Free software for the division of the relation between Y (response variable) and X (causal variable) into segments or ranges where X is influential and where X has no effect.

R.J. Oosterbaan, December 2021. https://www.waterlog.info

Abstract
Variables having influence on a the value of a certain phenomenon can be called causal or influential variables. Causal variables can often be divided into ranges (or segments) in which their effect on the response (dependent) variable are different. They can often be differentiated into zones with strong effects, weak effects and no effects (the plateau). Regressions of Y (the response or dependent variable) on X (the causal variable) may in such a case be of a curved type, for example the power function, the S-curve, the inverted (mirrored) S-curve, and the quadratic or cubic function. As an alternative one can use a segmented linear regression. In case of the presence of a segment (or range) of no effect (plateau), the regression coefficient should be statistically insignificant, that is the coefficient should statistically not be different from zero. The advantage of the use of segmented regression is that one can find a critical X value separating the range of no effect of X on Y from the range with obvious effect. The range of no effect (plateau) may found by a type of segmented regression called partial regression, that is only a regression over that part of the X values where no effect occurs. This paper discusses the different types of regression one can perform to detect the variation in degree of effect of X upon Y, with emphasis on the partial regression to see if a critical value of X can be determined. Free software is used for this purpose and various agricultural examples are given.

Contents

1. Introduction

The relation between the magnitude of an influential factor (X) on the rate of change of the response factor (Y) takes many different forms depending on the nature of the X and Y factors.

An example is given of the relation between yield (Y) of potato variety "927" and the soil salinity (X) expressed in electric conductivity of a water saturated soil extract (ECe in dS/m). The relation show in *Figure 1* has been made with the free software SegRegA (*Reference 1*). It is based on a cubic (3^{rd} degree) function after a transformation of the X values, which were raised to the power 0.73 before executing the cubic regression (reason why the cubic regression is called "generalized", *Reference 2*), resulting in the equations:

Yc = A*W^3 + B*W^2 + C*W + D, where
A = 5.25E-002   B = -7.68E-001   C = 2.65E+000   D = 3.49E+000, and
W = X^E, using E = 0.73

The symbol Yc stands for the Y value calculated according to the regression of Y upon X. It becomes

