Discontinuities in time series and probability distributions of temperature in the Netherlands as a result of global warming; analyses with SegReg and CumFreq models.

R.J.Oosterbaan, 24-09-2019. On www.waterlog.info public domain

Abstract

Temperatures were measured by the weather station of the Royal Dutch Meteorological Institute (KNMI) in de Bilt, Netherlands, from 1900 to present.
The annual average of daily means and the annual maximum of daily maxima of the temperature were used for analysis with segmented regressions of time series and composite probability distributions.
Both methods use the principle of minimization of the sum of the squares of deviations of the observed values from the values calculated by the models to optimize the values of the separation points or (break points).
Amazingly, the breakpoints in the segmented time series and the separation points of the probability distributions coincide.
It appears that both the segmented regression lines and the composite probability distributions confirm mutually the existence of a marked change of temperature characteristics in the same year and at the same value towards the end of the 20th century.
Also, in the graphs presented it can be seen that both methods confirm the effects of global warming.

Contents:

1. Segmented regression of time series

1a. Annual averages of daily mean temperatures

The SegReg model [Ref.1] enables regression of variable Y upon variable X whereby the X domain is divided into two parts, separated by a break point (BP). The optimal breakpoint is found by minimization of the sum of the squares of deviations of the observed Y values from the Y values calculated by the model.

The results of the application of SegReg to the annual averages of daily mean temperatures are shown in figure 1.

The break point (BP) here is found at year 1988. The regression line to the left of BP is written as Y = 0.014 X – 17.9 while that to the right as Y = 0.023 X – 35.2. Here, Y is temperature (in centigrades) and X is year. At BP there is a jump from Ypb = 9.3 to 10.2, a difference of 0.9 degrees.

