CSIRO Mathematical and Information Sciences

Analysing Trends in Groundwater Levels

Prepared for Agriculture Western Australia

As Part of a National Land and Water Resources Audit Project

By Q Shao[1], N A Campbell[1], R. Ferdowsian[2] and D. O'Connell[2]

February 1999

Report Number:
CMIS 99/37

[1]CSIRO Mathematical and Information Sciences
Leeuwin Centre
Floreat, WA 6014
[2]Agriculture WA
444 Albany Hwy
Albany, 6330

SUMMARY

This report presents some results from a new approach developed by CSIRO Mathematical and Information Sciences (CMIS) for determining trends in groundwater levels.

Inspection of a number of sets of borehole data indicates that there are two basic patterns. A linear trend (or trends) describes the medium- to long-term changes in the groundwater level, while a periodic response describes the seasonal variations. The observed trends in the groundwater level can sometimes be modelled by either one of these patterns and sometimes by a combination of both.

For many boreholes, different periods of time exhibit different (roughly) linear trend. These periods of time are referred to here as segments. The breaks in the trends at change points or thresholds are possibly related to hydrological events, such as rainfall or a change in land management.

The approach developed by CMIS allows for different linear trends in different segments over the duration of the records. A 12-month seasonal cycle can be superimposed on the segments; the amplitude of the cycle can also vary from segment to segment.

Three distinct types of change are allowed from segment to segment. Both the linear trend and periodic response can change and a discontinuity may occur. The linear trend can be continuous and the periodic responses can be the same for both segments. The periodic response can be continuous and the linear trends can be the same for both segments. Or the linear trend and periodic response can change at the same time, but in a continuous manner. These four types of threshold changes may happen for a single bore during the sampling period.

The suitability of these threshold models is examined by applying them to a series of bores from the south-west of Western Australia.

1. Introduction

The appropriate description of borehole data is an important component of the Dryland Salinity theme of the National Land and Water Resources Audit.

The usual approach is to fit a straight line to the data over time (after some editing of the values) or to the annual minima when a strong seasonal response is evident.

This report shows some results from a new approach developed by CSIRO Mathematical and Information Sciences (CMIS) for determining trends in groundwater levels.

The depth of the groundwater table is sampled for each borehole at irregular time intervals. The sampling frequencies can vary from one month to three months and sometimes six months.

The fitting of a simple linear regression of water level on time often provides questionable results, as the groundwater trends tend to follow different patterns over different periods.

The CMIS approach allows a consistent methodology to be applied to all borehole data. The approach recognises that a typical hydrograph may exhibit one or both of the following characteristics:

1. linear trends in water level over time - the level and nature of the trends may change as a result of significant rainfall events; and

2. a seasonal fluctuation in the water level.

Some examples are given in the next Section.

2. Variation in Borehole Response

Thousands of bores have been drilled and monitored throughout Western Australia. The sampling period varies from less than a year to over ten years. For each borehole, the sampling intervals are unequal and can vary from one month to three months, and sometimes as much as six months.

Different bores in the database show varying trends in groundwater level. Typical examples are shown in Figures 1a. and 1b. In most cases, a combination of both trend and seasonal response will be evident, as illustrated in Figure 1c. Many of the boreholes show changes in the trends at particular dates, attributable to external factors such as weather and changes in land management practices. Examples are shown in Figure 1d. and 1e.