$$Yc = 0.0525\ X^{2.19} - 0.768\ X^{1.46} + 2.65\ X^{0.46} + 3.48$$

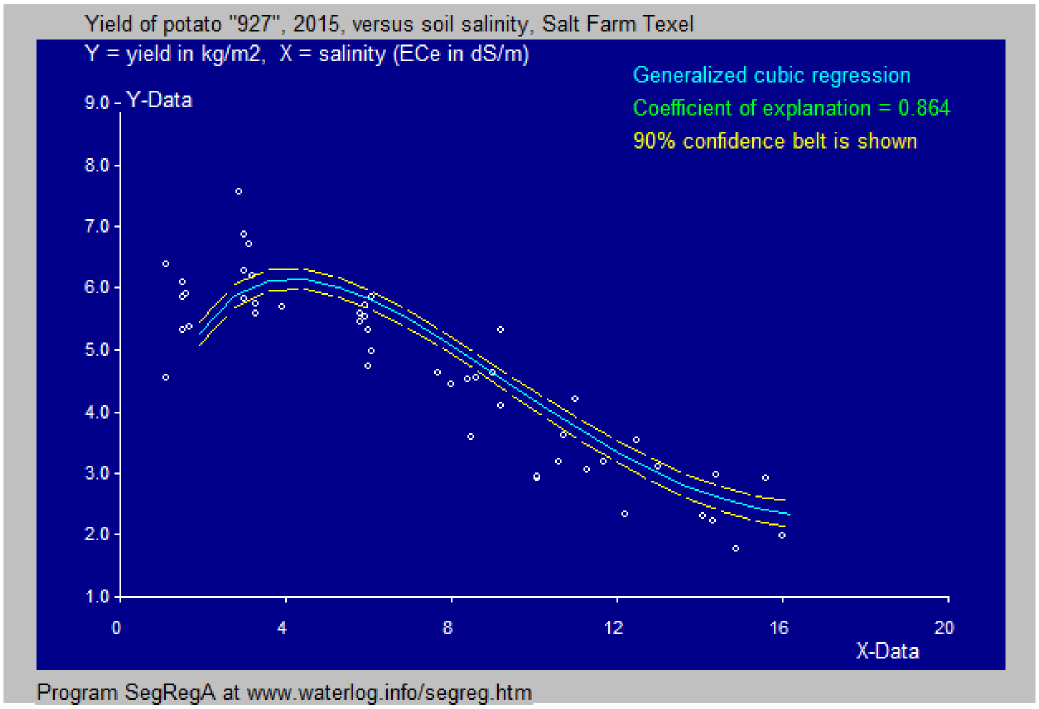

Figure 1. *Relation between yield (Y) of potato variety "927" and the soil salinity (X) expressed in electric conductivity of a water saturated soil extract (ECe in dS/m) obtained by generalized cubic regression through SegRegA*

The slope (S) of the generalized cubic regression line at X can be found from the derivative of the above Yc function:

$$S = 0.115\ X^{1.19} - 1.12\ X^{0.46} + 1.93^{-0.27}$$

A graph of the S function is depicted in *Figure 2*.

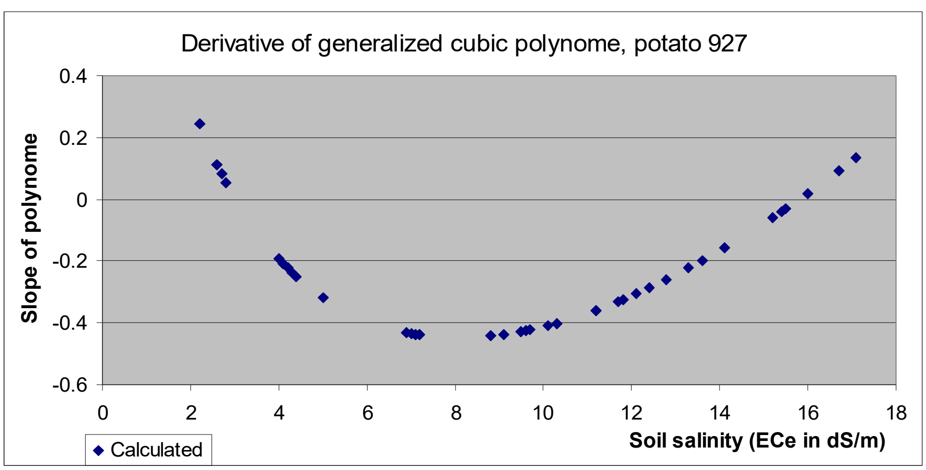

Figure 2. *The slope of the regression line in* figure 1 *initially is positive but decreasing until X=2. Between X=2 and X=8 the slope negative and decreasing. At X=8 the slope is minimal.*

Between X=8 and X=16 the slope is still negative but increasing. Thereafter the slope becomes positive.
Another example is given in *Figure 3*. It concerns an influential factor X that has a positive effect on the response (dependent) variable Y, though the response rate reduces at higher X values as the slope of the curve flattens here. The shape of the graph is an S-curve (*Reference 3*).

