

# SS2020 – Modelling Ground Water Depth From Bore Hole Data

Rob Dunne

CSIRO Mathematical and Information Sciences  
Leeuwin Centre for Earth Sensing Technologies  
Private Bag, P.O., Wembley, Western Australia, 6014

Mon Mar 27 11:25:31 WST 2000

# Contents

<b>1</b>	<b>Summary and Conclusion</b>	<b>3</b>
<b>2</b>	<b>Kent</b>	<b>5</b>
2.1	The <b>bores.kent</b> data set . . . . .	5
2.1.1	the standard analysis . . . . .	13
2.1.2	generalized additive models . . . . .	15
2.1.3	interactions . . . . .	19
2.1.4	robust linear models . . . . .	20
2.1.5	conclusions . . . . .	25
2.2	The <b>new.bores.kent</b> data set . . . . .	26
2.2.1	new.kent.model.1 . . . . .	26
2.2.2	new.kent.model.2 . . . . .	31
2.2.3	new.kent.model.3 . . . . .	33
2.3	Conclusions . . . . .	35
<b>3</b>	<b>Boscabel</b>	<b>36</b>
3.1	Combined Boscabel and Woodanilling . . . . .	36
3.2	Boscabel model . . . . .	37
3.2.1	solution 1 . . . . .	38
3.2.2	solution 7 . . . . .	44
3.2.3	solution 13 . . . . .	47
3.2.4	using only “trusted” bores . . . . .	47
3.3	The Kent model . . . . .	48
3.4	Conclusion . . . . .	48
<b>4</b>	<b>Woodanilling</b>	<b>52</b>
<b>5</b>	<b>Gnowangerup</b>	<b>56</b>
<b>6</b>	<b>Southern Stirlings</b>	<b>59</b>
<b>7</b>	<b>Toolibin</b>	<b>66</b>
7.1	Central Bores . . . . .	72
7.2	Eastern Bores . . . . .	75

<i>SS2020 – Modelling Ground Water Depth From Bore Hole Data</i>	2
7.3 Southern Bores . . . . .	78
7.4 Conclusion . . . . .	81
<b>A Questions of the DEM</b>	<b>82</b>
<b>B Robust Regression</b>	<b>87</b>

# 1

## Summary and Conclusion

The conclusions that emerge from attempts to model ground water depth using DEM-derived variables on the basis of the given data are as follows:

- for the Kent area, it is possible to derive a model that fits the bore-hole data ( $R^2=0.56$ )<sup>1</sup> such that the resulting hydraulic head is plausible;
- for the Boscabel scenario region:
  - the fit to the bore data yields  $R^2$  values as low as 0.003;
  - the resulting maps of hydraulic head surface (whether fitted by robust methods or not) make no real sense;
- for the other scenario areas, the results lie between these two extremes. In each of these cases (Woodanilling, Gnowangerup and the Southern Stirlings) the resulting hydraulic head surfaces have a number of plausible features but also have a number of problems such as excessively steep gradients.

How far these hydraulic head surfaces can be taken as predicting the depth of ground water is open to question. Two things should be considered:

- up to 1/3 of the bores have been effectively removed from the modeling process by the robust fitting procedure in some scenario areas. If the depths of these bores are accurate then clearly the hydraulic head surface will be erroneous at these points. If

---

<sup>1</sup>The least squares  $R^2$  values on the full data set are extremely low. The robust fitting techniques work by weighting each observation so that, for example, a “good” observation may get a weight of 1 (the maximum) and a “bad” observation may get a weight of 0 (the minimum). Some of the down-weighted bores can be seen, on examining the data-base, to be questionable; for others there is no apparent reason why they should be down-weighted. The  $R^2$  calculated with these weights will be higher (often much higher) than the  $R^2$  calculated on the raw values. However, either the raw or the robust  $R^2$  may give an unrealistic picture of the accuracy of the fit: the raw  $R^2$  because it includes bores that are questionable; and the robust  $R^2$  because it may be down-weighting “good” bores to make the data fit the model. The plausibility of the resulting hydraulic head appears to be the most satisfactory way to evaluate the models, despite the degree of subjectivity involved.

we can extrapolating from this, we would expect the surface to be accurate (within the uncertainty bounds) for 2/3 of the region and inaccurate (and perhaps wildly inaccurate) in 1/3 of the region;

- in some scenario areas (Boscabel is the worst instance) the fitted robust regression model is driven by a very small number of points. If these points are inaccurate then the model will likewise be inaccurate.

There are several reasons why modeling the hydraulic head may not be possible. These include:

- the explanatory variables may not reflect the physical processes that determine ground water depth. This is certainly at least partly true; ground water depth is determined by underground local features that are not reflected in the data;
- the poor quality of the bore data base may preclude effective modeling, see Hodgson (2000).

This leaves us with no obvious explanation as to why the technique gives some promise in Kent and none in Boscabel. We now know considerably more about the state of the bore-hole data than we did at the beginning of the project. None of the bores in the Kent region has a surveyed location and there is no compelling reason to suppose that the quality of the Kent data is superior to that of the other regions.

Figure 2.22 shows an hydraulic head surface for the Kent region. While this has a number of errors, it accords well with the general knowledge of the region and many of the problems with it could be rectified in a piecemeal fashion. Figure 4.3 shows an hydraulic head surface for Woodanilling. It appears that all of the bores in the Woodanilling area are quite shallow and there is a sampling bias in that the bores are located where there is an imminent danger of salinity (or to monitor a treatment in such a case), so there are no bores in the higher parts of the landscape. The result of this is that the fitted hydraulic head is shallow everywhere.

In the Toolibin area an attempt was made to use the DEM-derived variables to partition the landscape into units such that the depth to ground-water can be more effectively predicted with each unit. This was unsuccessful. An area with deep ground water (the eastern area) could not be separated from an area of shallow ground water (the southern area) on the basis of the DEM-derived variables. However the areas were clearly better fitted by separate regression models. It has been suggested that elapsed time since clearance is the variable that distinguishes these areas; unfortunately it is not available for most of the region.

## 2

# Kent

This section documents the analyses done on the Kent catchment, using two sets of data

- **bores.kent** – the augmented data for Kent, consisting of 127 bores;
- **new.bores.kent** – the trend bores for Kent plus the 40 bores from the study Salama et al. (1997), giving a total of 113 bores.

### 2.1 The bores.kent data set

The explanatory variable data set consists of 10 variables measured for each of the  $25 \times 25$  meter pixels in the Landsat TM image for the study area. Using the easting and northing for the 127 bores, a set of training data was extracted. The variables were as follows:

- depth to ground-water, **depth** – this is the response variable;
- average height of drainage area, **AvHgt**;
- average slope of drainage area, **AvSlp**;
- flow slope, **FISlope**;
- height above salt, **h.above.salt**;
- height above streams, **h.above.streams**;
- elevation, **elev**;
- water accumulation, **wa**, we used  $\log(\mathbf{wa} + 0.001)$  for this study;
- flow path length, **fplen**;
- percent cleared. **PerClr**. is derived from **wa** and **UpClr** via

$$\mathbf{PerClr} = 100\mathbf{UpClr}/\mathbf{wa};$$

- upland cleared, **UpClr**, we used  $\log(\text{UpClr} + 1.0)$  for this study.

First of all consider some scatterplots of the data in figures 1 to 11. It is apparent from these plots that no strong and obvious relationships exist between the explanatory variables and depth to ground water.

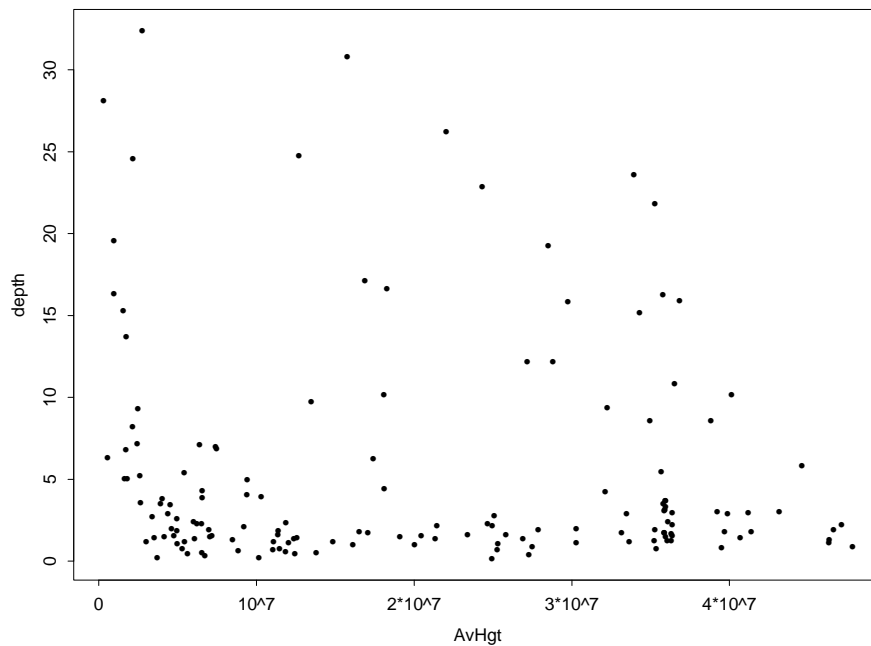


Figure 2.1: ***AvHgt*** versus ***depth***

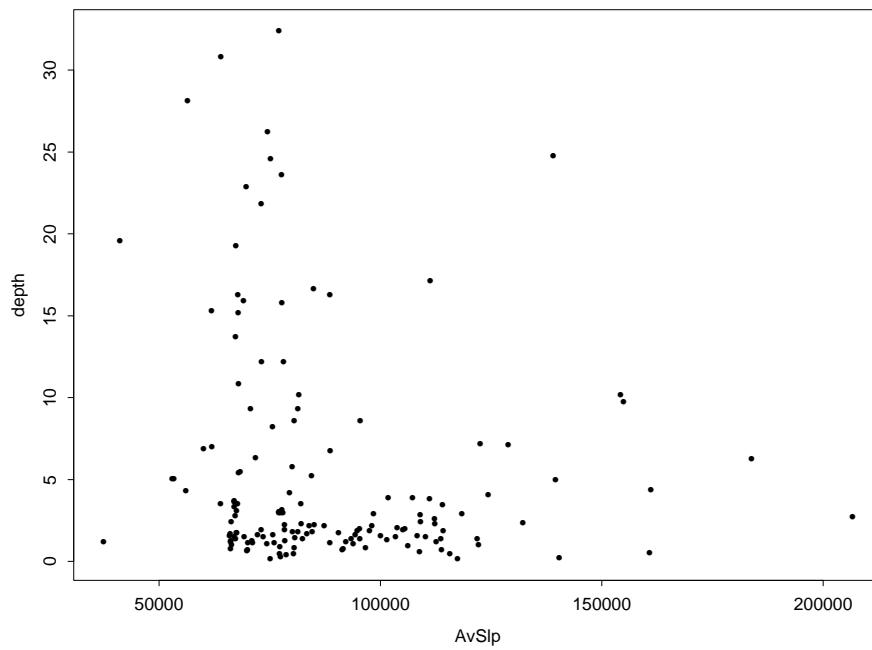


Figure 2.2: *AvSlp* versus *depth*

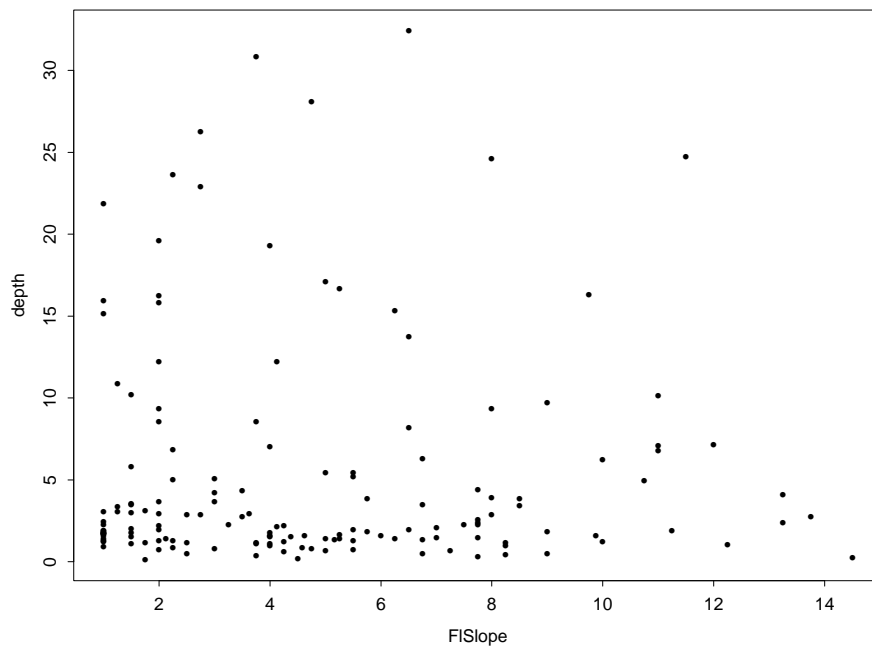


Figure 2.3: *FISlope* versus *depth*

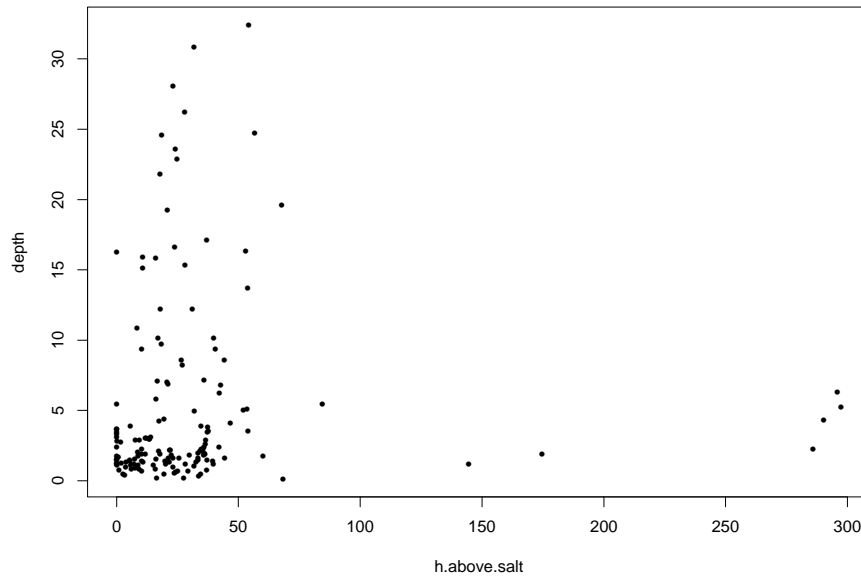


Figure 2.4: ***h.above.salt*** versus ***depth***

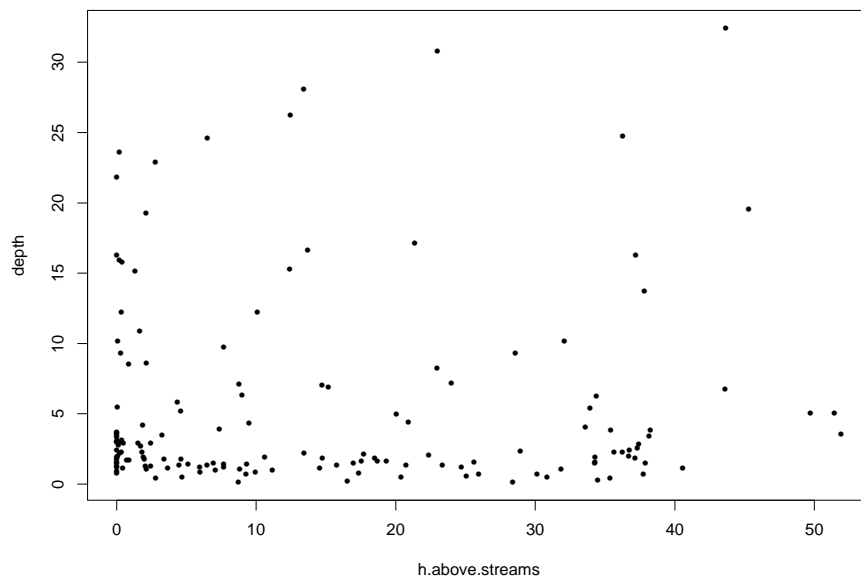


Figure 2.5: ***h.above.streams*** versus ***depth***

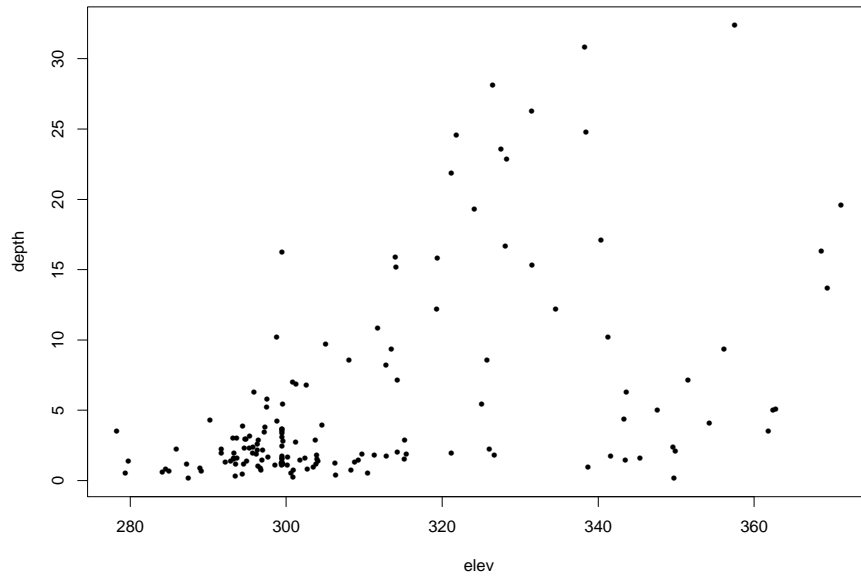


Figure 2.6: *elev* versus *depth*

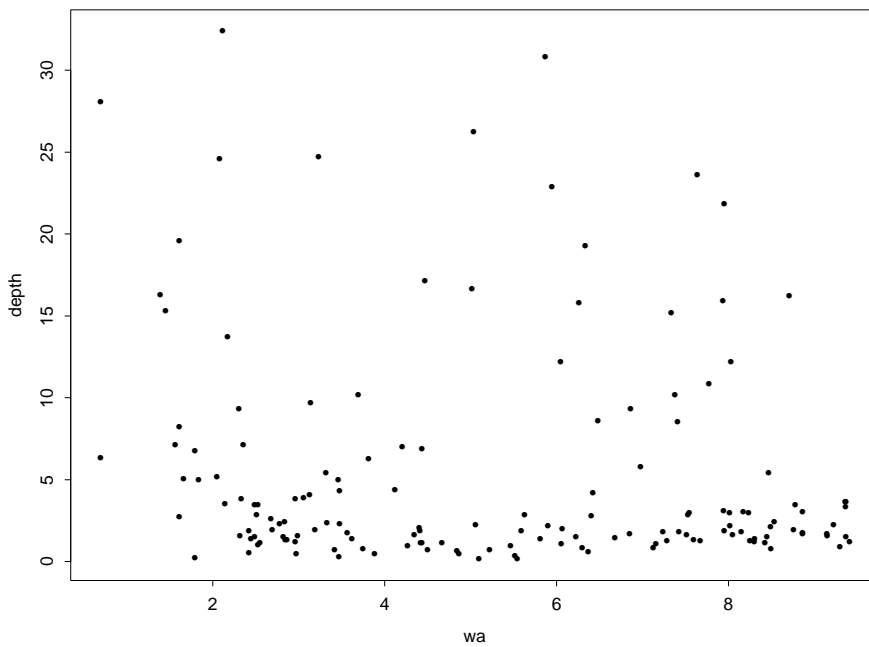


Figure 2.7: *wa* versus *depth*

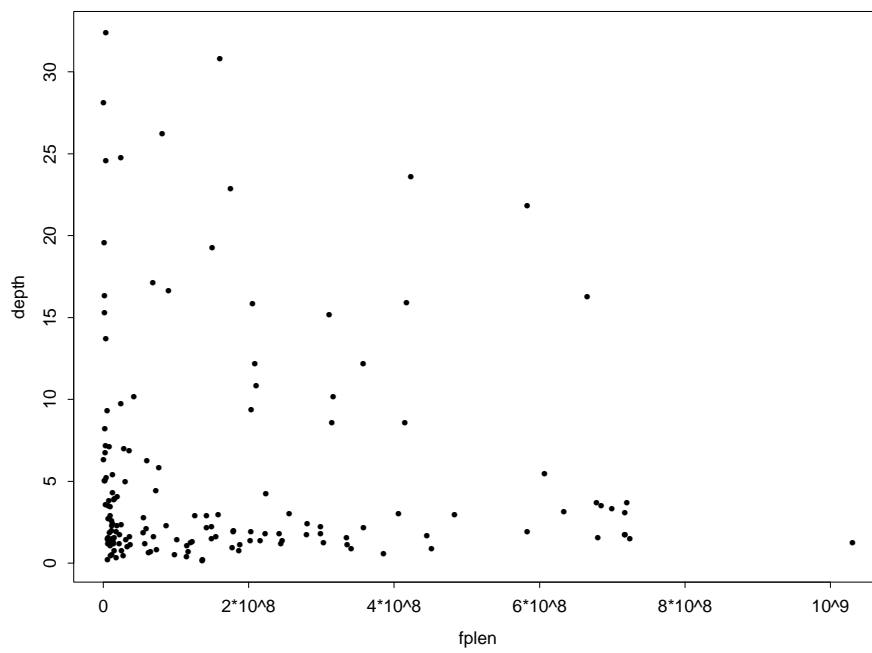


Figure 2.8: *fplen* versus *depth*

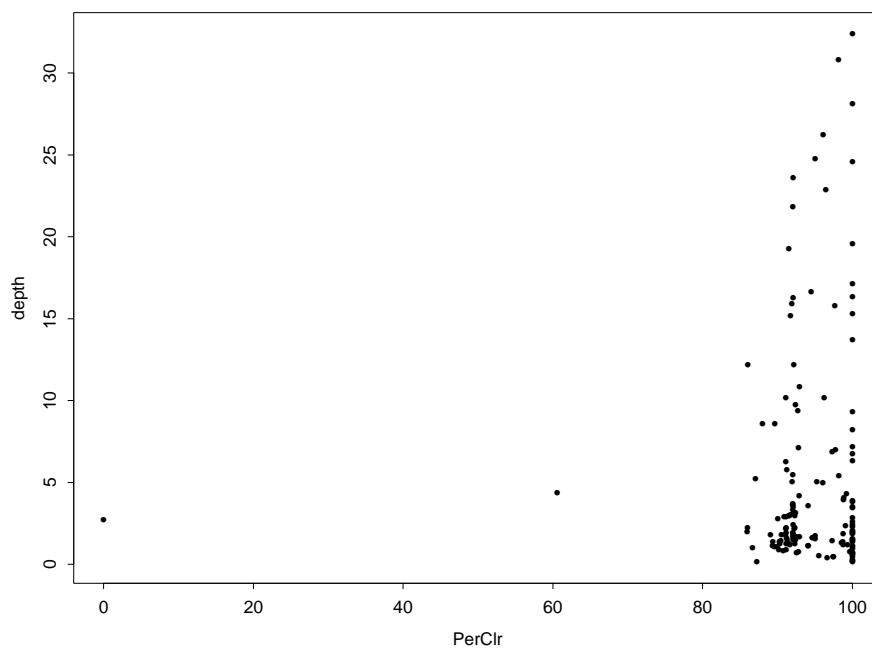


Figure 2.9: *PerClr* versus *depth*. The two outliers (97 and 130) were dropped from the analysis.