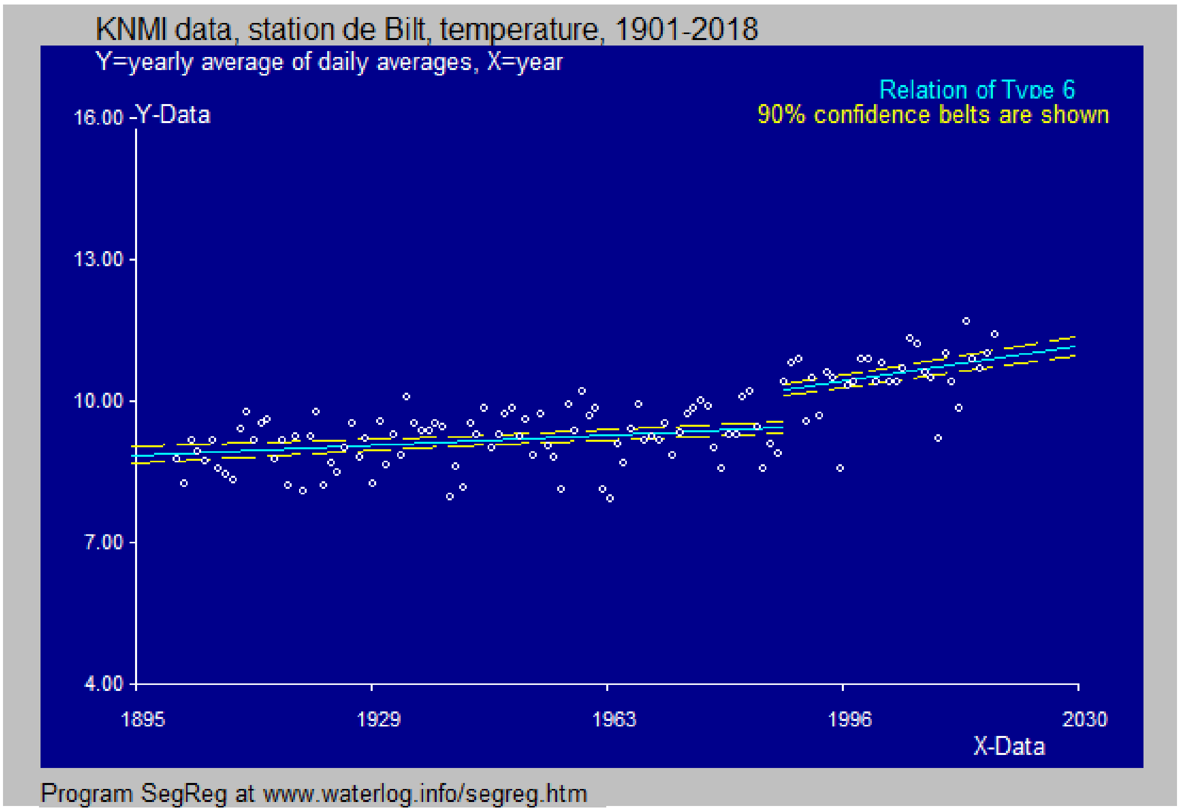


Figure 1. Initially the average temperature rises slowly up to year X=1988. Thereafter a jump can be seen followed by a faster rise. This may be the result of global warming.

A similar figure has been presented before [Ref. 2]

<u>1b. Annual maxima of daily maximum temperatures</u>

The results of the application of Segreg to the annual maxim of daily maximum temperatures are shown in figure 2.

In figure 2 there is a jump at the break point (BP) in year 1993. The regression line to the left of BP is written as Y = 0.046 X – 59.5 while that to the right as Y = 33.1. Here, Y is temperature (in centigrades) and X is year. At BP there is a jump from Ypb = 31.6 to 33.1, a difference of 1.5 degrees.

