Crop Yield versus Depth of the Ground Water Table, Statistical Analysis of Data Measured in Farm Lands Aiming at the Formulation of Drainage Needs

R.J. OOSTERBAAN

Retired from International Institute for Land Reclamation and Improvement (ILRI)
Wageningen, THE NETHERLANDS

https://www.waterlog.info

Abstract: It is generally found that agricultural crop yields decline at shallow groundwater tables, except hydrophilic (water loving) crops like rice. Most response functions of crop yield versus depth of water table (DWT) are based on a critical depth (CDW) below which the yield reduces while at deeper depth the yield is unaffected. As the water table usually fluctuates during the growing season one normally accepts the seasonal average depth as an indicator for DWT. In some, more complicated, cases one takes the frequency exceedance of DWT over a CDW and one assumes that the crop production is less as this frequency is higher. Such an approach needs a determination of the most indicative frequency (for example 10%, 20%, etc.) and the critical extreme value (for example 10 cm, 20 cm, etc.) need to be established, which complicates the analysis and is not done here, also because the required data are seldom available. In this article the data available in literature on the relation between yield and average DWT are used and concern (1) Banana (data from Surinam), (2) Cotton (data from Egypt), (3) Sugarcane (data from Australia) and (4) Winter Wheat (data from England). The methods of analysis are (a) visual inspection of envelope lines, (b) statistical segmented regression, (c) statistical determination of maximum horizontal stretches in the yield-DWT relationship, and (d) curved (non-linear) regression with the help of (d1) the Power function, (d2) the S-curve functions, and (d3) the quadratic as well as (d4) the cubic regressions. These functions use a generalization by transforming the DWT data by means of an optimized exponent before their application and a back transformation afterwards. Free software for this purpose is available. The generalized quadratic and cubic regressions need the method of matrices and determinants to find their parameters. After the analysis of the four data groups, the conclusions are formulated. The general conclusion is that for each case all types of analysis have to be done while the most appropriate method can be different from one case to the other.

Key words: - Crop yield, depth of the water table, data from farm lands, statistical analysis, drainage requirements.

1. Introduction

The relation between crop yield and depth of the ground water table (DWT) has been analyzed in first instance by Oosterbaan in 1988 [*Ref. 1*]. In that year the internet and corresponding software did not yet exist, so that the analysis was done by eye (visually) only, using graphs with envelope lines around the highest and lowest yield values.

The purpose of such an analysis is to determine criteria for the design of subsurface drainage systems in agricultural lands to control DWT [*Ref. 2*] .

A diagram of the role of the analysis in subsurface drainage design is given in *Appendix A*.

The data used in this article for the advanced analysis with free software [*Ref. 3*, *Ref. 4*] are related to the following crops (alphabetically):

- 1 - Banana cultivation in Surinam [Ref. 5]
- 2 - Cotton cultivation in Egypt [Ref. 6]
- 3 - Sugarcane cultivation in Australia [Ref. 7]
- 4 - Winter wheat in England [Ref. 8].

The analysis is done using the traditional visual method, which still is very valuable, and a number of modern segmented and generalized curved regressions to judge which one would be most appropriate in the light of purpose of the analysis.

Normally one employs the average depth of the water table during the cropping season. In exceptional cases one uses the frequency of extreme values. This method is seldom used for the following complicating reasons:

− One needs frequent observations

− The critical frequency has to be detected

− The critical extreme value has to be found

− Proof should be given that it leads to a better fit to the data

− The more complicated non-steady state drainage equation will have to be used for the design of the drainage system

Therefore, in this paper, it is the seasonal average depth of the groundwater table that is used, because it has the majority of data An impression of the use of extreme values is presented in *Appendix B*.

2. The Banana data from Surinam [Ref. 5]

The traditional data analysis using envelope lines made visually around the maximum and minimum yield (Y) values along the range of DWT values on the X-axis is shown in *figure 1*.

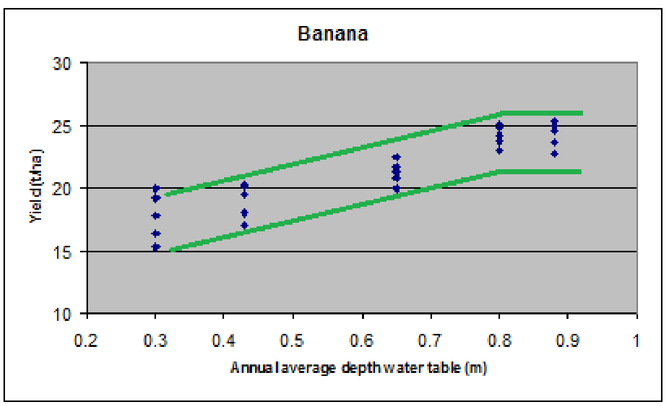

Figure 1. *Yield (t/ha) of banana on the Y-axis versus annual average value of DWT (m) on the X-axis using the traditional envelopes method estimated by eye.*

The banana plantations (banana is a perennial crop) investigated in Surinam show, in this envelope analysis, a clear yield decline at an average depth of the water table (ADWT) < 0.75 m while at ADWT values > 0.80 m there is no more yield reduction and the yield stays stable. This is the range of "no effect". The critical value of ADWT (called CDW) may be estimated at CDW = 0.8 m.

Instead of the envelopes method one could try to use the segmented regression method given as an option in the SegRegA program [*Ref. 3*]. The result is depicted in *figure 2*.

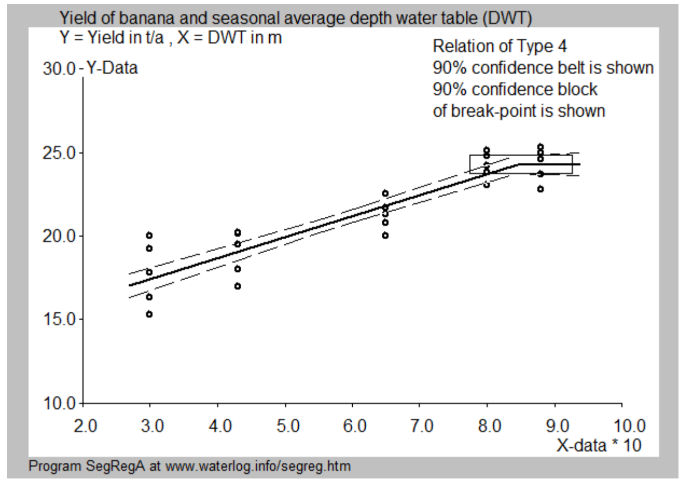

Figure 2. *Segmented regression of the banana yield (Y) on the annual average DWT value on the X-axis. There is a breakpoint (BP) at X = 0.85 m that may be taken as the CDW value. The coefficient of explanation (CE or R^2) equals 0.825. The values of BP and CE can be read from the output file of SegRegA (see* Appendix C*).*

The CDW value in *figure 2* (0.85) is slightly higher than that in figure 1, but according to the confidence belt of BP in *figure 2* the CDW value could be less than 0.80, so the difference of the outcomes in the two figures is statistically not significant. It would be up to the user, considering the banana cultivation conditions, to make the final choice of CDW somewhere between 0.75 and 0.95.

As an alternative one can use an S-curve option in SegRegA (see *Appendix D*) as shown in *figure 3*.