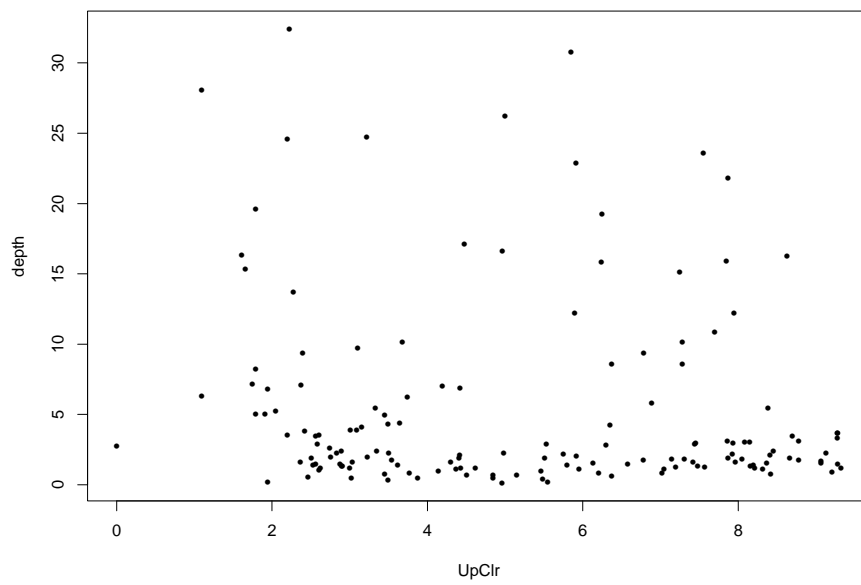


Figure 2.10: *UpClr* versus *depth*



Figure 2.11: *Pairwise plots of all variables. While some variables are highly correlated, none appear to have a high correlation with depth to ground-water.*

### 2.1.1 the standard analysis

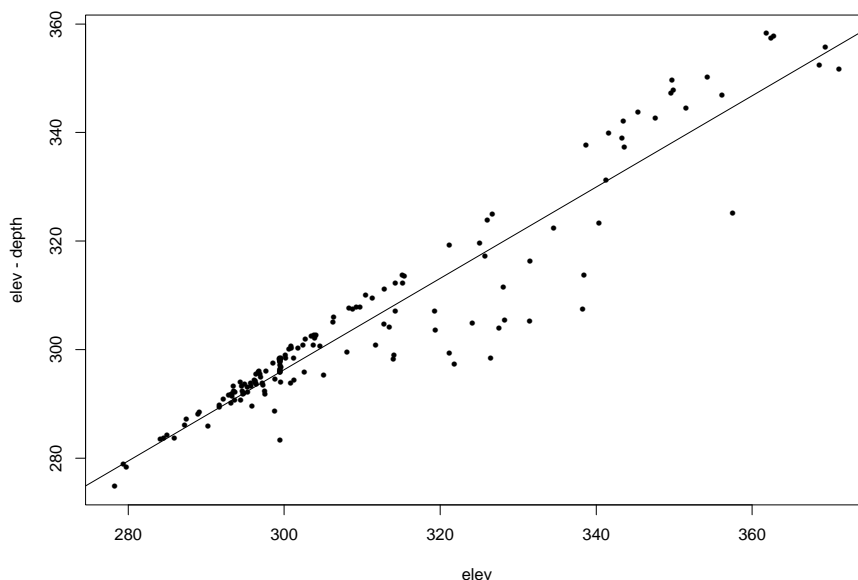


Figure 2.12: *elev - depth* versus *elev*

A standard procedure has been to regress **elev-depth** against **elev** (Salama et al., 1996). This is shown for the current data set in figure 2.12.

Now while regressing **elev-depth** on **elev** leads to the same estimates of **depth** as regressing **depth** on **elev**, it leads to a very false picture of how well the regression is fitting. If a straight line is fitted to this graph

$$\mathbf{elev - depth} = \beta_0 + \beta_1 \mathbf{elev}$$

the squared correlation  $\text{cor}(\widehat{\mathbf{depth}}, \mathbf{elev - depth})^2$  is 0.8994. Figure 2.13 shows a more realistic summary of the relationship; in this case,  $\text{cor}(\widehat{\mathbf{depth}}, \mathbf{depth})^2 = 0.2422865$ .

Figure 2.14 shows the 10-fold cross-validated estimates of observed versus predicted. That is, the data is divided into 10 randomly selected sets, the model is fitted on 9 of the sets combined and the predicted values are recorded for the remaining set. This is repeated for each of the 10 groups.

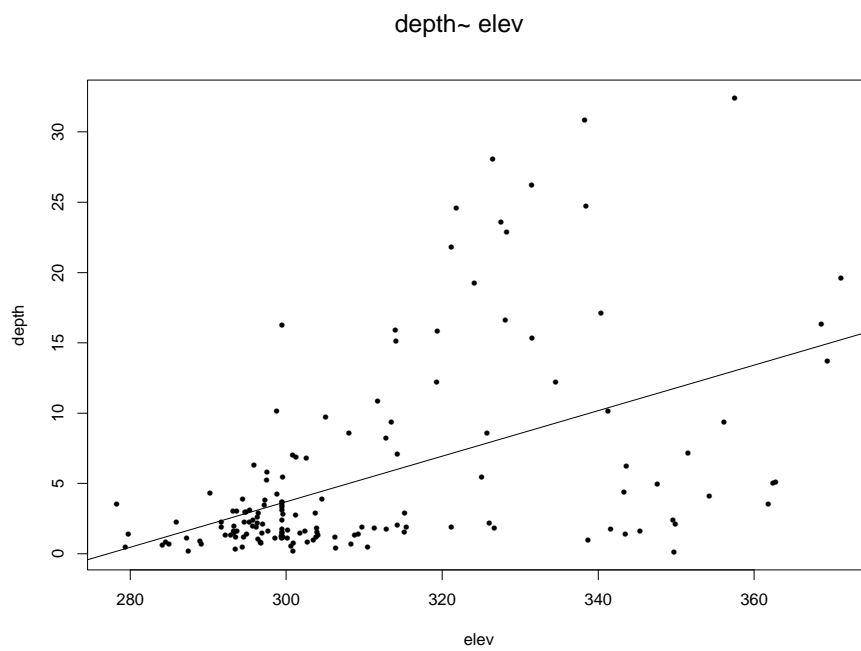


Figure 2.13: *depth* versus *elev*

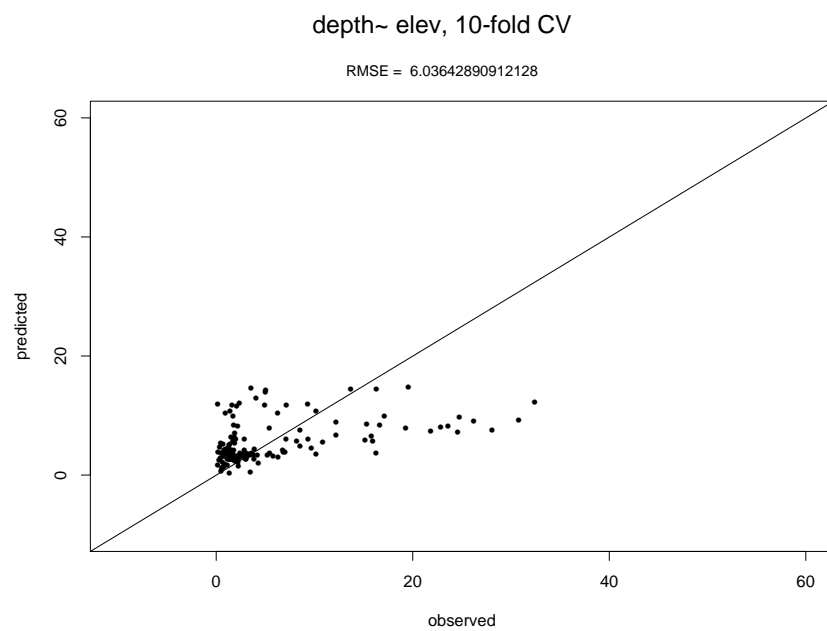


Figure 2.14: 10-fold cross-validation of *depth* versus *elev*.

### 2.1.2 generalized additive models

A generalized additive model (Hastie, 1992) was fitted to the 10 variables, and the resulting model, **gam1**, was subjected to pruning via forward and backward variable selection. The result of this process was **gam2**

```
gam2<-gam(formula = depth ~ FlSlope + s(h.above.streams) + s(elev) +
          wa + s(PerClr) +
          UpClr, data = bores.kent, control = gam.control(bf.maxit = 50),
          trace = F)
```

```
Degrees of Freedom: 127 total; 110.9972 Residual
Residual Deviance: 603.9297
```

**gam2** has a good fit to the data ( $R^2 = 0.678$ ). The observed versus fitted values, residuals versus fitted values and qq plots for **gam1** and **gam2** are given in figures 2.15 to 2.20. There is clearly some pattern left in the residual plots.

In addition, model (**gam2**) does not extrapolate well. Several regions of the high ground at the north-east edge of the catchment are predicted to have water table significantly above ground level. This is probably due to the non-linear terms in the model. To check this the smooth terms in **gam2** were replaced by linear terms in gam3.

```
gam3<-gam(formula = depth ~ FlSlope + h.above.streams + elev + wa +
          PerClr + UpClr, data = bores.kent,
          control = gam.control(bf.maxit = 50), trace = F)
```

**gam3** has an  $R^2 = 0.590$  and, as it has no smoothing terms in it, suggests that a linear model may do quite well.

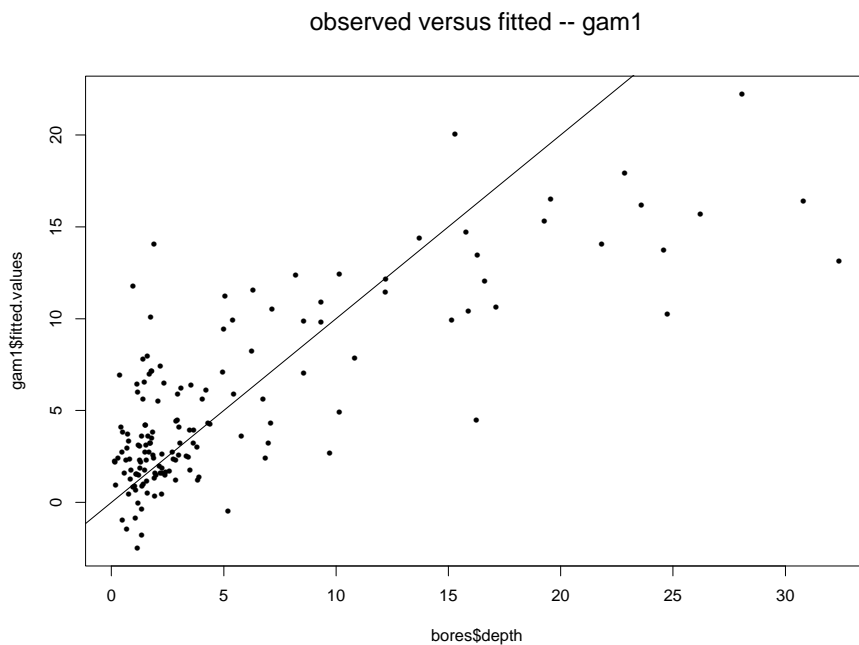


Figure 2.15: Observed versus fitted values for **gam1**. The line  $y = x$  is included for interpretation.

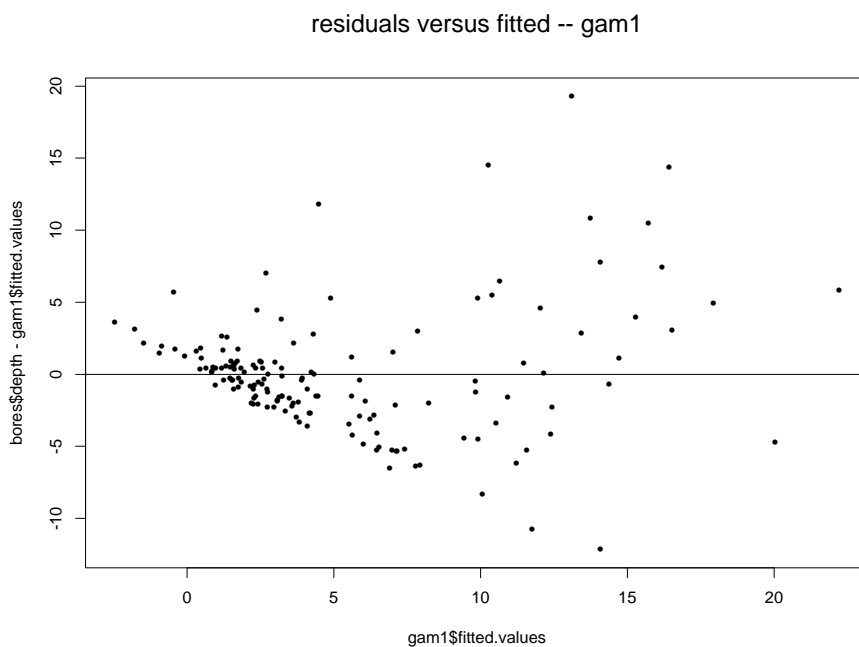


Figure 2.16: Residuals versus fitted values for **gam1**.

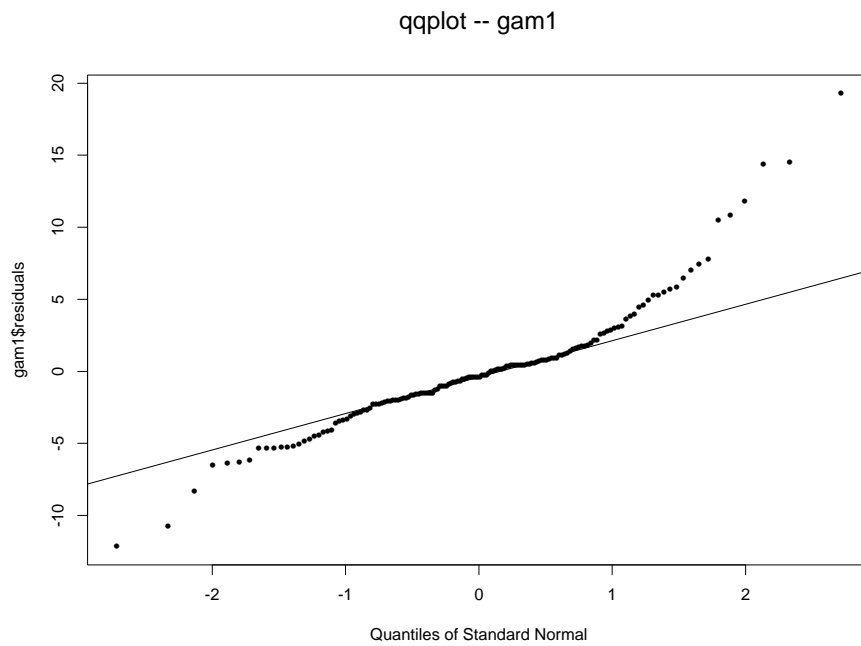


Figure 2.17: A *QQ plot* for **gam1**

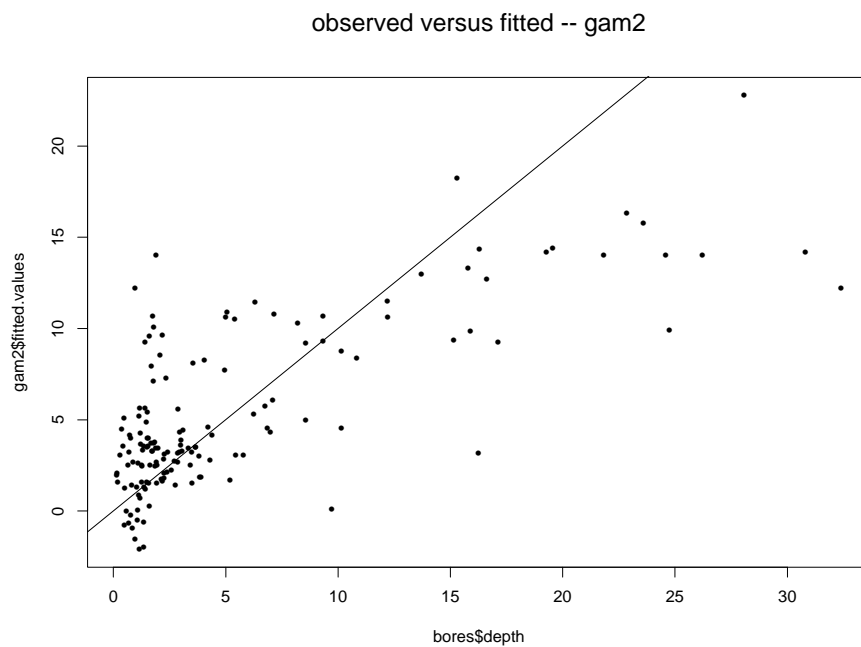


Figure 2.18: *Observed versus fitted values* for **gam2**.

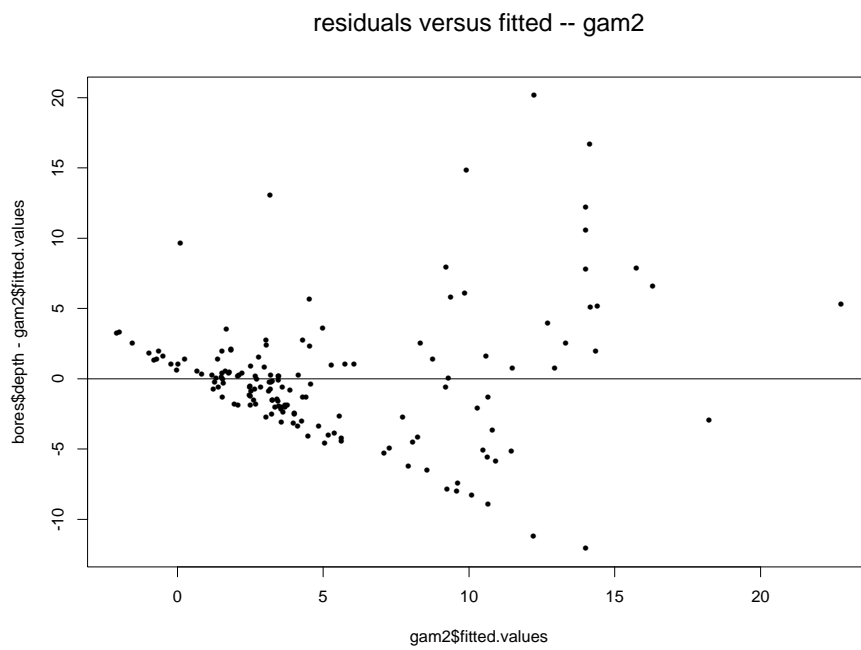


Figure 2.19: *Residuals versus fitted values for **gam2**.*

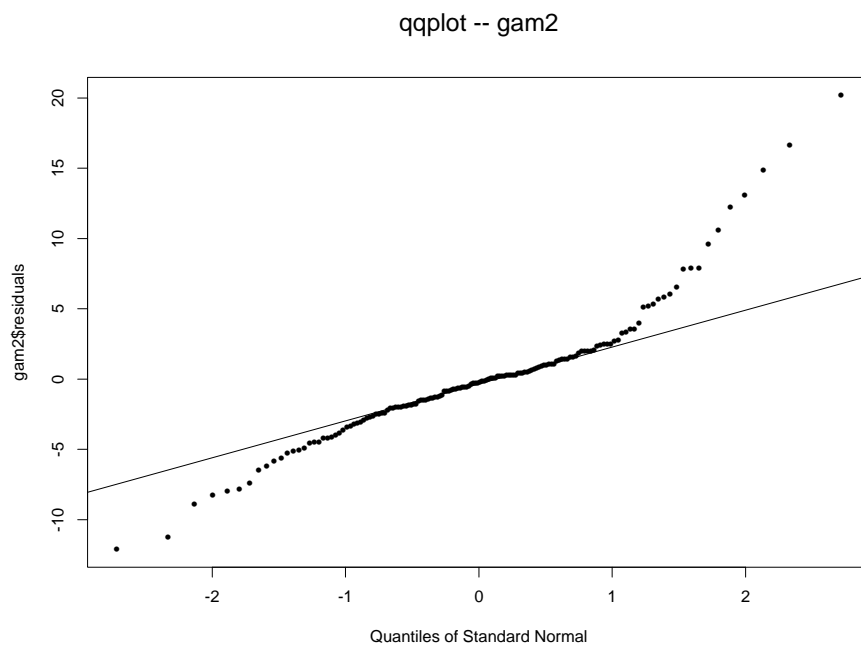


Figure 2.20: *A QQ plot for **gam2***

### 2.1.3 interactions

MARS (multivariate adaptive regression models) (Friedman, 1991a,b) allows flexible interactions to be modeled. The MARS algorithm fits the following model

$$\hat{f}(x_1, \dots, x_n) = \alpha_0 + \sum_{(i) \in V_m} \alpha_m B_m(x_i) + \sum_{(i,j) \in V_m} \alpha_m B_m(x_i, x_j) + \dots$$

where  $B_m$  is a multivariate spline basis function and  $V_m$  is the set of variables associated with the  $m^{\text{th}}$  basis function  $B_m$ .

Figure 2.21 shows a 10-fold cross-validation of observed versus predicted for the MARS model. The fit to the data is poor and the resulting hydraulic head is very poor.

The conclusion that can be drawn from this is that while there are likely to be interactions between the explanatory variables, the data are too noisy, or the regression technique not robust enough, to model them.

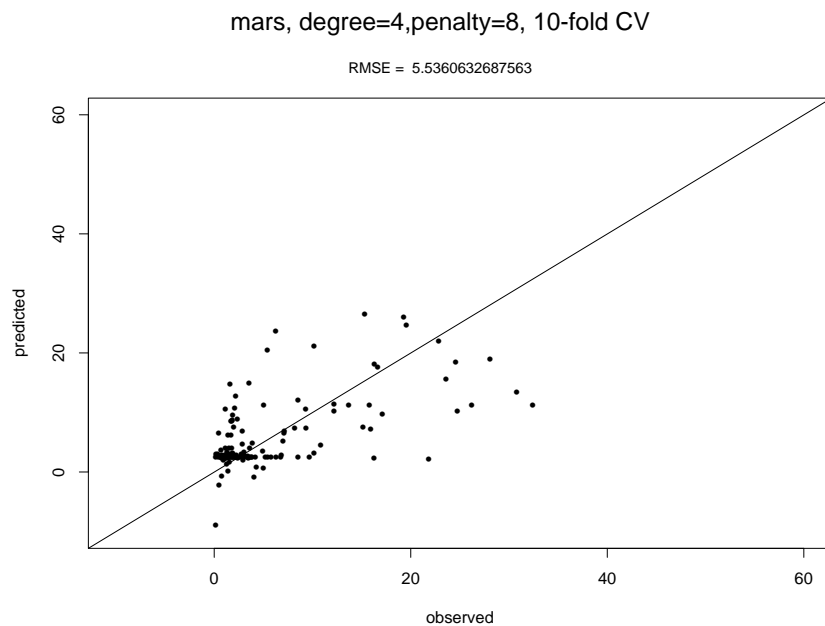


Figure 2.21: 10-fold cross-validation of observed versus predicted for the MARS model.

### 2.1.4 robust linear models

A robust linear model was fitted using the `rlm` option in *S-PLUS*. This uses Huber's M-estimator with a tuning parameter of  $c = 1.345$  (Venables and Ripley, 1994a).

```
lm1<-rlm(formula = depth ~ FlSlope + h.above.streams + elev +
          wa + PerClr + UpClr, data = bores.kent)
```

`lm1` has an  $R^2$  of 0.587. An hydraulic head surface was produced using this model (`kent_head.ers`); however, this had several large areas with a clearly anomalous prediction. In order to rectify this, the variables `elev` and `UpClr` were dropped. `UpClr` was dropped because it contains the same information as `wa`, and `PerClr` and `elev` on the suspicion that it only allows extrapolation within a sub-catchment, not between sub-catchments. This gave the model `lm2` with an  $R^2$  of 0.506. The hydraulic head produced with this model (`kent_head_2.ers`, see figure 2.22 for the hydraulic head.) is the most plausible of those produced so far.

