative, the parabolic function as in figure 4, may be studied.

9. Reference

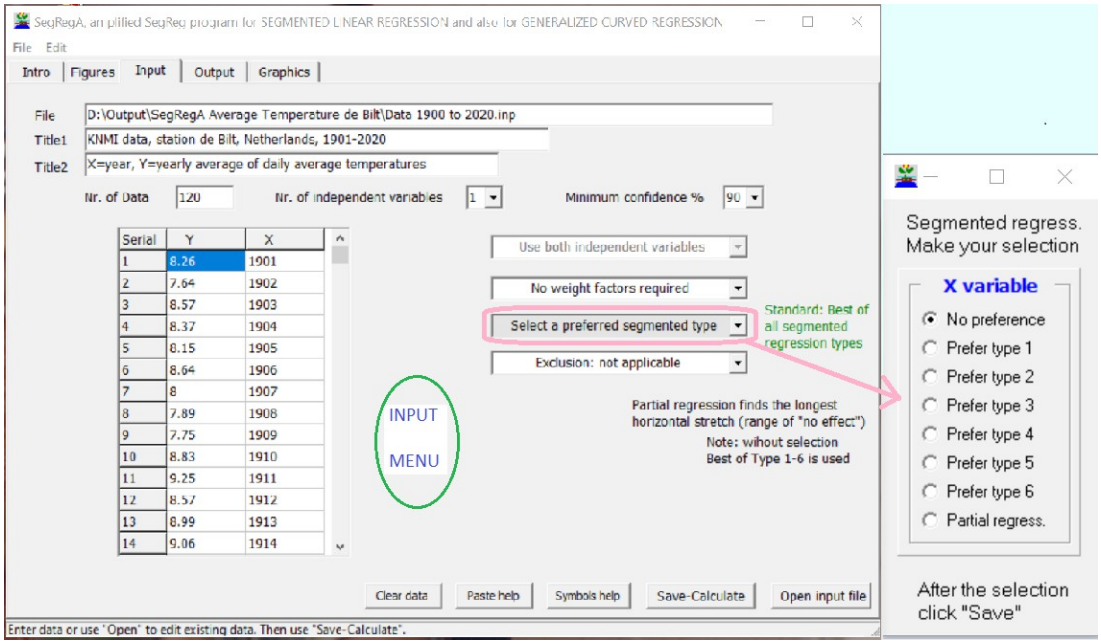
SegRegA, free software for segmented linear regression on the one hand and curved regression on the other. Download from:
https://www.waterlog.info/segreg.htm

List of publications in which SegReg is used:

https://www.waterlog.info/pdf/segreglist.pdf

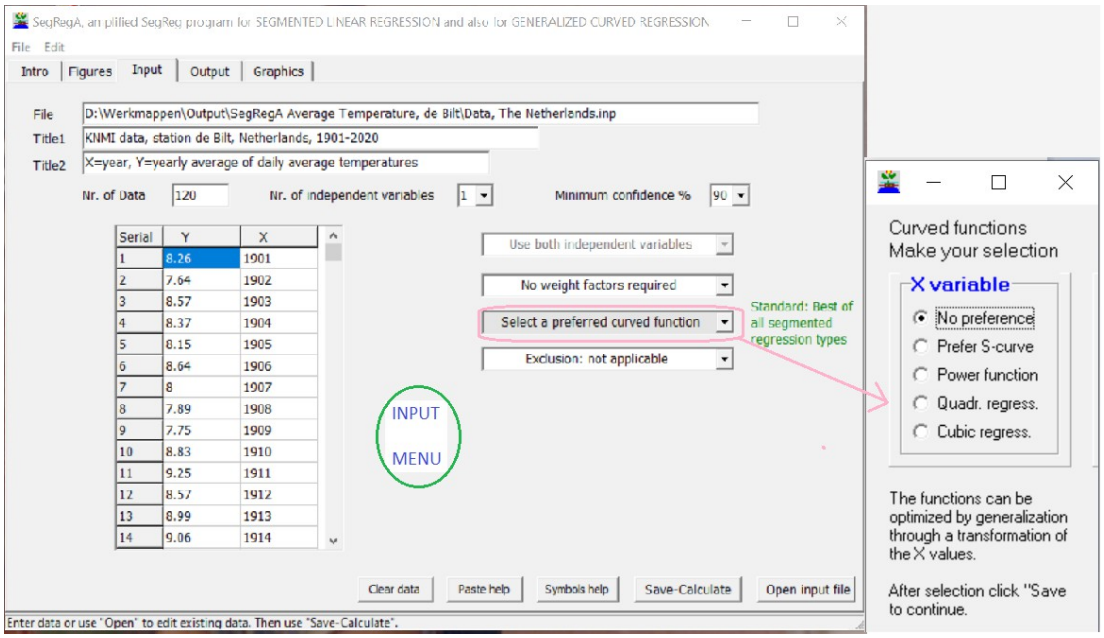
10. Appendix
(SegRegA input menu sheet, generalized cubic regression)

The input menu sheet of SegRegA with a selection table for segmented regressions is shown in figure 7 hereunder, while the same sheet with a selection table for curved regressions is shown in figure 8 below.



