Comparing the regressions of Y-X data by means of the amplified power function using Solver in Excel and SegRegA

R.J.Oosterbaan, april 2021-04-13 On www.waterlog.info public domain

Abstract

The amplified power function reads: $Y = A*(X-D)^E + B$ and represents a curved line. In statistical regression, to solve this equation in the Excel spreadsheet, one needs the help of the Solver application because a standard solution is not available. The required manual operations are tedious. SegRegA offers the same procedure automatically without manual operations, while it gives a solution with a better fit of the simulated to the observed Y values.

Contents

1. Introduction

In some cases it is advisable to use the power function for the regression of Y values upon X values. The function uses the parameters A, D , E and B (see the equation in the abstract), which are to be optimized in order to obtain the best possible fit of the calculated (simulated) Y values (Ysim) to the observed Y values (Yobs). The optimization is mostly done by minimizing the sum of squares of the differences Yobs-Ysim.

2. Example solution with Excel solver

The excel procedure for the regression with the power function is demonstrated in *figure 1* below while in the subscript of that figure the operations and results of the Solver application are described.

Microsoft Excel Power function

	M	N	O	P	Q	R	S	T	U	V	W
1											
2		***Simulation of Y using the power function: Ysim = A*(X-D)^E + B = A*Xv + B***									
3	best										
4		**Parameters optimized using Solver**									
5		Deduction	Factor	Addition	exponent						
6		D	A	B	E						
7		1513.6	107.9	-167.5	0.150						
8						***Ysimulated***					
9	Serial	***Y observed***			Xv =	Ysim =					
10	number	Yobs (%)	X (year)	X - D	(X-D)^E	A*Xv+B	Y-Ysim	(Y-Ysim)^2		(Yobs-Yav)^2	
11	1	19.5	1560	46.37	1.78	24.45	-4.95	24.46		1177.96	
12	2	12.0	1572	58.37	1.84	31.19	-19.19	368.42		1749.03	
13	3	35.0	1580	66.37	1.88	35.06	-0.06	0.00		354.25	
14	4	23.0	1580	66.37	1.88	35.06	-12.06	145.52		949.96	
15	5	55.0	1600	86.37	1.95	43.23	11.77	138.43		1.39	
16	6	54.0	1600	86.37	1.95	43.23	10.77	115.90		0.03	
17	7	61.0	1600	86.37	1.95	43.23	17.77	315.62		51.53	
18	8	46.0	1600	86.37	1.95	43.23	2.77	7.65		61.17	
19	9	64.0	1660	146.37	2.11	60.60	3.40	11.53		103.60	
20	10	84.0	1660	146.37	2.11	60.60	23.40	547.39		910.75	
21	11	51.0	1660	146.37	2.11	60.60	-9.60	92.23		7.96	
22	12	82.0	1820	306.37	2.36	87.36	-5.36	28.72		794.03	
23	13	89.0	1820	306.37	2.36	87.36	1.64	2.69		1237.53	
24	14	78.0	1820	306.37	2.36	87.36	-9.36	87.58		584.60	
25		53.82						1886		7984	
26		Yav=average of Yobs						Sum (Y-Ysim)^2		Sum (Yobs-Yav)^2	
27								**(S1) to be minimized**		**(S2)**	
28											
29								R^2 = 1 - S1/S2 =		0.76	
30											

Figure 1*. Solution of regression of by a power function in Excel. In columns N and O one finds the 14 data values Yobs and X. In the handmade columns P, Q and R one finds respectively the reduced X values by subtracting the D (deduction) value shown in cell N7, the transformed Xvalue (Xv) by raising X-D to the power E shown in cell Q7, and the simulated Y values (Ysim) by multiplying Xv with factor A in cell O7 and adding the addition B found in cell P7.*

The sum of squares of the differences Yobs-Ysim (S1) is seen in cell T25. The solver application in Excel has minimized S1 by determining the optimal values of D, A, B and E which are found to be respectively 1513.6, 107.9, -167.5 and 0.150 (see cells N7, O7, P7 and Q7) so that the power equation reads Ysim = $(X-1513.6)^{0.150} - 167.5$*. The* R^2 *value is calculated in cell V29 and found to be 0.76 so that the goodness of fit is 76%*

The Excel graphics show the results as in *figure 2* hereunder.