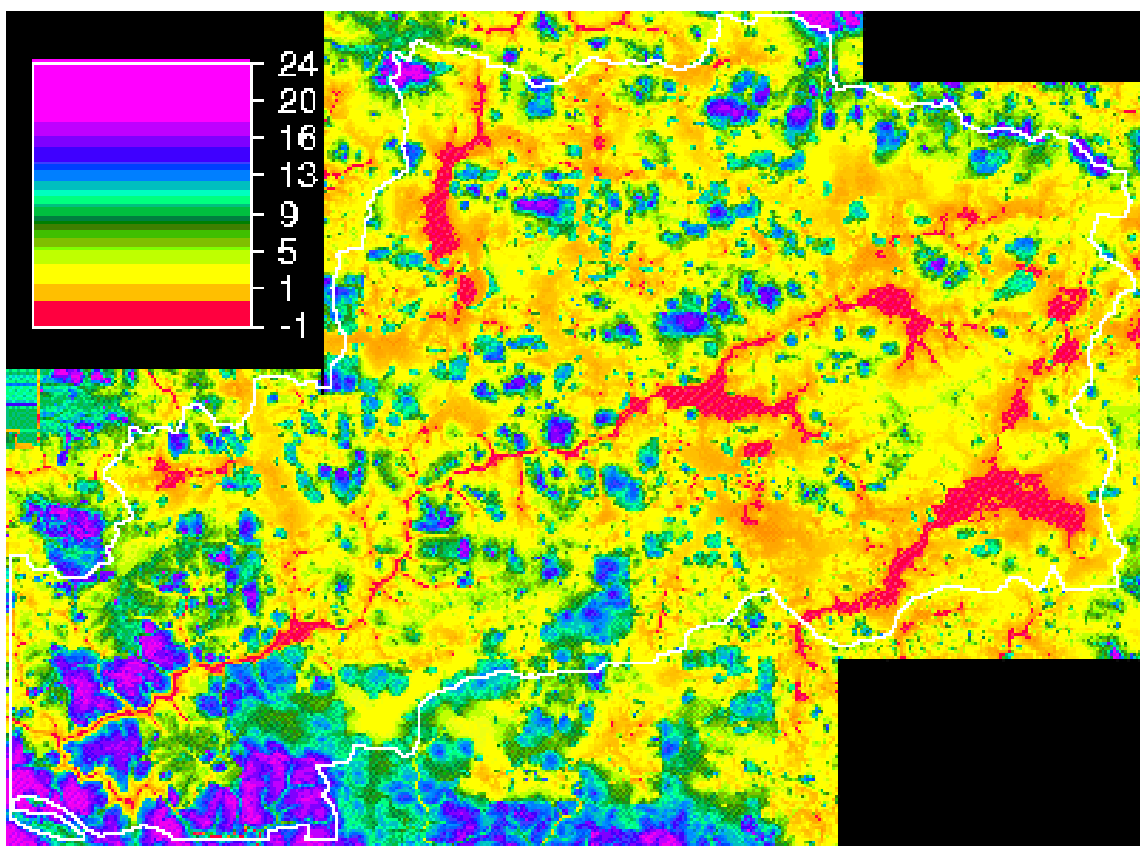


Figure 2.22: *The hydraulic head for `lm2`*

```
> summary(lm2)
```

```
Call: rlm(formula = depth ~ FlSlope + h.above.streams + wa + PerClr,
          data = bores.kent)
```

```
Residuals:
```

```
      Min       1Q   Median       3Q      Max
-9.511 -0.9302 -0.07253  1.178  17.4
```

```
Coefficients:
```

	Value	Std. Error	t value
(Intercept)	9.2062	1.1063	8.3216
FlSlope	0.1483	0.0543	2.7312
h.above.streams	0.1292	0.0207	6.2406
wa	-0.3749	0.1556	-2.4102
PerClr	-0.0911	0.0090	-10.0715

```
Residual standard error: 1.554 on 122 degrees of freedom
```

```
Multiple R-Squared: NA
```

```
F-statistic: on and degrees of freedom, the p-value is NA
```

The success of **lm2** suggests fitting a robust linear regression on all the variables and doing forward selection of the variables. This was done with the following results:

```
rlm1<-rlm(formula = depth ~ AvHgt + AvSlp + FlSlope + h.above.streams
+ h.above.salt+ wa + elev + fplen + PerClr , data = bores.kent)
cor(bores.kent$depth,rlm1$fitted.values)^2 0.5422214
```

```
rlm2<-rlm(formula = depth ~ AvSlp + FlSlope + h.above.streams +
h.above.salt+ wa + elev + fplen + PerClr , data = bores.kent)
cor(bores.kent$depth,rlm2$fitted.values)^2 0.5425003
```

```
rlm3<-rlm(formula = depth ~ AvSlp + FlSlope + h.above.streams +
          wa + elev + fplen + PerClr , data = bores.kent)
cor(bores.kent$depth,rlm3$fitted.values)^2
0.542309
```

```
rlm4<-rlm(formula = depth ~ AvSlp + FlSlope + h.above.streams +
          wa + elev + PerClr , data = bores.kent)
cor(bores.kent$depth,rlm4$fitted.values)^2
0.5375192
```

```
rlm5<-rlm(formula = depth ~ AvSlp + FlSlope + h.above.streams +
          wa + PerClr , data = bores.kent)
cor(bores.kent$depth,rlm5$fitted.values)^2
```

```
0.5251078
```

```
> summary(rlm5)
```

```
Call: rlm(formula = depth ~ AvSlp + FlSlope + h.above.streams + wa
          + PerClr, data = bores.kent)
```

```
Residuals:
```

```
      Min       1Q   Median       3Q      Max
-10.23  -1.036  -0.178   1.159   16.73
```

```
Coefficients:
```

	Value	Std. Error	t value
(Intercept)	10.6980	1.1784	9.0785
AvSlp	0.0000	0.0000	-2.3602
FlSlope	0.2329	0.0691	3.3692
h.above.streams	0.1135	0.0209	5.4260
wa	-0.4561	0.1571	-2.9031
PerClr	-0.0934	0.0091	-10.2800

```
Residual standard error: 1.622 on 121 degrees of freedom
```

```
Multiple R-Squared: NA
```

```
F-statistic: on and degrees of freedom, the p-value is NA
```

```
rlm6<-rlm(formula = depth ~ FlSlope + h.above.streams +
           wa + PerClr , data = bores.kent)
cor(bores.kent$depth,rlm6$fitted.values)^2
0.5058309
```

At each stage, the variable with the smallest absolute  $t$  value was removed until **rlm5** was achieved. This model differs from **lm2** in that it includes the variable **FlSlope**.

The robust method of Campbell et al. (1998) was used to generate model **r1**. Figure 2.23 shows a comparison of the weights for the **lm2** model and for the **r1** model. It can be seen that both models downweight the same observations, but that in general the **r1** model downweights them more severely. Figure 2.24 shows the observed values versus the predicted values for the model **r1** as well as the downweighting of the bores. Figure 2.25 shows the resulting hydraulic head surface.

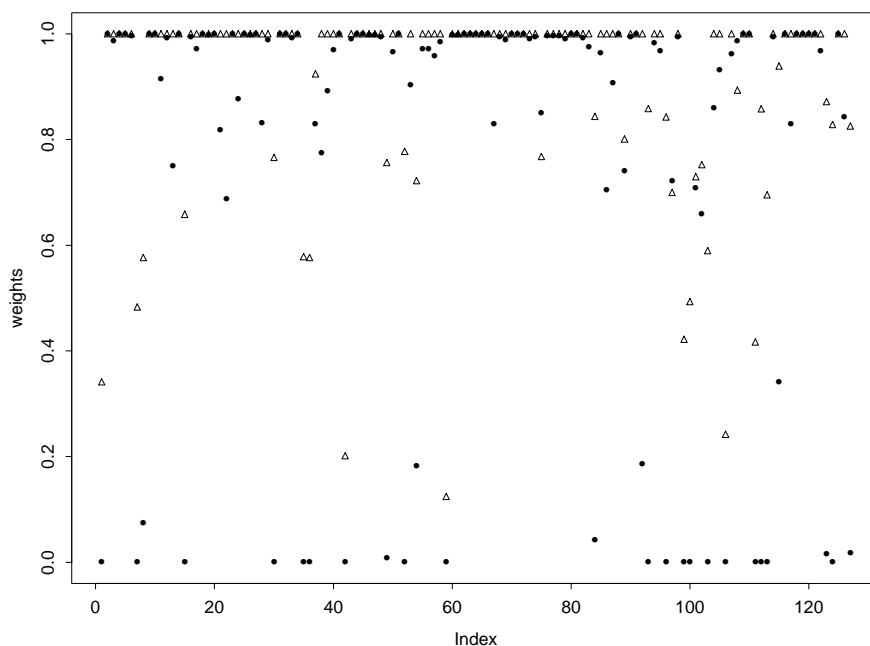


Figure 2.23: A comparison of the weights for the **lm2** model and for the **r1** model. The weights for the **lm2** are shown as triangles and the weights for the **r1** model are shown as stars.

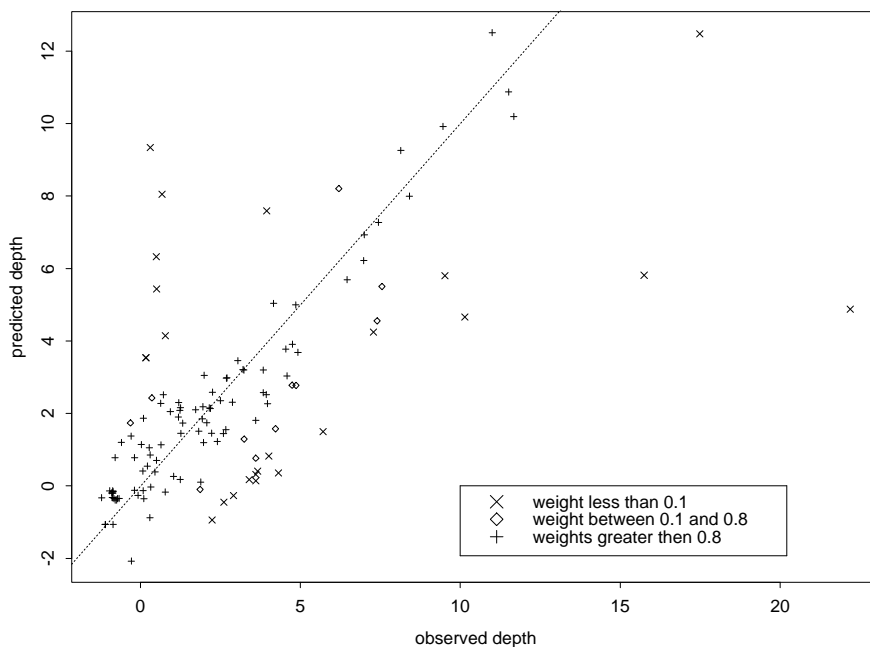


Figure 2.24: The observed versus the predicted values for the model **r1**.

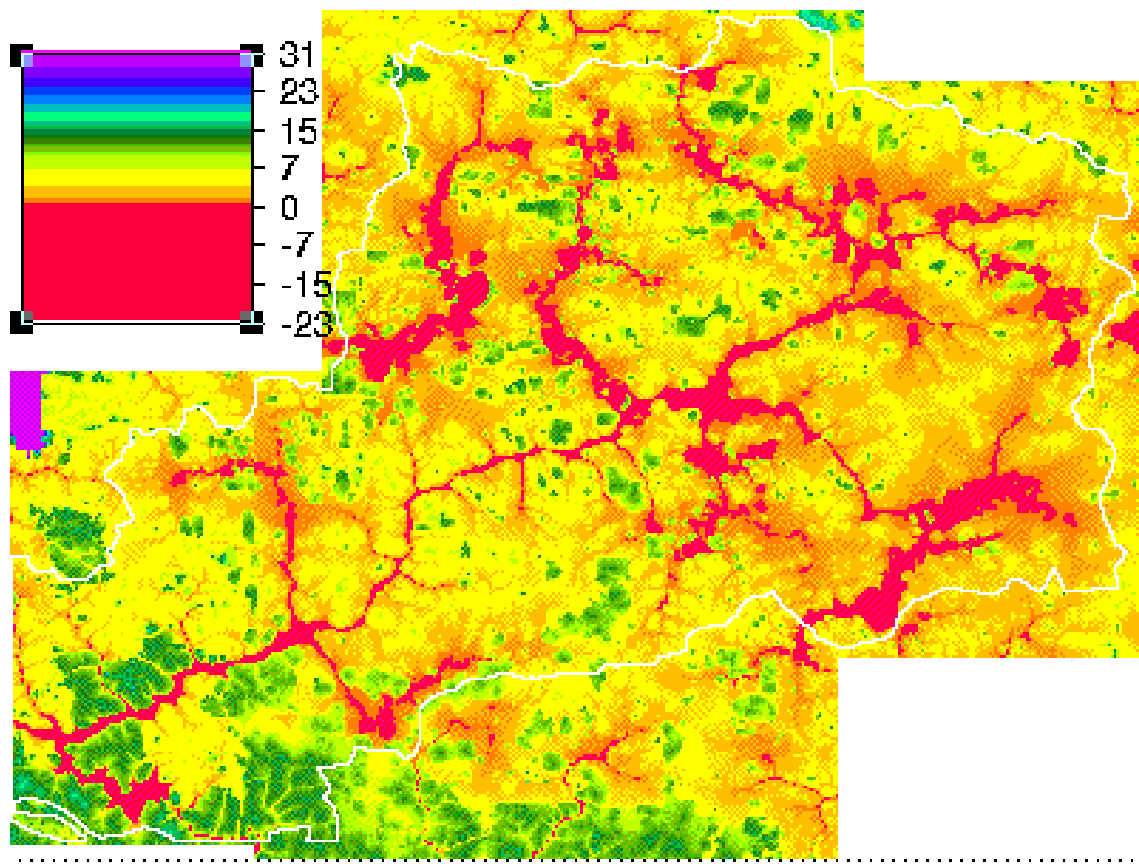


Figure 2.25: *The hydraulic head for r1*

### 2.1.5 conclusions

model	$R^2$	MSEP	
elev	0.25	5.95	
gam1	0.722	3.196337	
gam2	0.678	2.68698	
gam3	0.591	2.782963	image unsatisfactory
lm1	0.587	2.832347	image unsatisfactory
lm2	0.506	2.812388	image sent out for evaluation

Table 2.1: The mean squared error of prediction is a leave-one-out estimate.

Lessons learned from these analyses include:

- from prior knowledge of the processes that govern ground water depth, it would be expected that the relationship between the explanatory variables and the depth to ground water would be:
  - monotonic;
  - non-linear (and probably approaching an asymptote).

However, it was found that flexible models such as generalized additive models were not usable, due to the poor nature of the data (such models are fitted by least squares which is very susceptible to outliers);

- the sampled range of the variables is sometimes much less than their range in the image. However, in order to get a reasonable coverage of the image, some extrapolation of the fitted model is essential. This also argues for less flexible models.

Models **lm2** and **rl** are both plausible models and have been sent to hydrologists for evaluation.

## 2.2 The new.bores.kent data set

This analysis was undertaken in the belief that this set of bores was more accurate in terms of position and elevation since they are surveyed, and give a good coverage of the landscape. However, on this issue, see appendix A.

For the 113 bores, 10 variables were extracted from the DEM-derived variable files. Rather than the log transformations described in section 2.1, A cube root transform was applied to all variables to make the histograms more symmetric.

The variables **wa**, **UpClr** and **PerClr** are functionally related,  $\text{PerClr} = 100\text{UpClr}/\text{wa}$  and **wa** and **UpClr** are very highly correlated ( $\rho = 0.98$ , see table 2.2). It does not seem advisable to include all three variables in the regression<sup>1</sup>. Each of the variables was dropped in turn for the robust analysis.

The method described in Campbell et al. (1998) for the robust estimation of a covariance matrix was adapted to a regression problem (see appendix B) and applied here.

### 2.2.1 new.kent.model.1

This includes the variables **AvHgt**, **AvSlp**, **FlSlope**, **h.above.streams**, **wa**, **fplen**, and **UpClr**. The final model was **new.kent.model.1**

```
> summary(new.kent.model.1)
Call: lm(formula = depth ~ AvHgt + AvSlp + h.above.streams + wa + UpClr,
          data = new.bores.kent, weights = c(w[1:56], 0, w[57:112]))

Residuals:
    Min       1Q   Median       3Q      Max
-0.3134 -0.08411 -0.005727  0.07613  0.3168

Coefficients:
                Value Std. Error  t value Pr(>|t|)
(Intercept)    2.0956    0.1299   16.1338  0.0000
      AvHgt     0.0078    0.0027    2.9407  0.0040
      AvSlp    -0.0015    0.0004   -3.6942  0.0004
h.above.streams 0.0038    0.0009    4.3811  0.0000
           wa    0.1751    0.0299    5.8513  0.0000
           UpClr -0.1751    0.0323   -5.4217  0.0000
```

Residual standard error: 0.1272 on 106 degrees of freedom

Multiple R-Squared: 0.5648

F-statistic: 27.51 on 5 and 106 degrees of freedom, the p-value is 0

For the “good” bores, the  $R^2$  value is 0.71289. The observed versus predicted values for the model **new.kent.model.1** are shown in figure 2.26, the fitted hydraulic head surface

---

<sup>1</sup>Note: ridge regression can be used in cases where variables are collinear (or nearly so) – this could be tried here.

	AvHgt	AvSlp	FISlope	h.above.salt	h.above.streams	elev	wa	fplen	PerClr	UpClr
AvHgt	1.00	-0.13	0.69	0.36	0.15	-0.27	-0.43	-0.45	-0.20	-0.46
AvSlp		1.00	-0.43	-0.11	-0.39	-0.28	0.74	0.74	0.40	0.76
FISlope			1.00	0.30	0.38	-0.06	-0.54	-0.66	-0.29	-0.58
h.above.salt				1.00	0.40	-0.11	-0.13	-0.28	-0.01	-0.16
h.above.streams					1.00	0.35	-0.43	-0.56	-0.01	-0.41
elev						1.00	-0.21	-0.14	0.16	-0.14
wa							1.00	0.79	0.24	0.98
fplen								1.00	0.21	0.79
PerClr									1.00	0.41
UpClr										1.00

Table 2.2: Correlations between the derived variables.

in figure 2.27 (the hydraulic head file is **kent\_head\_3.ers**), and the downweighted bores (plus brief comments from Geoff Hodgson) are shown in table 2.3.

The stream line in the lower left of figure 2.27 has a level of 30+ metres. This is due to the positive coefficient of **wa** – which itself is (probably) due to the high correlation of **wa** and **UpClr**.

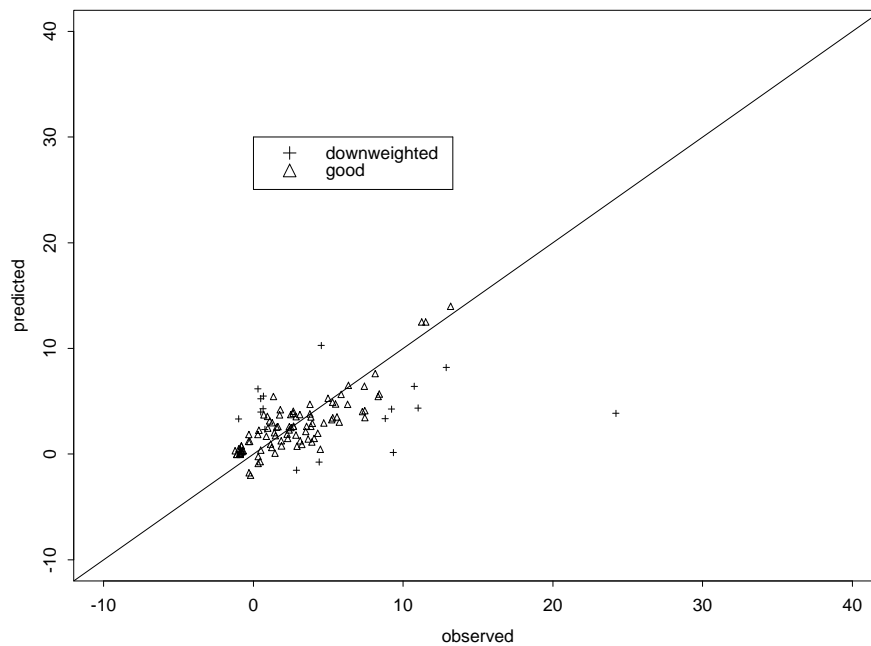


Figure 2.26: *Observed versus predicted values for the model **new.kent.model.1**.*

bore	comment
JC8I98	Don't know. Bore is drilled to 10m but consistently shallow water levels.
kw1	We already know whats going on with this one - steep slopes
mw1	Both these are high in the upper slopes and could be due to positional accuracy
mw2	"
t3	not sure why this one doesn't fit, as T2 is very close, same elevation and depth
AH20I	This is an intermediate depth bore drilled to about 8m. It has always had a water level less than 1m deep which suggests this could be a perched system. It quite high in the landscape on steepish slopes but has accurate location data. It does appear to lie in a drainage line but high up, so height above streams may not get this if it is above the threshold set for streams. I would probably drop this one.
dsh3d92	Definitely ignore - according to the database, this bore has casing 0.0m above ground thus how can a positive head of 0.98m be measured? Having said that, this area was mapped as discharge by the 1995 study so again I don't know why this one doesn't fit.
dsh4i92	This one may fit better with the omission of above
ke16d85	try -1.4
n3	this is the deepest bore. Maybe try -15 or so. I suspect that most of the catchment will have bedrock at 15m therefore this bore is a true anomaly. The water table that is here is very shallow and probably meaningless.
P1	try -2.25
sy5	try -3.75 also -3.4 for sy6
bu10i84	ignore. No depth drilled in database. Lies in same drainage line as below. Other bores nearby suggest water at about 7m
bu5d84	This one lies on a drainage line and could be real. Again height above streams?
kw3	These kw3,4, and 5 can be omitted as they don't actually fall in the Kent Catchment.
kw5	
RG1d93	location? maybe try with 3m shallower ie -8m

Table 2.3: The downweighted bores for **new.kent.model.1** (plus brief comments from Geoff Hodgson).

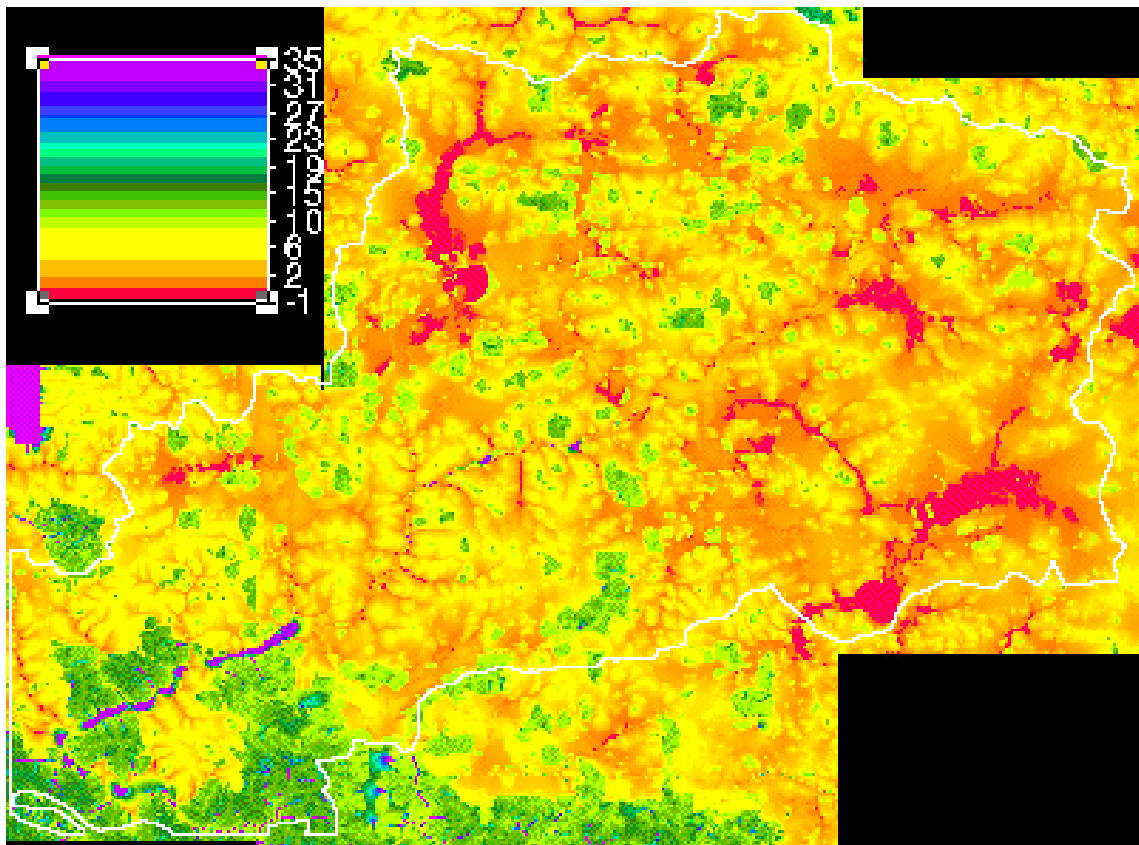


Figure 2.27: The hydraulic head for **kent\_head\_3.ers**. The stream line in the lower left of the image has a level of 30+ metres.

## 2.2.2 new.kent.model.2

This includes the variables **AvHgt**, **AvSlp**, **FISlope**, **h.above.streams**, **wa**, **fplen**, and **PerClr**.