The S-curve option in *figure 3* has a coefficient of explanation CE = R^2 = 0.826 which is practically the same as that in *figure 2* (0.825). Hence from point of view of goodness of fit it is not possible to decide which method is preferable.

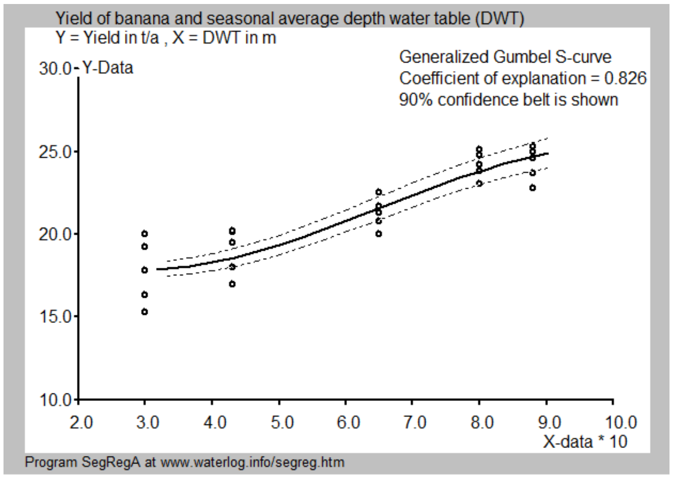

Figure 3. *S-curve regression of the banana yield (Y) on the yearly average annual DWT on the X-axis.*

Figure 4 suggests that the CDW value is beyond the maximum X-value measured so that CDW>0.90 m. Given the results of *figure 1 and 2* it may be decided that application of the S-curve in figure 3 is not advisable. Instead one could try a generalized quadratic regression to detect the maximum Y value as in figure 4.

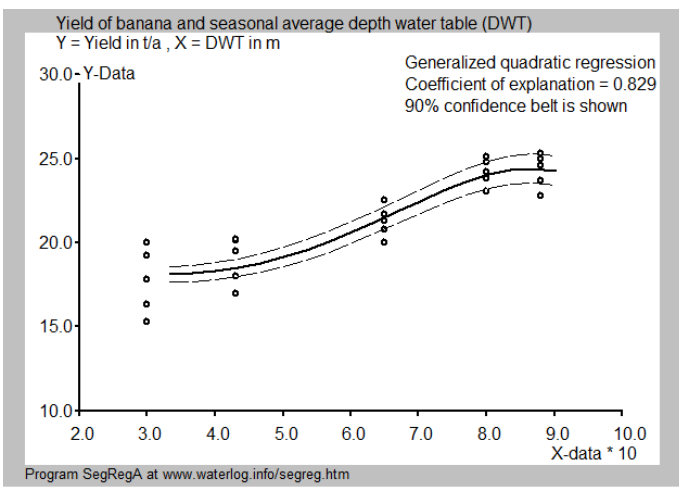

Figure 4. *Generalized quadratic regression of the banana yield (Y) on the yearly average annual DWT on the X-axis. The coefficient of explanation (0.829) is only slightly higher than in the previous two cases, so it does not indicate a much better goodness of fit.*

According to the output file of SegRegA the best fitting generalized quadratic equation reads:

Yc = A*W^2 + B*W + C
where W = (X - Xmin)^E, Xmin = 0.268,
E = 2.36, A = −9.23E+001, B = 4.83E+001 and
C = 1.81E+001

From the above quadratic function it can be seen that the X values have been raised to the power E =2.36 before executing the quadratic regression with the aim to obtain the best possible goodness of fit.

The overall deduction from figure 4 is that the methods used in figure 1 and 2 are reliable as the trend in figure 4 for X values > 0.8 m is horizonal.

As the quadratic function has a higher number of parameters and the coefficient of explanation is not much more, it would, for drainage design, even be preferable to use the segmented regression given in figure 2.

However, if the interest of the user is in the plant physiological processes, then the use of the quadratic function may be preferable as it demonstrates that initially, at low X values, the Y increase is small with increasing X, while later the increase rate gets higher.

A cubic regression as demonstrated in *figure 5* reveals practically the same picture as in *figure 4* dealing with a quadratic regression. Also the coefficient of explanation (0.823) is only little higher than that in *figure 4*. Hence, the conclusions arrived at in the discussion about *figure 4* applies equally here.

Note: In SegRegA, the quadratic and cubic regression use the method of (inversed) matrices, determinants and Cramer's rule [*Ref. 6*].

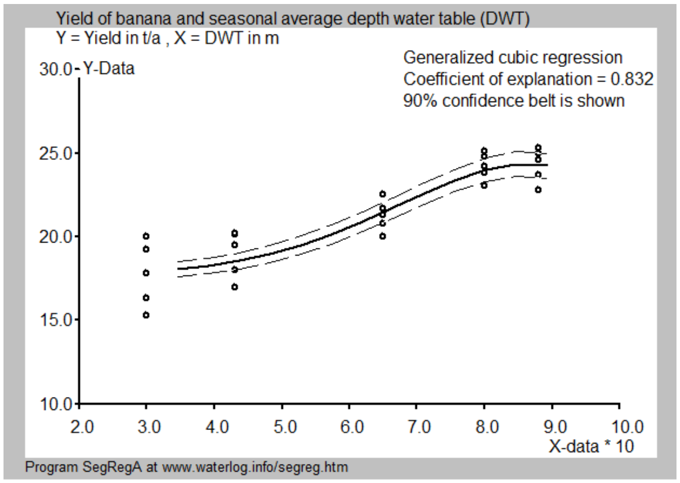

Figure 5. *Generalized cubic regression of the banana yield (Y) on the yearly average annual DWT on the X-axis.*

3. The Cotton data from Egypt [Ref. 7]

The traditional data analysis using envelope lines made visually around the maximum and minimum yield (Y) values along the range of DWT values on the X-axis is shown in *figure 6*.

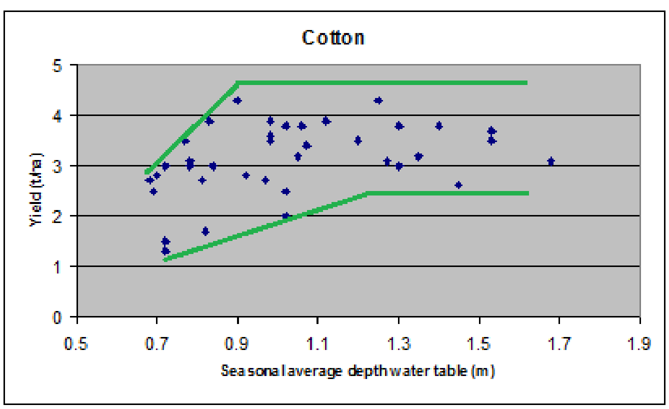

Figure 6. *Yield of cotton (Y, t/ha lint and seed) plotted against seasonal average depth of the water table (DWT, on the X-axis, m).*

The upper envelope as a visual estimate shows a breakpoint at a DWT value of about 0.9 m, whereas the lower envelope breaks at DWT = 1.2 m. The critical depth of the water table (CDW) will then be between 0.9 and 1.2 m, say 1.05 m. Anyway it is sure that the cotton yield decreases sharply at water depths below 0.9 m and that the yield is relatively high and, apart from the random fluctuations, fairly constant at DWT values above 1.2 m, although the yields are getting down somewhat at DWT values above 1.4 m.