Serial	Y	X
1	8.26	1901
2	7.64	1902
3	8.57	1903
4	8.37	1904
5	8.15	1905
6	8.64	1906
7	8	1907
8	7.89	1908
9	7.75	1909
10	8.83	1910
11	9.25	1911
12	8.57	1912
13	8.99	1913
14	9.06	1914

Figure 7. Input menu of SegRegA with selection options of segmented linear regression types. These types (see figures 2 and 3) are found by optimizing the position of the breakpoint to obtain the best fit and by testing their statistical significance.



Serial	Y	X
1	8.26	1901
2	7.64	1902
3	8.57	1903
4	8.37	1904
5	8.15	1905
6	8.64	1906
7	8	1907
8	7.89	1908
9	7.75	1909
10	8.83	1910
11	9.25	1911
12	8.57	1912
13	8.99	1913
14	9.06	1914

Figure 8. Input menu of SegRegA with selection options of curved regression types.

The appendix is closed by demonstrating a generalized cubic regression in figure 9.

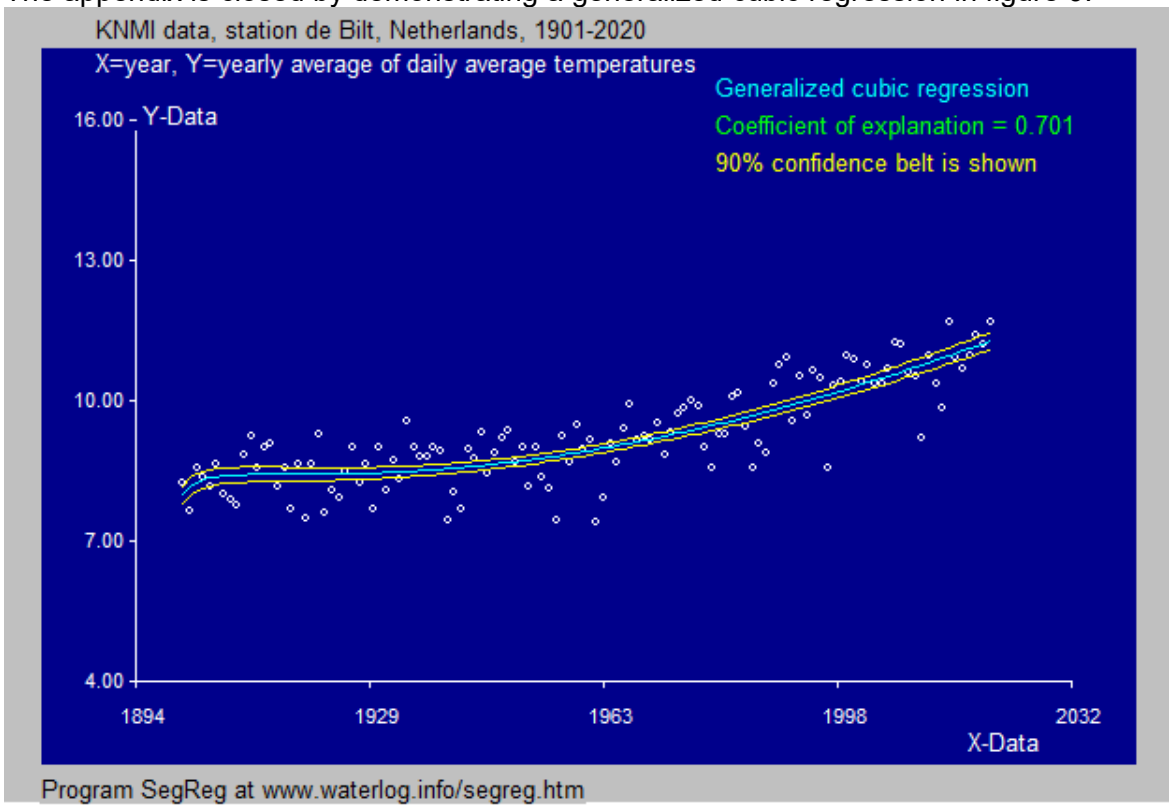


Figure 9 Generalized cubic regression with the highest coefficient of explanation of all. The curve shows a small wave in the first years, but thereafter it looks similar to the quadratic curve (figure 4) and the power curve (figure 5).

The regression is called generalized because the X-data are transformed to Xt data by raising them to a power E that is optimized to obtain the best fit.

Here, the expression is:

$$Y = A * Xt^{3} + B * Xt^{2} + C * Xt + D$$

where

$A = 0.038$ $B = -0.37$ $C = 1.21$ and $D = 7.12$ while $Xt = (X - X_{minim})^{E}$ with the value of power E amounting to 0.42.

Since the value of E is smaller than 2/3, the order of the equation actually becomes less than quadratic.