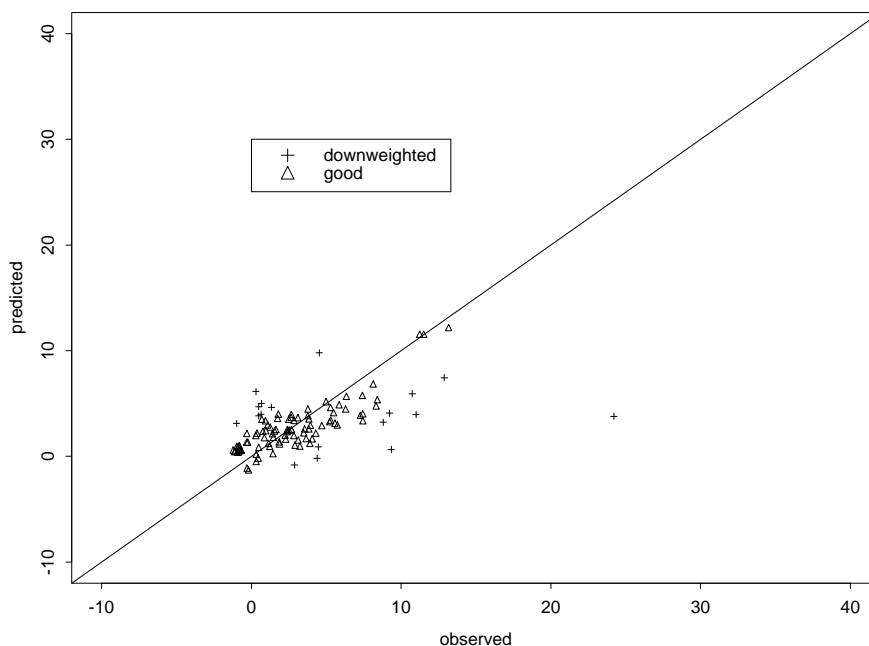


Figure 2.28: *observed versus predicted values for the model **new.kent.model.2**.*

For the “good” bores, the  $R^2$  value is 0.741911. The observed versus predicted values for the model **new.kent.model.2** are shown in figure 2.28, the fitted head in figure 2.29 (the hydraulic head file is **kent\_head\_4.ers**).

```
> summary(new.kent.model.2)
```

```
Call: lm(formula = depth ~ AvHgt + AvSlp + h.above.streams + PerClr, data =
  new.bores.kent, weights = weights)
```

```
Residuals:
```

```
   Min       1Q   Median       3Q      Max
-0.183 -0.07335 -0.007864  0.07781  0.2054
```

```
Coefficients:
```

	Value	Std. Error	t value	Pr(> t )
(Intercept)	2.2615	0.1184	19.1070	0.0000
AvHgt	0.0088	0.0019	4.5733	0.0000
AvSlp	-0.0010	0.0002	-4.0835	0.0001

```
h.above.streams    0.0051    0.0008    6.8011    0.0000
                   PerClr  -0.0165    0.0029   -5.7710    0.0000
```

Residual standard error: 0.09572 on 90 degrees of freedom

Multiple R-Squared: 0.7238

F-statistic: 58.95 on 4 and 90 degrees of freedom, the p-value is 0

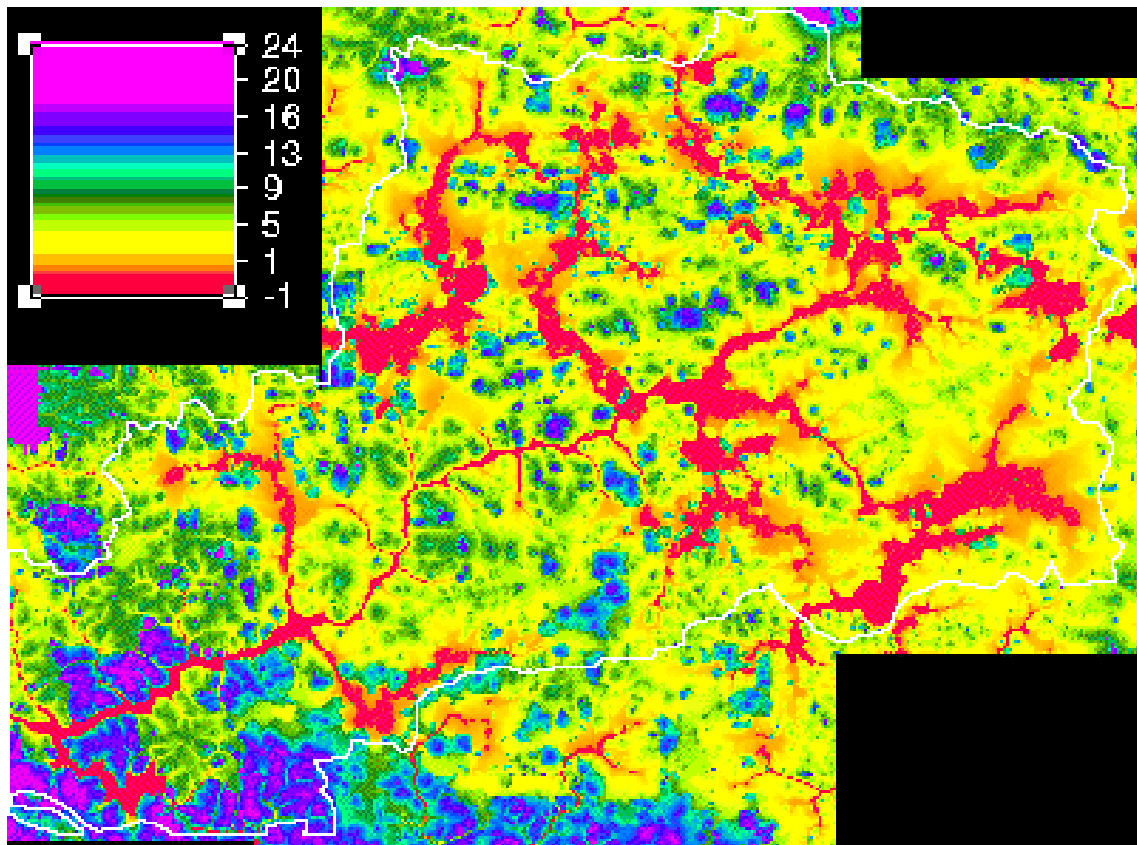


Figure 2.29: The hydraulic head surface for *kent\_head\_4.ers*, which was generated from *new.kent.model.2*.

### 2.2.3 new.kent.model.3

This includes the variables **AvHgt**, **AvSlp**, **FISlope**, **h.above.streams**, **fplen**, **PerClr** and **UpClr**.

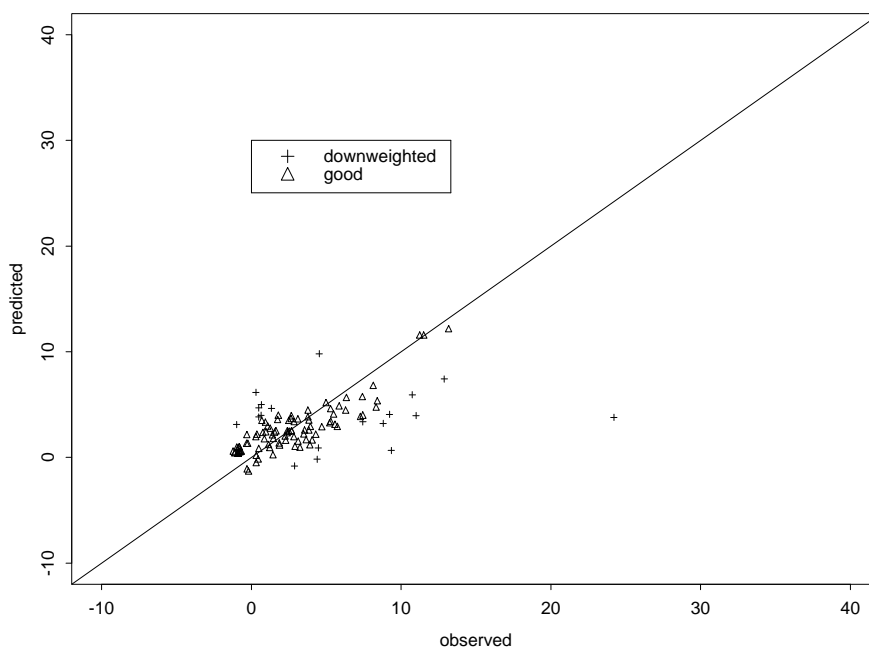


Figure 2.30: *The observed versus the predicted values for the model **new.kent.model.3**.*

For the “good” bores, the  $R^2$  value is 0.7523367. The observed versus the predicted values for the model **new.kent.model.3** are shown in figure 2.30, the fitted head in figure 2.31 (the hydraulic head file is **kent\_head\_5.ers**). The downweighted bores were

```
[1] "JC8I98" "kw1" "mw1" "mw2" "t3" "AH20I" "dsh3d92"
[8] "dsh4i92" "ke16d85" "n3" "P1" "sy5" "sy6" "bu10i84"
[15] "bu13i84" "kw3" "kw5" "s10" "RG1d93"
```

```
> summary(new.kent.model.3)
```

```
Call: lm(formula = depth ~ AvHgt + AvSlp + h.above.streams + PerClr, data =
new.bores.kent, weights = c(w[1:56], 0, w[57:112]))
```

```
Residuals:
```

```
      Min       1Q   Median       3Q      Max
-0.2935 -0.09565 -0.003194  0.08358  0.3695
```

```
Coefficients:
```

	Value	Std. Error	t value	Pr(> t )
(Intercept)	2.3596	0.1377	17.1341	0.0000
AvHgt	0.0068	0.0024	2.8642	0.0050
AvSlp	-0.0010	0.0003	-3.0074	0.0033
h.above.streams	0.0041	0.0009	4.4723	0.0000
PerClr	-0.0150	0.0035	-4.3344	0.0000

Residual standard error: 0.1337 on 107 degrees of freedom

Multiple R-Squared: 0.5146

F-statistic: 28.36 on 4 and 107 degrees of freedom, the p-value is 4.441e-16

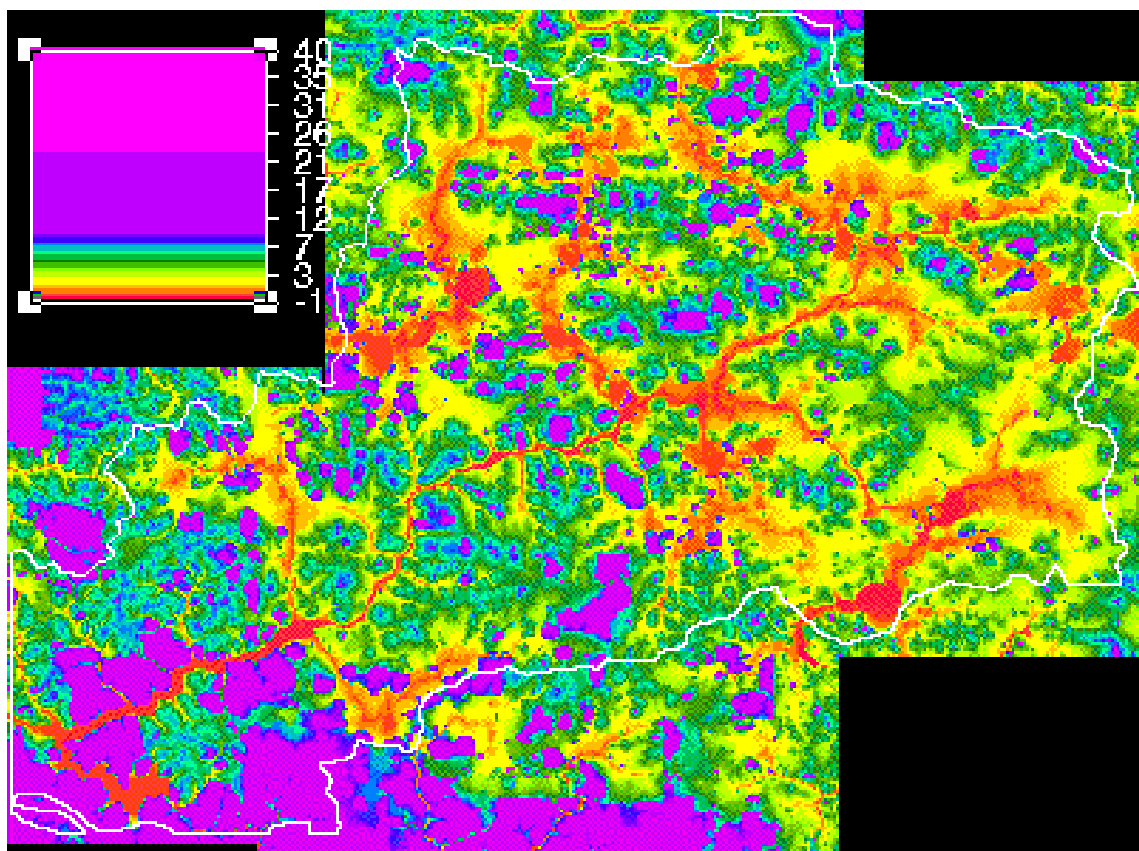


Figure 2.31: The hydraulic head surface for *kent\_head\_5.ers*, which was generated from *new.kent.model.3*.

## 2.3 Conclusions

A number of different models have been fitted and it appears that a robust linear regression gives the best results. However, it is plausible that the true relationship is non-linear in that the influence of any of the explanatory variables on depth to ground water is likely to asymptotically approach a limiting value.

The evaluation of the hydraulic head surfaces for models **lm2** and **r1** as well as **hydraulic\_head\_3**, **hydraulic\_head\_4** and **hydraulic\_head\_5** should be done independently of the bore-hole data.

# 3

## Boscabel

This section documents the analyses carried out on data from the Boscabel study region.

### 3.1 Combined Boscabel and Woodanilling

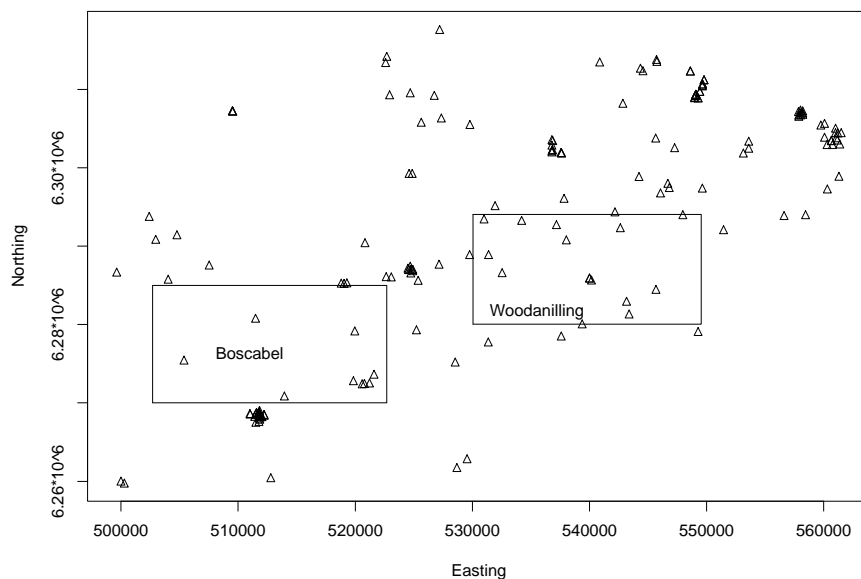


Figure 3.1: *The Boscabel and Woodanilling areas with the locations of the bores*

The first analysis was done on a set of bores covering the combined Boscabel and Woodanilling areas, as shown in figure 3.1. The data set consisted of 245 bores with easting and northing coordinates and depth measurements taken during the period from 1990 to 1999. A generalized additive model (Hastie, 1992) fitted on on all 10 variables gave an  $R^2$  value of 0.253.

Due to the very poor fit, this approach was then dropped in favor of restricting the bores to those in the immediate regions of interest.

## 3.2 Boscabel model

74 bores in the immediate vicinity of Boscabel were selected, as shown in figure 3.2. The bores were a subset of the augmented bores, plus three extra bores suggested by Richard George. The cube root was taken of the 10 variables in order to make the histograms more symmetric, and several models were fitted. The model  $\text{depth} \sim \text{elev}$  gave an  $R^2$  value of 0.0004 while

$$\text{depth} \sim \text{AvHgt} + \text{AvSlp} + \text{FlSlope} + \text{h.above.streams} + \text{h.above.salt} + \text{wa} + \text{fplen} + \text{PerClr} + \text{UpClr}$$

gave an  $R^2$  value of 0.26<sup>1</sup>. These are very low values and it would appear that one of the following is occurring:

- the location data provided with the water levels may be less accurate than needed to extract useful explanatory variables;
- there is a poor relationship between the explanatory variables and depth to ground water;
- the relationship between the explanatory variables and depth to ground water may be non-linear;
- there may be a large number of outliers.

We proceeded to examine whether there are a large number of outliers in the data. In order to consider this it is proposed to

- fit a number of models using the regression version of Campbell et al. (1998);
- evaluate these models by examining the plausibility of the fitted hydraulic heads.

The robust regression procedure randomly selects “elemental subsets” to start the process. The variables

AvHgt	AvSlp	FlSlope
h.above.streams	wa	fplen
PerClr		

were used for the analysis. 20 iterations produce the following set of standard deviations

---

<sup>1</sup>fitted robustly using the **rlm** model in *S-PLUS* (Venables and Ripley, 1994b, §8), the  $R^2$  value fell to 0.211.

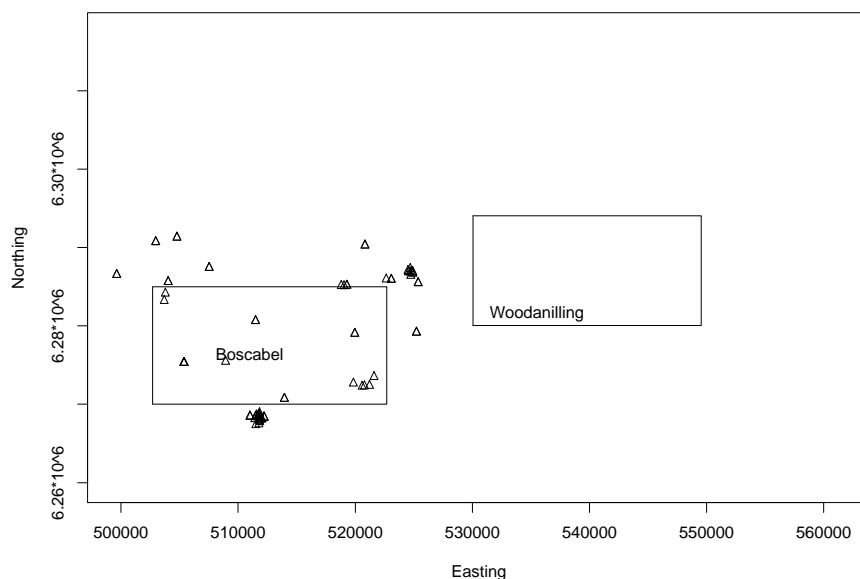


Figure 3.2: The 74 bores used for the Boscabel models. Note that three extra bores have been added.

```
[1] 0.05640192 0.05640192 0.04994067 0.05367536 0.05555792 0.05681216
[7] 0.07008392 0.04959007 0.04959007 0.05164116 0.05164108 0.04994066
[13] 0.04959007 0.04994066 0.05182924 0.05164116 0.04994067 0.05182925
[19] 0.05164116 0.04959010
```

In previous experience with this robust technique (see appendix B, page 87), it has been our experience that generally there are two solutions found, one fitting the outlying points and one fitting the good points. In this instance, there were a large number of different solutions. This perhaps indicates that the outlying points are not grouped together. Some of these standard deviations will undoubtedly correspond to very similar solutions. Three, numbers 1, 7 and 13, were selected; the largest, the smallest, and one in the middle, were selected for further investigation. The usual method would be to choose number 7, the smallest.

### 3.2.1 solution 1

Solution 1 gave a standard deviation of 0.056 and, after the variables `F1Slope` and `wa` were dropped, was summarized as:

```
> summary(model.1)
```

```
Call: lm(formula = depth ~ AvHgt + AvSlp + h.above.streams + fplen +
```

```
PerClr, data = bores.boscabel, weights = weights)
```

```
Residuals:
```

```
      Min      1Q      Median      3Q      Max
-0.05072 -0.01822 -0.005047  0.02027  0.04058
```

```
Coefficients:
```

	Value	Std. Error	t value	Pr(> t )
(Intercept)	0.5780	0.0949	6.0928	0.0000
AvHgt	0.0073	0.0013	5.4517	0.0000
AvSlp	-0.0005	0.0002	-2.7308	0.0090
h.above.streams	0.0014	0.0004	3.6371	0.0007
fplen	0.0001	0.0000	4.9943	0.0000
PerClr	0.0550	0.0043	12.6540	0.0000

```
Residual standard error: 0.02761 on 45 degrees of freedom
```

```
Multiple R-Squared: 0.8833
```

```
F-statistic: 68.12 on 5 and 45 degrees of freedom, the p-value is 0
23 rows with zero weights not counted in: summary(model.1)
```

Table 3.1 and figure 3.3 show the weights for solution 1, while figure 3.4 shows the observed values versus the predicted values. There are two questions that should be raised about this plot:

- how much is the perceived relationship driven by the two outlying points, bores "83DP07" and "b37"?
- is there any spatial pattern in the bores that have been down-weighted.

To answer the first question, the 74 values on the diagonal of the projection or "hat" matrix were plotted in figure 3.5. These show the influence that individual observations have on the fitting process. It is clear that bore "b37" (observation 74) is very influential in determining the final fit (as can be seen from the plot 3.4). Plots of Cook's distance for each term in the regression confirm the fact that "b37" is highly influential<sup>2</sup>.

This is a cause for concern. It appears that the regression relationship is being driven by one or two points, and thus is highly dependent on the accuracy of these points.

In order to test for any spatial pattern in the weights, they are plotted in figure 3.6. No pronounced pattern is evident. Figure 3.7 shows the fitted hydraulic head surface for solution 1.

---

<sup>2</sup>These diagnostic tools confirm what can be seen from figure 3.4 itself – the isolated and extreme points are determining the regression relationship.

bore names	weights
78DP03	0
85CN01	0.04
86DP04	0
93DP05	0
93DP06	0
93KM05	0
93MB03	0
93MB04	0
94BB06	0.76
94BK01	0
94WW11	0
94WW12	0
94WW13	0.01
97BK05	0
97WW32	0
97WW33	0.73
98CT03	0
98CT04	0
98CT07	0
WW03D	0.39
WW07D	0
WW09D	0
WW12D	0
WW13D	0
WW17D	0.8
WW61D	0.02
WW71D	0
WW74D	0
WW83D	0
b66	0

Table 3.1: The down-weighted bores from solution 1.

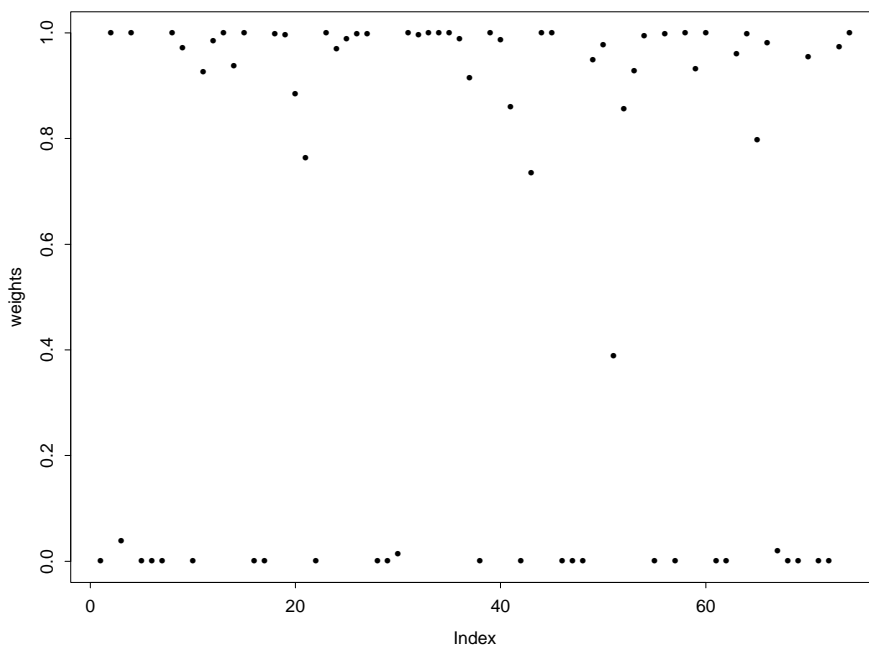


Figure 3.3: *The weights for the bores from solution 1.*

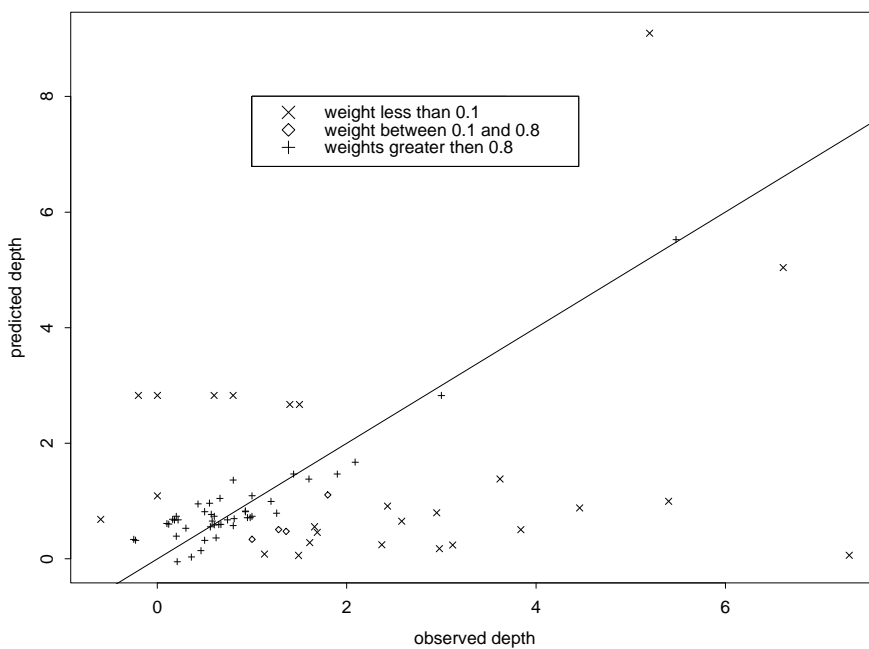


Figure 3.4: *Observed values versus predicted values for solution 1. The plot indicates which bores have been down-weighted.*

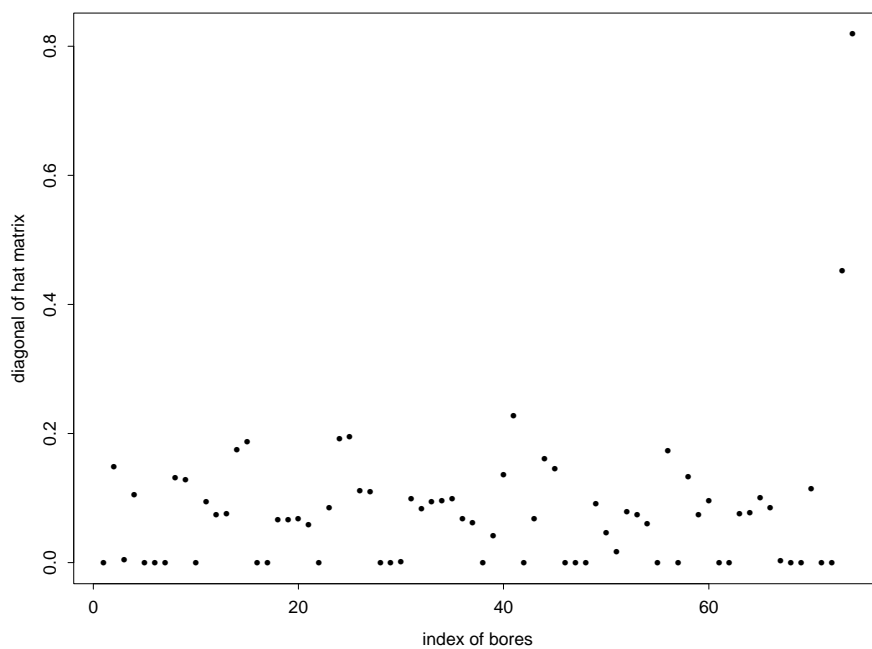


Figure 3.5: The diagonals of the projection matrix (often termed the “hat” matrix) for solution 1.