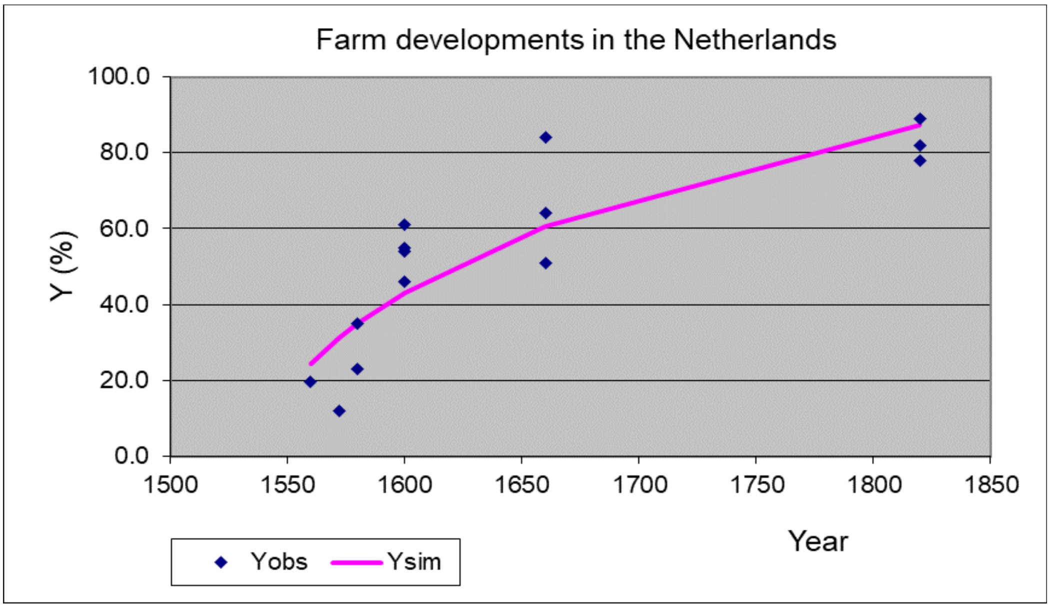


Figure 2*. Graph of the Excel solver results dealt with in* ***figure 1****.*

3. Example solution with SegRegA using the same data

SegRegA (the A stands for the amplified SegReg calculator as it does not only do Segmented Regressions but also curved regressions, *Reference 1*) has an input, output and graphic user menus. The input menu is shown in *figure 3*. I uses the same data as employed in Excel (*figure 1*).