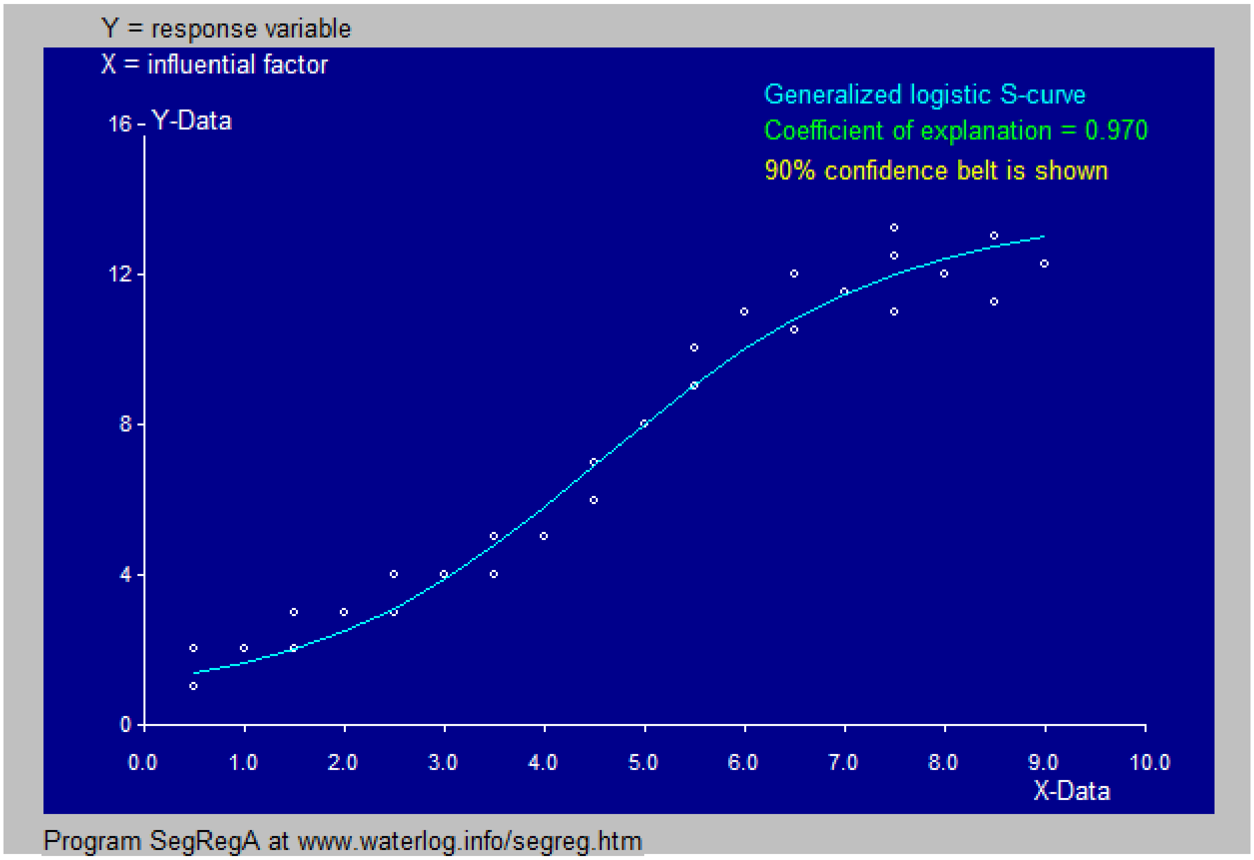

Figure 3. *Generalized logistic S-curve for the relation between an influential factor (X) and the response variable (Y). Y = C/[1+(X/A)^B] where C = 10.800, A = 3.900 B = 3.650.*

Some times the S=curve is abused (*Reference 4*), so a careful application is required.

Figure 3 does not help to find a critical value of X separating the range of clear effect of X on Y from the range with no effect (plateau). If one is interested to find a critical value, the free software PartReg can be used (*Reference 5*). The result obtained from PartReg is depicted in the next section in *Figure 4*.