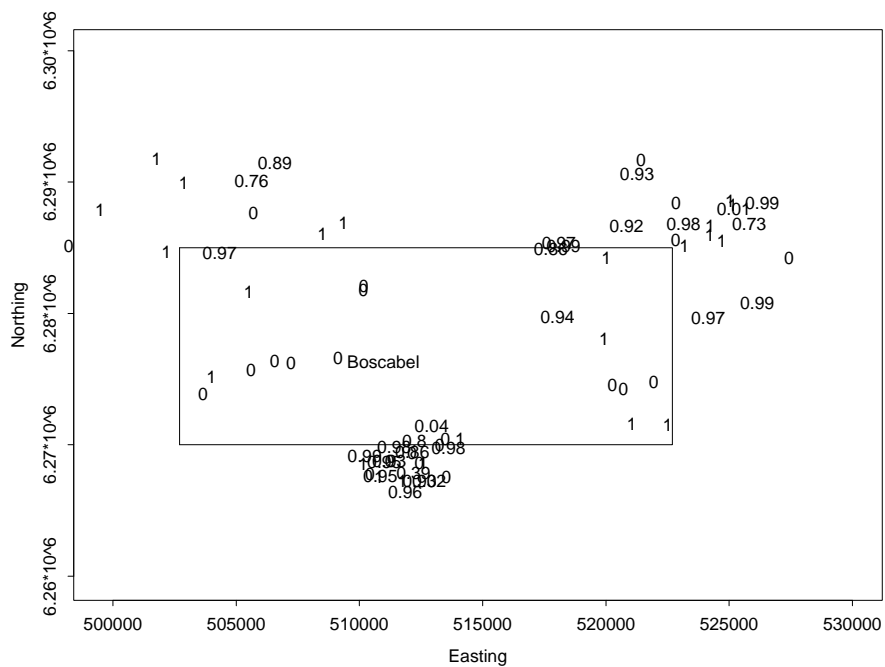


Figure 3.6: The spatial distribution of the weights ascribed to the bores in solution 1.

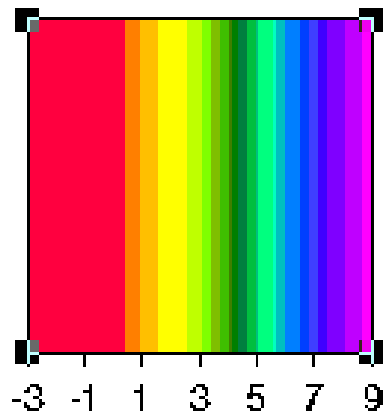
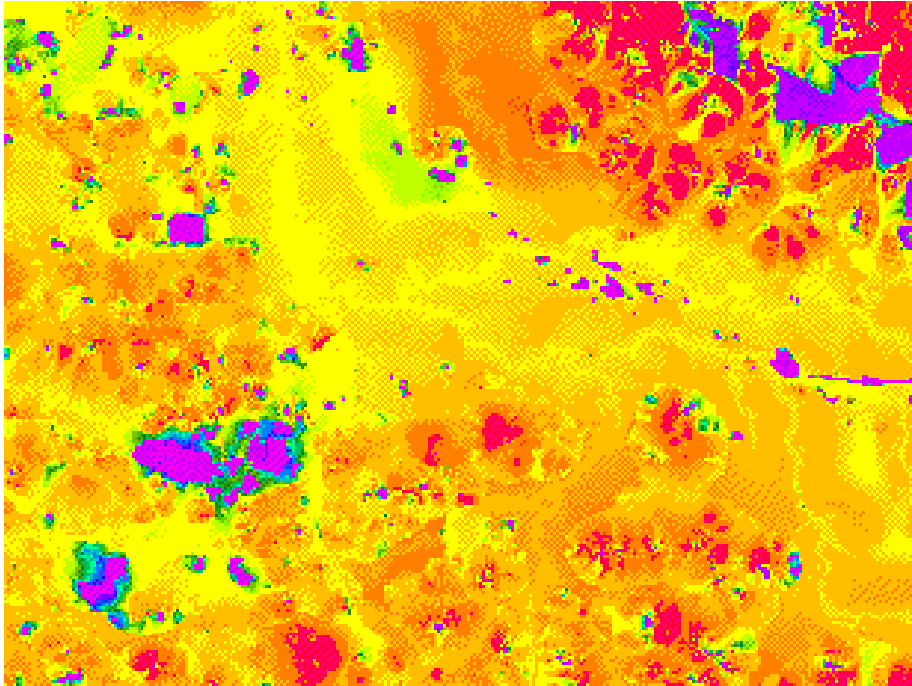


Figure 3.7: *The fitted hydraulic head for solution 1.*

### 3.2.2 solution 7

Solution 7 gave a standard deviation of 0.07 and, after the variables `F1Slope` and `h.above.streams` were dropped, is summarized as:

```
> summary(model.7)
```

```
Call: lm(formula = depth ~ AvSlp + wa + fplen + PerClr, data = bores.boscabel,
         weights = weights)
```

```
Residuals:
```

```
      Min       1Q   Median       3Q      Max
-0.06045 -0.02279  0.008052  0.02229  0.05762
```

```
Coefficients:
```

	Value	Std. Error	t value	Pr(> t )
(Intercept)	2.7099	0.1351	20.0598	0.0000
AvSlp	-0.0005	0.0001	-3.4814	0.0012
wa	0.0134	0.0018	7.4255	0.0000
fplen	-0.0002	0.0000	-6.3964	0.0000
PerClr	-0.0233	0.0064	-3.6566	0.0007

```
Residual standard error: 0.03311 on 42 degrees of freedom
```

```
Multiple R-Squared: 0.7562
```

```
F-statistic: 32.56 on 4 and 42 degrees of freedom, the p-value is 2.273e-12
```

```
27 rows with zero weights not counted in: summary(model.7)
```

Figure 3.8 shows which bores were downweighted for solution 7 and figure 3.9 shows the fitted hydraulic head surface.

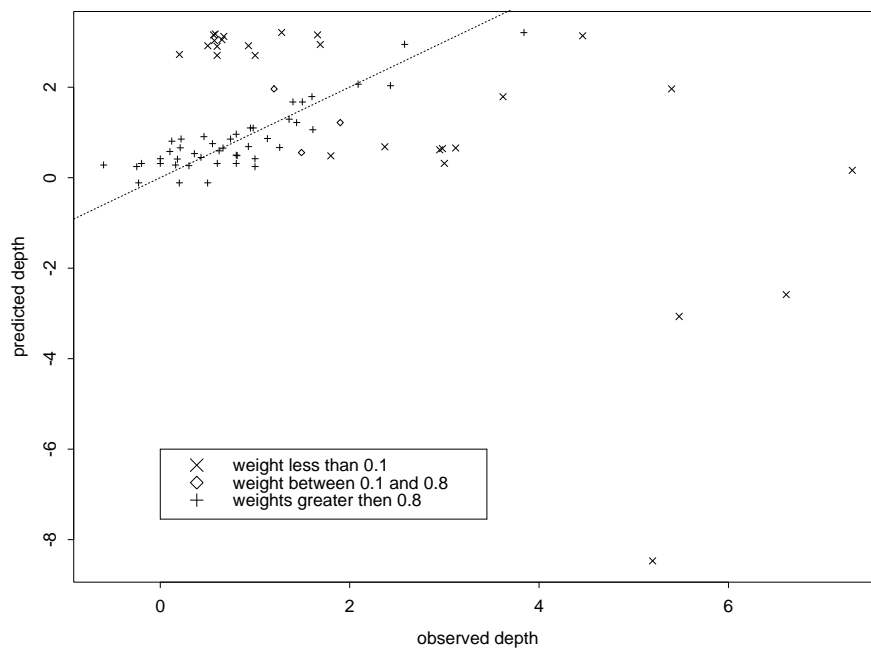


Figure 3.8: Observed values versus predicted values for solution 2. The plot indicates which bores have been down-weighted.

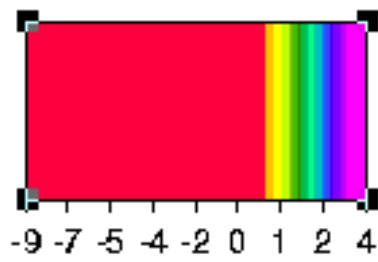
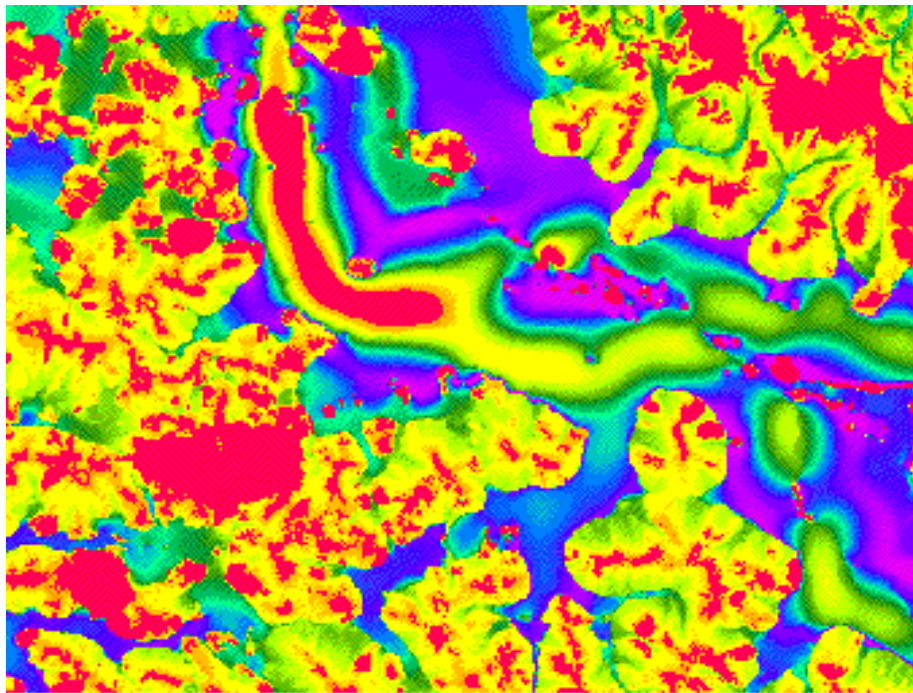


Figure 3.9: *The hydraulic head for Boscabel derived from solution 7.*

### 3.2.3 solution 13

Solution 13 gave a standard deviation of 0.049 and, after the variables `F1Slope` and `h.above.streams` were dropped, is summarized as:

```
> summary(model.13)
```

```
Call: lm(formula = depth ~ AvHgt + AvSlp + wa + fplen + PerClr,
         data = bores.boscabel, weights = weights)
```

```
Residuals:
```

```
      Min       1Q   Median       3Q      Max
-0.04146 -0.01948 -0.001119  0.01791  0.04246
```

```
Coefficients:
```

	Value	Std. Error	t value	Pr(> t )
(Intercept)	1.5664	0.0511	30.6672	0.0000
AvHgt	0.0048	0.0011	4.2076	0.0001
AvSlp	-0.0012	0.0002	-5.5744	0.0000
wa	0.0027	0.0012	2.3109	0.0254
fplen	0.0001	0.0000	5.1530	0.0000
PerClr	0.0232	0.0014	17.0832	0.0000

```
Residual standard error: 0.02538 on 46 degrees of freedom
```

```
Multiple R-Squared: 0.8691
```

```
F-statistic: 61.09 on 5 and 46 degrees of freedom, the p-value is 0
```

Figure 3.10 shows which bores were downweighted for solution 13. In this instance the plot regression relationship is driven by a single outlying point. Figure 3.11 shows the fitted hydraulic head surface.

### 3.2.4 using only “trusted” bores

Instead of selecting the elemental sample randomly from the 74 bores, they were selected from the group of bores with a “WW” prefix plus the three extra bores. These, in the opinion of Richard George, are all good bores in the sense of the data being reliable (but perhaps not necessarily representative of a significant range of the explanatory variables). This was done and 20 samples gave the following standard deviations:

```
[1] 0.04959010 0.06384267 0.05861289 0.05861288 0.06384270 0.05164116
[7] 0.05164116 0.05640190 0.06384270 0.05640192 0.04959010 0.05164116
[13] 0.05164116 0.04959010 0.04959007 0.04959007 0.05861288 0.05164116
[19] 0.04959007 0.05681216
```

Solution 15 was investigated further and was found to down-weight the bores as follows:

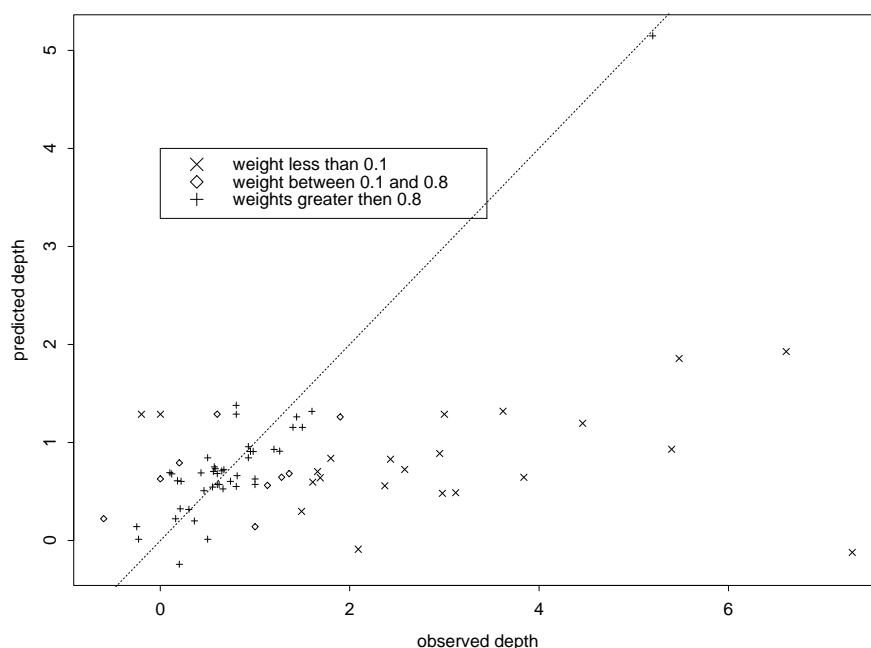


Figure 3.10: Observed values versus predicted values for solution 13. The plot indicates which bores have been down-weighted.

```
[1] "78DP03" "83DP07" "85CN01" "86DP04" "93DP05" "93KM04" "93KM05" "94BB06"
[9] "94BK01" "94WW11" "94WW12" "94WW13" "95KM22" "97BK05" "97WW32" "97WW33"
[17] "98CT03" "98CT07" "WW03D" "WW07D" "WW09D" "WW12D" "WW13D" "WW17D"
[25] "WW61D" "WW71D" "WW74D" "WW83D" "b66" "b38" "b37"
```

which includes many of the “trusted” bores. In addition, the final model:

```
(Intercept)      AvHgt      AvSlp      wa      fplen      PerClr
  1.566408  0.004795768 -0.001155049  0.002682367  0.0001324093  0.02319613
```

is essentially the same as the model from solution 13. This approach was not considered further.

### 3.3 The Kent model

The Kent model was applied to the Boscabel data to produce the hydraulic head surface shown in figure 3.12.

### 3.4 Conclusion

The images produced by solutions 1, 7 and 13 on the Boscabel data are very poor.

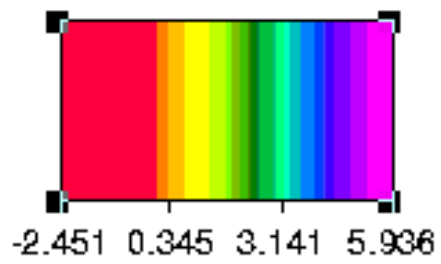
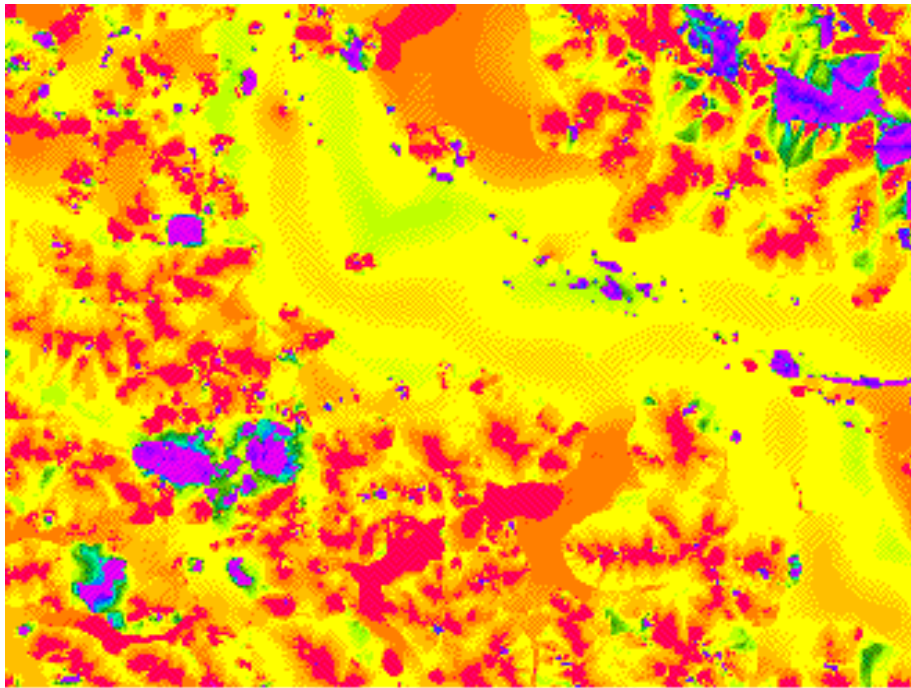


Figure 3.11: *The hydraulic head for Boscabel derived from solution 13.*

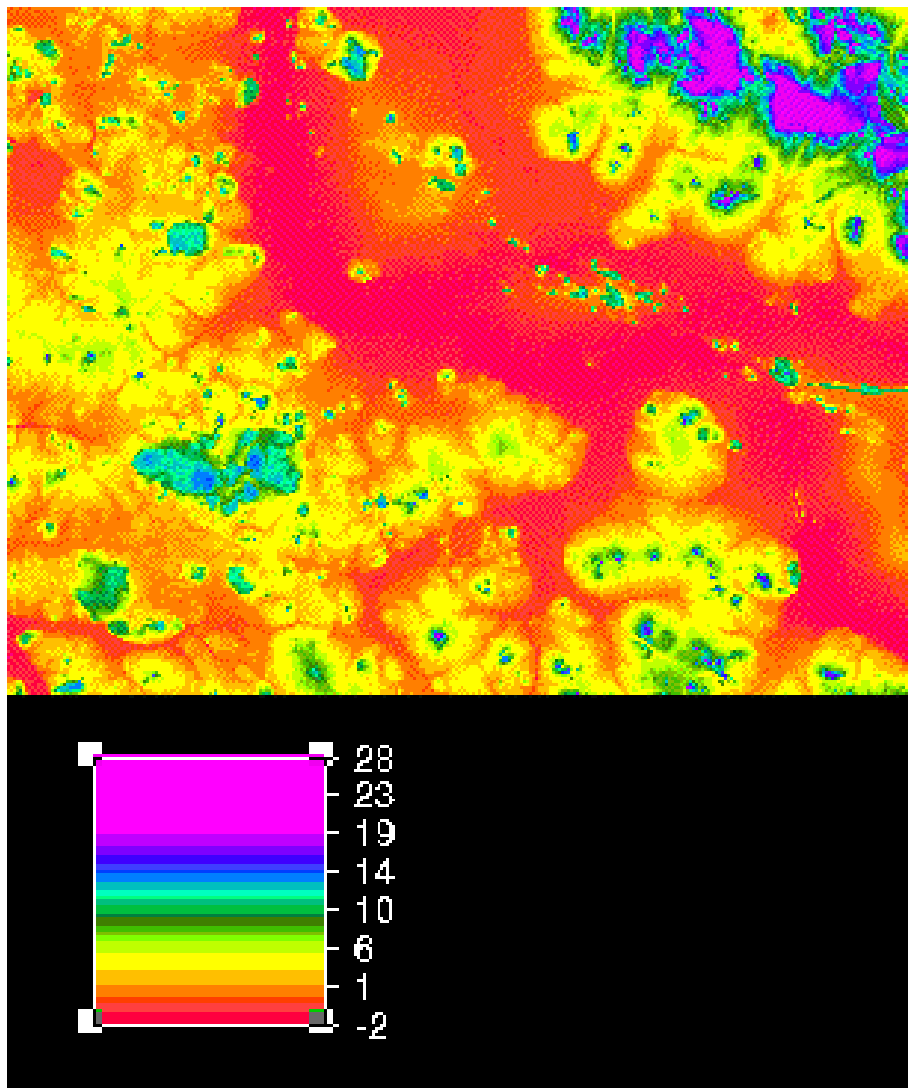


Figure 3.12: The hydraulic head surface for Boscabel derived from the Kent model.

In the judgment of hydrologist Richard George, the Kent model applied to the Boscabel data (figure 3.12) produces an image that, while it is inaccurate, is consistently inaccurate. That is, if the scale was changed to change the water level by 2 metres, the resulting image would be close to the true hydraulic head. In addition, there are several areas where the hydraulic head predicts a ground water depth that conflicts with the known depth of the bedrock. Errors of this sort are unavoidable and could, if necessary, be moderated by cropping the depth to ground water at some given value.

# 4

## Woodanilling

Figure 4.1 shows the bores used to fit the hydraulic head for Woodanilling. Figure 4.2 shows the result of fitting a robust model to this data, while figure 4.3 shows the resulting hydraulic head surface. It appears that all of the bores in the Woodanilling area are quite shallow. There is a sampling bias here in that the bores are located where there is an imminent danger of salinity (or to monitor a treatment in such a case), so there are no bores in the higher parts of the landscape. The result of this is that the fitted hydraulic head is shallow everywhere.

Figure 4.4 shows the result of applying the Kent model to Woodanilling.

The images in figures 4.3 and 4.4 are offered for evaluation.

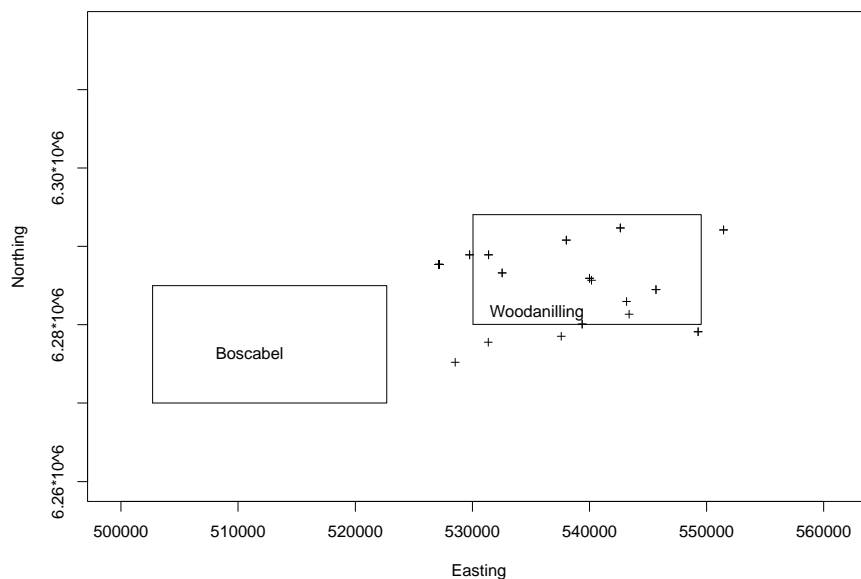


Figure 4.1: *The set of bores used for Woodanilling.*

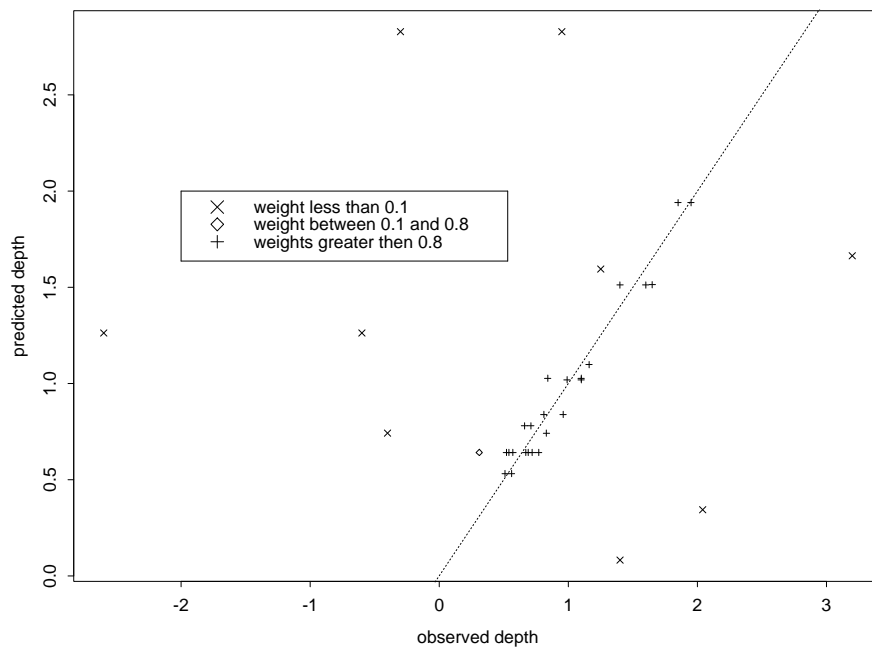


Figure 4.2: Observed versus predicted values for the robust Woodanilling model.