Instead of the envelopes method one could try to use the segmented regression method expressing a preference for Type 4 given as an option in the SegRegA program [*Ref. 3*]. The result is depicted in *figure 7*.

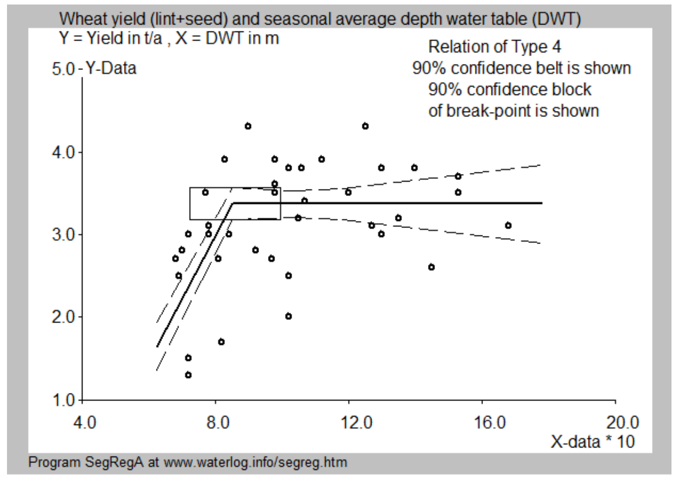

Figure 7. *Segmented regression of the cotton yield (Y) on the seasonal average DWT value on the X-axis. There is a breakpoint (BP) at X = 0.85 m that may be taken as the CDW value. The coefficient of explanation (CE or R^2) equals 0.23, which is low owing to the large random variation.*

The breakpoint in figure 7 (0.85 m) is less then determined in figure 1 (1.2 m), although its confidence block ranges from 0.7 m to 1.05 m. It may even be still lower when trying the partial regression method [*Ref 4*] with which one searches for the statistically largest horizontal stretch in a data set (*figure 8*).

According to *figure 8* the trend of the yield beyond X = 0.785 m is horizontal and the tendency of lower yields to descend beyond X=1.4 m, as mentioned in the discussion of *figure 6*, is not recognized,

In order to obtain a more detailed impression of the trend beyond X = 1.4 m, the generalized cubic regression may be tried, see *figure 9*.

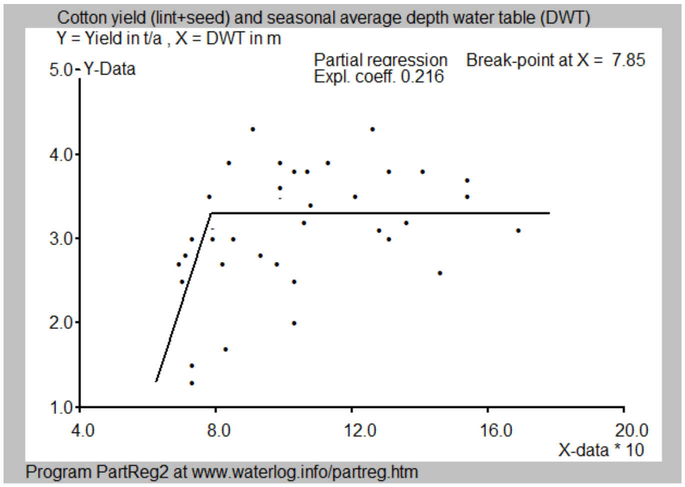

Figure 8. *Partial regression of the cotton yield (Y) on the seasonal average DWT value on the X-axis. There is a breakpoint (BP) at X = 0.785 m. Statistically, the maximum horizontal stretch ranges from X = 0.785 m until X = 1.7 m.*

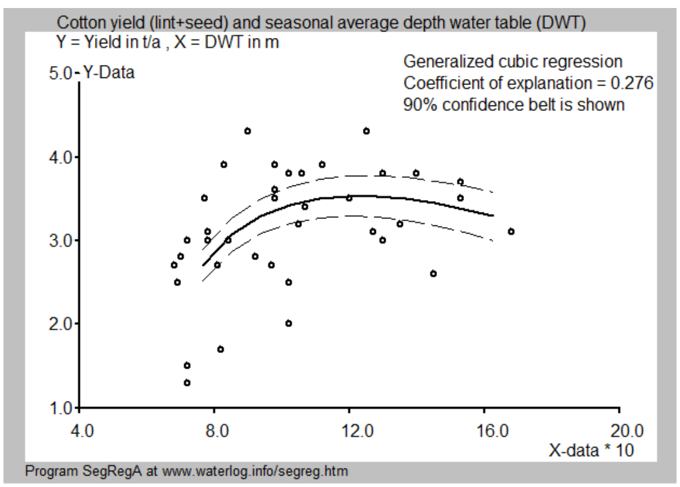

Figure 9. *Generalized cubic regression of the cotton yield (Y) on the seasonal average DWT value on the X-axis. The coefficient of explanation or R^2, being 0.276, is higher than in all previous cases and the descending trend beyond X=1.4 m becomes more clear. The mathematical expression of the generalized cubic function is:*

*Y = A*W^3 + B*W^2 + C*W + D, where*
A = −2.15E+001, B = 4.06E+001,
C = −2.18E+001, D = 5.93E+000,
W = (X - Xmin)^E with Xmin = 0.625 and
E = 0.23

In figure 9, the maximum yield is obtained at DWT = 1.2 m, which could be taken as the critical value (CDW). This value, however, is much higher than in the previous cases and, if subsurface drainage is considered, it would lead to a more expensive drainage system.

When it can be established that a deep water table corresponds to scarcity of irrigation water, the cubic regression would give the best possible solution. Otherwise, preference should be given the results of the segmented regression of Type 4 (*figure 7*) or the partial regression (*figure 8*).

As the partial regression leads to the lowest of all CDW values, it would help in the design of a subsurface drainage system at the lowest cost.

4. The Sugarcane data from Australia [Ref. 8]

The traditional data analysis using envelope lines made visually around the maximum and minimum yield (Y) values along the range of DWT values on the X-axis is shown in *figure 10*.

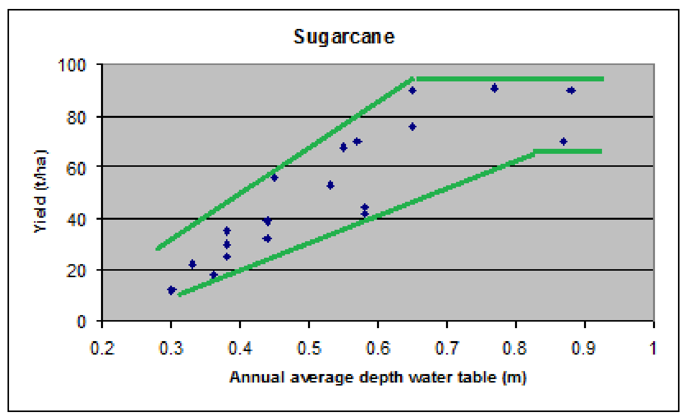

Figure 10. *Yield of sugarcane (Y, t/ha) plotted against seasonal average depth of the water table (DWT, on the X-axis, m).*

The upper envelope as a visual estimate shows a breakpoint at a DWT value of about 0.65 m, whereas the lower envelope breaks at DWT = 0.85 m. The critical depth of the water table (CDW) will then be between 0.65 and 0.85 m, say 0.70 m. Anyway it is sure that the sugarcane yield decreases sharply at water depths below 0.65 m and that the yield is relatively high and, apart from the random fluctuations, fairly constant at DWT values above 0.85 m.