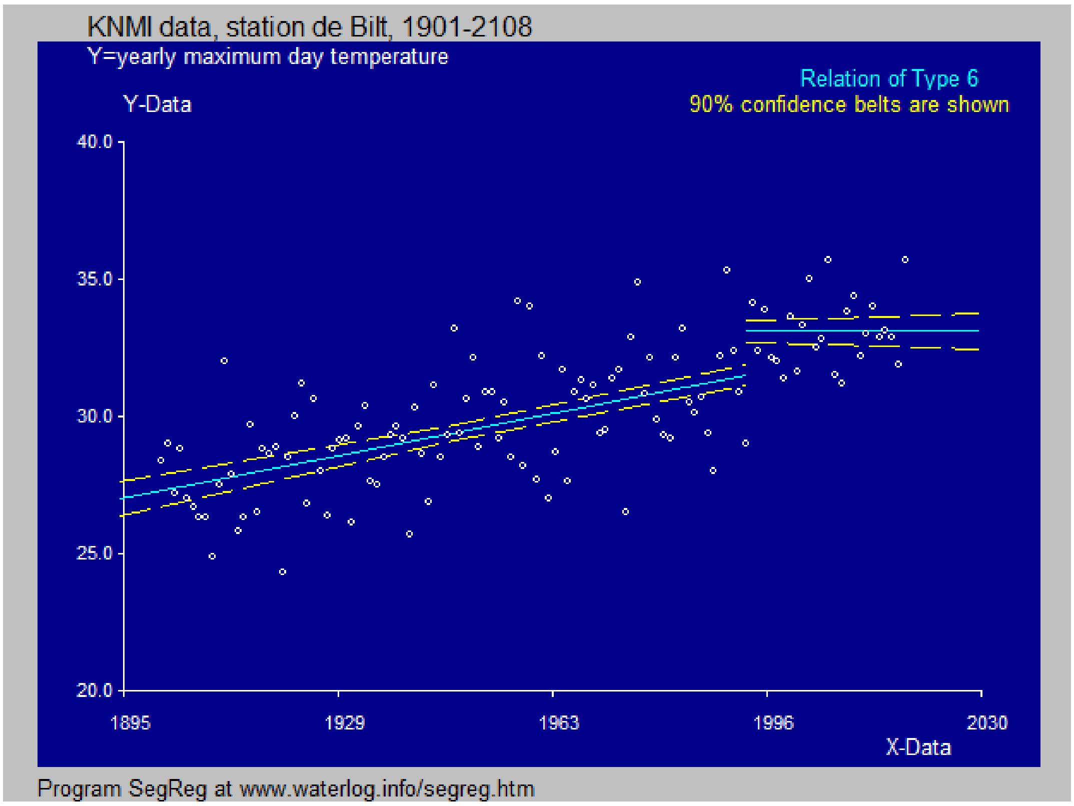


Figure 2. Initially the maximum temperature rises relatively rapidly up to year X=1993. Thereafter a jump can be seen. This may be the result of global warming. The record temperature of 2019 (37.5 degrees) has not yet been included in this analysis.

A similar figure has been presented before [Ref. 3].

2. Composite probability distributions

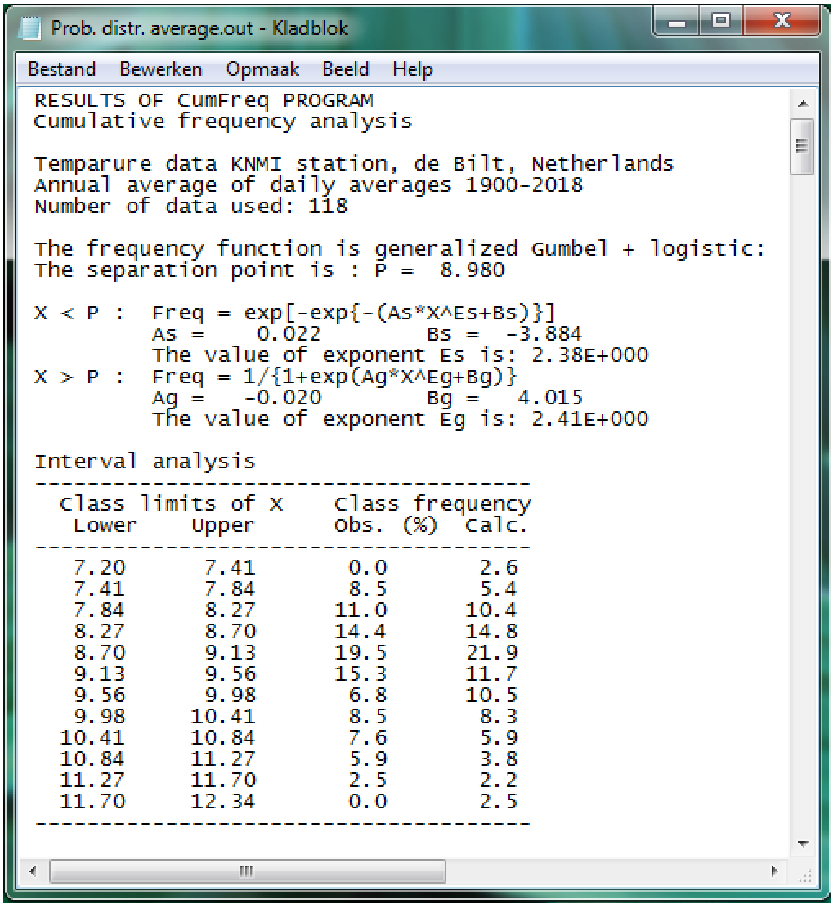
<u>2a. Annual averages of daily mean temperatures</u>

The CumFreqA model [Ref.4] enables the fitting of composite probability distributions to a set of data X whereby the X domain is divided into two parts, separated by a break point (P). The optimal breakpoint is found by minimization of the sum of the squares of deviations of the observed Y values from the Y values calculated by the model (the least squares or LS method).

The results of the application of CumFreqA to the annual averages of daily mean temperatures are shown in figures 3 and 4.

Figure 3 demonstrates that the break point (P) is found at temperature X = 9.0, which is surprisingly close to the point X = 9.3 of the jump discussed in section 1a.