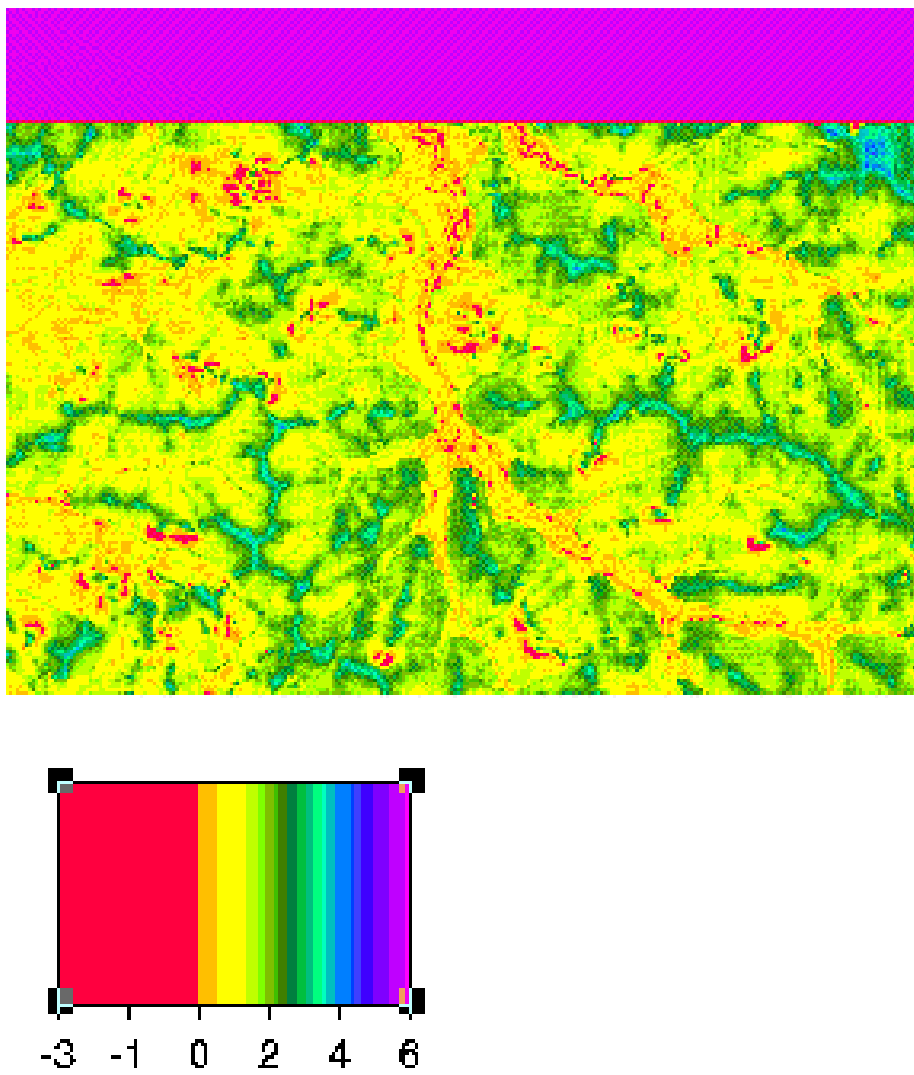


Figure 4.3: *The hydraulic head derived from the robust model for Woodanilling.*

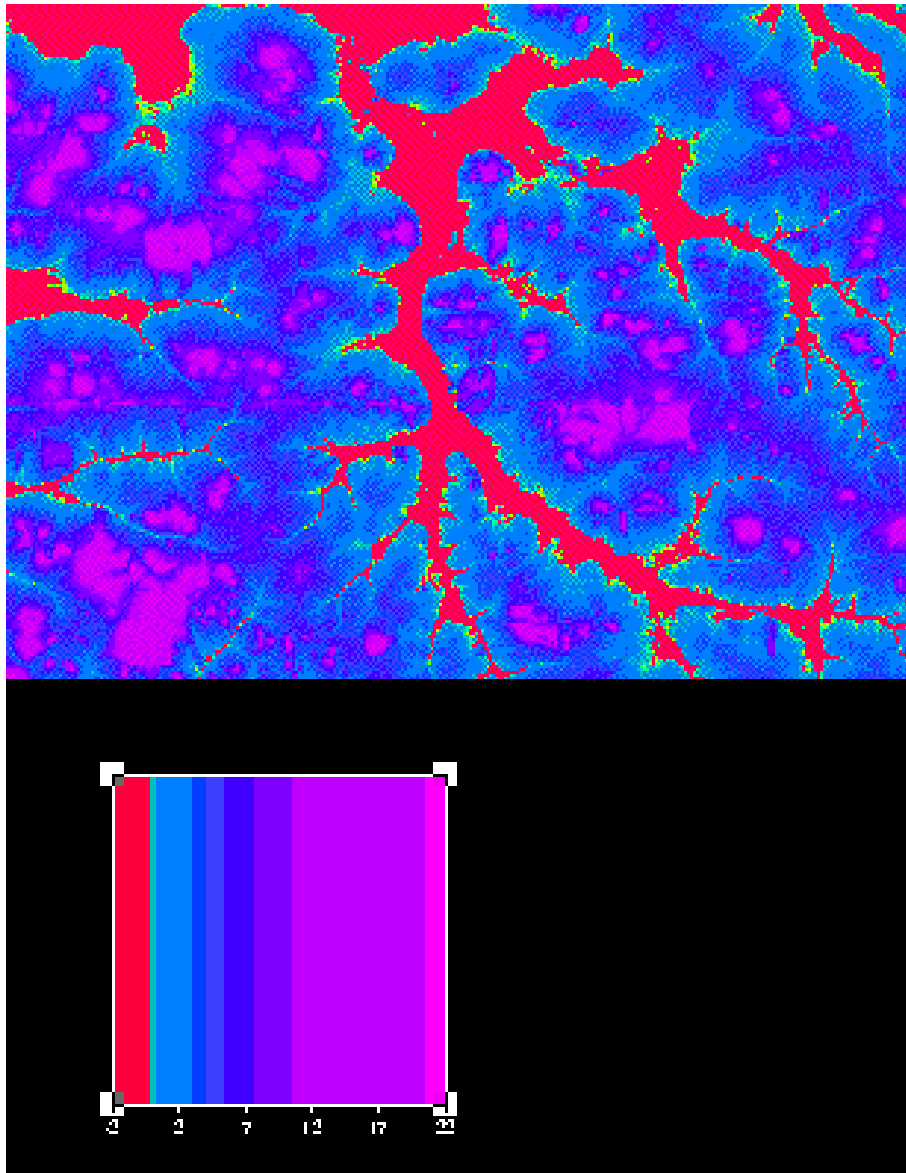


Figure 4.4: *The Kent model applied to Woodanilling.*

# 5

## Gnowangerup

Figure 5.1 shows the result of fitting a robust model to this data while figure 5.2 shows the resulting hydraulic head.

Figure 5.3 shows the result of applying the Kent model to Gnowangerup also the elevation variable. As the anomaly on the right of the image appears in the DEM, it is reflected in all of the DEM derived variables.

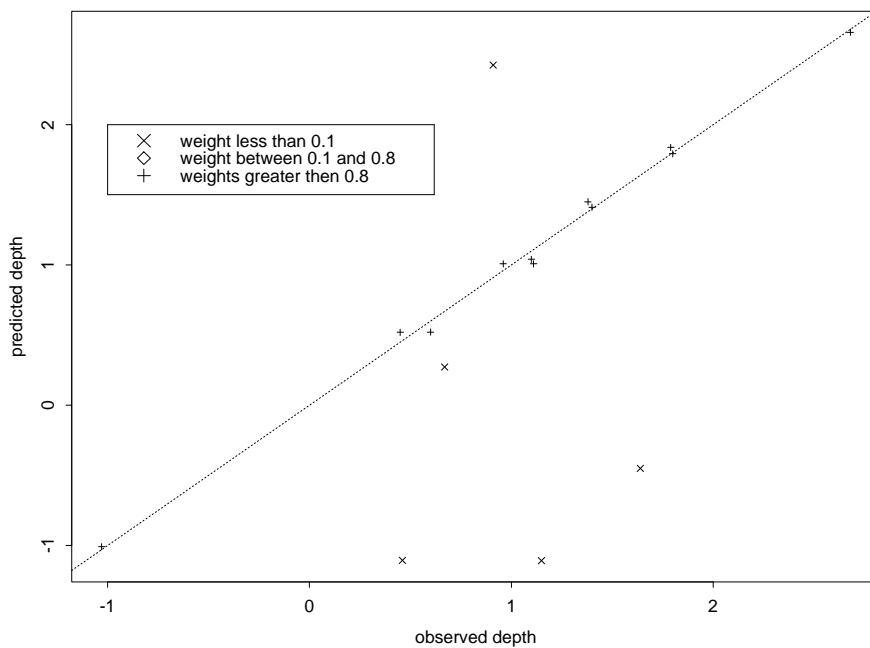


Figure 5.1: *The observed versus the predicted values for the robust Gnowangerup model.*

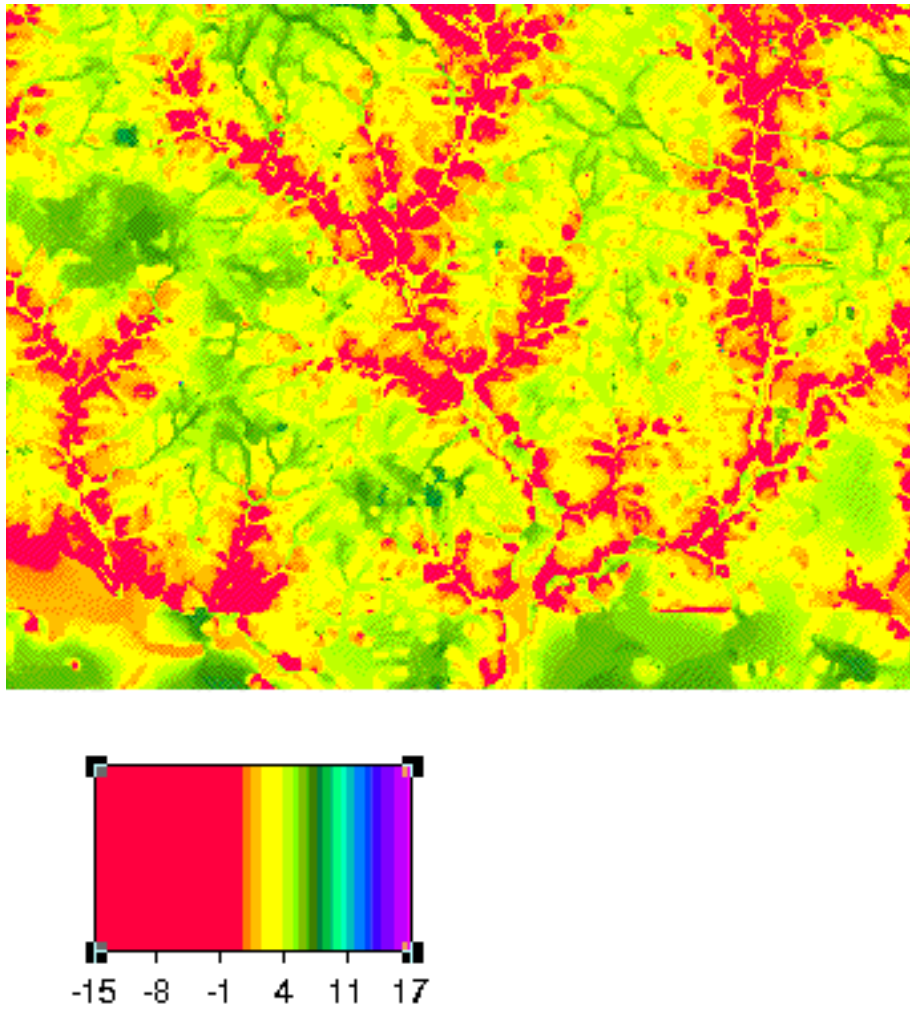


Figure 5.2: The hydraulic head derived from the robust model for Gnowangerup.

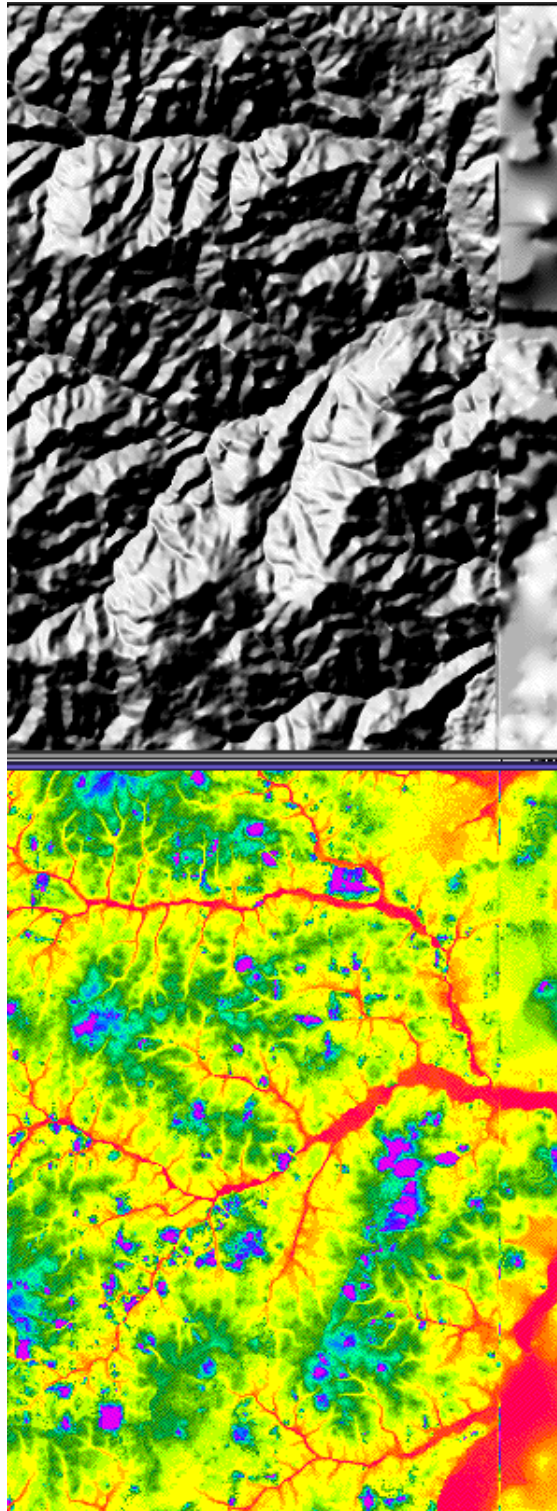


Figure 5.3: *The Kent model applied to Gnowangerup, beside the elevation variable. As the anomaly on the right of the image appears in the DEM, it is reflected in all of the DEM derived variables.*

# 6

## Southern Stirlings

We consider two sets of bores for this area. The data set, **bores**, is taken from the augmented set of bores and consists of 69 bores. The locations are shown in figure 6.1. We also consider a set of 29 trend bores in **trendbores**. The locations of these bores is shown in figure 6.2.

Figure 6.3 shows the observed values versus the predicted values (and the down-weighting of the bores) for a robust model applied to the **bores** data set. Figure 6.4 shows the resulting hydraulic head. Figure 6.5 and figure 6.6 show the results for the trend bores. Both models appear plausible, but require the judgment of hydrologists experienced in the area. Note that (as commonly happens) the range of the hydraulic head is too great. It is unlikely that the head is ever 21 meters above ground. We have simply included the full image here for discussion purposes, if this was to be used in any way the it would be truncated to reasonable values.

Figure 6.7 shows the result of fitting the Kent model to this area. It is generally agreed that this is a considerable overestimate of the current discharge area.

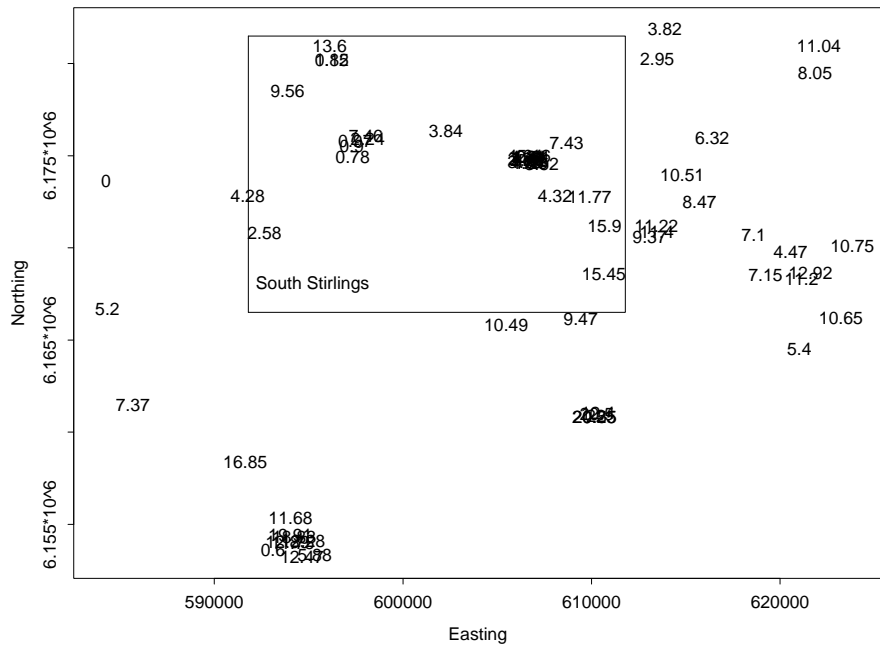


Figure 6.1: The distribution of the **bores** data set.

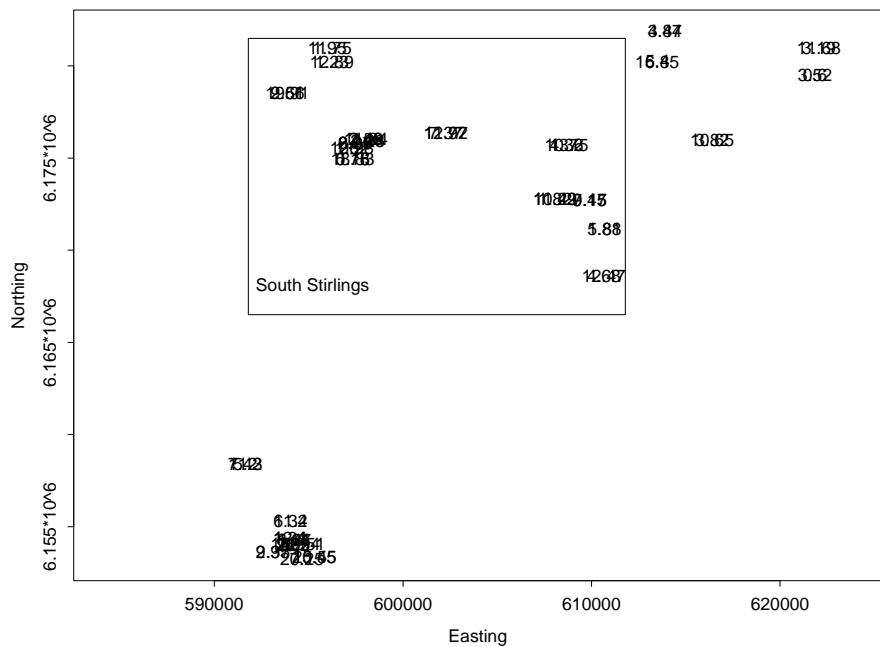


Figure 6.2: The distribution of the **trendbores** data set.

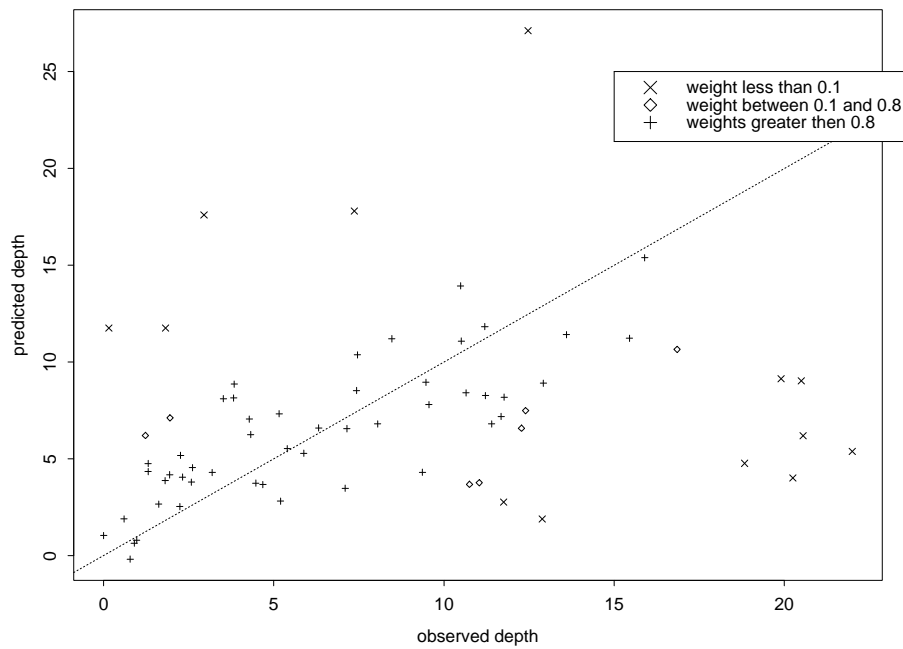


Figure 6.3: The observed versus the predicted values for the robust model applied to the **bores** data set.

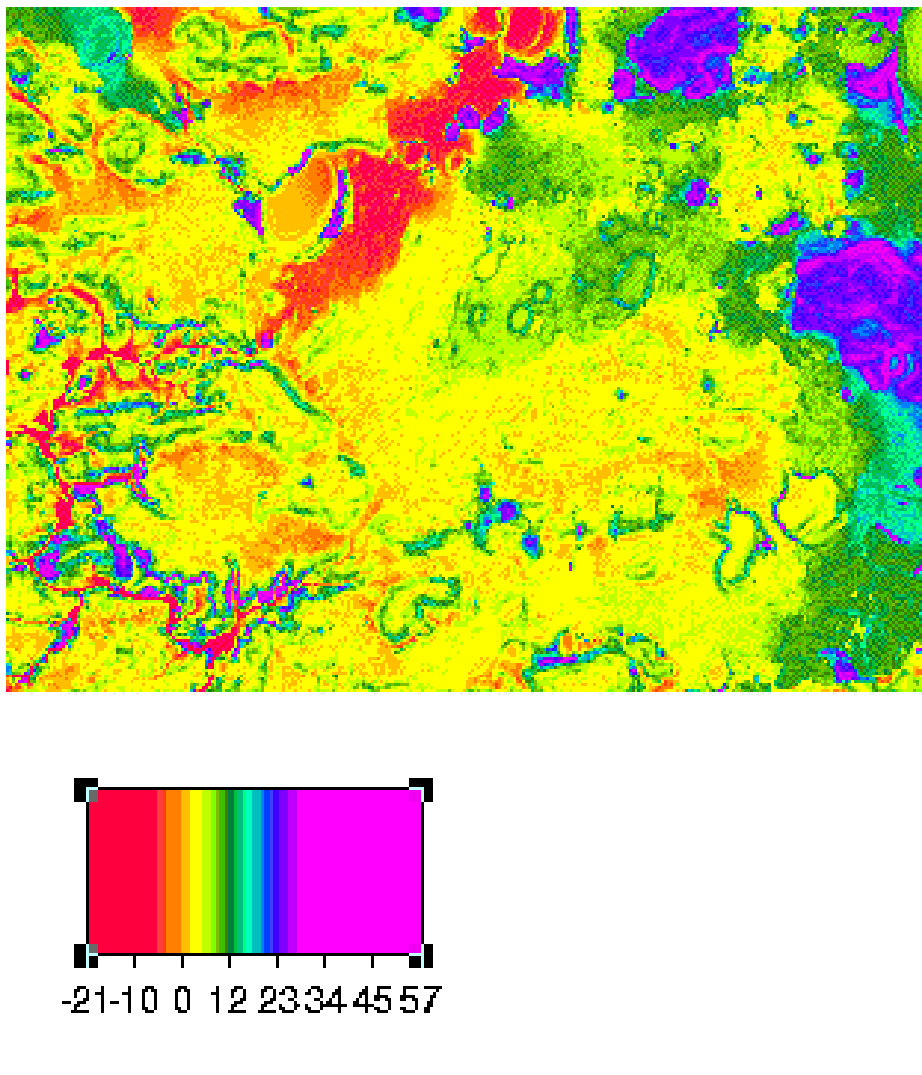


Figure 6.4: The hydraulic head derived from the robust model applied to the **bores** data set.