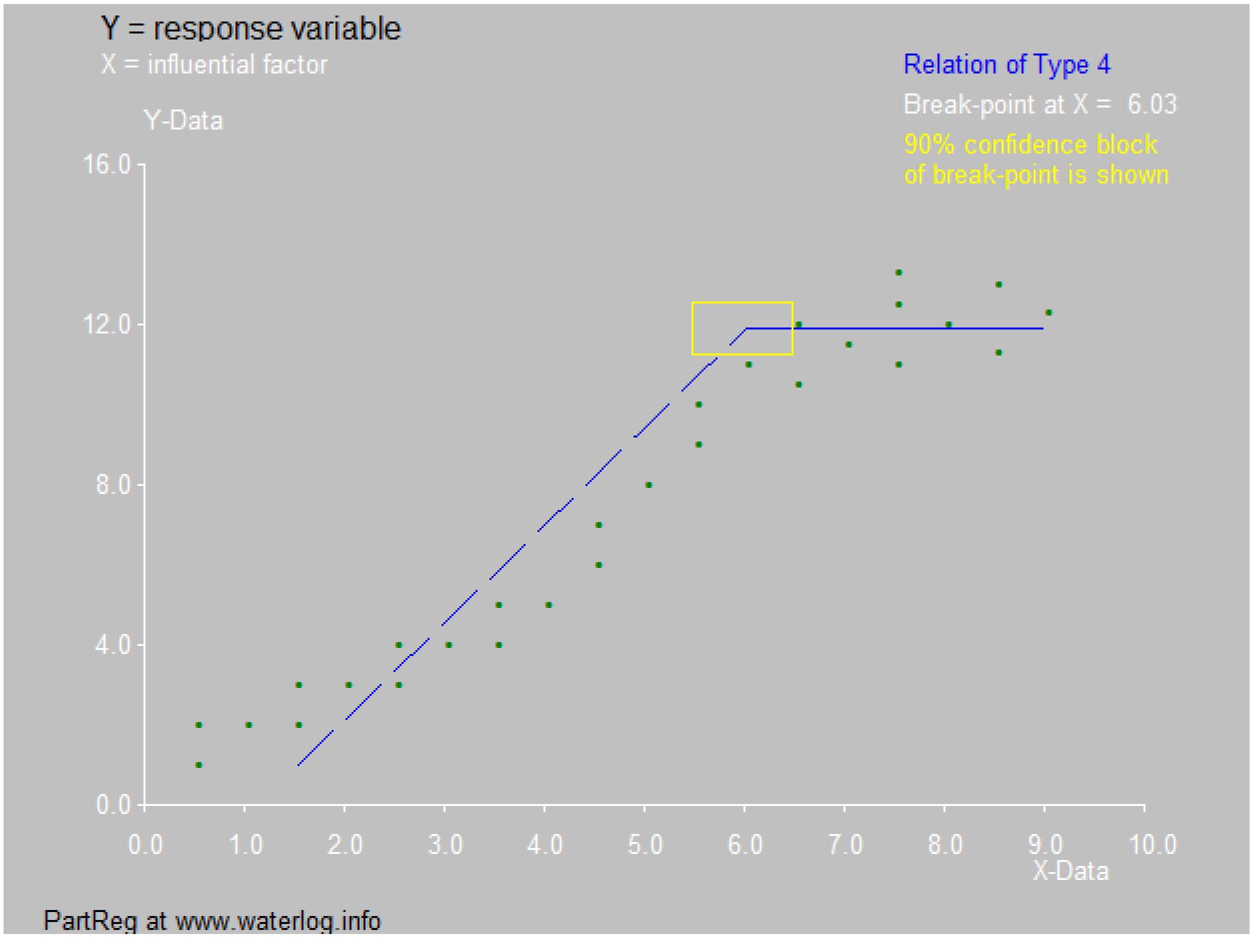

Figure 4. *The same as data used in Figure 3 this time treated with PartReg for partial regression to find the range of X values over which there is no effect (the plateau) of X on and the regression line has zero slope. The critical value (breakpoint) is found at X = DWT = 0.603 (say 6) dm. For values of DWT beyond the breakpoint the influential factor has no effect on the response (dependent) variable, but below that value the response variable declines.*

2. Agricultural examples range of no effect

2.1 Type 3. Field data measured in Sampla, India

In the next figure (*figure 5*) the data of barley yield (Y) have been plotted against soil salinity (X) and treated by the PartReg program to find the largest stretch over which the Y-X relation can be considered horizontal (the plateau), that is: the slope of the partial regression line over this stretch can be taken equal to zero (*Reference 6*).