There are different probability distributions to the left and right of P: a generalized *Gumbel distribution* and a generalized *logistic distribution*. The generalization consists of raising the X values to an exponent E before applying the distribution fit. The optimal value of E is, like the one of P, found with the LS method. Use of the optimized E value enhances the goodness of fit. The same holds for the large number of optional combinations from which the best fitting is selected. The use of the Gumbel distribution has been discussed before [Ref. 5]


Prob. distr. average.out - Kladblok

Bestand Bewerken Opmaak Beeld Help

RESULTS OF CumFreq PROGRAM
Cumulative frequency analysis

Temparure data KNMI station, de Bilt, Netherlands
Annual average of daily averages 1900-2018
Number of data used: 118

The frequency function is generalized Gumbel + logistic:
The separation point is : P = 8.980

```
X < P :  Freq = exp[-exp{-(As*X^Es+Bs)}]
         As =     0.022          Bs =  -3.884
         The value of exponent Es is: 2.38E+000
X > P :  Freq = 1/{1+exp(Ag*X^Eg+Bg)}
         Ag =    -0.020          Bg =   4.015
         The value of exponent Eg is: 2.41E+000
```

Interval analysis

Class limits of X Lower	Upper	Class frequency Obs. (%)	Calc.
7.20	7.41	0.0	2.6
7.41	7.84	8.5	5.4
7.84	8.27	11.0	10.4
8.27	8.70	14.4	14.8
8.70	9.13	19.5	21.9
9.13	9.56	15.3	11.7
9.56	9.98	6.8	10.5
9.98	10.41	8.5	8.3
10.41	10.84	7.6	5.9
10.84	11.27	5.9	3.8
11.27	11.70	2.5	2.2
11.70	12.34	0.0	2.5

Figure 3. Part of the output file of CumFreqA for the case of the average temperature showing the parameters of the two probability distributions used