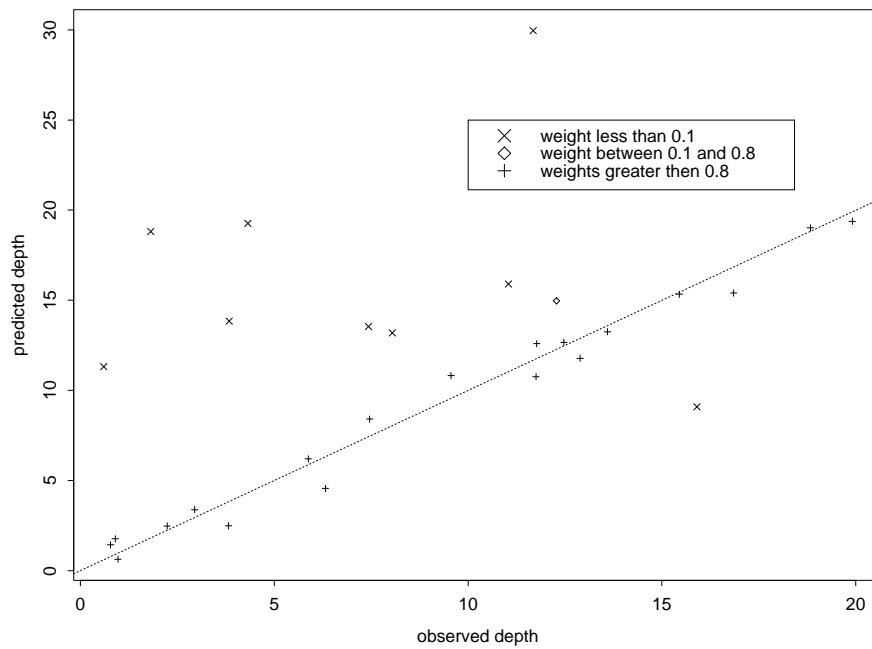


Figure 6.5: The observed versus the predicted values for the robust model applied to the **trendbores** data set.

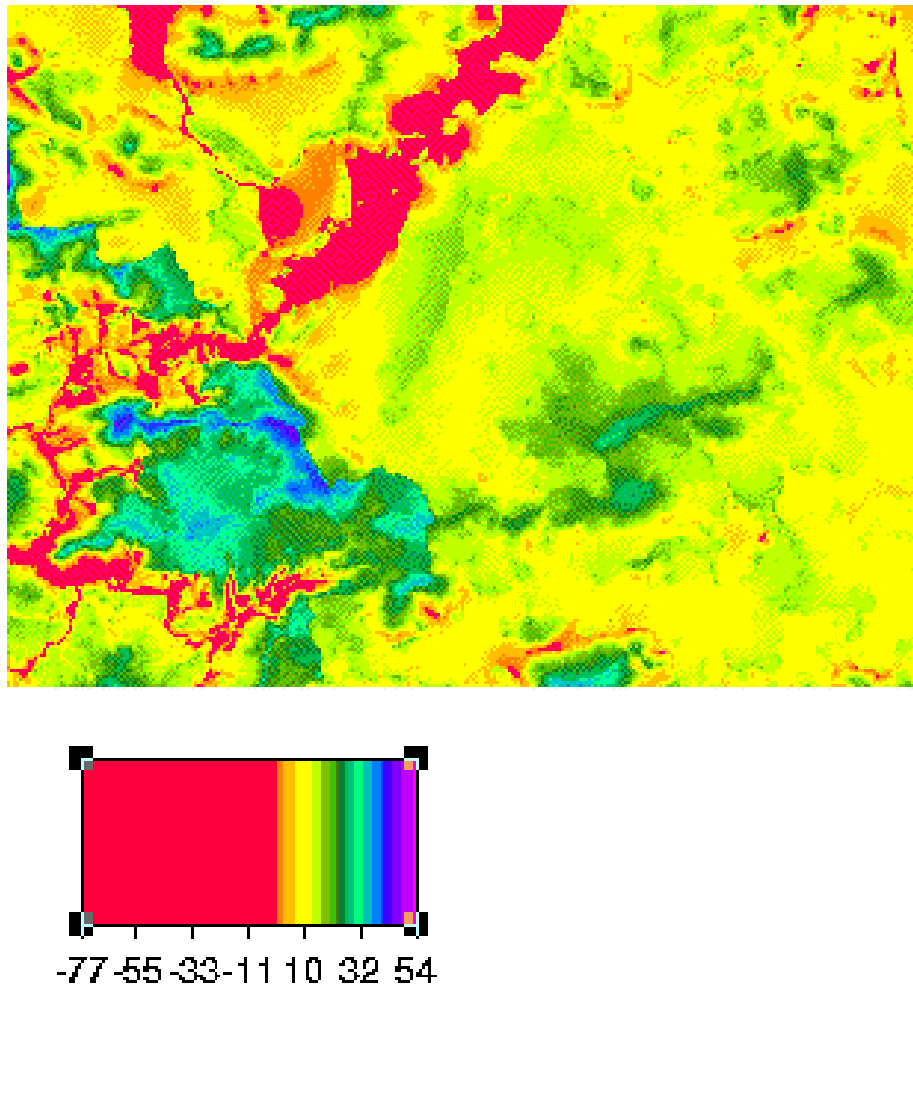


Figure 6.6: The hydraulic head derived from the robust model applied to the **bores** data set.

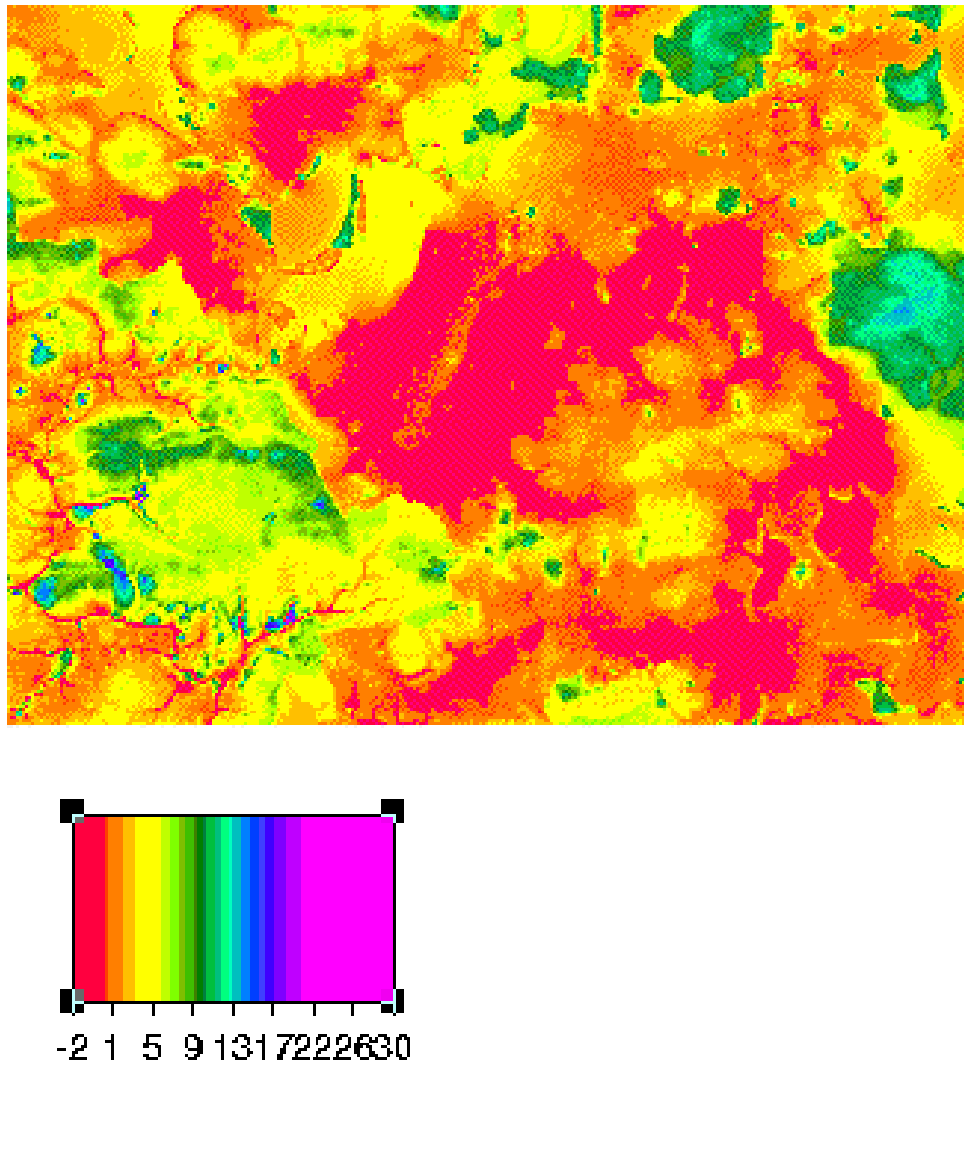


Figure 6.7: The hydraulic head for the South Stirlings derived from the Kent model.

# 7

## Toolibin

This section documents the analyses done on the Toolibin catchment. Toolibin has four groups of bores, which we designate the “central”, “northern”, “eastern” and “southern”, as shown in figure 7.1.

Our aim here is to see if we can use the given variables to partition the landscape in units such that the depth to ground-water can be more effectively predicted with each unit. We drop the three “northern” bores and consider the other three groups.

Figures 7.2 and 7.3 shows canonical variate analyses of the three groups of bores. It appears that the central bores are separable from the others but that the separability of the eastern and southern bores is poor.

As we hope to partition the image data using these canonical variate vectors and fit a different model to each partition, this is disappointing. This approach would require a partition such that contiguous regions of the image were assigned to a single unit. If this is not possible, the resulting hydraulic head surface is likely to be unacceptably rough.

A robust model was fitted to the Toolibin data (summarized below). Figure 7.4 shows the weights and it can be seen that the eastern group of bores have all been down-weighted. Figure 7.5 and 7.6 show the observed versus predicted values for the robust model.

```
> summary(robust.model.toolibin)
```

```
Call: lm(formula = depth ~ AvSlp + FlSlope + h.above.streams + wa + fplen +  
        PerClr, data = bores.t, weights = weights.t)
```

```
Residuals:
```

```
    Min      1Q  Median     3Q     Max  
-2.378 -0.6648  0.1166  0.5465  2.098
```

```
Coefficients:
```

```
                Value Std. Error t value Pr(>|t|)  
(Intercept) -4.0253   6.7317   -0.5980  0.5530  
      AvSlp   -0.1539   0.0657   -2.3408  0.0239  
      FlSlope -2.3824   0.7127   -3.3427  0.0017
```

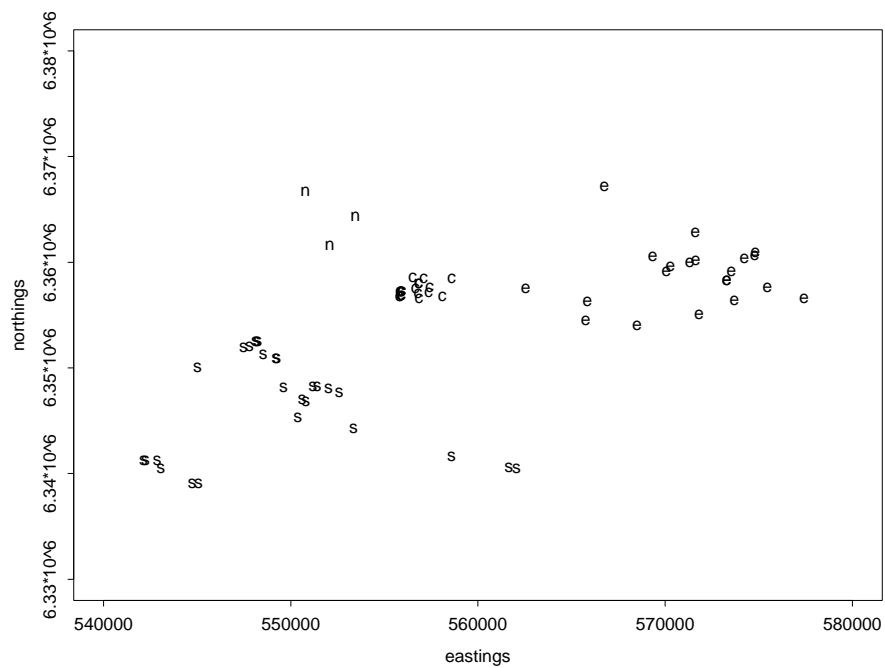


Figure 7.1: The distribution of bores in the Toolibin region. The “central”, “northern”, “eastern” and “southern” bores are respectively designated “c”, “n”, “e” and “s.”

h.above.streams	0.0142	0.0149	0.9475	0.3487
wa	-0.0480	0.0642	-0.7489	0.4580
fplen	-0.0015	0.0018	-0.7898	0.4340
PerClr	2.8907	1.4658	1.9720	0.0551

Residual standard error: 0.9998 on 43 degrees of freedom

Multiple R-Squared: 0.446

F-statistic: 5.769 on 6 and 43 degrees of freedom, the p-value is 0.0001798

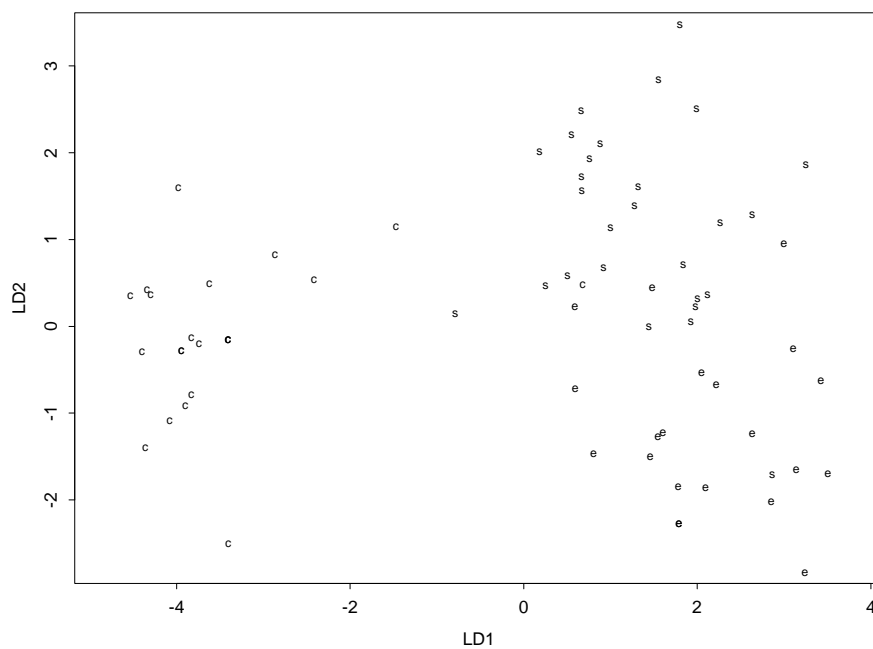


Figure 7.2: The eastern, central and southern Toolibin bores in Canonical Variate space.

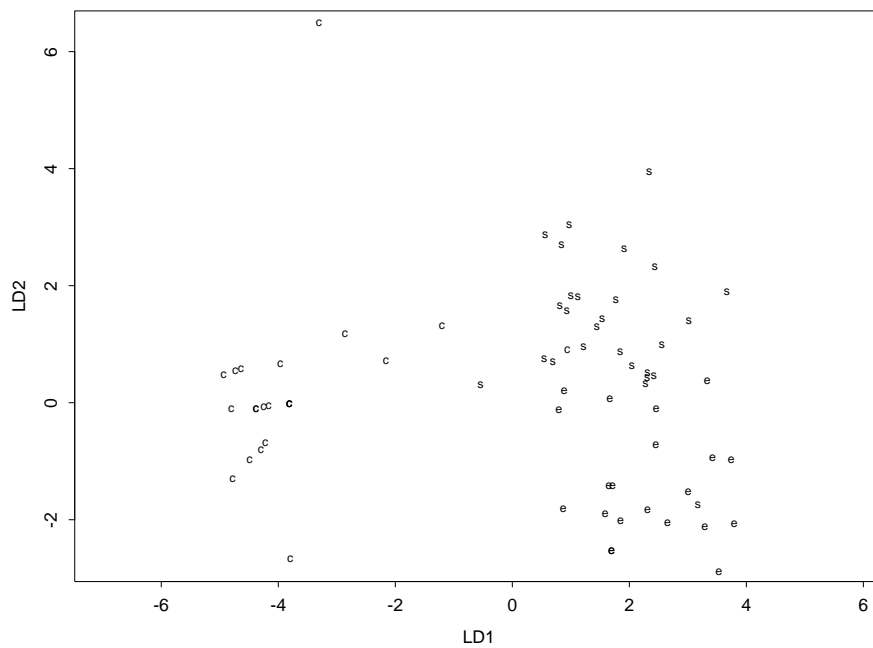


Figure 7.3: The eastern, central and southern Toolibin bores in Canonical Variate space where the covariance matrices are calculated assuming a  $t$ -distribution with 20 degrees of freedom.

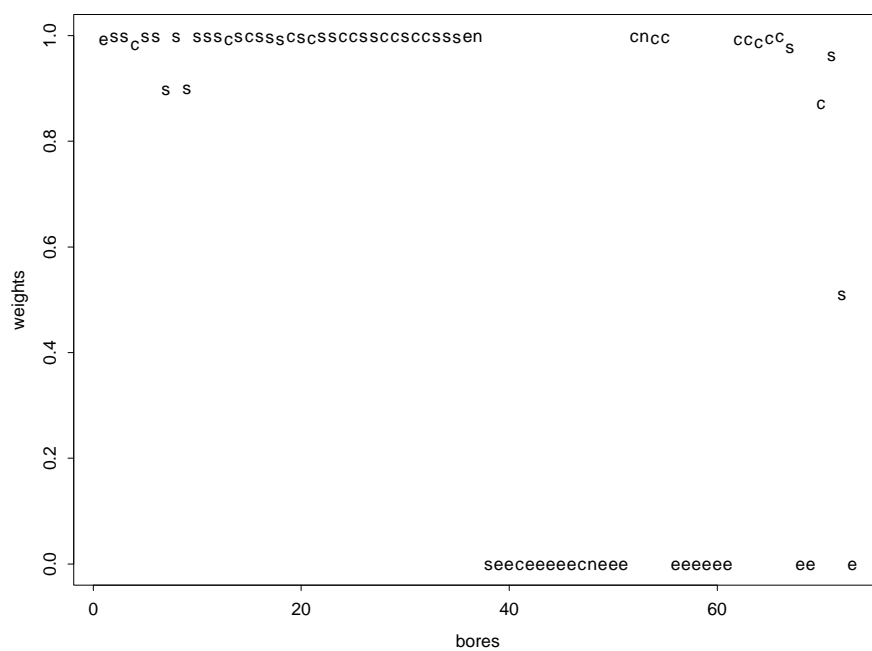


Figure 7.4: Weights for the robust model applied to the whole Toolibin data set. The “central”, “northern”, “eastern” and “southern” bores are respectively designated “c”, “n”, “e” and “s.”

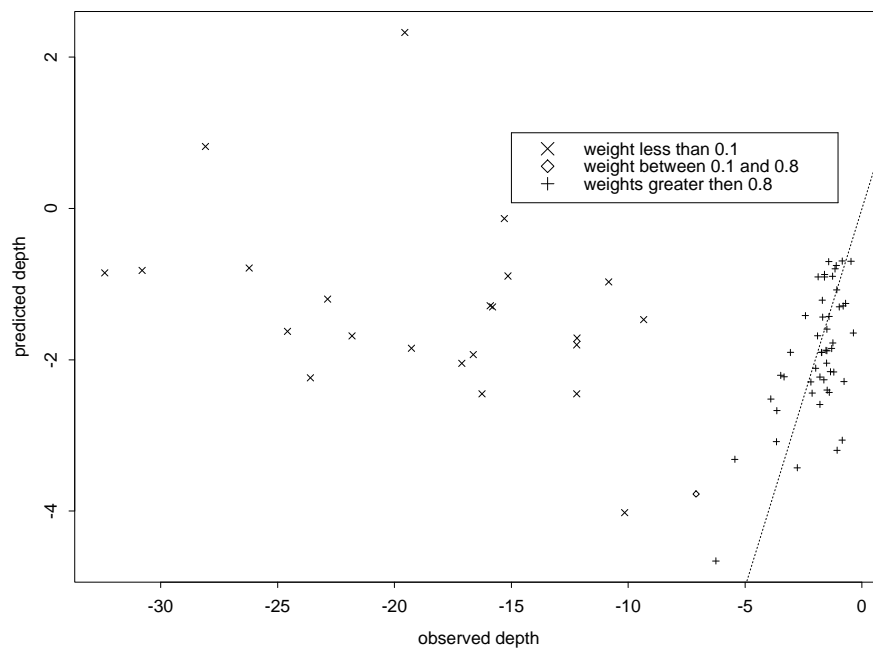


Figure 7.5: Observed versus predicted values for the robust model applied to the whole Toolibin data set. The plot indicates which bores have been down-weighted.

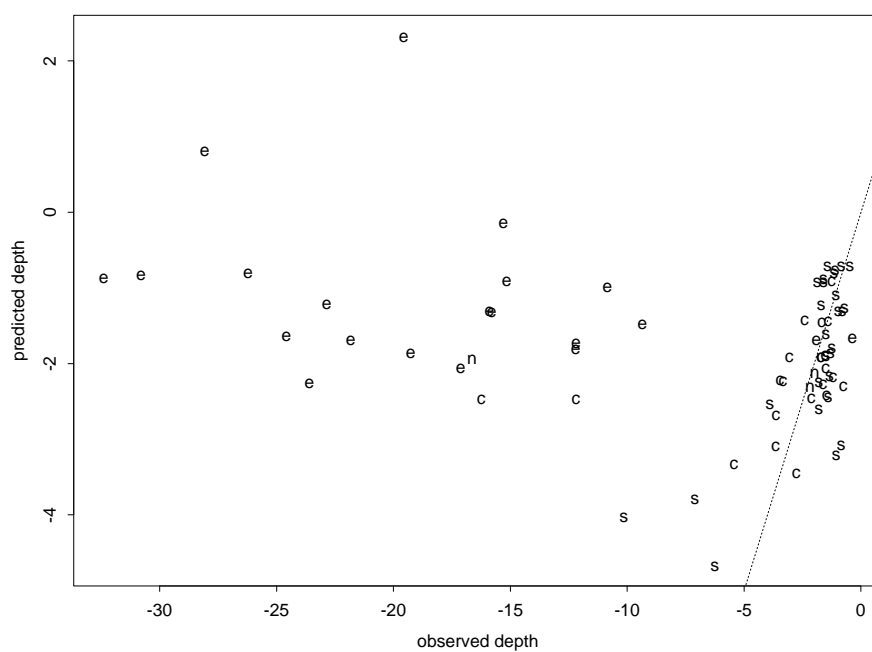


Figure 7.6: Observed versus predicted values for the robust model applied to the whole Toolibin data set. The “central”, “northern”, “eastern” and “southern” bores are respectively designated “c”, “n”, “e” and “s.”