Figure 1 - Some typical groundwater trends

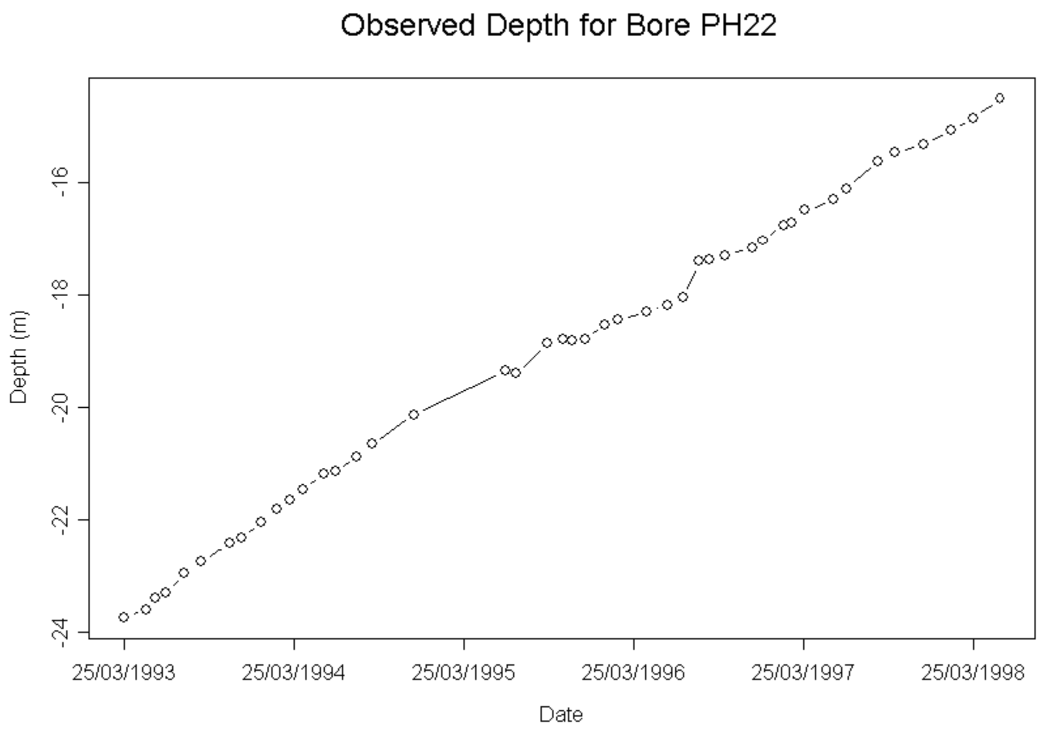


Figure 1(a) - Linear trend

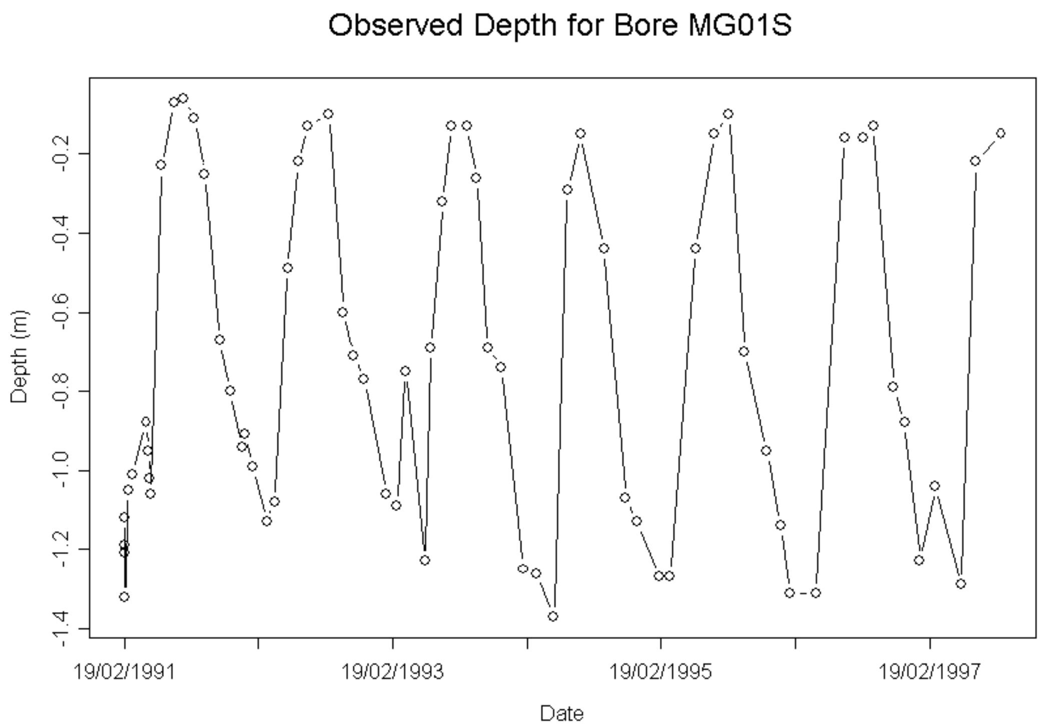


Figure 1(b) - Periodic response

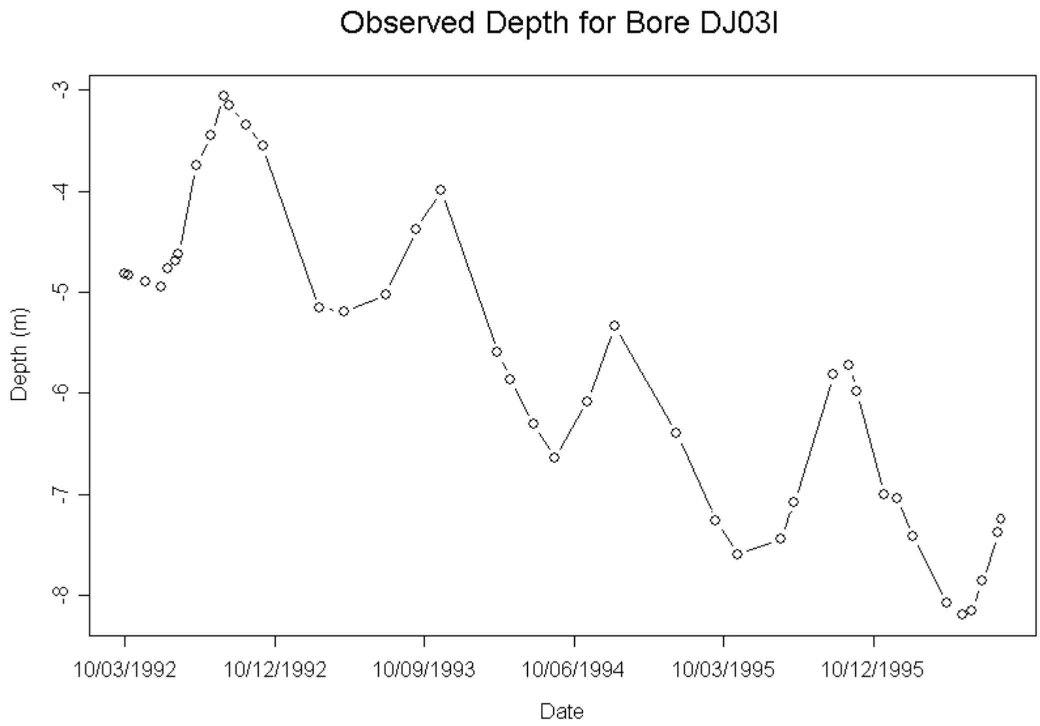


Figure 1(c) - Both linear trend and periodic response

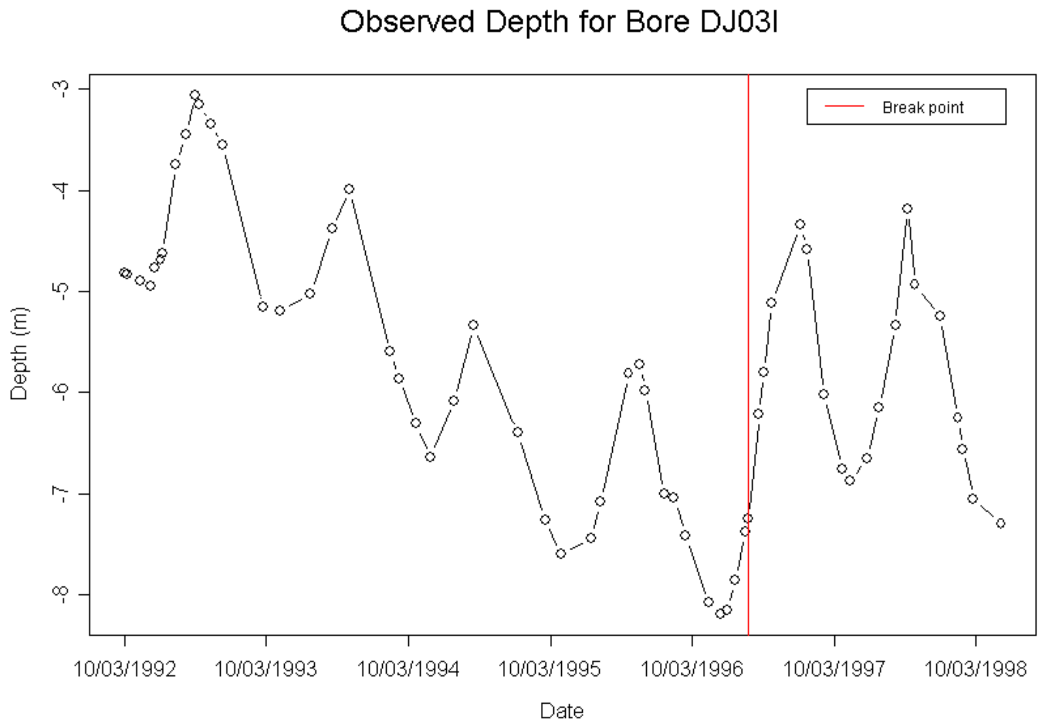


Figure 1(d) - Two segments with different linear trends and periodic responses

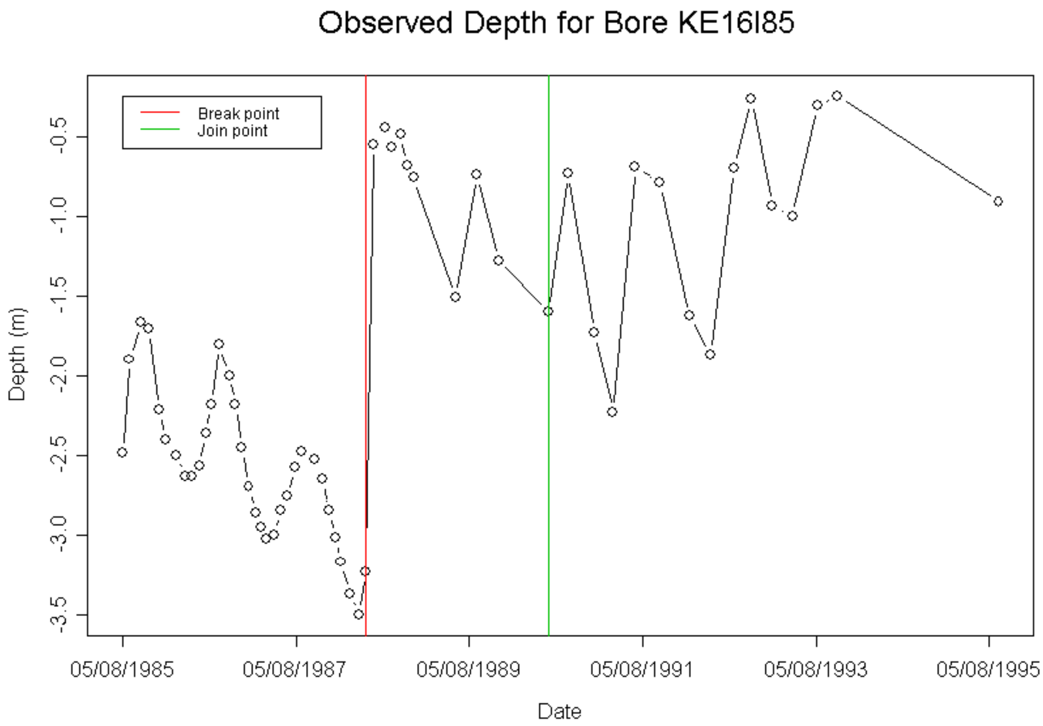


Figure 1(e) - Three segments with discontinuous linear trend

3. Statistical Approach to Modelling Borehole Data

The approach developed by CMIS allows for different linear trends in different segments over the duration of the records. A 12-month seasonal cycle can be superimposed on the segments; the amplitude of the cycle can also vary from segment to segment.

Three distinct types of change are allowed from segment to segment. Both the linear trend and periodic response can change and a discontinuity may occur. The linear trend can be continuous and the periodic responses can be the same for both segments. The periodic response can be continuous and the linear trends can be the same for both segments. Or the linear trend and periodic response can change at the same time, but in a continuous manner. These four types of threshold changes may happen for a single bore during the sampling period; examples are shown in Figures 2a and 2b.