Instead of the envelopes method one could try to use the segmented regression method expressing a preference for Type 4 given as an option in the SegRegA program [*Ref. 3*]. The result is depicted in *figure 11*.

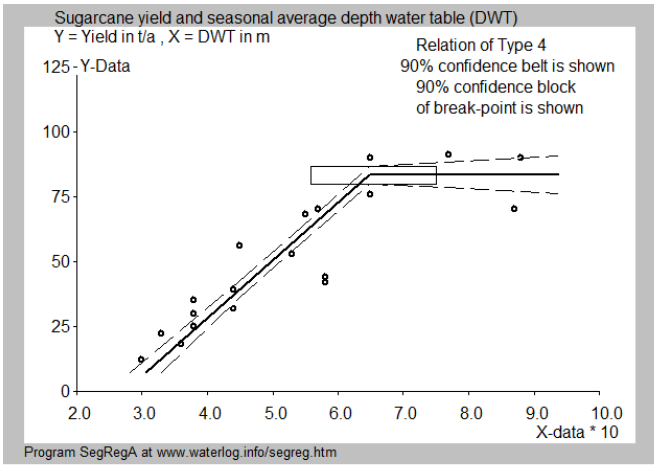

Figure 11. *Segmented regression of the sugarcane yield (Y) on the seasonal average DWT value on the X-axis. There is a breakpoint (BP) at X = 0.65 m that may be taken as the CDW value. This value is almost equal to the value of 0.70 m determined in* figure 10.

The largest horizontal stretch, over which the regression coefficient (or slope of the line) is statistically not different from zero, can be found using the ParReg2 software [*Ref. 4*] and is depicted in *figure 12*.

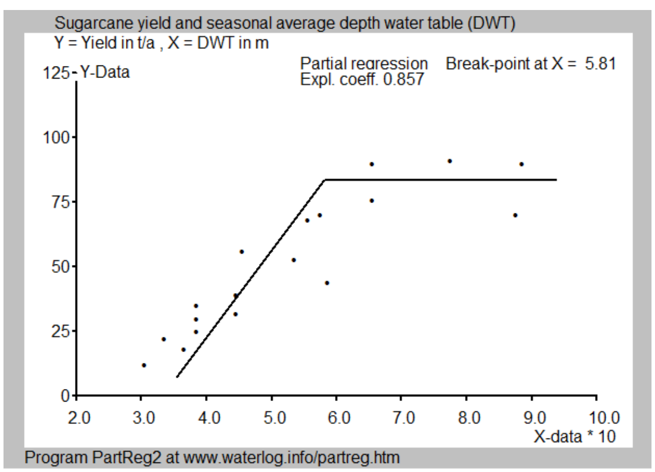

Figure 12. *Partial regression of the sugarcane yield (Y) on the seasonal average DWT value on the X-axis. There is a breakpoint (BP) at X=0.58 m. Statistically, the maximum horizontal stretch ranges from X = 0.58 m until X = 0.9 m.*

Owing to the relative large and significant horizontal stretch in figures 10, 11 and 12, it is of no use to try curved, nonlinear regressions. The breakpoint (or CDW value), important for the design of subsurface drainage systems, can reliably fixed at 0.70 m.

5. The Wheat data from England [Ref. 9]

Winter wheat is sown in autumn. It develops slowly in winter, faster in spring time, and in early summer it is harvested.

The depth of the water table is less in winter then in summer owing to the negligible evaporation in winter when temperatures are low. Therefore the indicative depth is taken as an average during the winter season.

The traditional data analysis using envelope lines made visually around the maximum and minimum yield (Y) values along the range of DWT values on the X-axis is shown in *figure 13*.

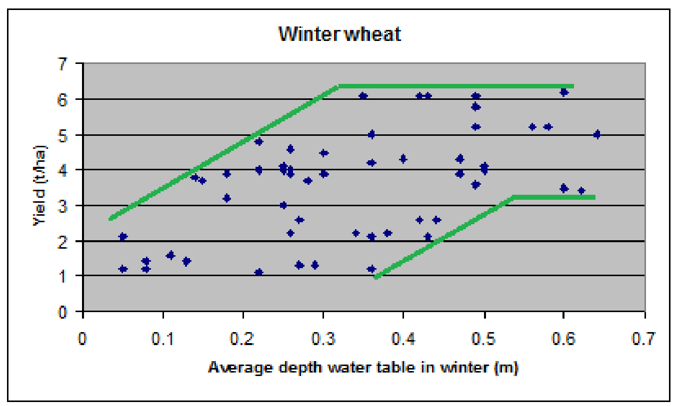

Figure 13. *Yield (t/ha) of winter wheat on the Y-axis versus average value of DWT (m) in winter on the X-axis using the traditional envelopes method estimated by eye.*

The upper envelope as a visual estimate shows a breakpoint at a DWT value of about 0.33 m, whereas the lower envelope breaks at DWT = 0.53 m. The critical depth of the water table (CDW) will then be between 0.33 and 0.53 m, say 0.43 m. Anyway it is sure that the wheat yield decreases sharply at water depths below 0.33 m and that the yield is relatively high and, apart from the random fluctuations, fairly constant at DWT values above 0.53 m.

Instead of the envelopes method one could try to use the segmented regression method expressing a preference for Type 4 given as an option in the SegRegA program [*Ref. 3*]. The result is depicted in *figure 14*.

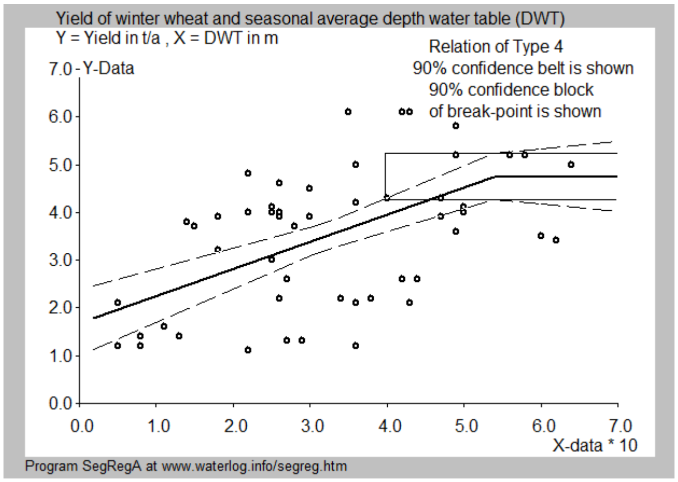

Figure 14. *Segmented regression of the wheat yield (Y) on the average DWT value in winter on the X-axis. There is a breakpoint (BP) at X = 0.53 m that may be taken as the CDW value. This value corresponds to the breakpoint of the lower envelope determined in* figure 13.

The confidence block of BP in *figure 14* extends beyond the maximum measured X-value (DWT = 0.65 m). Apparently there are insufficient data in the range of X-values greater than 0.65 m to establish the BP = CDW value reliably.