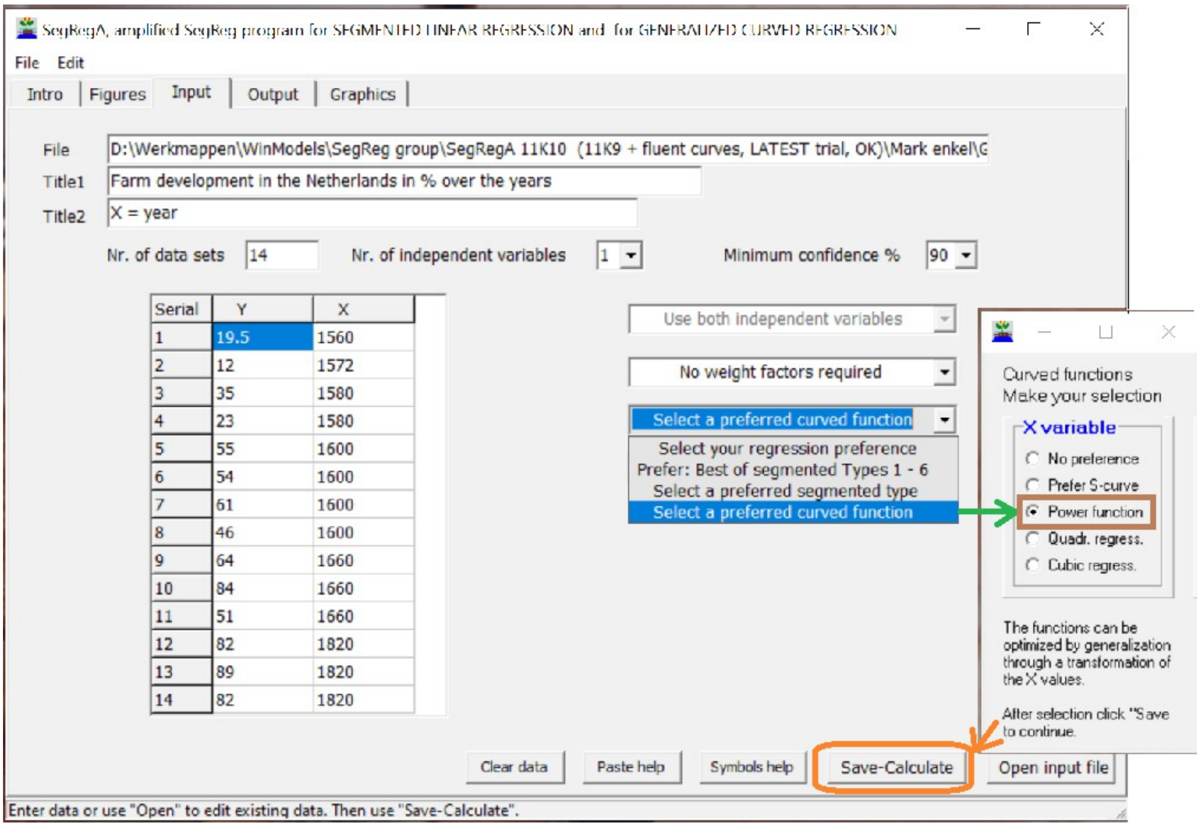


Figure 3. *Input user interface of SegRegA. The input data are entered in the table. The selection box has been set to "Select a preferred curved function, blue rectangle), whereafter the option "Power function" has been selected (brown rectangle). The calculator gives the possibility to use more than 1 independent variable as well as the determination of the confidence level (here 90%). The calculations are done clicking the "Save-Calculate" button (orange rectangle).*

The top part of the output table is replicated in ***table 1***.

***Table 1**. Top part of the output table. It can be seen that the power function reads:*
Yc =323 (1554)^0.60^ − 365*

$Yc = 323 * (1554)^{0.60} - 365$

*Yc corresponds to Ysim in **figure 1 and 2**, while C corresponds to D. The power E in Excel correspond to P in SegRegA.*
*The R^2 in **figure 1** corresponds to ExplCoeff in this table and it equals 0.827 so that the goodness of fit is 82.7%, which is higher than the 76% in the Excel case.*

```
Farm developments in the Netherlands in % (Y)
X = year

1st Transformation  :  Xt = X - C            C = 1544.400
2nd Transformation  :  Xv = Xt ^ P
Linearization       : Y  = A * Xv + B
Regression result   : Power P = 0.060    A =  3.23E+002    B =  -3.65E+002
Ycalculated according to power curve :
                                         Yc = B + A * Xt ^ P
Parameters:
NrOfData =    14          AvXv  = 1.29E+000          AvY = 54.107
 StDevXv =  96.441     StDevY  = 14.470       StErr(Y-Yc) = 10.445
Explanations by regression:
 CorrCoeffSq(Y,X) =  0.668    ExplCoeff(Yc,X) =  0.823
```

Part of the output sheet of SegRegA

The graph produced by SegRegA is copied to *figure 4*.