Figure 2 - Some more complicated groundwater trends

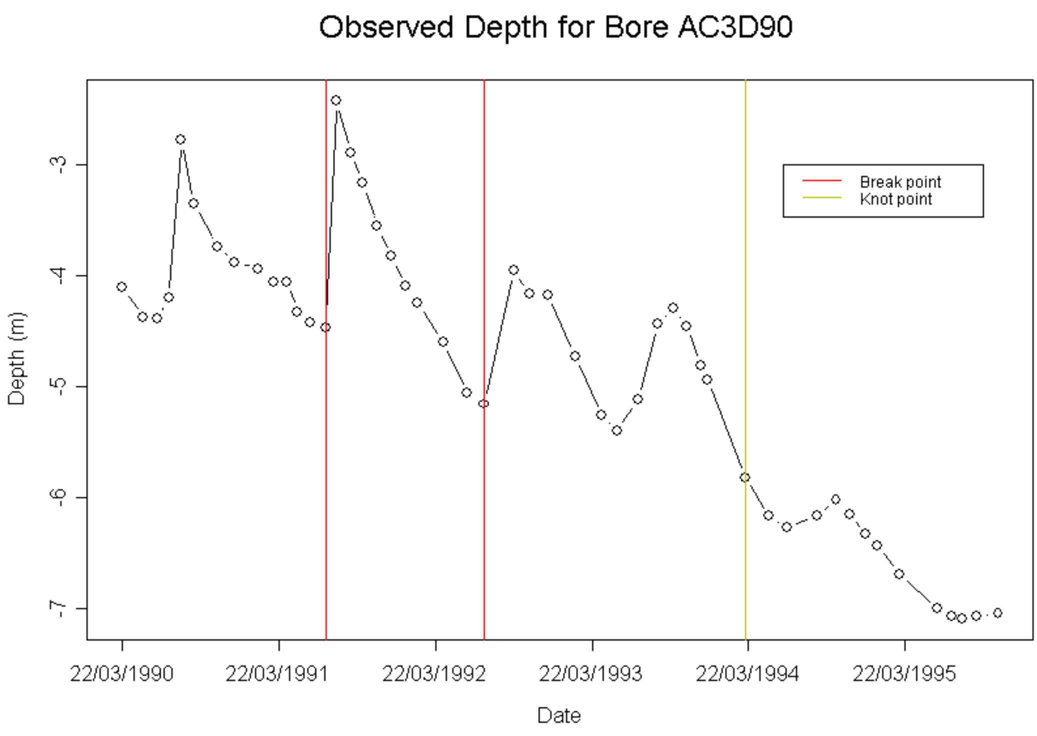


Figure 2(a) - Four segments with discontinuous and continuous linear trends

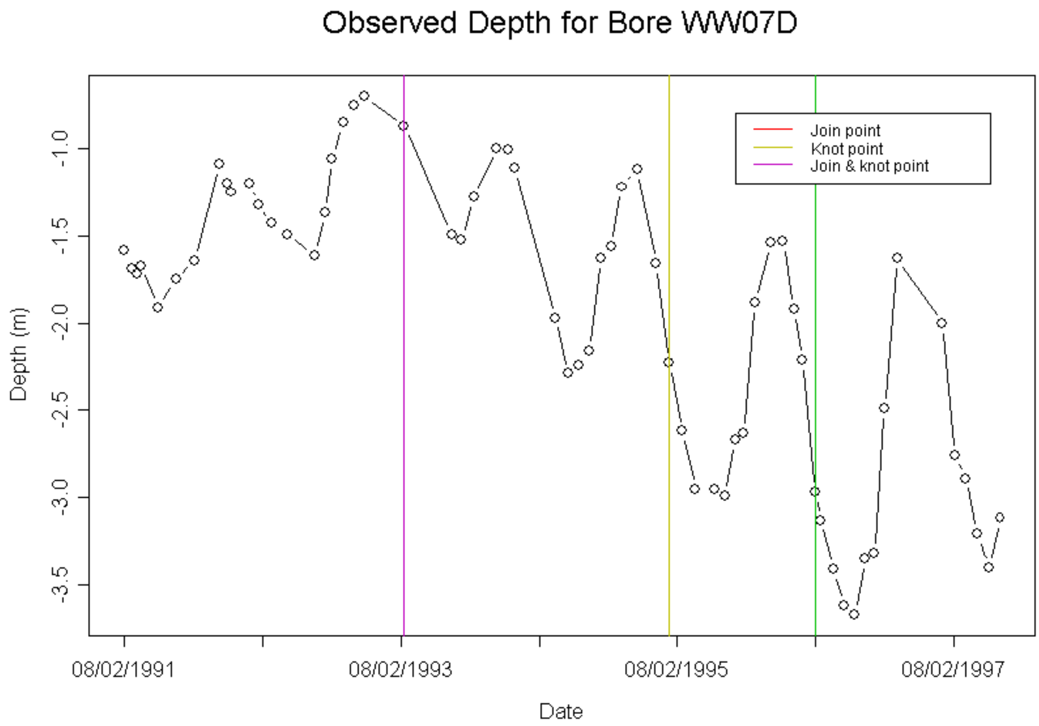


Figure 2(b) - Four segments with different linear trends and seasonal responses

The detection of thresholds or change points is important in practice because it encourages an investigation into their causes, such as variation in weather (rainfall and temperature) and land management. This information can then potentially be integrated into the modelling and subsequent spatial analysis.

This section outlines in stages a statistical model that is capable of modelling the trends that have been identified

When fitting the models discussed in this report, the set of candidate segment break points are chosen interactively by the expert.

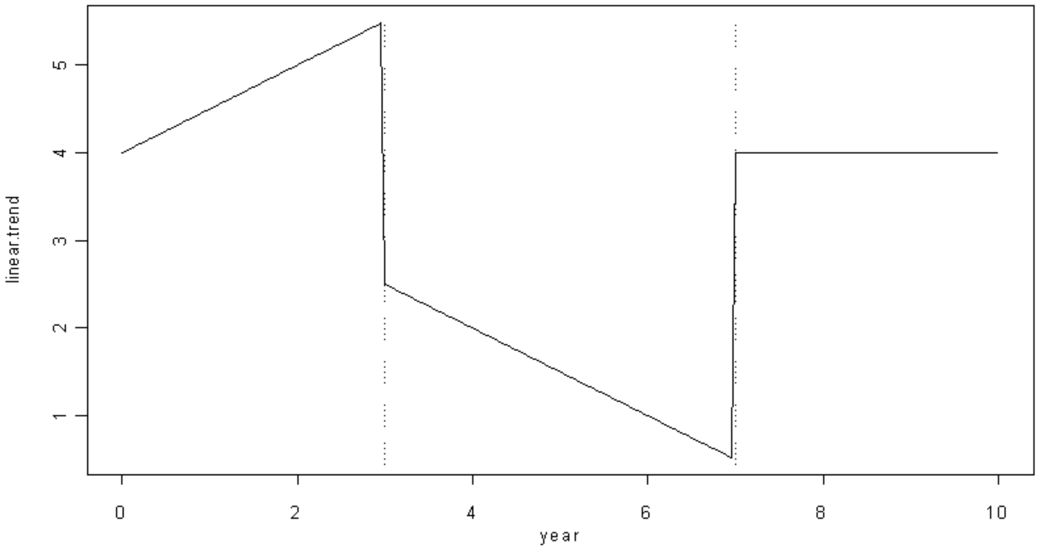
A challenging issue is the selection of an optimal model. In regression analysis, a model fit can be measured by its sum of squared errors, which decreases as the number of parameters increases. A criterion is needed so that not too many parameters are used in selecting competing models. A modified Akaike Information Criterion is adopted here, which penalises the sum of squared errors by a function of the number of free parameters.

3.1. Simple Threshold Regression

A simple threshold regression assumes different intercepts and linear trends for each segment. The model is characterised by the intercepts and slopes for each segment, and the threshold values which define the segments. An idealised example with three segments is illustrated in Figure 3.

Figure 3 - An illustration of a simple threshold model on time with segments

The solid lines denote the linear trends; the vertical dotted lines delineate the segments.



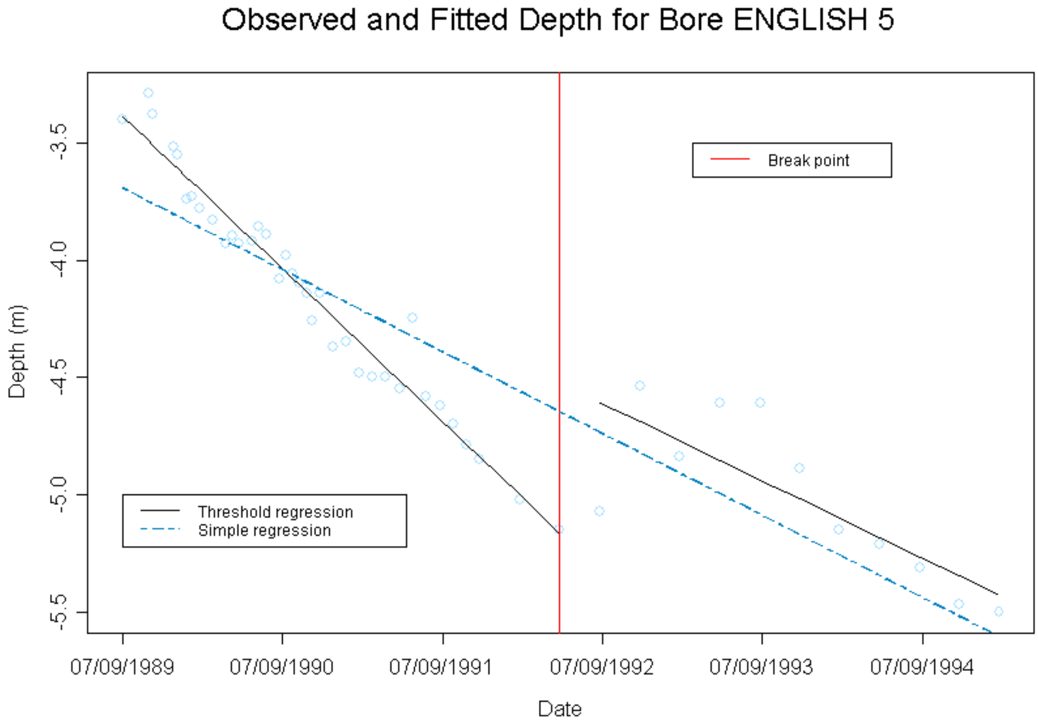
English Bore #5 is used to illustrate the approach. Potential candidate change points are (1-Aug-90, 1-Jun-91, 1-Jun-92, 1-Dec-93), giving possibly five linear segments. 01-Jan-85 is taken as day 1, and the unit of time is a year. The optimum model has two segments, giving as the fitted model

$$Y_t = \begin{cases} -0.34 - 0.650t & \text{if} \quad t \le 7.42 \\ -2.09 - 0.328t & \text{if} \quad t > 7.42 \end{cases}$$

where the break point is at 1-Jun-92. The model fit and comparison with a simple linear regression are plotted in Figure 4. For this threshold model, the estimated rates of change are 65cm/year and 32.81cm/year respectively, for the two segments.