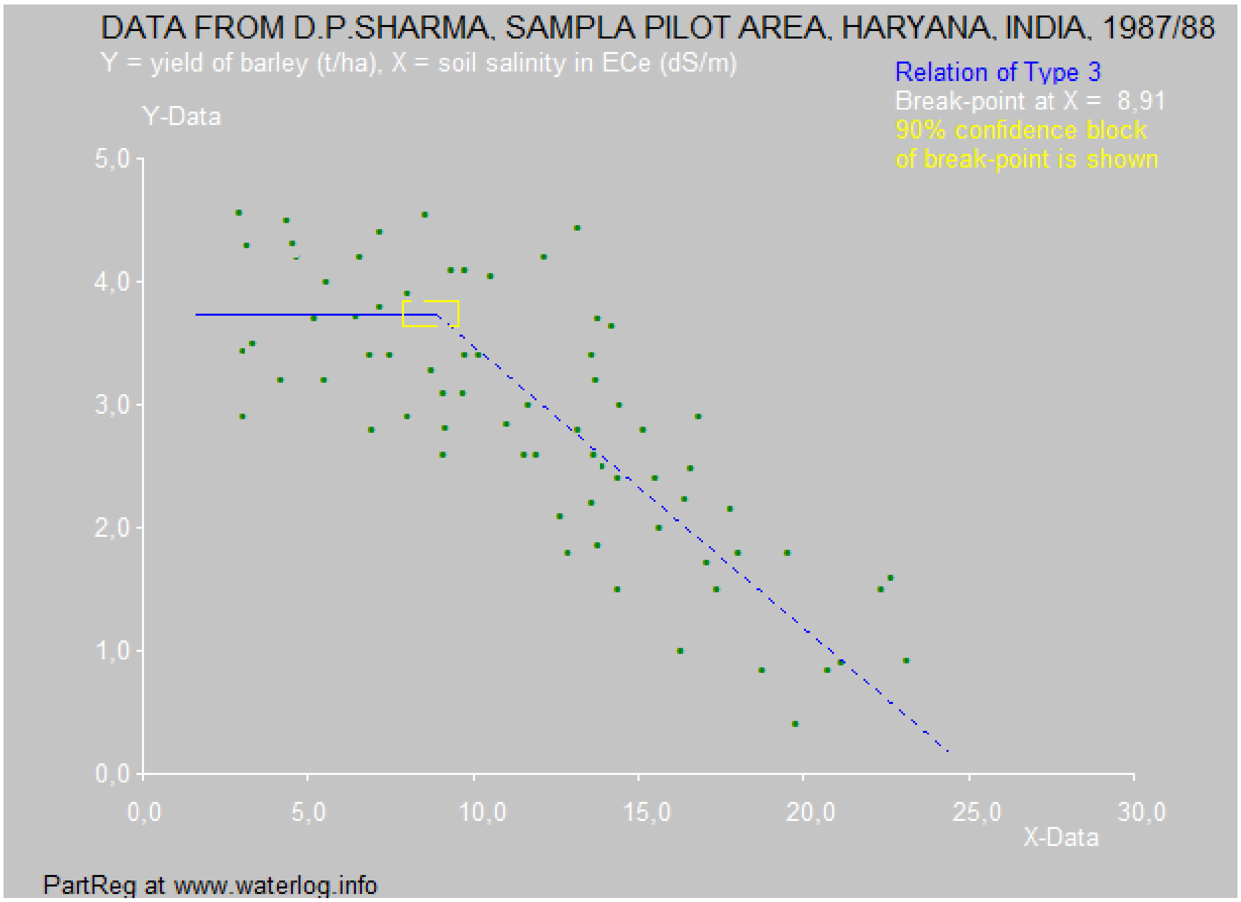

Figure 5. *Relation between yield (Y) of barley and the soil salinity (X) expressed in electric conductivity of a water saturated soil extract (ECe in dS/m) obtained by PartReg through partial regression. The Y=X relation is of type 3, meaning the it is the first part of the relation that is horizontal (the plateau). The salt tolerance is 8.91 (say 9) dS/m. Higher salinity values lead to yield decline.*

2.2 Type 4. Field data measured in Australia

In the next figure (*figure 6*) the data of sugarcane yield (Y) have been plotted against seasonal average depth of the groundwater table (X) and treated by the PartReg program to find the largest stretch over which the Y-X relation can be considered horizontal (the plateau, (*Reference 7*).