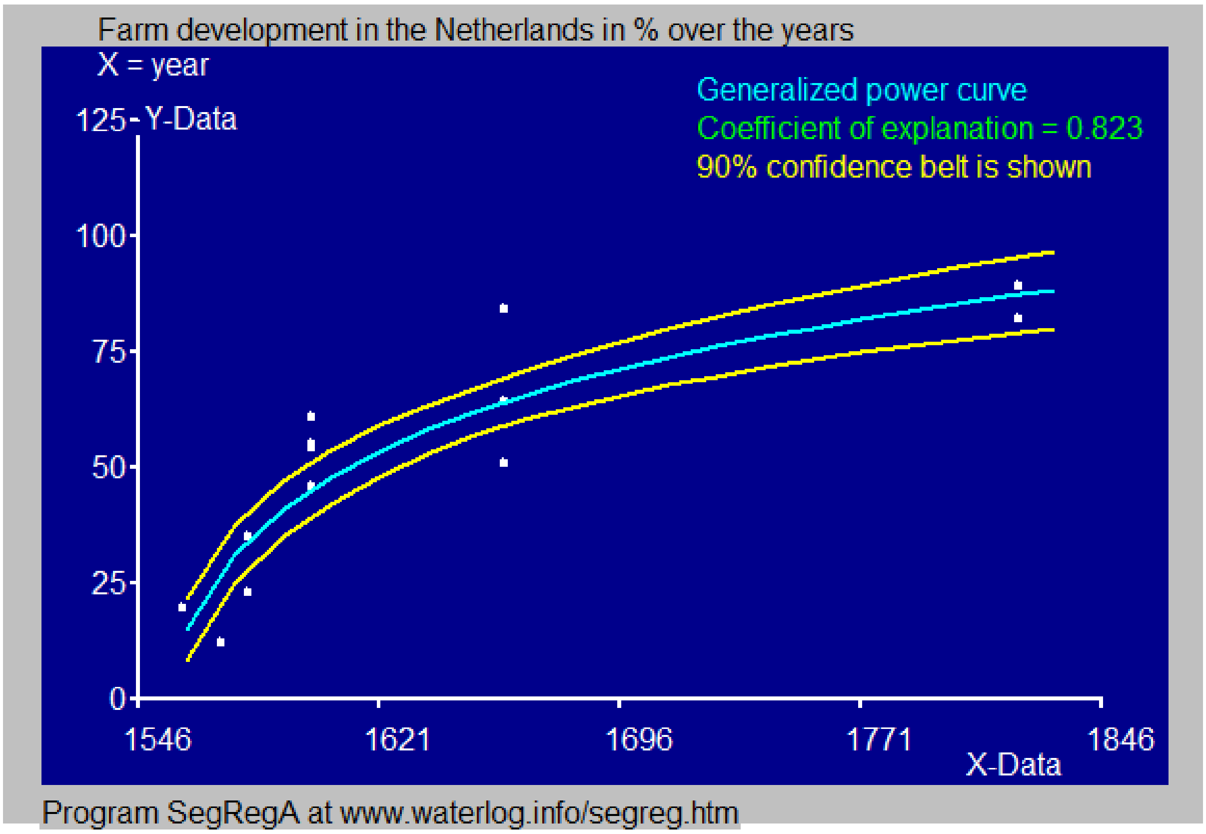
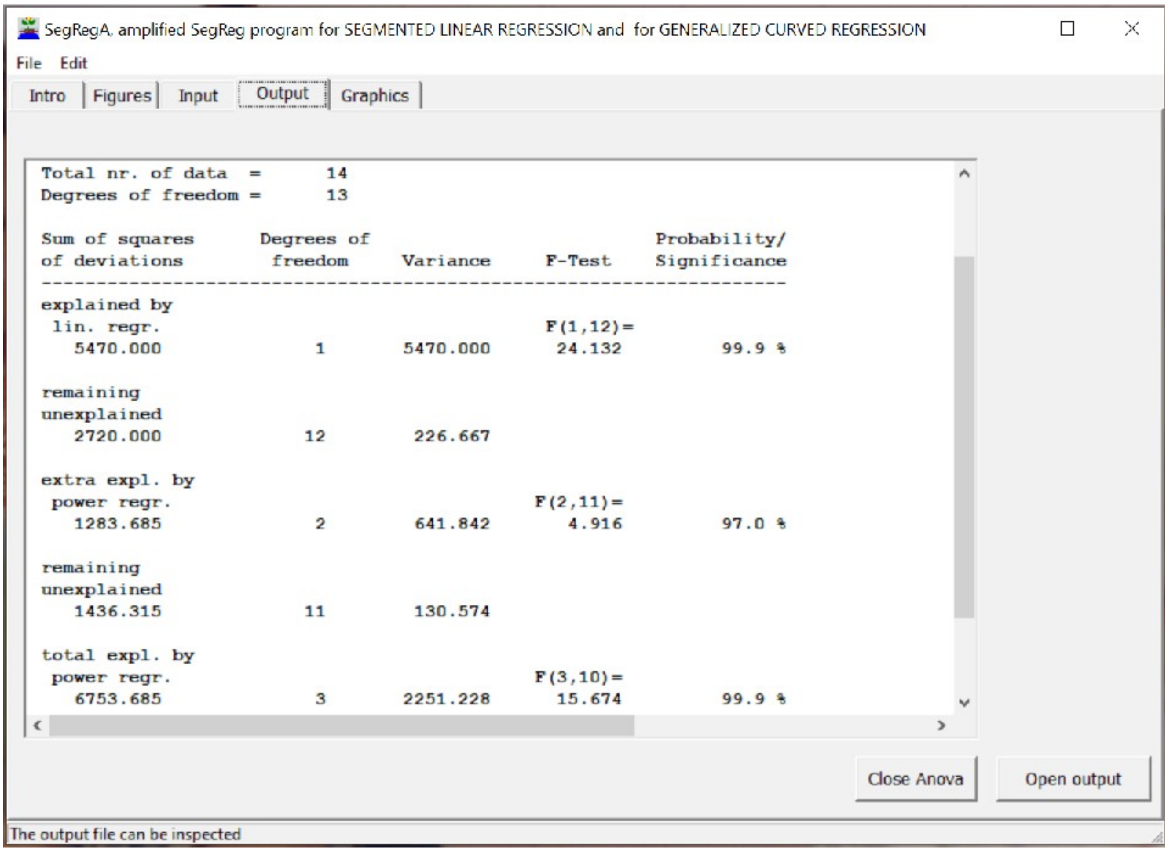