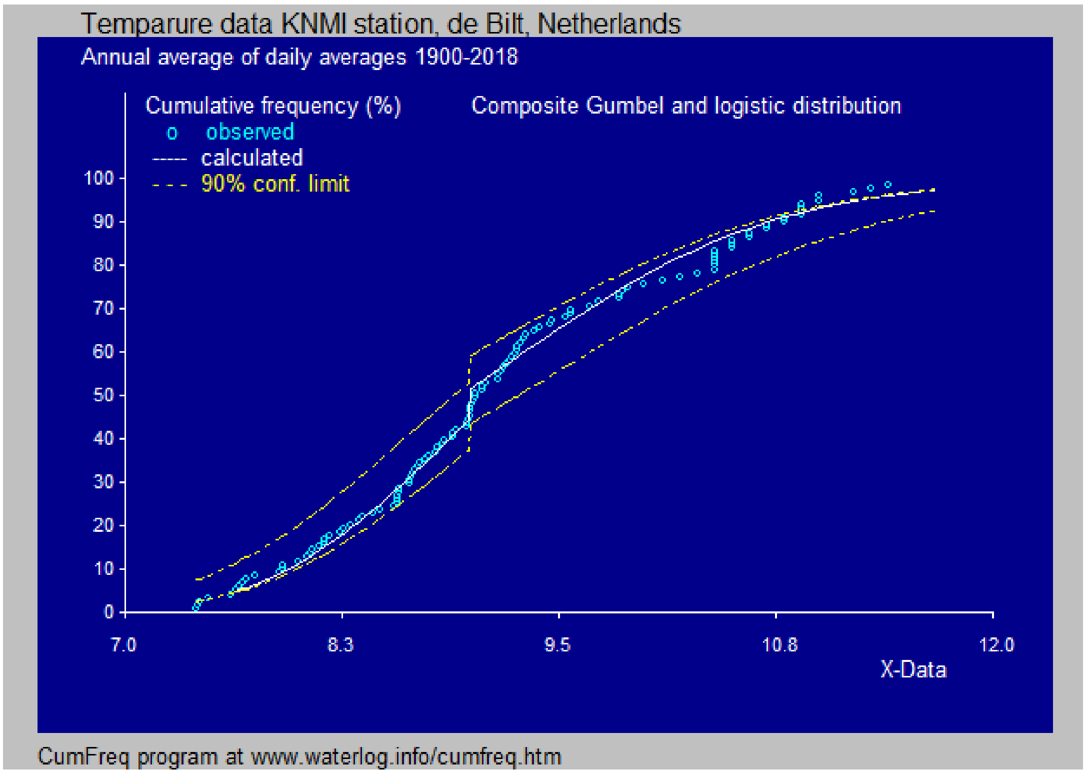


Figure 4. Graph of the two probability distributions described in figure 3 fitted to the average temperature data. The jump at X = 9 is clearly visible.

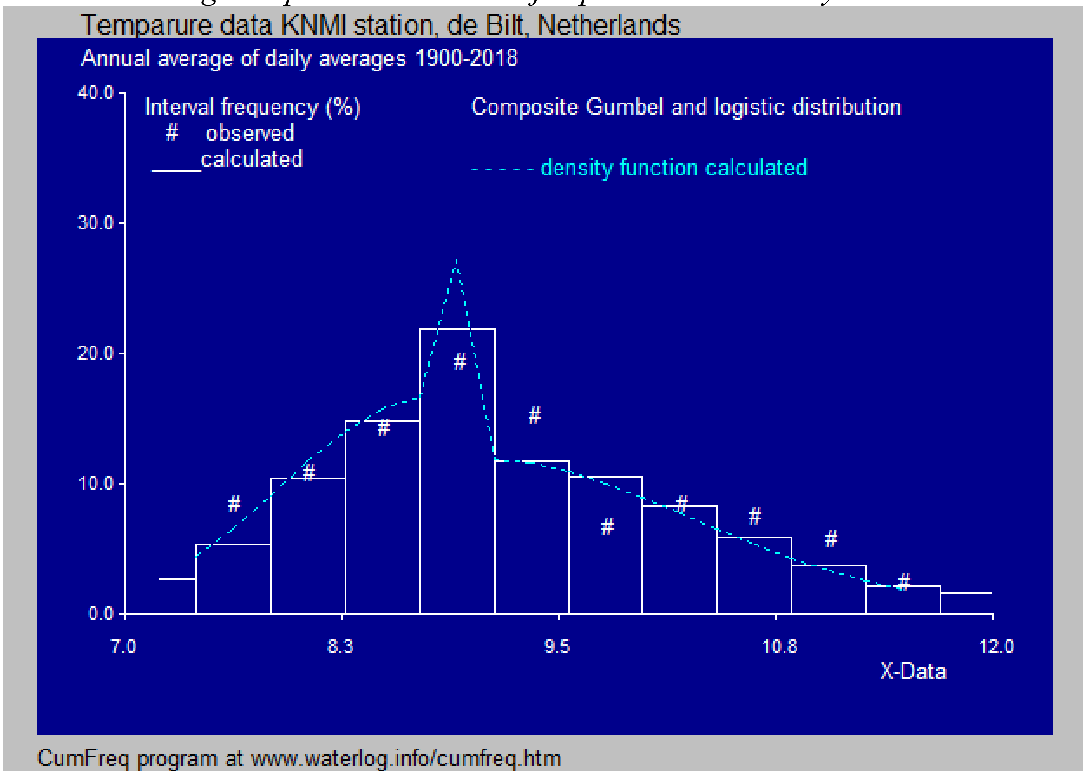


Figure 5. Histogram and probability density function for the cumulative distribution function depicted in figure 4 and the intervals shown in figure 3.