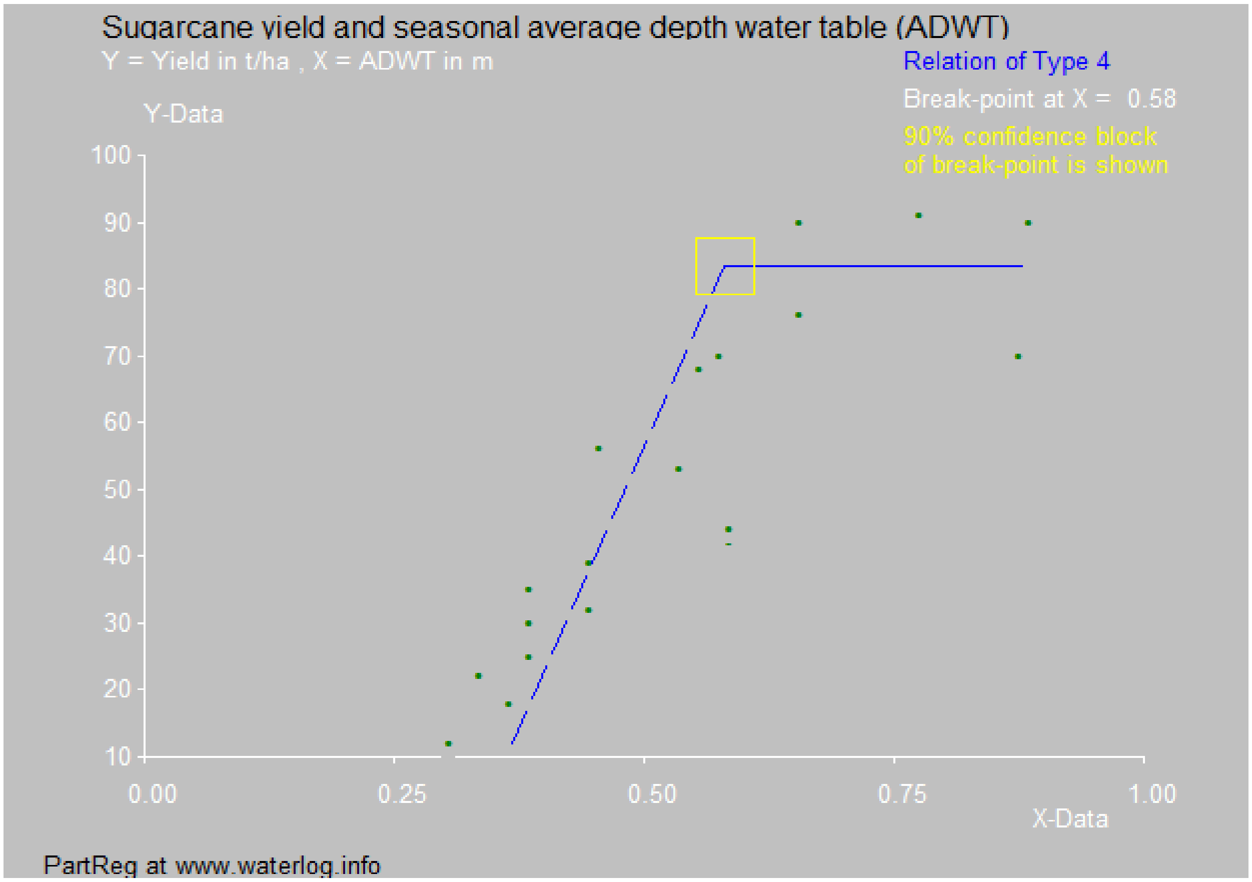

Figure 6. *Relation between yield (Y) of sugarcane and the seasonal average depth of the groundwater table (X in m) obtained by PartReg through partial regression. The Y=X relation is of type 4, meaning that it is the last part of the relation that is horizontal (the plateau). The critical value of X is 0.58 (say 0.6) m. Lower values of X lead to yield decline.*

3. References

Reference 1.
Free software for the calculation of segmented and curved functions (SegRegA). On line:
https://www.waterlog.info/cumfreq.htm

Reference 2.
Article preprint published in Researchgate on Sept. 2019. On line:
The potato variety "927" tested at the Salt Farm Texel,The Netherlands, proved to be highly salt tolerant
or:
https://www.waterlog.info/pdf/Potato 927

Reference 3.
Technical report published in Researchgate on Febr. 2019. On line:
Free calculator for the determination of positive and inverted S-curves for the response function of influential treatments or conditions with examples of crop yield versus soil salinity and depth of the water table
or:
https://www.waterlog.info/pdf/S-curves.pdf