Figure 4. *Graph made by SegRegA which corresponds to the Excel graph in* ***figure 2****,*
but it is more smooth and shows a confidence belt while the fit is better.

In addition SegRegA presents an ANOVA (Analysis of Variance) table to test if the function used gives a statistically significant improvement over the regression to a linear (straight line) function. The test is done with Fisher's F-test [*Reference 2*] The table is shown in *figure 5*.




```
Total nr. of data   =      14
Degrees of freedom =      13

Sum of squares      Degrees of                       Probability/
of deviations        freedom     Variance    F-Test   Significance
-------------------------------------------------------------------
explained by
 lin. regr.                                  F(1,12)=
   5470.000            1        5470.000     24.132        99.9 %

remaining
unexplained
   2720.000           12         226.667

extra expl. by
 power regr.                                F(2,11)=
   1283.685            2         641.842      4.916        97.0 %

remaining
unexplained
   1436.315           11         130.574

total expl. by
 power regr.                                F(3,10)=
   6753.685            3        2251.228     15.674        99.9 %
```

Figure 5. *ANOVA table prepared by SegRegA. The improvement of the power function compared to a linear regression is statistically highly significant as there is 97% chance that the improvement is true.*

4. Conclusion

The user interface for the input data in SegReg requires a cut and paste of the data into the table and a selection of the type of regression analysis desired. That is all.