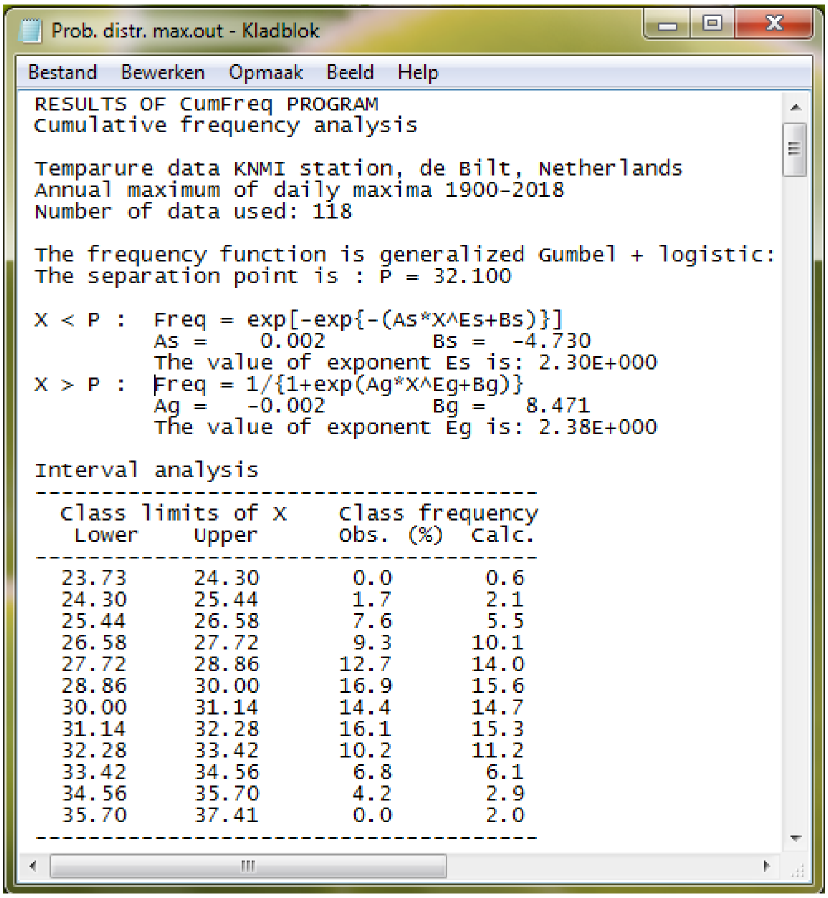
2b. Annual maxima of daily maximum temperatures

The results of the application of CumFreqA to the annual maxima of daily maximum temperatures are shown in figures 6 and 7.

Figure 6 demonstrates that the break point (P) is found at temperature X = 32.1, which is surprisingly close to the points X = 31.6 - 33.1 of the jump discussed in section 1b.

There are different probability distributions to the left and right of P: a generalized *Gumbel distribution* and a generalized *logistic distribution*. It is remarkable that this is the same combination of distributions as seen in figure 3 for the average values, they are selected from a large number of options.

The versatility of the logistic distribution has been discussed before [Ref. 6] .



```
Prob. distr. max.out - Kladblok
Bestand  Bewerken  Opmaak  Beeld  Help

RESULTS OF CumFreq PROGRAM
Cumulative frequency analysis

Temparure data KNMI station, de Bilt, Netherlands
Annual maximum of daily maxima 1900-2018
Number of data used: 118

The frequency function is generalized Gumbel + logistic:
The separation point is : P = 32.100

X < P :  Freq = exp[-exp{-(As*X^Es+Bs)}]
         As =     0.002        Bs =   -4.730
         The value of exponent Es is: 2.30E+000
X > P :  Freq = 1/{1+exp(Ag*X^Eg+Bg)}
         Ag =    -0.002        Bg =    8.471
         The value of exponent Eg is: 2.38E+000

Interval analysis
--------------------------------------
  Class limits of X     Class frequency
   Lower      Upper     Obs. (%)  Calc.
--------------------------------------
  23.73      24.30       0.0       0.6
  24.30      25.44       1.7       2.1
  25.44      26.58       7.6       5.5
  26.58      27.72       9.3      10.1
  27.72      28.86      12.7      14.0
  28.86      30.00      16.9      15.6
  30.00      31.14      14.4      14.7
  31.14      32.28      16.1      15.3
  32.28      33.42      10.2      11.2
  33.42      34.56       6.8       6.1
  34.56      35.70       4.2       2.9
  35.70      37.41       0.0       2.0
--------------------------------------
```