Reference 4.
Technical report published in Researchgate. On line:
Questionable mirrored S-curves used in literature on crop yield relations with soil salinity to determine salt tolerance of crops
or:
https://www.waterlog.info/pdf/Strange S-curves.pdf

Reference 5.
Free software to find a critical value of X separating the range of clear effect of X on Y from the range with no effect (PartReg). On line:
https://www.waterlog.info/partreg.htm

Reference 6.
Researchgate technical report. On line:
Using the Free Partial Regression Software (PartReg) to Determine the Maximum Tolerance of Crops to Soil Salinity as Measured in Farm Lands
or:
https://www.waterlog.info/pdf/Partial Regression.pdf

Reference 7.
Preprint in Research gate. On line:
Crop yield and depth of water table, statistical analysis of data measured in farm lands
or:
https://www.waterlog.info/pdf/Crop yield and depth of watertable.pdf

4. APPENDIX

4a. SegRegA operation

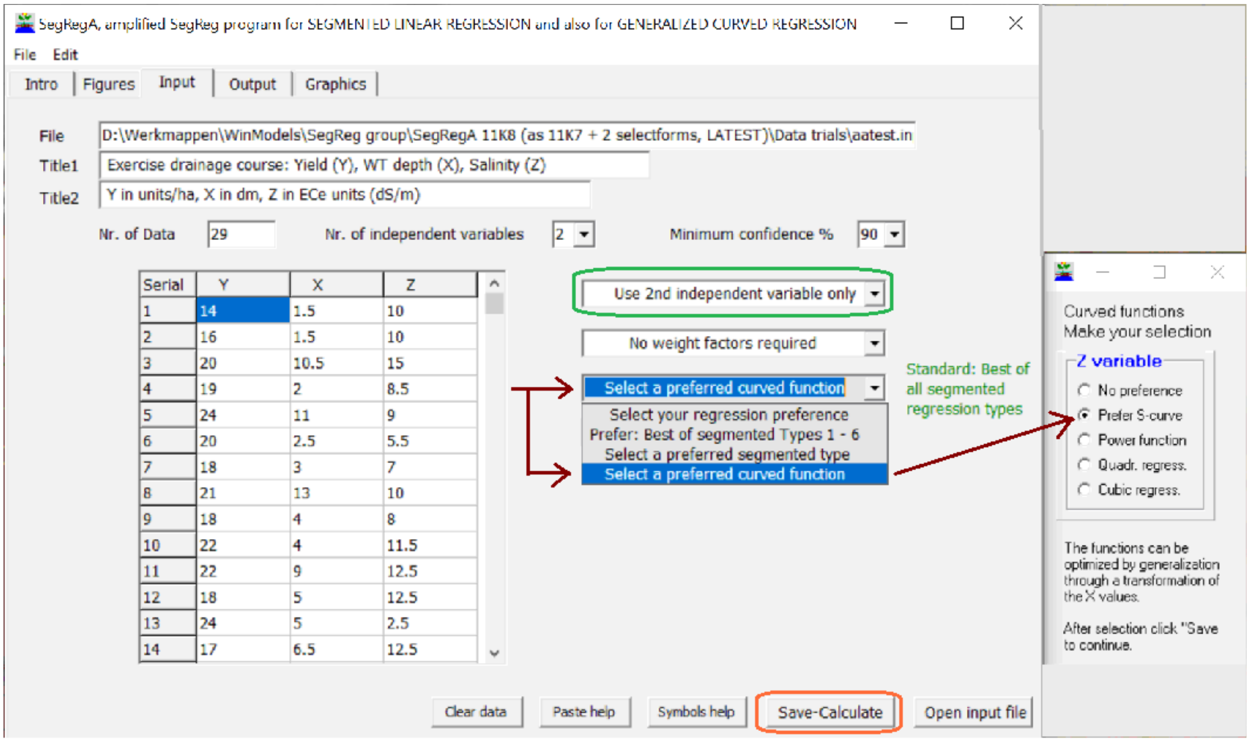

***Figure** 7. The input menu of SegRegA. The input data have been pasted from Excel into the input table. The second independent variable (Z) has been selected for further analysis (green rectangle, the options are both variables, the first variable only or the second variable only). In the selection box, the group "curved functions" has been chosen while in the relevant decision box the preference for the "S-curve" has been fixed. Use "Save-Calculate"(red box) to continue.*