Comparing the results of figures 13 and 14, it can be said that the breakpoint of the upper envelope in figure 13 does not seem to be a proper index for the critical depth of the water table (CDW).

However, this observation is contradicted when using the PartReg2 method [*Ref. 4*] for the determination of the longest possible horizontal segment (*figure 15*). Here it appears that the BP (or CDW) value is only 0.4 m, much less than that found in figure 14 (0.65 m). Also, the PartReg method does not indicate a shortage of observations beyond the critical X-value as does the segmented regression. The advantage of this method is that the trend at lower DWT values is not considered, where as in the segmented regression (*figure 14*) it is. In the latter case, the trend at lower DWT values can disturb the trend at higher values, as here the parameters are optimized by minimizing the sum of deviations of observed and simulated values over the entire range of X-values.

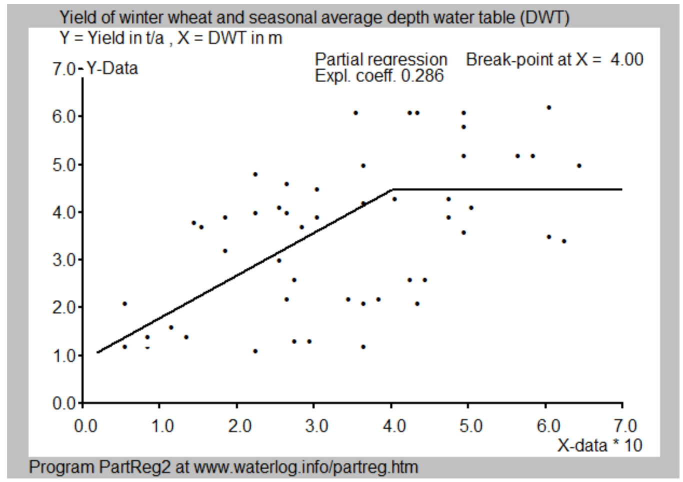

Figure 15. *Partial regression of the wheat yield (Y) on the average DWT value in winter on the X-axis. There is a breakpoint (BP) at X = 0.4 m. Statistically, the maximum horizontal stretch ranges from X = 0.4 m until X = 0.65 m.*

A problem arises when one uses a curved regression, like with the S-curve function (figure 16). The horizontal trend at higher DWT values is not detectable at all. The reason is the same as that given for the segmented regression in *figure 14*.

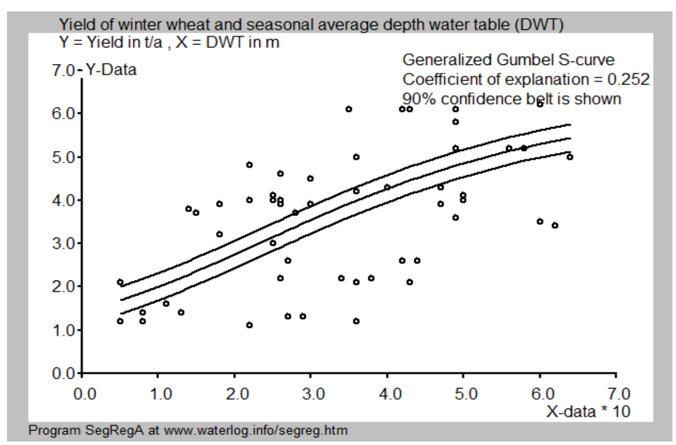

Figure 16. *S-curve regression of the wheat yield (Y) on the average DWT in winter time on the X-axis.*

All in all, it appears that in this case the PartReg method is the most appropriate

6. Summary of the results of the 4 cases dealt with

Table 1. *Summary of CDW values by crop and by type of regression*

Crop	Critical Depth Water table (CDW, BP, m)				Comments
	Envelopes method	Segmented regression	Partial regression	Curved regression	
Banana, Surinam	0.80	0.85	n. a. 1)	0.83	Three methods are OK, use CDW=0.85 a)
Cotton, Egypt	1.05	0.85	0.79	1.20	Large variation use CDW = 0.80 b)
Sugarcane, Australia	0.70	0.65	0.58	n. a. 2)	Some variation Use CDW = 0.60 b)
winter Wheat, England	0.40	0.53	0.40	n. a. 3)	Some variation Use CDW = 0.40 c)

1) The partial regression could not be done because the number of different X-values is too small since several X-values were used repeatedly.
2) and 3) The number of data beyond at deeper water tables is limited while the variation in Y values is very large so that curved regressions do not give decisive outcomes, as clarified in figure 16 for example.

From *table 1* it can be seen that:

a) For Banana, the three methods that could be applied give similar results

b) The partial regression yields logical CDW values in the cases of Cotton, Sugarcane, and winter Wheat

c) For winter Wheat the envelopes method and the partial regression give the same result, while the segmented regression method gives a much higher value. The reason is that, according to the analysis of variance, the segmented regression does not give a sufficient augmentation of the goodness of fit compared to that of a straight line obtained by straightforward linear regression. Hence the segmented regression is not reliable. As SegRegA automatically produces an analysis of variance, this feature can be checked in the user menu for the output in which the calculation results are presented.

In all cases it is clear that very shallow water tables depress the yield and need to be cured.

7. Conclusions

Viewing the variation in results of the different methods it follows that there is no particular method to be recommended.

While analyzing statistically the relation between yield and depth of water table (DWT) in data sets collected in farmers' fields, which by nature reveal many serious irregularities, it can be recommended to apply all possible types regression to reach a decision about the most appropriate one.

Quite often the PartReg method yields acceptable results. However, it would be good to keep in mind the environmental conditions that may influence the trend of the crop yield versus the DWT as for example discussed under *figure 9*. This figure shows a cubic regression with a maximum yield at DWT=1.2 m while at higher DWT values the cotton yield comes down, possibly due to a shortage of irrigation water, resulting in both a yield decline and a deeper water-table.

References

[1] R. J. Oosterbaan, 1988. *Agricultural Criteria for Subsurface Drainage, a Systems Analysis.* In the journal: Agricultural Water Management, 14 (1988) 79-90. On line: https://www.waterlog.info/pdf/AgriCrit.doc

[2] *Agricultural Drainage Criteria.* Chapter 17 in: H.P.Ritzema (Ed.), Drainage Principles and Applications. International Institute for Land Reclamation and Improvement (ILRI), Publication 16, second revised edition, 1994, Wageningen, The Netherlands. ISBN 90 70754 3 39. On line: https://www.waterlog.info/pdf/chap17.pdf

[3] SegRegA, Free software for segmented and curved regression. On line : https://www.waterlog.info/segreg.htm

[4] PartReg2, free software for partial regression analysis to detect a horizontal segment in the Y-X relation. Download from: https://www.waterlog.info/partreg.htm

[5] Lenselink, K.J. 1972. Drainage requirements for banana in the coastal plain (in Dutch, title translated by author). In the journal: De Surinaamse Landbouw, Vol. 20, pp. 22-36.

[6] Neutrium. *Fitting of a Polynomial using Least Squares Method*. On line: https://neutrium.net/mathematics/least-squares-fitting-of-a-polynomial/

[7] Nijland, H.J. and S. El Guindy 1984. Crop yields, soil salinity and water table depth in the Nile Delta. In: ILRI Annual Report 1983, Wageningen, pp. 19-29. On line: https://www.waterlog.info/pdf/egypt.pdf

[8] Rudd, A.V. and C.W Chardon 1977. The effects of drainage on cane yields as measured by water table height in the Machnade Mill area. In: Proceedings of the 44th Conference of the Queensland Society of Sugar Cane Technology, Australia.