The Excel procedure requires the completion of several columns based on the type of regression equation used. Further it needs separate cells for the parameters needed and a summation of the sum of squares of Yobs-Ysim. Thereafter the Solver application can be used with a definition of the parameter cells for optimization and the sum of squares cell for minimization. The preparation in Excel are far more elaborate than for SegRegA.

SegRegA and Excel give different results, summarized in the following table.

Parameter	A	B	C resp. D	E resp. P	Goodness of fit (%)
Excel	108	-168	1514	0.150	76
SegRegA	323	-365	1544	0.060	82

$Y = A\,(X-C)^E + B$

SegRegA produces a considerably higher goodness of fit.

5. References

[**Reference 1**] SegRegA, free calculator for segmented linear and curved regression. On line: https://www.waterlog.info/zip/CumFreqA.zip

See also for the cubic regression:

The potato variety "927" tested at the Salt Farm Texel, The Netherlands, proved to be highly salt tolerant, see:
https://www.researchgate.net/publication/335789831_The_potato_variety_927_tested_at_the_Salt_Farm_TexelThe_Netherlands_proved_to_be_highly_salt_tolerant

[**Reference 2**] A free calculator for Fisher's F-test. On line: https://www.waterlog.info/f-test.htm

ANNEX: Convexity, Concavity, Ascension, Descension

The previous case (figure 4) showed an ascending convex curve. Concave curves are also possible, see next figure (A).

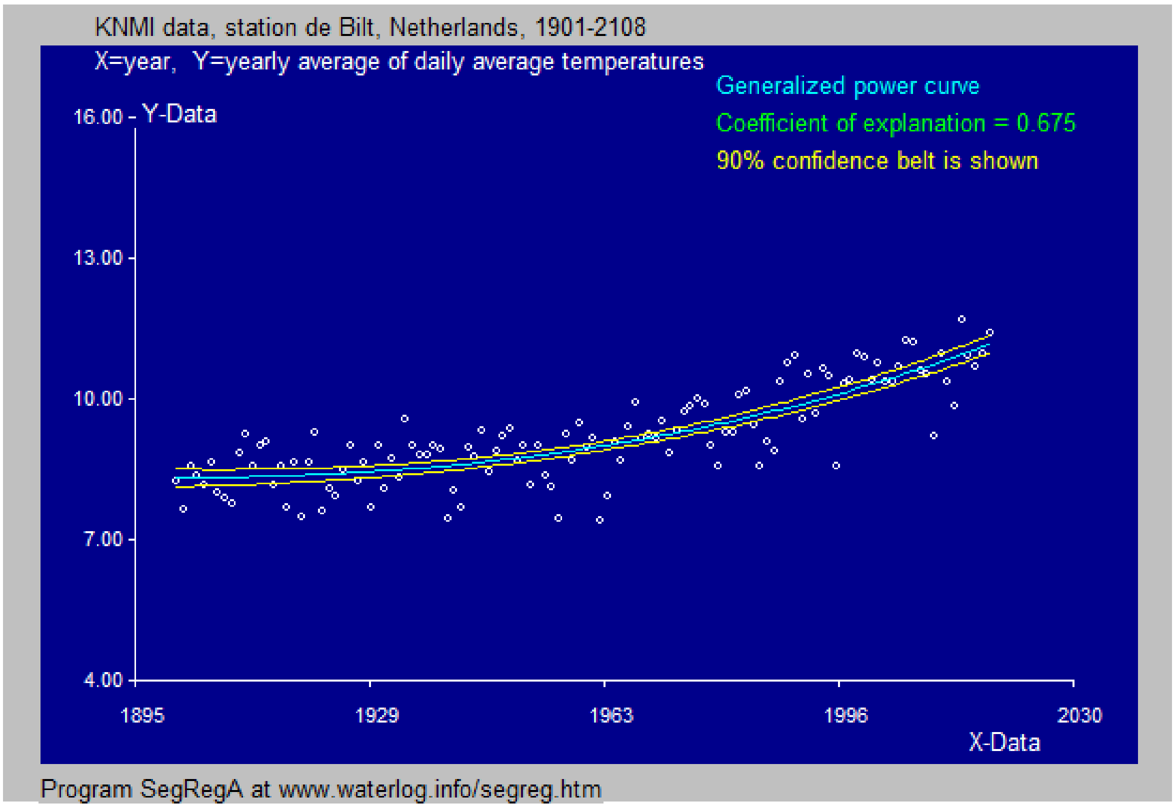


Figure A. The corresponding output data are given in figure B hereunder.