The simple linear regression is $Y_t = -2.06 - 0.349t$, giving an estimated rate of decrease of 34.88cm/year.

Figure 4 - Observed and fitted depths for G. English Bore 5

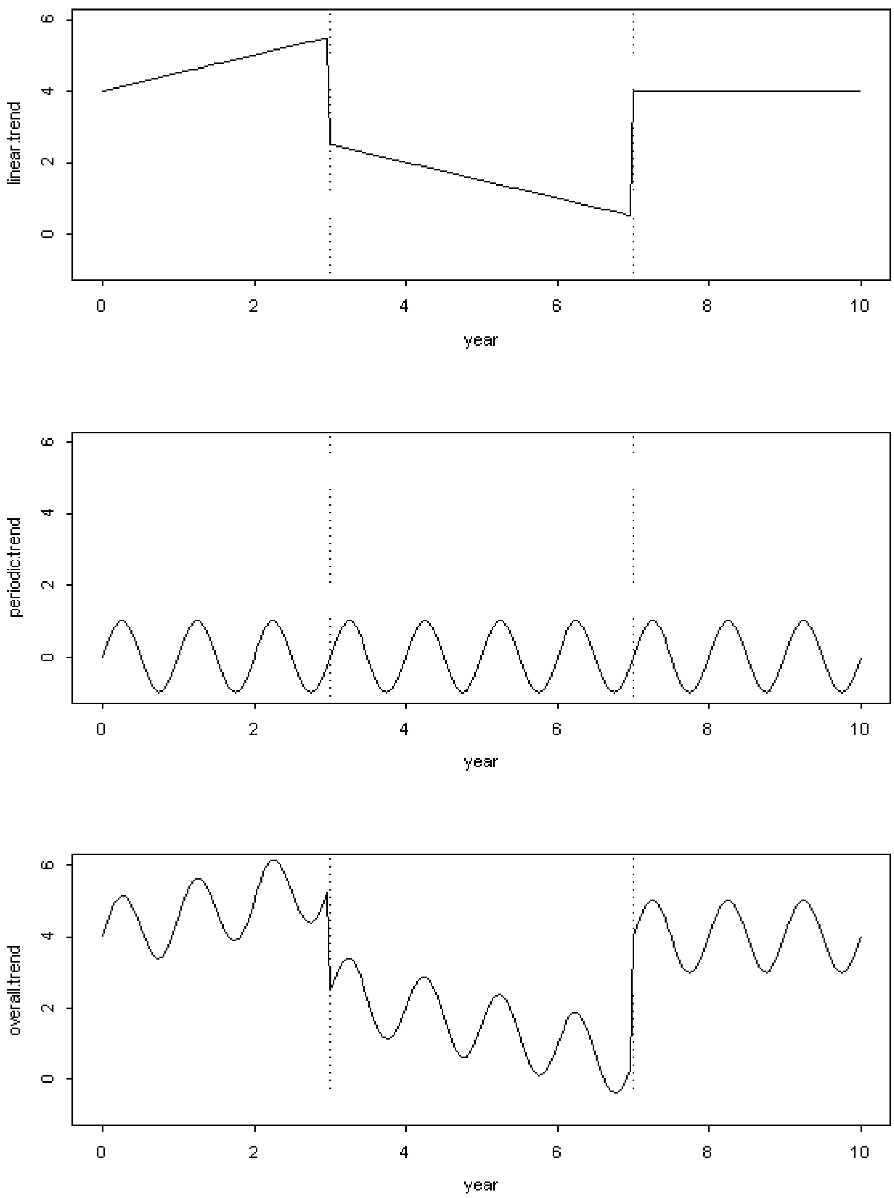


3.4 An Extension to a Special Periodic Response

Assume now that a periodic response is superimposed on the linear trends. The interaction between the two components is illustrated in Figure 5.

Figure 5 - Combination of linear trends and periodic response for threshold model with three segments

The vertical dotted lines indicate the segment break points. The top plot gives the underlying linear trends. The middle plot gives the overall periodic response. The bottom plot gives the overall trends. In real applications, the plots will be subject to random fluctuations.



In our applications, the period is assumed to be known and equal to one year, representing the annual cycle.

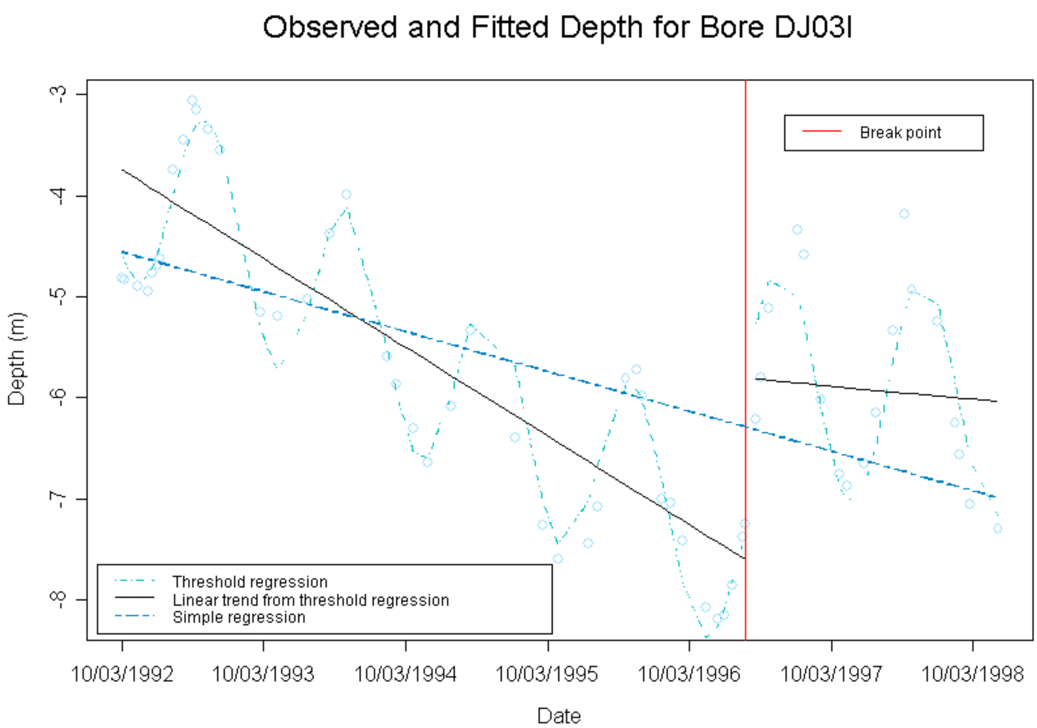
This model is applied to Bore DJ03I, which is the example given in Figure 1c. The potential change points are selected as (24-Aug-94, 31-Jul-96, 14-Aug-97), giving possibly four linear segments. The optimal model is given by

$$Y_t = \begin{cases} 2.57 - 0.878t + 1.025\sin(2\pi t - 0.30) & \text{if} \quad t \le 11.59 \\ -4.27 - 0.133t + 1.152\sin(2\pi t - 0.59) & \text{if} \quad t > 11.59 \end{cases}$$

where the break point is at 31-Jul-96. The model fit and comparison with a simple linear regression are plotted in Figure 6. In the threshold model, the estimated rates of decrease are 87.78cm/year and 13.26cm/year respectively.

The simple linear regression is $Y_t = -1.72 - 0.394t$, giving an estimated rate of decrease of 39.41cm/year.