4b. PartReg operation

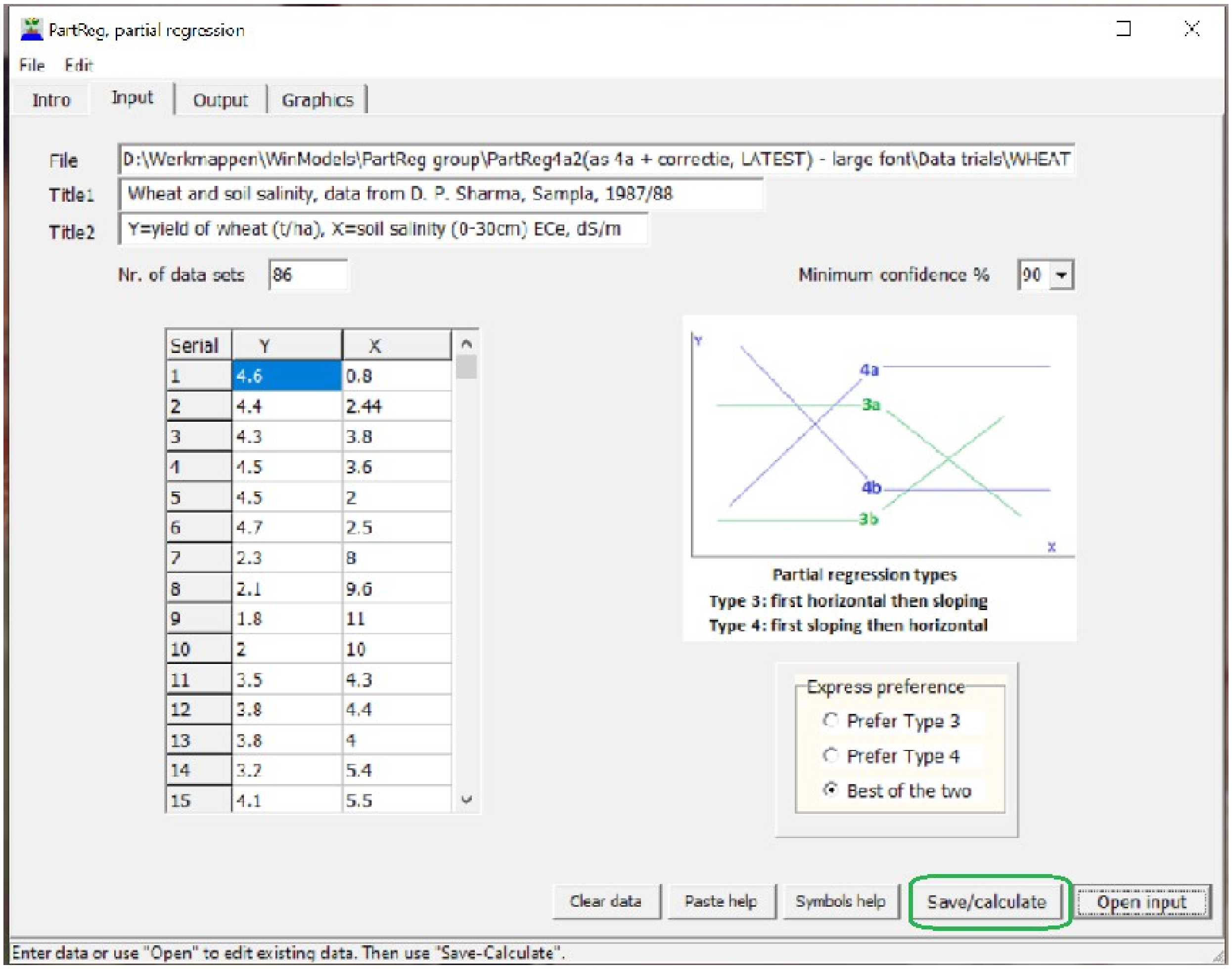

Figure 8. *The input menu of PartReg. The input data have been pasted from Excel into the input table. The difference between type 3 and 4 is illustrated and a selection can be made. Click the "Save/calculate" button (green rectangle) to proceed.*