SegRegA, amplified SegReg program for SEGMENTED LINEAR REGRESSION and for GENERALIZED CURVED REGRESSION

File Edit

Intro | Figures | Input | Output | Graphics

```
Results of program SEGREG for the power curve regression of Y upon X.
Output filename: D:\SegReg group\Power data\KNMI.out

KNMI data, station de Bilt, Netherlands, 1901-2108
X=year,  Y=yearly average of daily average temperatures

1st Transformation   : Xt = X - C         C = 1881.990
2nd Transformation   : Xv = Xt ^ P
Linearization        : Y  = A * Xv + B
Regression result    : Power P = 2.640   A =    6.71E-006   B =    8.28E+000
Ycalculated according to power curve :
                       Yc = B + A * Xt ^ P
Parameters:
NrOfData =     118          AvXv   = 1.38E+005         AvY =  9.205
 StDevXv =  34.208       StDevY  =  0.636   StErr(Y-Yc) =  0.000
Explanations by regression:
  CorrCoeffSq(Y,X)  =  0.616     ExplCoeff(Yc,X)  =  0.675
```

Figure B. $Y = 82.8 + 0.0000067\,(X-1882)^{2.64}$

The reference for figures A and B is:
https://www.researchgate.net/publication/347495001_Trend_of_annual_averages_of_daily_average_temperatures_in_the_Netherlands_since_1900_first_showing_slow_and_then_fast_increases
or:
https://www.waterlog.info/pdf/average temperature.pdf

In figures 4 and A the curves are ascending, Below an example of a descending concave function is given.