[9] FDEU 1972. *Annual Report*. Field Drainage Experimental Unit, Ministry of Agriculture, Cambridge, UK.

Appendix A (referred to in Section 1, Introduction)

Figure A depicts a diagram with the procedures to be followed when designing a subsurface drainage system for control of the groundwater table.

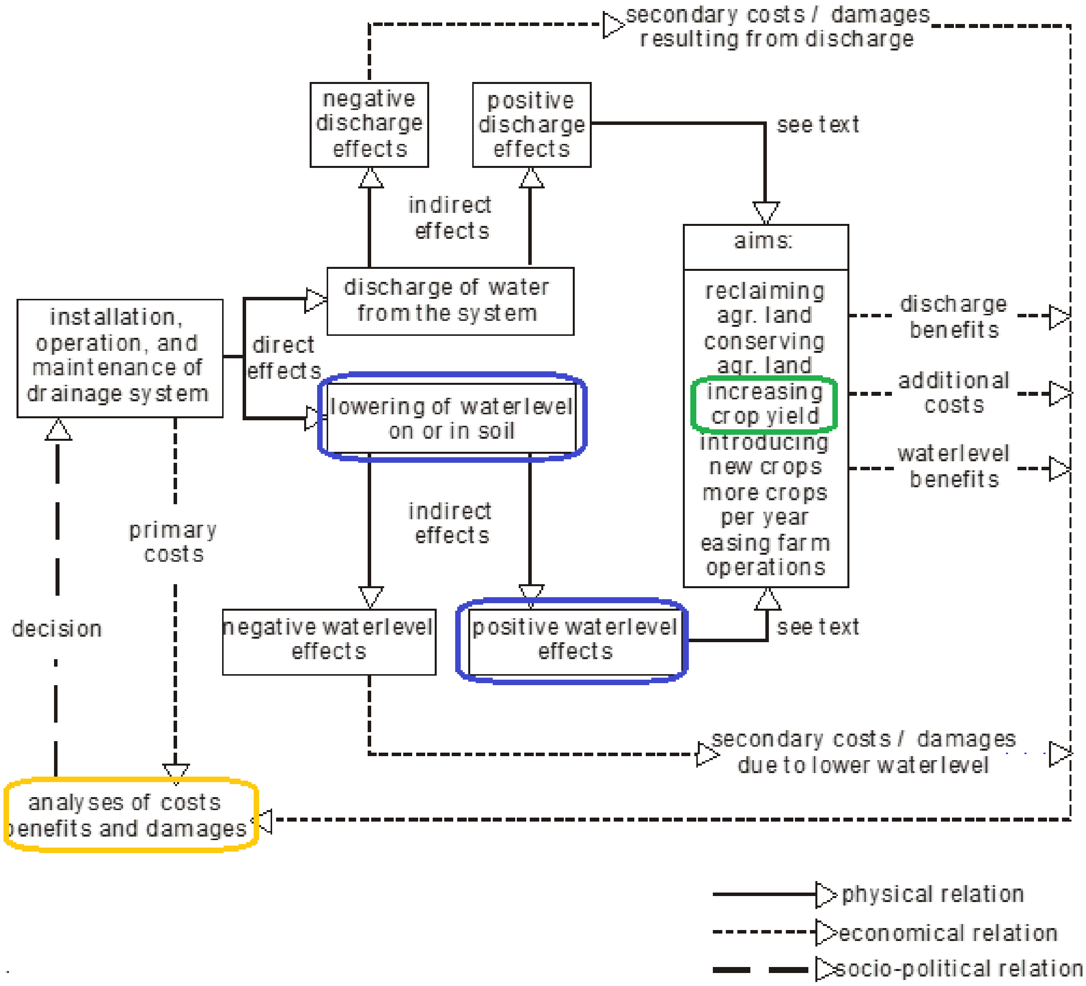

Figure A. *Generalized diagram of effects of subsurface drainage on agriculture and their economic evaluation.*

Note that in figure A the role of the lowering of the water level in the soil and its positive effects leading to increased crop yields have been highlighted with blue rectangles while the analysis of benefits (for example the increase in crop yield less the cost of the subsurface drainage) has been earmarked with an orange rectangle.

In general it can be stated that subsurface drainage system is costlier as the drawdown of the water table is larger, because the drain spacing should be narrowed and the drain depth enlarged, so that the drainage system becomes larger. That explains the importance to find the critical depth of the water table (CDW) as shallow as possible. In this respect the PartReg method can be helpful when data on crop yield and depth of the groundwater table are being analyzed.

Appendix B (referred to in Section 1, Introduction)

The following considerations apply to the use of frequency distributions in the determination of the critical depth of the water table as a permissible value of extremes.

Figure B1 provides an example of the fluctuation of the water table over time as normally happens under natural conditions.

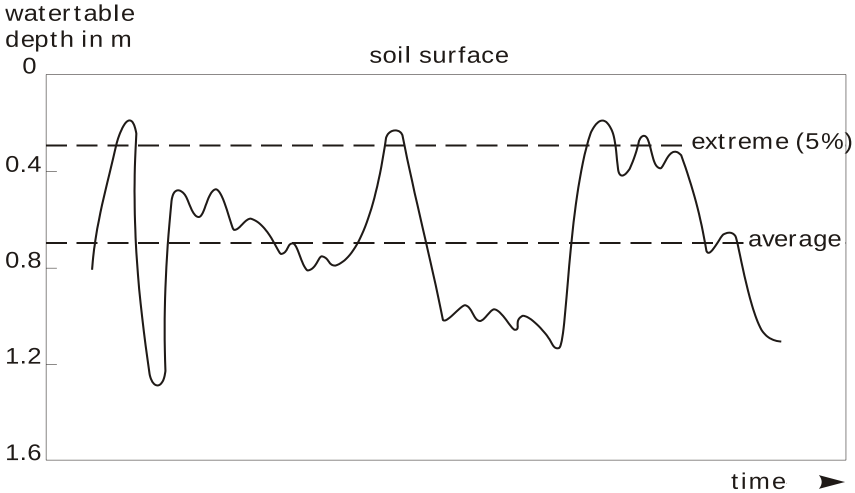

Figure B1. *A fluctuating water table with an indication of the average depth of the water table and the 5% extreme value pointing to an extremely shallow depth indicating that the depth is lower than that in only in 5% of the time.*

A fluctuating water table helps the breathing of the soil: at a rising water table the CO2 produced by the roots is expelled and at a descending water table fresh air, including oxygen, is inhaled.

The relation between the average depth and the extreme value is often linear. A deeper average depth corresponds to a deeper extreme value so that in the relation with crop yield both indices will produce almost similar tendencies. In such a case, the extreme does not offer an added value, the more so as the short term extreme values will do no harm to the crop.

An example of the relation between an extreme value an crop production is shown in *figure B2* (reference Rudd, A.V. and C.W Chardon 1977. The effects of drainage on cane yields as measured by water table height in the Machnade Mill area. In: Proceedings of the 44th Conference of the Queensland Society of Sugar Cane Technology, Australia).