Figure 6 - Observed and fitted depths for Bore DJ03I



3.5. Threshold Model with Constraints on Changes in Trend

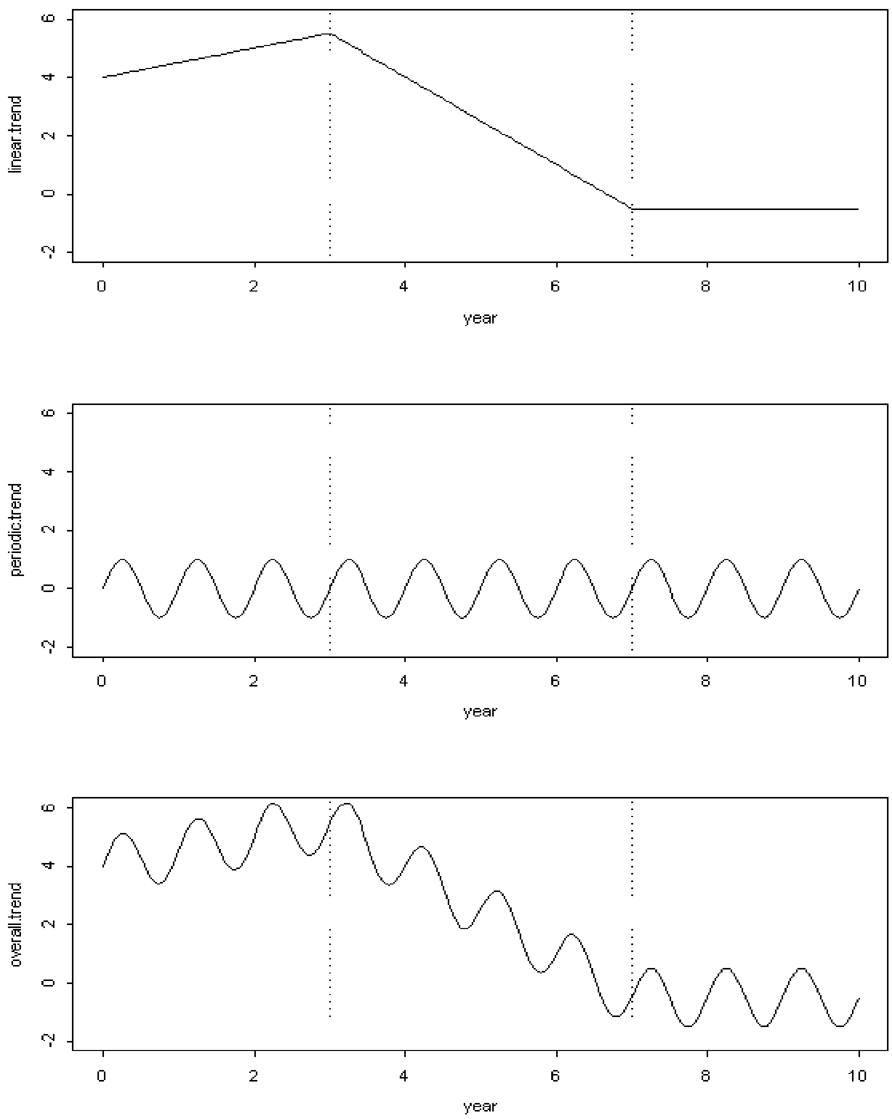
In the last two sub-sections, the linear trends and periodic responses were allowed to change freely. In this sub-section, continuity constraints are imposed on the segments.This requires simultaneous model fitting across segments, because either the linear trend or periodic response may be unchanged at the join points.

Threshold Model with Common Periodic Response

For this model, the periodic response remains unchanged at a join point, but the slopes are changed. Figure 7 illustrates an idealised plot.

Figure 7 - Threshold model with common periodic response

The vertical dotted lines indicate the join points. The top plot gives the underlying linear trend. The middle plot gives the periodic response. The bottom plot gives the overall trend. In real applications, the plots will be subject to random fluctuations.



This model is applied to Bore KE16I85 (see Figure 8), which is the example given in Figure 1e. There is a clear break point at 25-May-1988; this is fixed during the model selection. The possible join points in the second segment are (04-Jul-1990, 25-Aug-1992). The optimal model is given by

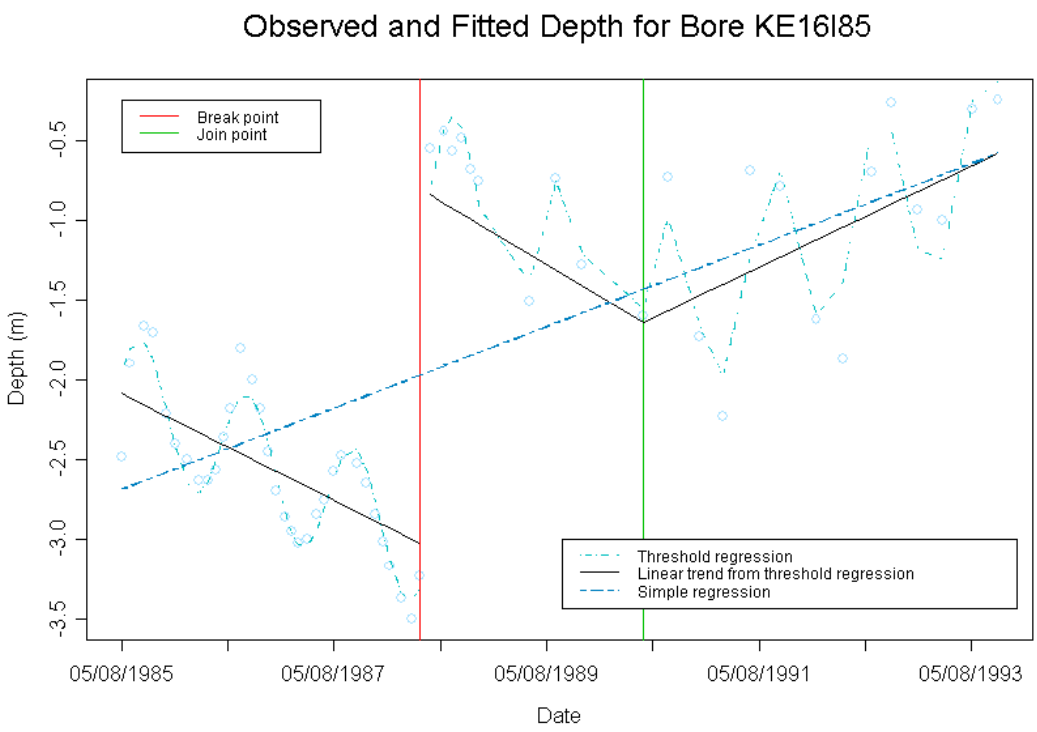
$$Y_t = \begin{cases} -1.89 - 0.335t + 0.396\sin(2\pi t - 0.20) & \text{if} \quad t \le 3.40 \\ 0.54 - 0.395t + 0.711(t-5.51)_+ & \\ \qquad + 0.584\sin(2\pi t - 0.10) & \text{if} \quad t > 3.40 \end{cases}$$

where the break point is at 31-Jul-96, and the + notation is $u_+ = u$ if $u > 0$, 0 if $u \le 0$.

The model fit and comparison with a simple linear regression are plotted in Figure 8. In the threshold model, the estimated rates of decrease are 33.47cm/year and 39.53cm/year respectively, for the first two segments and the estimated rate of increase is 30.57cm/year for the third segment.

The simple linear regression is $Y_t = -2.84 + 0.255t$, giving an estimated rate of increase of 25.52cm/year.