Figure 6. Part of the output file of CumFreqA for the case of the maximum temperatures showing the parameters of the two probability distributions used

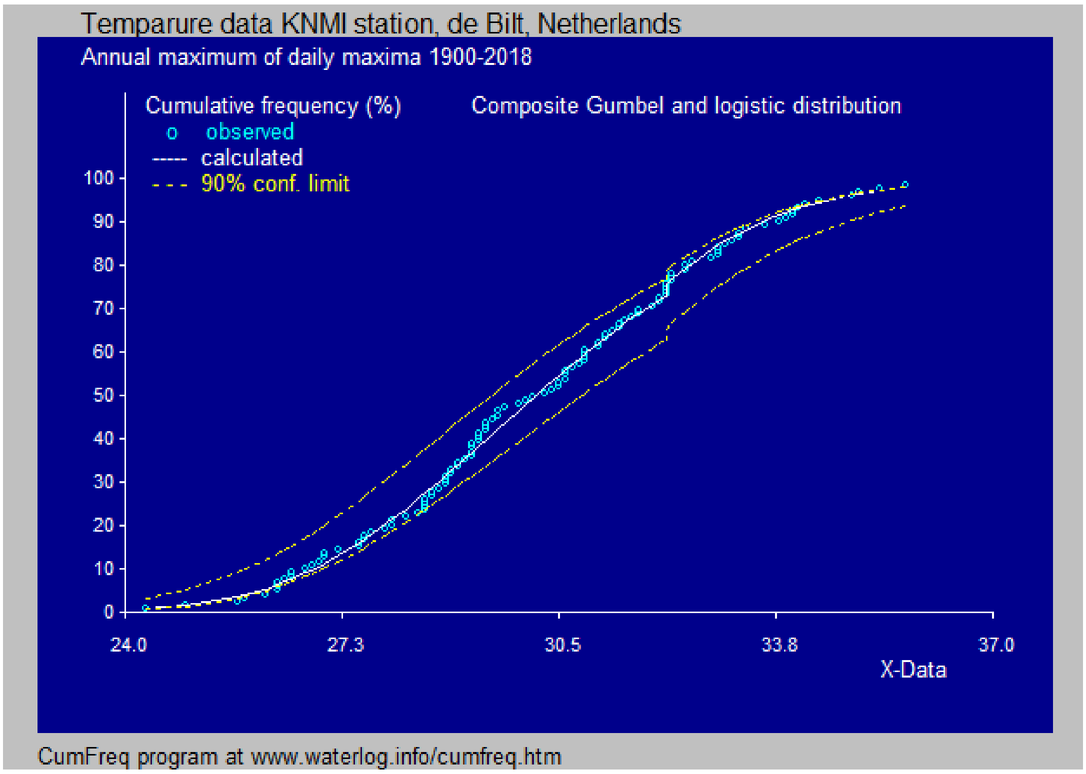


Figure 7. Graph of the two probability distributions described in figure 6 fitted to the maximum temperature data. The jump at X = 32.1 is visible.