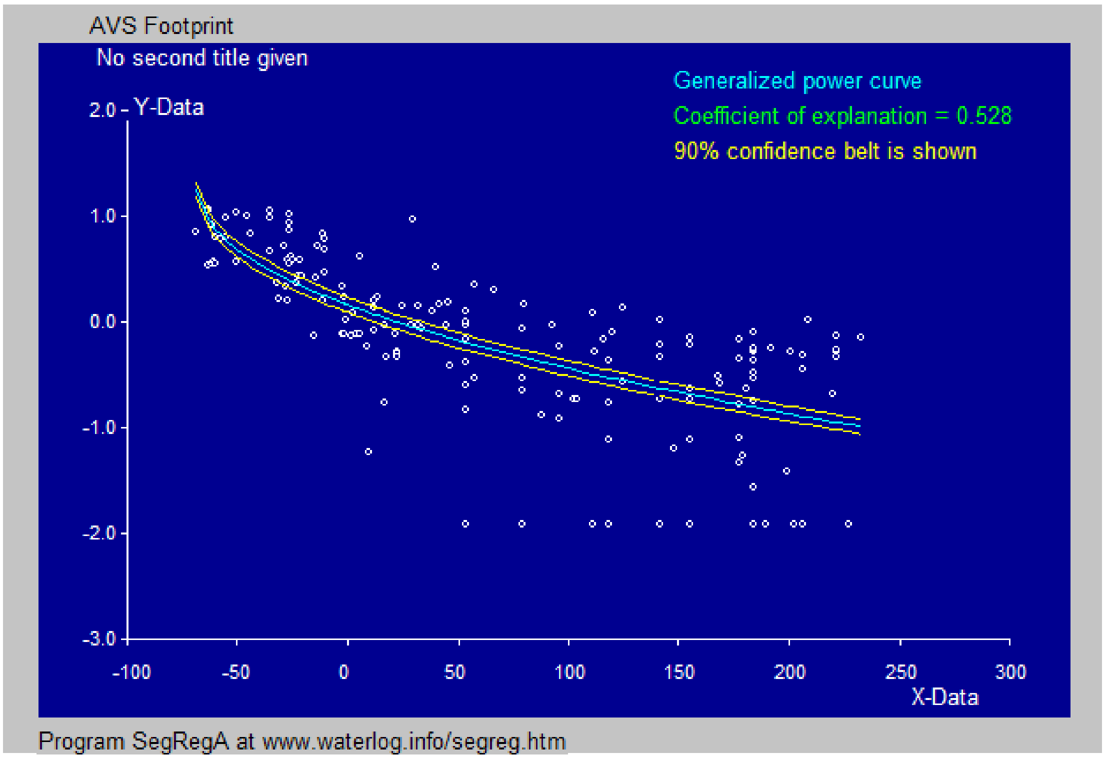


Figure C. $Y = -\ 0.145\ (X+67.4)^{0.48} + 1.26$

Finally an example of a descending convex power curve is shown in figure D.

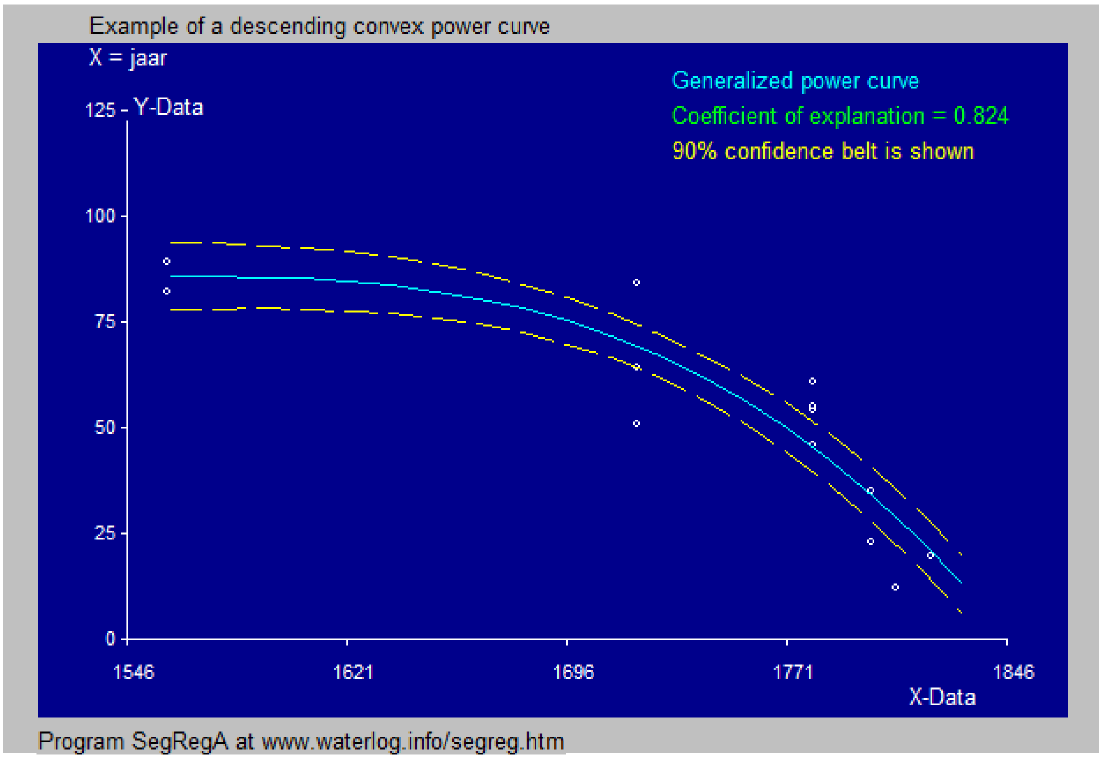


Figure D. Convex descending power curve $Y = -\ 0.0000032\ (X-1554)^{3.0} + 85.9$