Figure 8 - Observed and fitted depths for Bore KE16I85

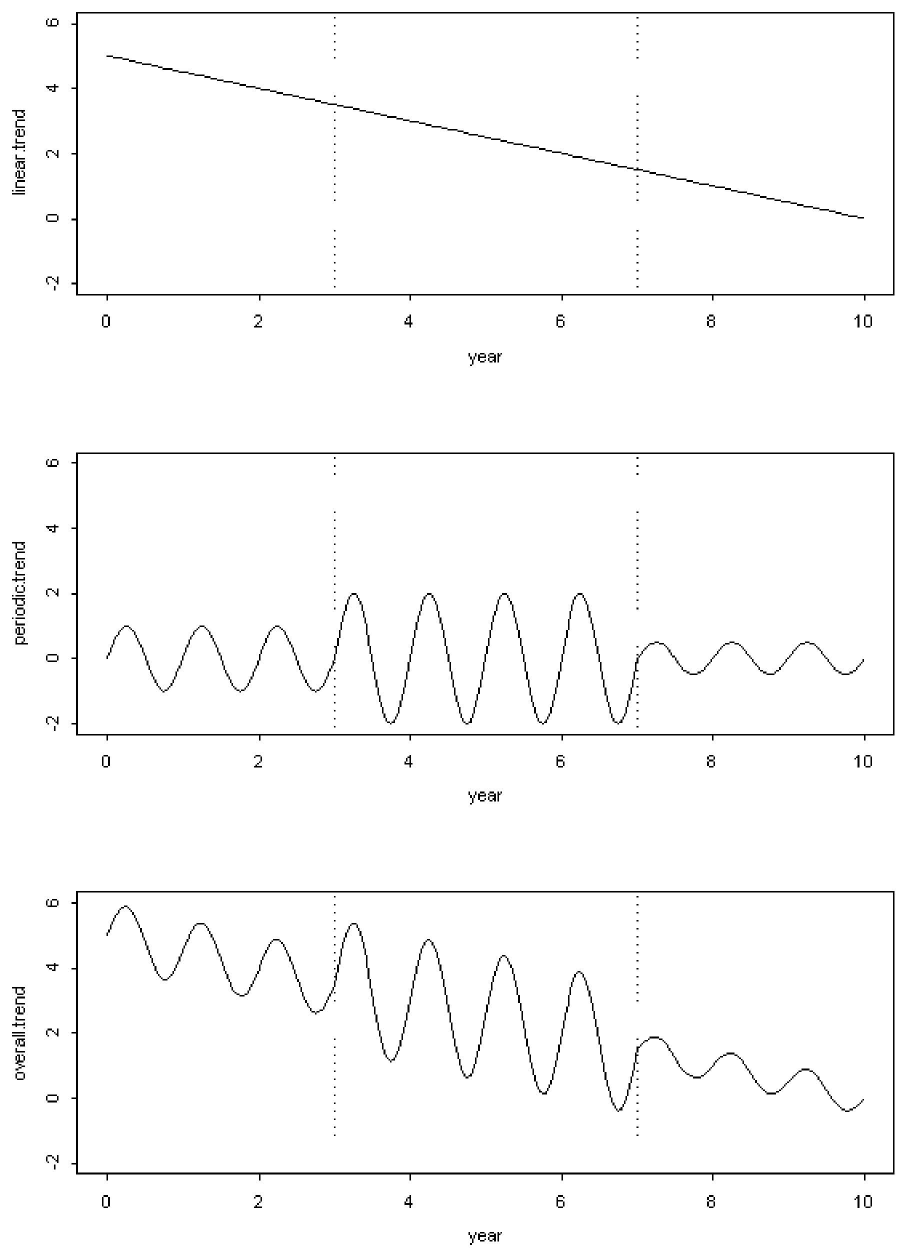


Threshold Model with Common Linear Trend

In this model, the linear trend remains unchanged at a change point, but the amplitudes and phases change. In order to make the model easier to interpret, it is assumed that all the phases are the same. Figure 9 shows an idealised plot.

Figure 9 - Model with common linear trend and different periodic responses

The vertical dotted lines indicate the join points. The top plot gives the underlying linear trend. The middle plot gives the periodic responses. The bottom plot gives the overall trend. In real applications, the plots will be subject to random fluctuations.



In order to ensure the continuity of the regression curve at the join points, it is assumed that the join points always occur at integer values. This assumption implies that the amplitudes can only change at the completion of a full cycle, which is a reasonable assumption in practice.

This model is applied to part of Bore AC3D90 (from September 1992). The possible join points are in February each year, which correspond to (11-Feb-1993, 15-Mar-1994, 10-Mar-1995). The optimal model is given by

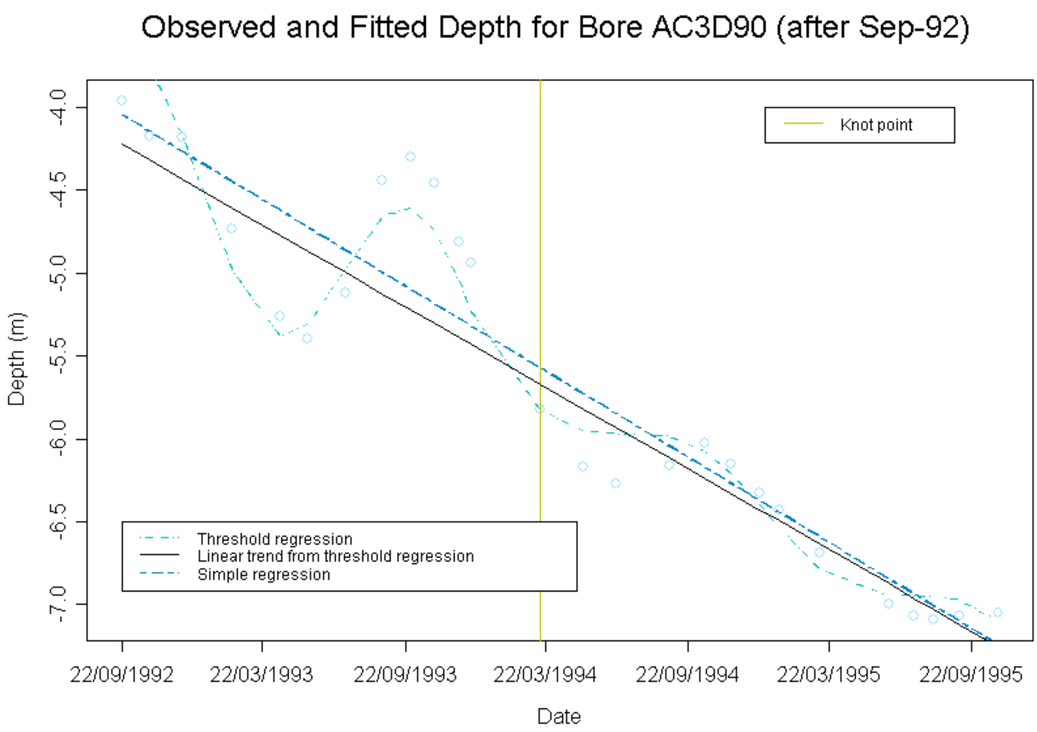
$$Y_t = 3.34 - 0.979\,t - 0.618\sin(2\pi t - 30/365)$$
$$+0.454\,\delta_{(t-9.21)}\sin(2\pi t - 30/365)$$

where the knot point is at 15-Mar-94, and $\delta_u = 1$ if $u > 0$, 0 if $u \le 0$.

The model fit and comparison with a simple linear regression are plotted in Figure 10. The estimated rate of decrease is 97.92cm/year. The estimated amplitudes are 61.84cm and 16.41cm ($= 61.84 - 45.39$) respectively.

The simple linear regression is $Y_t = 3.95 - 1.034\,t$, giving an estimated rate of decrease of 103.40cm/year.

Figure 10 - Observed and fitted depths for Bore AC3D90 (after September 1992)



4. More Applications to Borehole Data

Model selection is now demonstrated using Bore AC3D90.

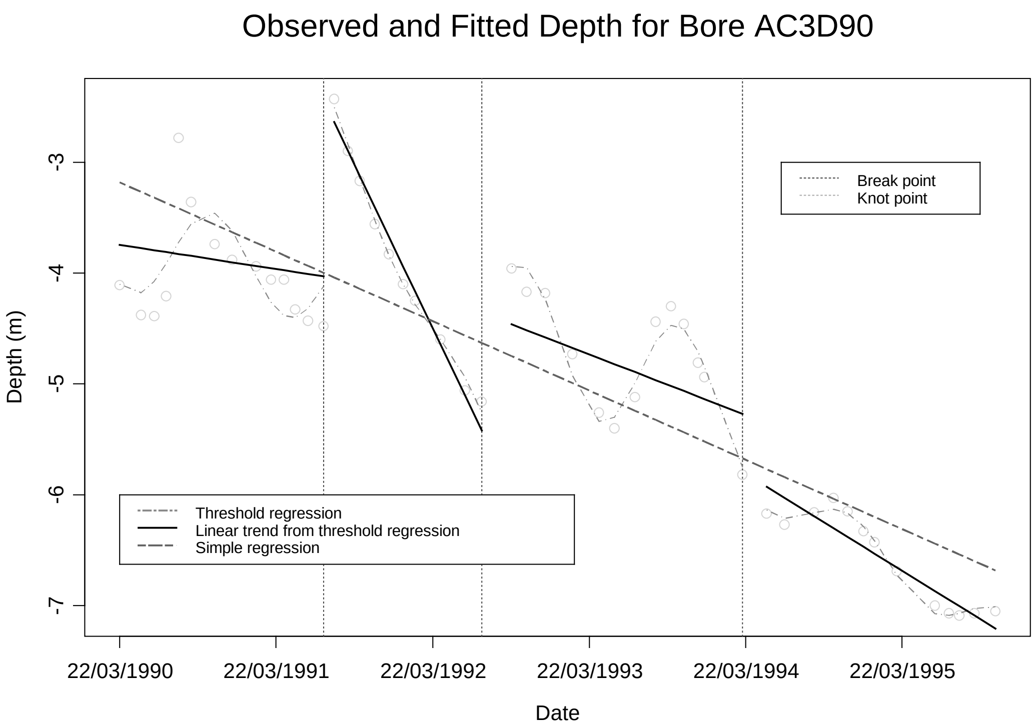
The simple linear regression is given by $Y_t = 0.08 - 0.625t$, giving an estimated rate of groundwater decrease of 62.52cm/year. Clearly a simple linear regression model is not appropriate for this dataset.