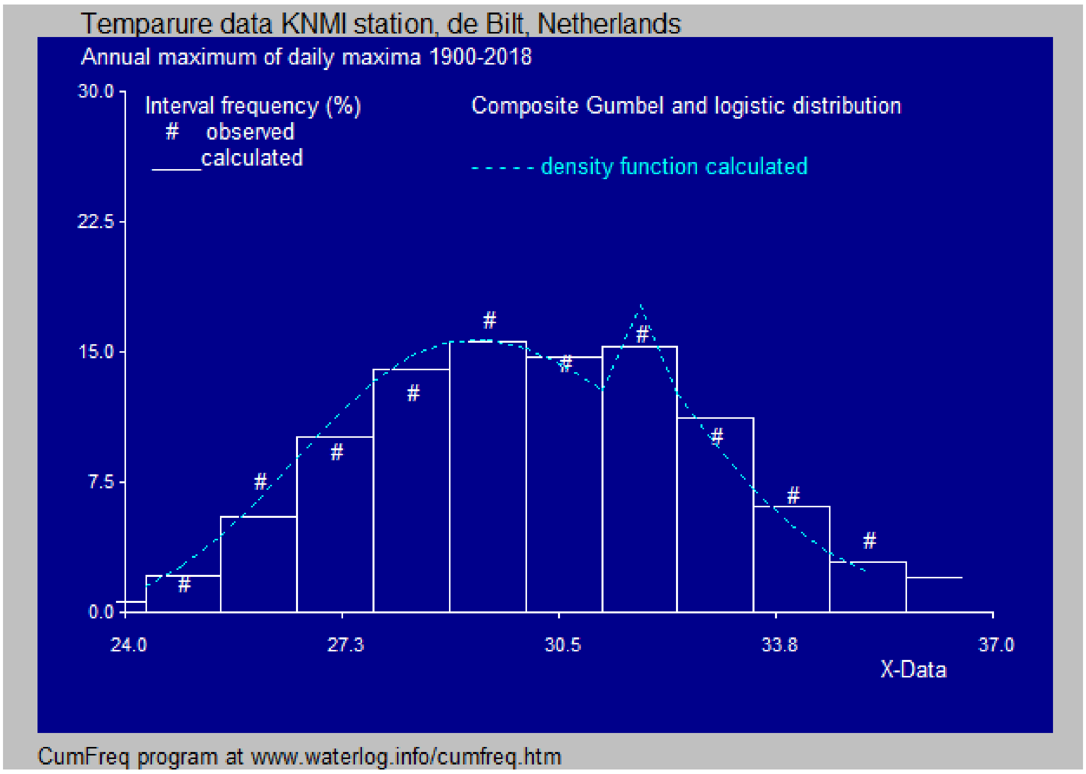


Figure 8. Histogram and probability density function for the cumulative distribution function depicted in figure 7 and the intervals shown in figure 6. The change at X = 32 degrees is clearly visible.

3. Comparison

The SegReg cases have shown an increasing trend of both average and maximum temperatures from 1900 to 2018. Sudden temperature increases occurred in respectively 1988 and 1993, say around 1990. The reason for these jumps is yet unclear.

The CumFreq cases also showed a change of type of probability distribution. The points of change for both the average and maximum temperatures corresponded with the points of change found with the SegReg analysis, respectively at around 9 and *32* degrees centigrades.

4. Conclusion

Both the segmented regression and the fitting of composite probability yield the same result for the sudden changes in temperature towards the end of the twentieth century. They confirm the the temperatures are rising continuously, probably owing to global warming.

5. References

[Ref. 1] SegReg, free software for segmented regression analysis. Download from: https://www.waterlog.info/segreg.htm

[Ref. 2] On line:
https://www.researchgate.net/publication/335541155_Trend_of_annual_averages_of_daily_average_temperatures_in_the_Netherlands_since_1900_first_showing_slow_and_then_fast_increases

[Ref. 3] On line:
https://www.researchgate.net/publication/335757889_Trend_of_the_annually_maximum_temperatures_in_the_Netherlands_since_1900_first_showing_slow_and_after_1988_faster_increases

[Ref. 4] CumFreqA, free software for composite probability distribution fitting.
Download from: https://www.waterlog.info/segreg.htm

[Ref.5] On line:
https://www.researchgate.net/publication/332466331_SOFTWARE_FOR_GENERALIZED_AND_COMPOSITE_PROBABILITY_DISTRIBUTIONS

[Ref. 6] On line:
https://www.researchgate.net/publication/335022301_FITTING_THE_VERSATILE_LINEARIZED_COMPOSITE_AND_GENERALIZED_LOGISTIC_PROBABILITY_DISTRIBUTION_TO_A_DATA_SET