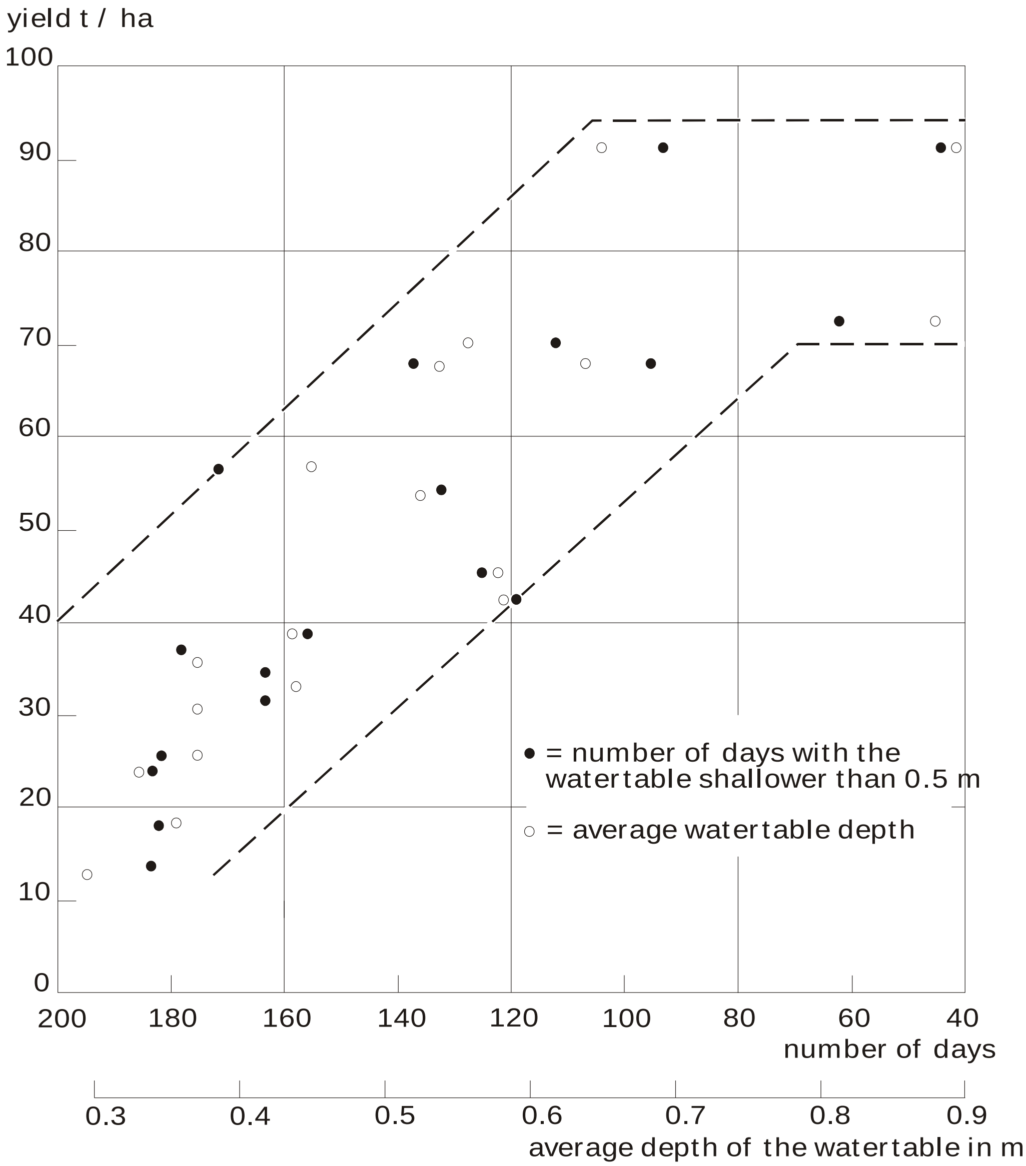

Figure B2. *A plot of yield data of sugarcane versus average depth of the water table and number of days (NrD) with a water table shallower than 0.5 m during the growing season from December to June in N. Queensland, The envelope lines are added.*

Dividing NrD by the total number of days (TND, 200) will provide the frequency of exceedance and dividing 100*(TND-NrD)/TND will yield the extreme percentage used in figure B1.

Figure B2 shows the production of sugarcane as a function of the average depth of the water table during the growing season from December to June (indicated by circles), and the number of days during which the water table is shallower than 0.5 m below the soil surface in the same period (indicated by dots). The function shows that both indices give the same result, because the long-term average depth and the number of extremely shallow depths are apparently strongly correlated. This is logical because, when the average depth is great, a shallow depth is relatively infrequent, and vice versa. Therefore, if one employs either of these indices, the other will not provide any new insight.

Appendix C (referred to in figure 2, section 2)

A part of the output file of SegReg for the Banana case, dealt with in *section 2* and referring to *figure 2*.