The potential thresholds are (11-Jul-1991, 14-Jul-1992, 15-Mar-1994). A threshold model is given by

$$Y_t = \begin{cases} -2.60 - 0.218t + 0.421\sin(2\pi t - 0.40) & \text{if} \quad t \le 6.65 \\ 16.83 - 2.951t + 0.174\sin(2\pi t - 1.48) & \text{if} \quad 6.65 < t \le 7.64 \\ 0.24 - 0.546t + 0.560\sin(2\pi t - 0.29) & \text{if} \quad 7.54 < t \le 9.21 \\ 2.25 - 0.874t + 0.214\sin(2\pi t - 0.88) & \text{if} \quad t > 9.21 \end{cases}$$

The estimated rates of decrease are 21.84cm/year, 295.12cm/year, 54.6084cm/year and 87.3884cm/year respectively, for the four segments. As can be seen, the rates of fall of the groundwater levels increase as time increases, except for the second segment. The amplitudes of the periodic responses tend to decrease. The model fit and residuals are given in Figure 13.

Figure 13 - Observed and fitted depths for Bore AC3D90

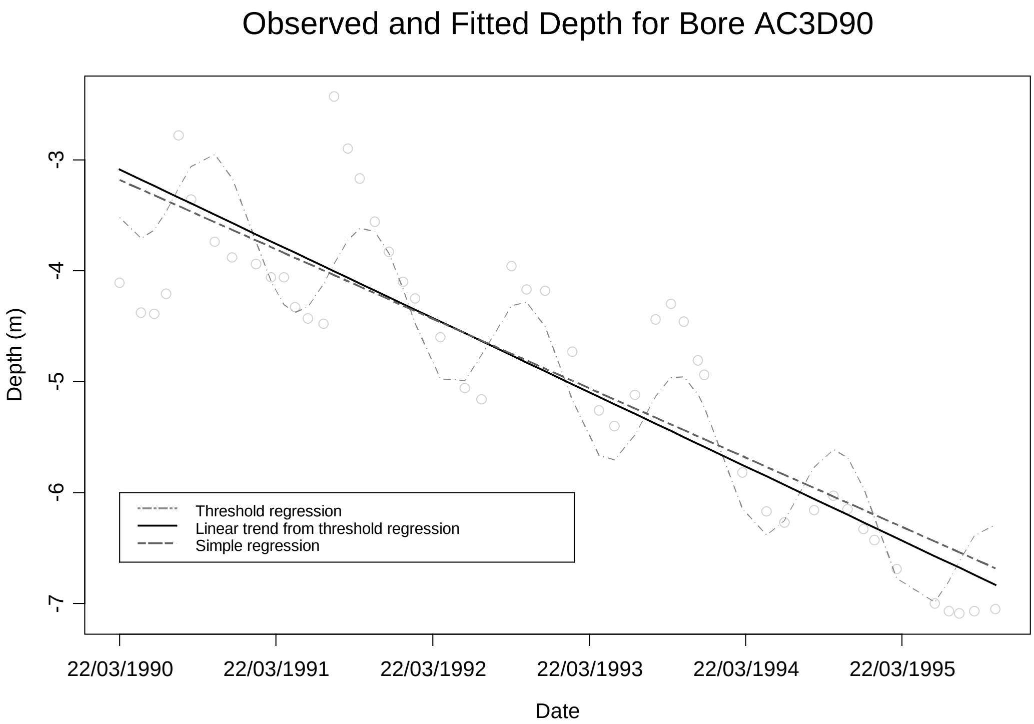


A regression model with linear trend and periodic response is given by

$$Y_t = 0.41 - 0.669\,t + 0.543\sin(2\pi t - 0.468)$$

The model fit and residuals are given in Figure 14. The estimated rate of decrease is 66.87cm/year. There are obvious patterns in the residuals.

Figure 14 - Observed and fitted depths for Bore AC3D90



Assume now that the full cycle starts at the beginning of February of each year. A threshold model with common linear trend is given by

$$\begin{aligned} Y_t = 0.34 - 0.666t \quad & -0.663\sin(2\pi t - 30/365) \\ & -0.063\sin(2\pi t - 30/365)\delta_{(t-t_1)} \\ & +0.039\sin(2\pi t - 30/365)\delta_{(t-t_2)} \\ & +0.569\sin(2\pi t - 30/365)\delta_{(t-t_3)} \end{aligned}$$

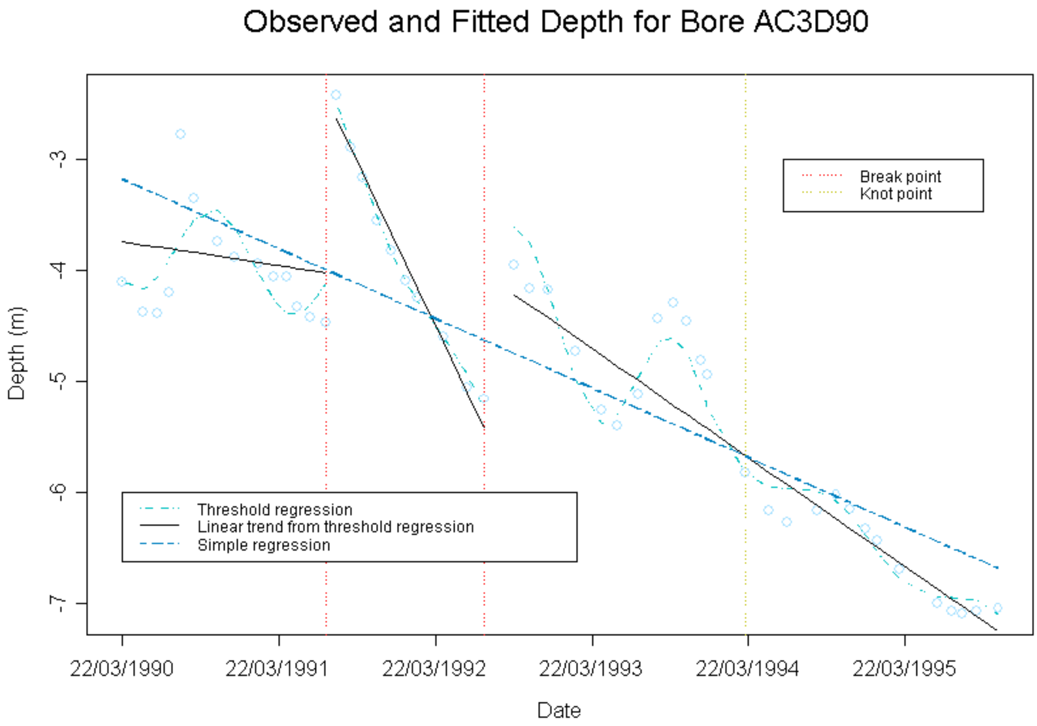
where the join points are $(t_1, t_2, t_3) = (6 - 30/365, 7 - 30/365, 9 - 30/365)$, which fall in the years 1991, 1992 and 1994 respectively. This model provides a better fit that the model with a single periodic trend.

The best-fitting model for these data is given by

$$x_t = \begin{cases} -2.60 - 0.218t + 0.421\sin(2\pi t - 0.397) & \text{if} \quad t \le 6.65 \\ 16.82 - 2.951t + 0.174\sin(2\pi t - 1.481) & \text{if} \quad 6.65 < t \le 7.64 \\ 3.45 - 0.988t - 0.554\sin(2\pi t - 30/365) & \\ \qquad + 0.394\sin(2\pi t - 30/365)\delta_{(t-t_0)} & \text{if} \quad t > 7.64 \end{cases}$$

The model fit and residuals are given in Figure 15.

Figure 15 - Observed and fitted depths for Bore AC3D90



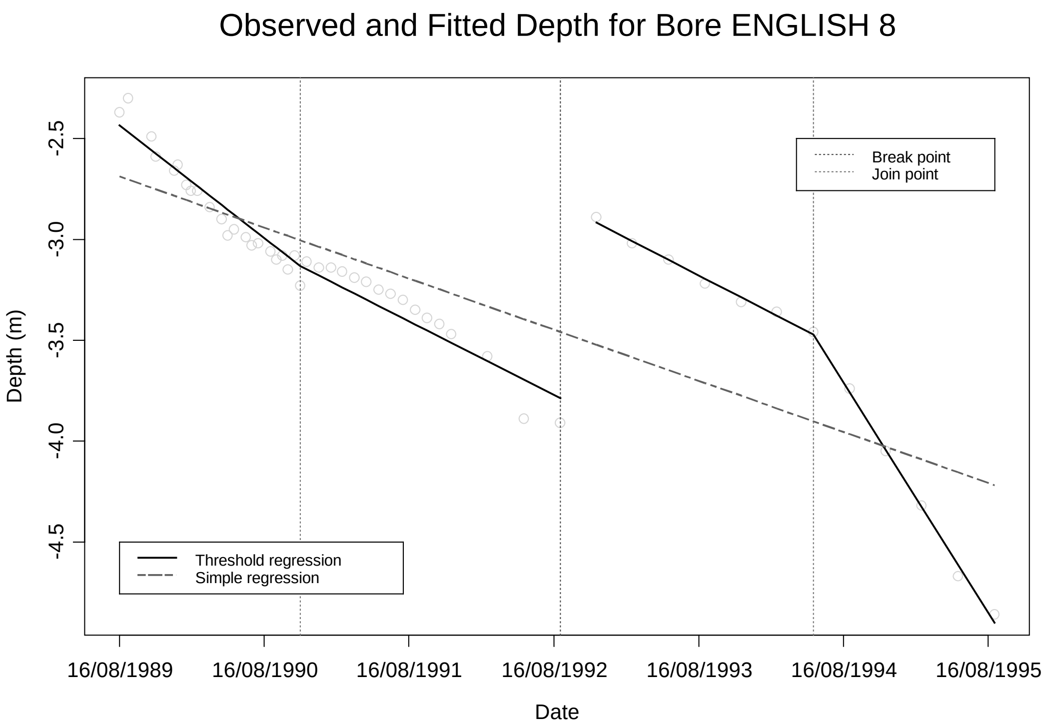
The balance between a statistically optimum fit and expert judgement is now demonstrated using G. English Bore #8. The candidate change points are (15-Nov-1990, 01-Sep-1992, 01-Jun-1994). The optimal model is given by

$$Y_t = \begin{cases} 0.15 - 0.558t + 0.194(t - t_1)_+ & \text{if} \quad t \le 7.67 \\ 0.02 - 0.371t - 0.771(t - t_2)_+ & \text{if} \quad t > 7.67 \end{cases}$$

where $t_1 = 5.874$ and $t_2 = 9.419$ correspond to 15-Nov-1990 and 01-Jun-94 respectively. The break point is 01-Sep-1992. The fit and comparison with a simple linear regression are plotted in Figure 16. The estimated rates of decrease are 55.80cm/year, 36.45cm/year, 37.05cm/year and 114.13cm/year respectively.

The simple linear regression is $Y_t = -1.52 - 0.253t$, giving an estimated rate of decrease of 25.32cm/year.

Figure 16 - Observed and fitted depths for Bore English 8



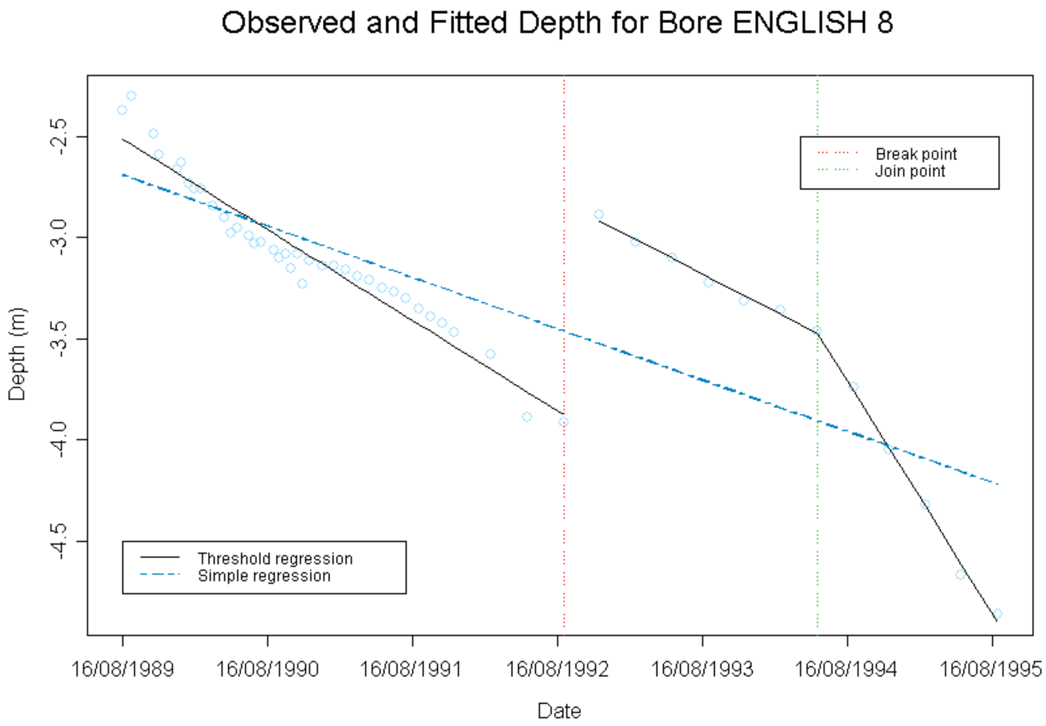
For this bore, expert advice suggests that 15-Nov-1990 is not a true threshold. After removing this threshold, the model is

$$Y_t = \begin{cases} -0.44 - 0.449\,t & \text{if} \quad t \le 7.67 \\ 0.02 - 0.371t - 0.771(t - t_2)_+ & \text{if} \quad t > 7.67 \end{cases}$$

The model fit is plotted in Figure 17. The estimated rates of decrease are 44.85cm/year, 37.05/year and 114.13cm/year respectively.

Figure 17 - Observed and fitted depths for Bore English 8

(removing the first join point from the previous model)



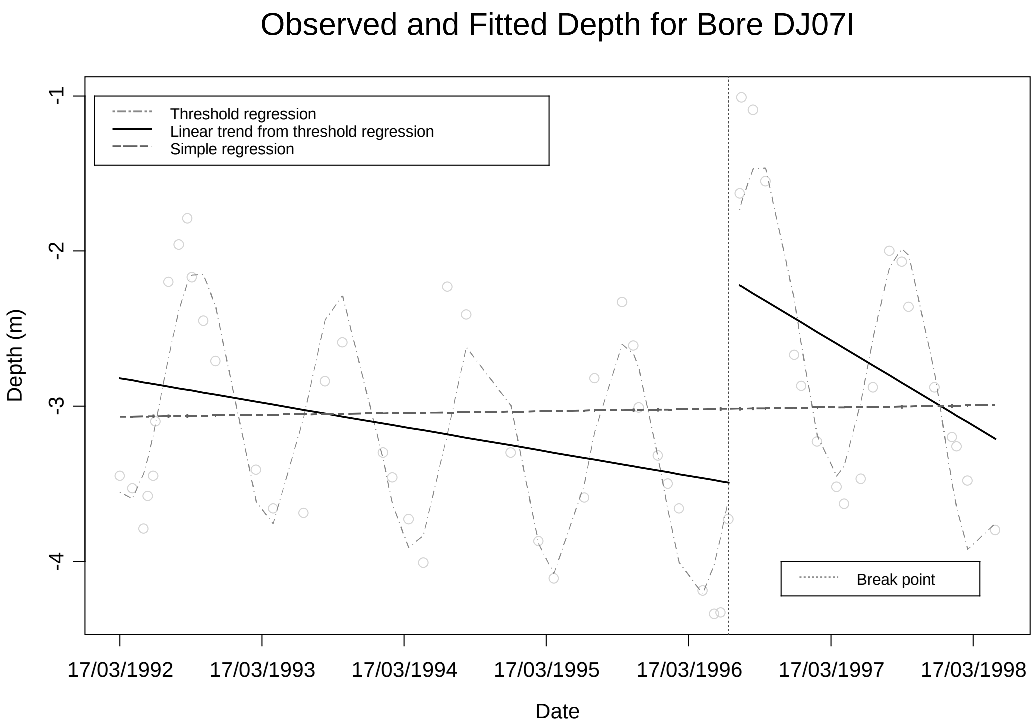
The approach is now applied to Bore DJ07I. The potential thresholds are (06-May-1994, 27-Jun-1996). The optimal model is

$$Y_t = \begin{cases} -1.69 - 0.157\,t + 0.780\sin(2\pi t - 0.096) & \text{if} \quad t \le 11.49 \\ 4.15 - 0.550\,t + 0.867\sin(2\pi t + 0.139) & \text{if} \quad t > 11.49 \end{cases}$$

where the break point is at 27-Jun-96. The model fit and comparison with a simple linear regression are plotted in Figure 18. The estimated rates of decrease are 15.70cm/year and 86.67cm/year respectively.

The simple linear regression is $Y_t = -3.16 + 0.012\,t$, giving an estimated rate of ***increase*** of 1.21cm/year.

Figure 18 - Observed and fitted depths for Bore DJ07I



5. Discussion

The examples presented above illustrate the power and flexibility of the segmented regression approach for describing borehole data developed by CMIS. The approach provides different linear trends for different segments of the data, and at the same time estimates the amplitude(s) of the seasonal response(s) in the data.

Both the slopes (or fitted values at a nominated time) and amplitudes can be used to characterise the borehole data, and these values can then be related spatially to appropriate characteristics of the landscape.