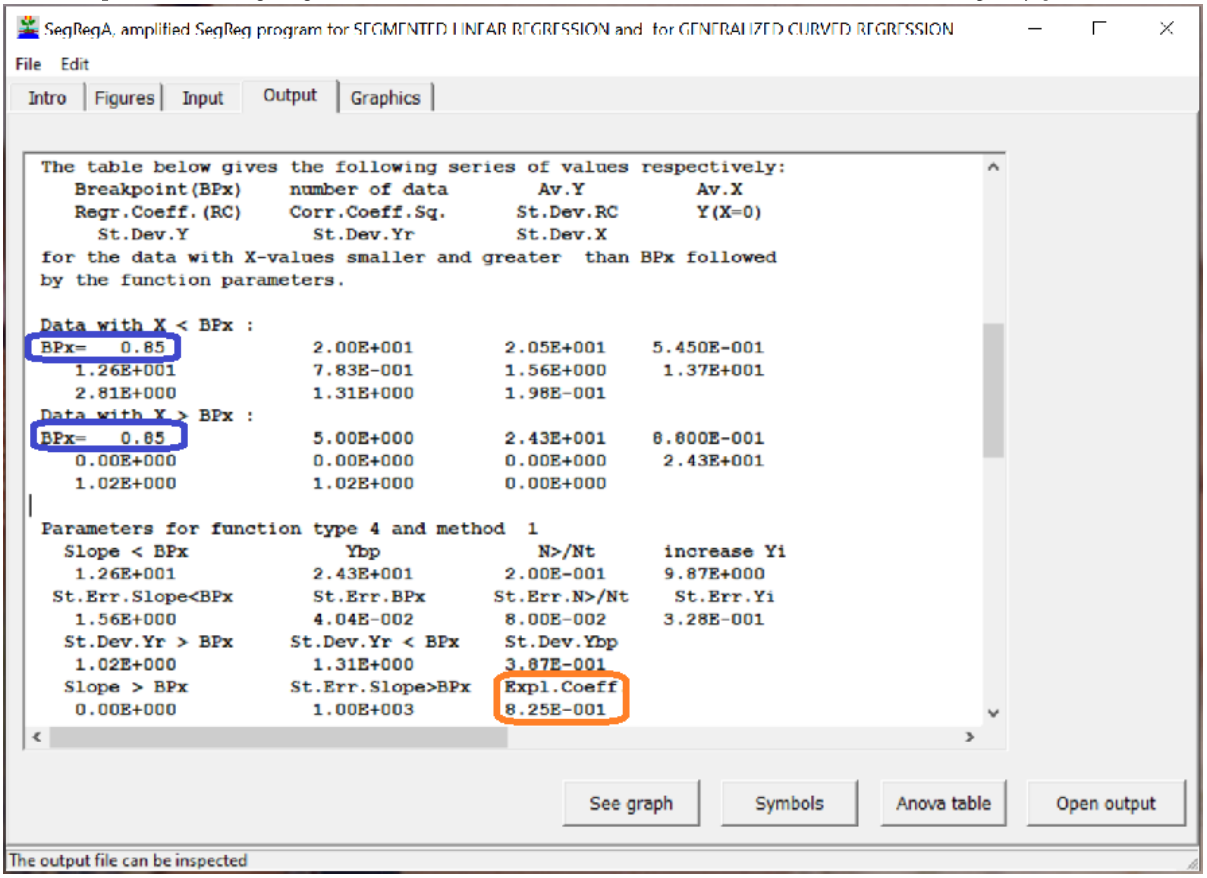

Figure C. *Part of the output file of SegReg (screen print) for the banana data dealt with in section 2 and referring to* figure 2. *The value of the breakpoint (BP or CDW) is encircled in blue colour and the coefficient of explanation (CE or* R^2*) in orange.*

For clarity a repetition of *figure 2* is given hereunder

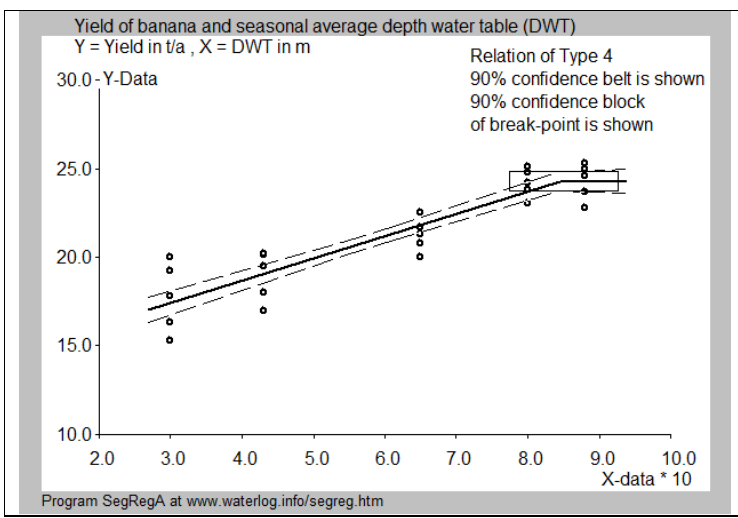

Figure 2. *Segmented regression of the banana yield (Y) on the annual average DWT value on the X-axis. There is a breakpoint (BP) at X = 0.85 m that may be taken as the CDW value. The coefficient of explanation (CE or* R^2*) equals 0.825. The values of BP and CE can be read from the output file of SegRegA (see* Appendix C*).*

Appendix D (referred to in Section 2)

The next figure illustrates the input menu of SegReg in the case of Banana and how the selection of the application of an S-curve function for regression is realized.

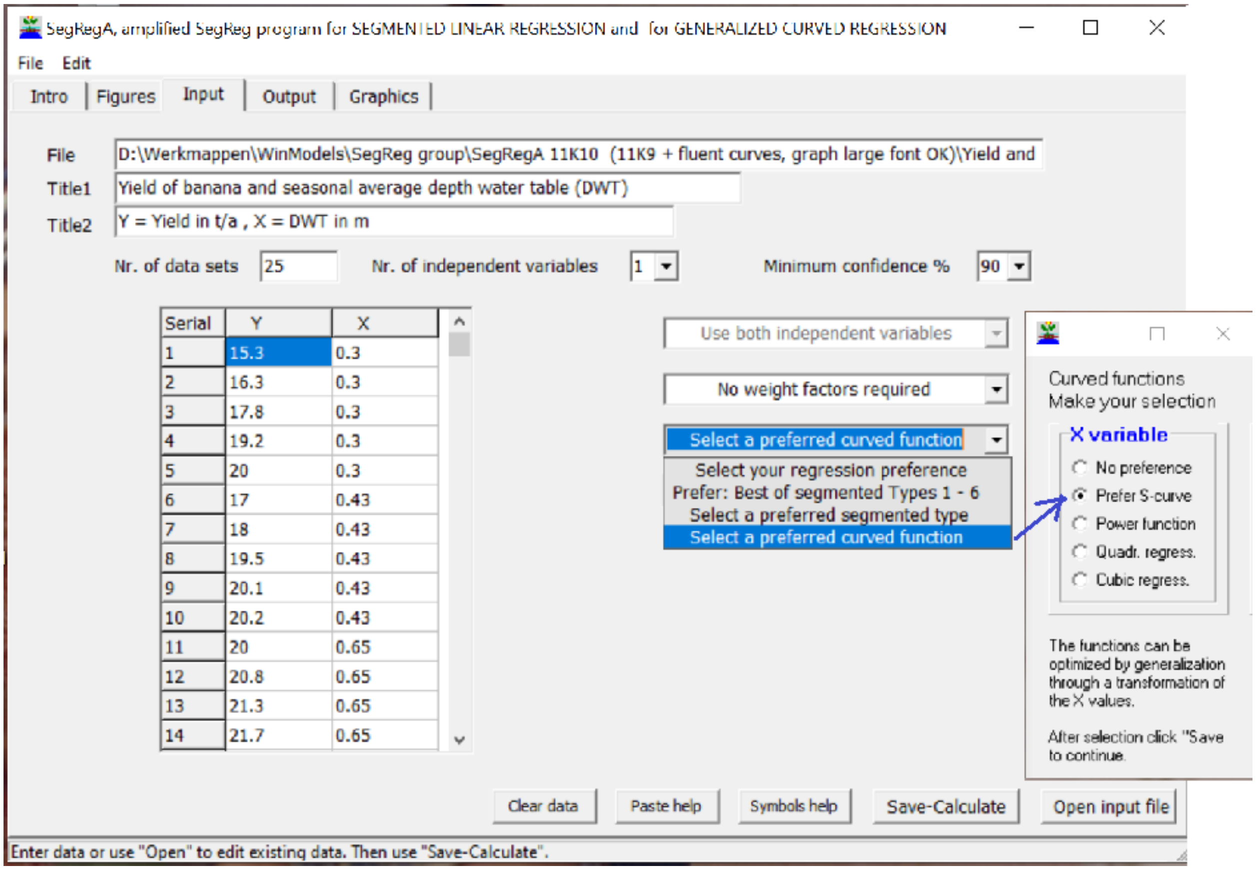

Figure D. *Screen print of the SegReg input menu. The first selection is for the application of regression using curved functions (blue rectangles) and the second selection specifies the S-curve (blue arrow).*

There is a similar option to select preferred segmented regression types, either as the best of all Types 1-6, or as a specific one. Further a large number of other options can be seen in the action buttons at the bottom like "Symbols help", "Save and Calculate", and "Open input". At the top one finds the possibilities to see the "Introduction", the "Figure" (depicting the various segmented regression types), the "Output", and the "Graphics".