## 7.1 Central Bores

Figure 7.7 shows the observed versus predicted values for a robust model applied to the central Toolibin data set. Figure 7.8 shows the result of applying this model to the whole set of data.

```
> summary(model)
```

```
Call: lm(formula = depth ~ AvHgt + AvSlp + h.above.streams + wa +  
fplen + PerClr, data = tt.central, weights = weights.central)
```

```
Residuals: Min 1Q Median 3Q Max -0.3206 -0.1281 0.032 0.1314 0.2855
```

```
Coefficients:
```

	Value	Std. Error	t value	Pr(> t )
(Intercept)	268.2833	32.3791	8.2857	0.0000
AvHgt	-0.7620	0.0975	-7.8154	0.0001
AvSlp	-6.1003	0.7243	-8.4228	0.0000
h.above.streams	0.0795	0.0129	6.1627	0.0003
wa	-0.5684	0.0543	-10.4666	0.0000
fplen	0.0120	0.0013	9.1597	0.0000
PerClr	51.0889	6.0673	8.4204	0.0000

```
Residual standard error: 0.2431 on 8 degrees of freedom
```

```
Multiple R-Squared: 0.9583
```

```
F-statistic: 30.61 on 6 and 8 degrees of freedom, the p-value is 4.254e-05
```



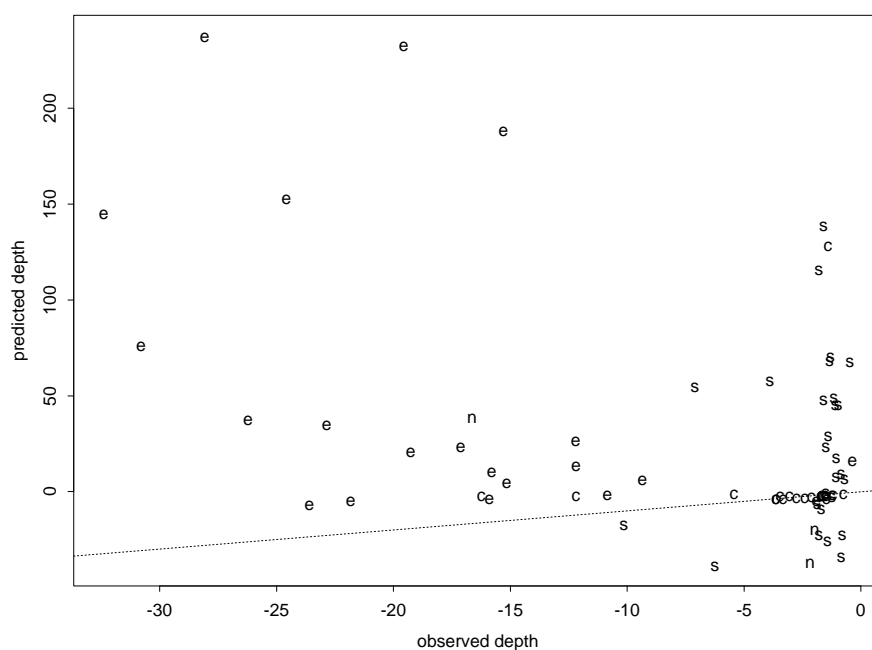


Figure 7.8: Observed versus predicted values for the robust model for the central bores applied to the whole Toolibin data set. The “central, “northern”, “eastern” and “southern” bores are respectively designated “c”, “n”, “e” and “s.”

## 7.2 Eastern Bores

Figure 7.9 shows the observed versus predicted values for a robust model applied to the central Toolibin data set. Figure 7.10 shows the result of applying this model to the whole set of data.

```
> summary(model)
```

```
Call: lm(formula = depth ~ AvHgt + FlSlope + h.above.streams + wa +
fplen + PerClr, data = tt.eastern, weights = weights.eastern)
```

```
Residuals: Min 1Q Median 3Q Max -1.992 -0.3432 0.04199 0.3464 1.365
```

```
Coefficients:
```

	Value	Std. Error	t value	Pr(> t )
(Intercept)	-322.8531	38.9057	-8.2983	0.0001
AvHgt	0.2213	0.0116	19.0346	0.0000
FlSlope	-15.3057	2.0305	-7.5378	0.0001
h.above.streams	-0.3006	0.0274	-10.9532	0.0000
wa	-0.9966	0.3806	-2.6187	0.0345
fplen	-0.0643	0.0056	-11.4127	0.0000
PerClr	69.5859	7.8459	8.8691	0.0000

```
Residual standard error: 1.123 on 7 degrees of freedom
```

```
Multiple R-Squared: 0.9859
```

```
F-statistic: 81.45 on 6 and 7 degrees of freedom, the p-value is 4.051e-06
```

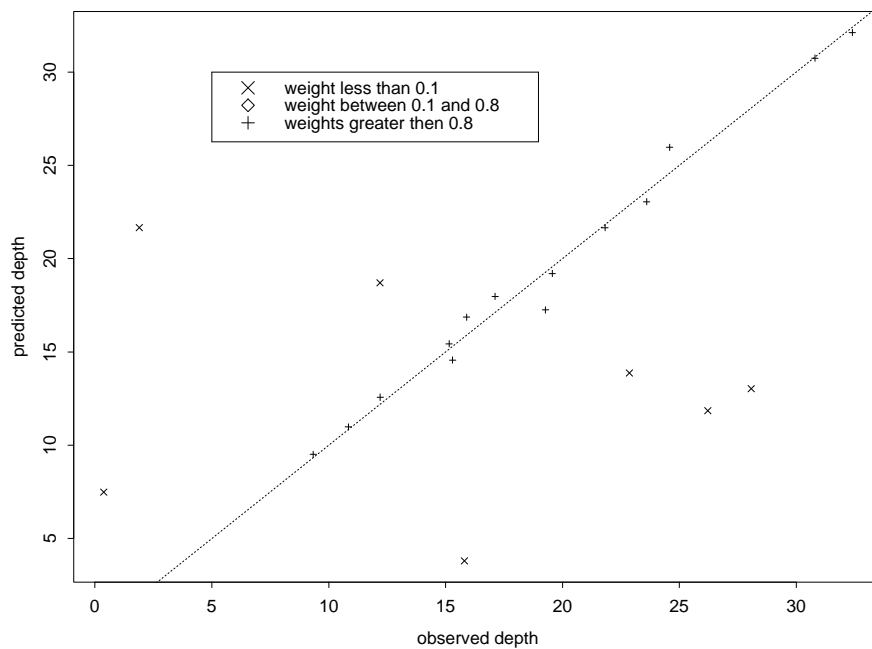


Figure 7.9: Observed versus predicted values for the robust model applied to the eastern Toolibin data set. The plot indicates which bores have been down-weighted.

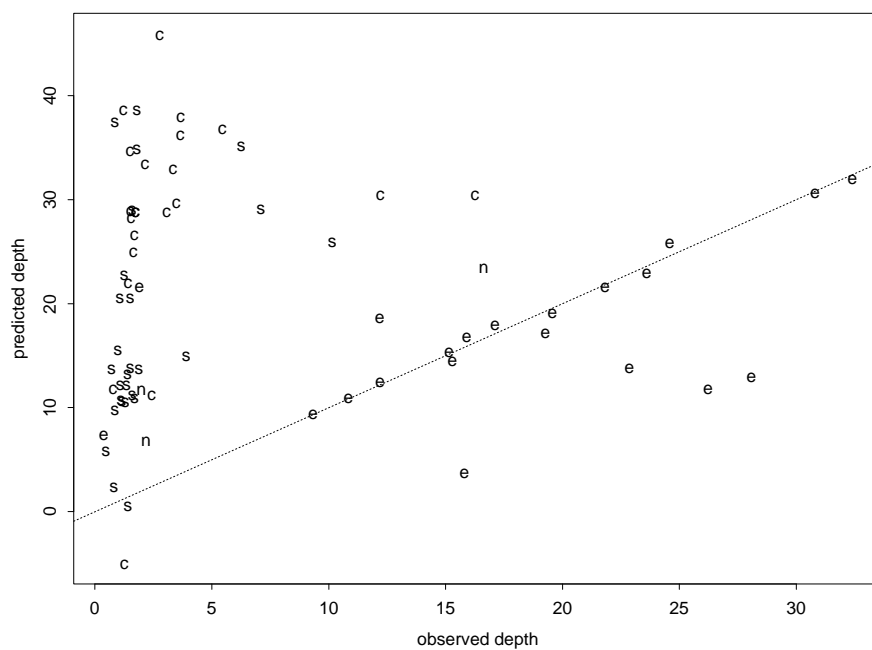


Figure 7.10: Observed versus predicted values for the robust model for the eastern bores applied to the whole Toolibin data set. The “central, “northern”, “eastern” and “southern” bores are respectively designated “c”, “n”, “e” and “s.”

### 7.3 Southern Bores

Figure 7.11 shows the observed versus predicted values for a robust model applied to the central Toolibin data set. Figure 7.12 shows the result of applying this model to the whole set of data.

```
> summary(model)
```

```
Call: lm(formula = depth ~ AvHgt + AvSlp + FlSlope + wa, data =  
tt.southern, weights = weights.southern)
```

```
Residuals:
```

```
Min      1Q      Median 3Q      Max  
-0.2663 -0.1437 -0.005669 0.0858 0.2778
```

```
Coefficients:
```

```
                Value Std. Error t value Pr(>|t|)  
(Intercept) -4.7806  0.9092    -5.2583  0.0001  
      AvHgt   0.0120  0.0019     6.4585  0.0000  
      AvSlp  -0.0378  0.0161    -2.3544  0.0326  
      FlSlope 1.6065  0.2869     5.5996  0.0001  
      wa    -0.0740  0.0150    -4.9473  0.0002
```

```
Residual standard error: 0.1809 on 15 degrees of freedom
```

```
Multiple R-Squared: 0.7413
```

```
F-statistic: 10.75 on 4 and 15 degrees of freedom, the p-value is 0.0002587
```

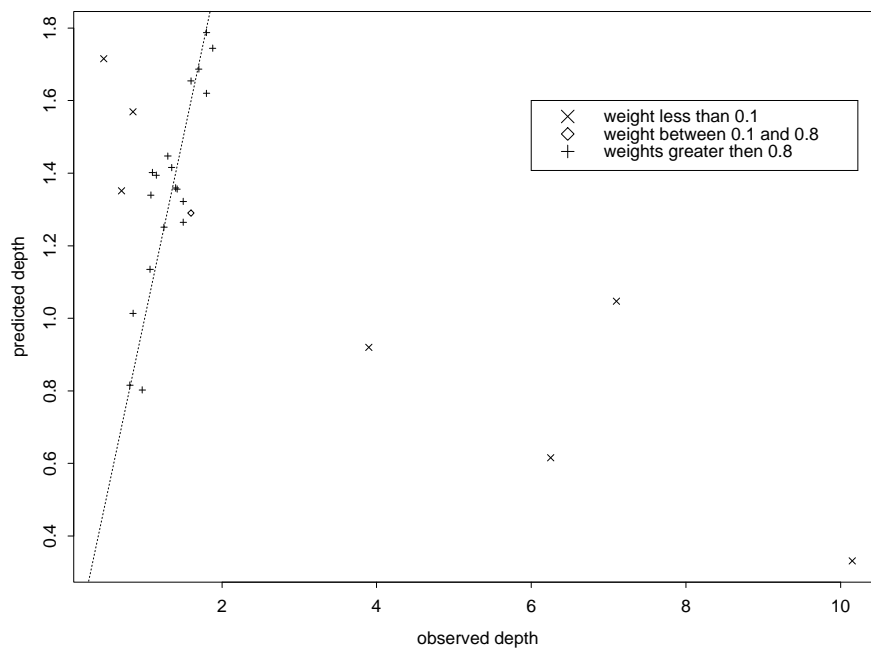


Figure 7.11: Observed versus predicted values for the robust model applied to the southern Toolibin data set. The plot indicates which bores have been down-weighted.

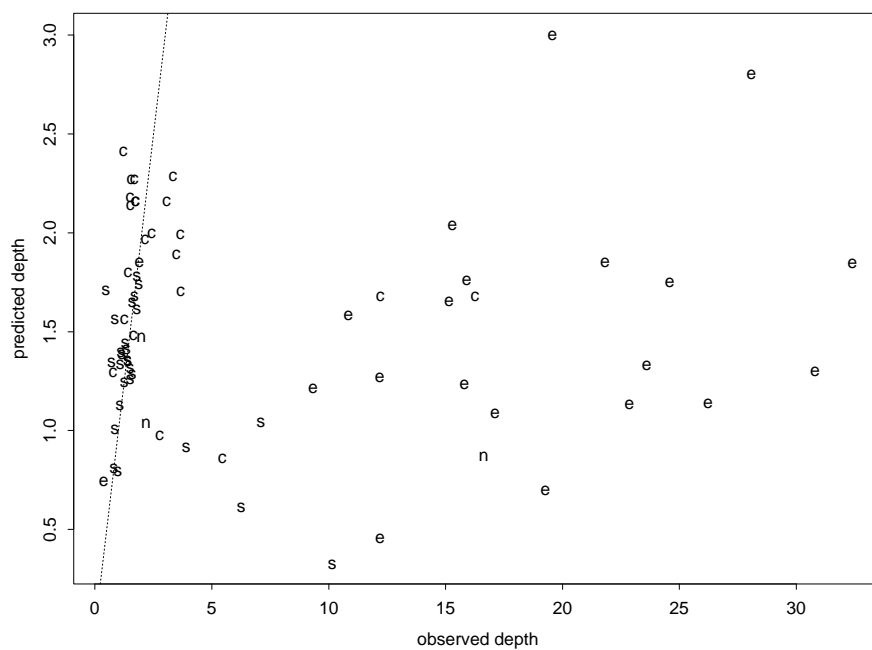


Figure 7.12: Observed versus predicted values for the robust model for the southern bores applied to the whole Toolibin data set. The “central”, “northern”, “eastern” and “southern” bores are respectively designated “c”, “n”, “e” and “s.”

## 7.4 Conclusion

It appears from figure 7.11 that the model that fits the southern bores also models the central bores. However the southern (figure 7.12) and eastern bores (figure 7.10) require separate models.

Hence, we have the situation where groups of bores that can not be separated require different models, but two groups that can be separated are described by the same model. The central bores are in a drainage channel and should be quite distinguishable on the DEM-derived variables. It has been suggested that elapsed time since clearance is the variable that distinguishes the eastern and southern areas; unfortunately it is not available for most of the region.

We conclude that it is not possible to effectively partition the landscape for the purpose of modeling ground water depth using the DEM-derived variables

# Appendix A

## Questions of the DEM

The analysis was repeated with **new.bores.kent** because it was believed that this data set was more accurate. There now seems little reason for this optimism. Table A.1 shows data for 40 bores from the Kent catchment<sup>1</sup> and gives the surveyed water level and elevation and the LandMonitor DEM height together with the discrepancies in elevations. Bore **kw1** has a 5.8 meter discrepancy between the surveyed elevation and the LandMonitor DEM height. The surveyed elevation should be accurate to 20cm, so this suggests that the LandMonitor DEM is only accurate to 6 metres.

Following up the original study, it was found<sup>2</sup> that “surveyed” in the context of these bores only means surveyed for elevation. The eastings and northings were estimated off a map. In the case of **kw1**, if both the surveyed elevation and the LandMonitor DEM are correct, then the position of the bore must be incorrect by at least 100 metres.

Figures A.1, A.2, and A.3 show the position of bore **kw1** in the unsmoothed LandMonitor DEM, the smoothed LandMonitor DEM and the elevation variable in the derived variables file.

Given that none of the 113 bores in this data set is surveyed for position, it does not appear that this data set will provide any advantage over **kent.bores** (see Hodgson, 2000).

---

<sup>1</sup>These bores are from the study Salama et al. (1997)

<sup>2</sup>By Geoff Hodgson.

bore name	Easting	Northing	water level	Surveyed Elevation	DEM	difference
P3d93	520052.7	6193646	-3.24	230.824	231.37	0.546
RG1d93	509856.8	6184230	-11.00	234.564	231.56	-3.004
RG3d93	510690.6	6184162	-1.86	218.249	217.31	-0.939
k3	525074.4	6183500	-1.55	211.417	213.24	1.823
k4	525083.8	6183645	-6.35	223.056	221.75	-1.306
k5	525077.5	6183886	-13.17	231.680	231.44	-0.240
kw1	538439.6	6175904	-9.35	239.586	233.72	-5.866
kw7	528600.4	6176308	-3.94	243.926	244.06	0.134
mw1	542818.7	6184746	-8.80	264.316	262.77	-1.546
mw2	540620.8	6184753	-9.23	263.873	265.04	1.167
mw4	535945.7	6184791	-1.75	231.659	229.07	-2.589
mw5	535716.0	6184803	-2.84	227.695	226.98	-0.715
rg2	510569.4	6184230	-7.41	224.966	221.94	-3.026
rg4	511379.4	6184246	-1.47	216.895	217.62	0.725
rg5	512327.7	6184180	-1.40	216.405	216.58	0.175
s1	527317.9	6180834	-4.06	229.389	228.91	-0.479
s10	517838.0	6181693	-1.34	206.506	208.22	1.714
s3	524964.4	6181538	-5.49	237.367	239.97	2.603
s5	522610.5	6181724	-5.86	222.562	223.57	1.008
s6	521312.9	6181718	-6.30	223.747	224.73	0.983
sy2	525627.9	6198335	-8.33	253.665	253.85	0.185
sy3	525660.1	6198202	-5.58	241.086	244.59	3.504
sy4	526976.5	6197725	-5.29	236.014	237.27	1.256
sy5	527725.2	6197169	-4.39	235.834	236.04	0.206
sy6	527918.9	6197092	-4.47	235.768	236.03	0.262
t1	548603.0	6189530	-3.78	254.170	251.58	-2.590
t2	548923.1	6190396	-4.99	253.406	251.00	-2.406
t3	548918.8	6190328	-4.53	252.701	251.59	-1.111
tu1	521547.6	6192099	-4.30	223.483	224.25	0.767
tu2	522513.6	6192074	-3.12	221.858	222.04	0.182
tu4	524762.0	6190932	-3.91	222.207	221.28	-0.927
tu5	525508.5	6190578	-0.70	228.055	226.35	-1.705
p1	520669.4	6192408	-2.88	226.410	223.58	-2.830
p2	520071.1	6192908	-1.44	227.560	227.84	0.280
m2	543986.5	6188660	-4.71	252.874	250.91	-1.964
m3	543312.4	6189122	-1.22	243.497	242.87	-0.627
n1	536605.8	6191684	-3.84	241.768	239.97	-1.798
n2	535159.1	6191668	-0.37	240.472	240.19	-0.282
n3	533397.9	6191788	-24.21	265.192	262.74	-2.452
k2a	525073.8	6182328	0.90	221.869	222.07	0.201

Table A.1: Data for 40 bores from the Kent catchment. The table gives the surveyed water level and elevation and the LandMonitor DEM height together with the discrepancies in elevations.

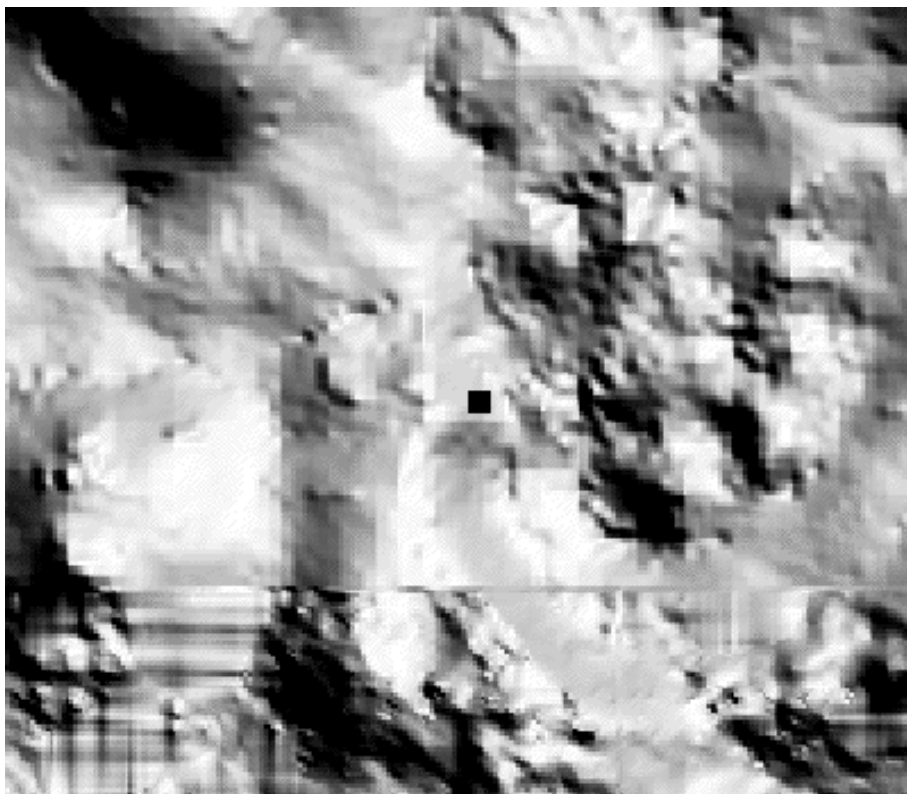


Figure A.1: *The unsmoothed LandMonitor DEM. The bore **kw1** is shown at an elevation of 232.25 meters.*

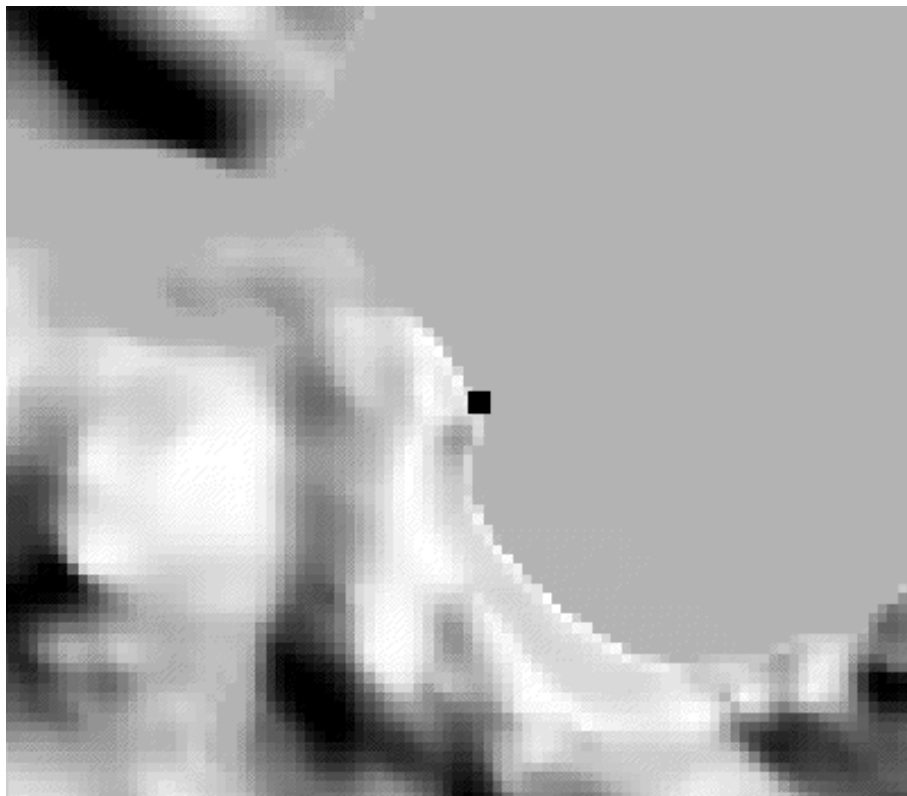


Figure A.2: *The smoothed and pit-filled LandMonitor DEM. The bore **kw1** is shown at an elevation of 233.9 meters.*

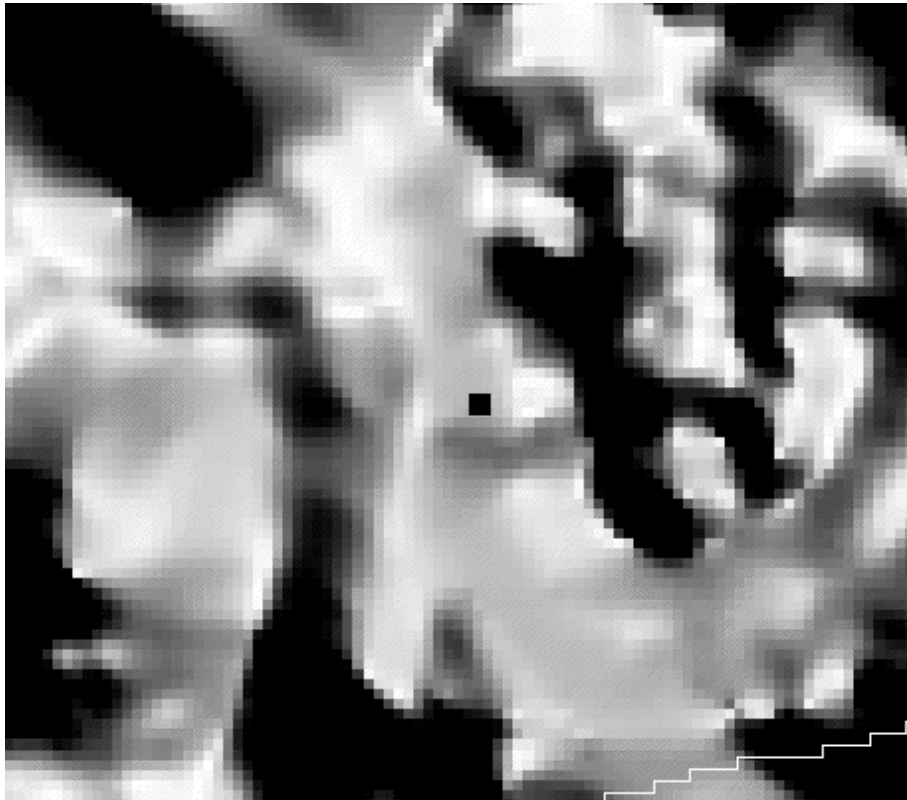


Figure A.3: *The elevation variable from the derived variables file. This is smoothed but not pit-filled. The bore **kw1** is shown at an elevation of 232.09 meters.*

# Appendix B

## Robust Regression

The robust method of Campbell et al. (1998) is illustrated on the Belgian telephone data discussed in Rousseeuw and Leroy (1987). It has been our experience that generally there are two solutions found, one fitting the outlying points and one fitting the good points. Choosing the solution with the minimum (weighted) standard deviation leads to the good solution.

Figure B.1 shows a solution with a large standard deviation (solution 1), a solution with a small standard deviation (solution 2) and the least squares solution for comparison.

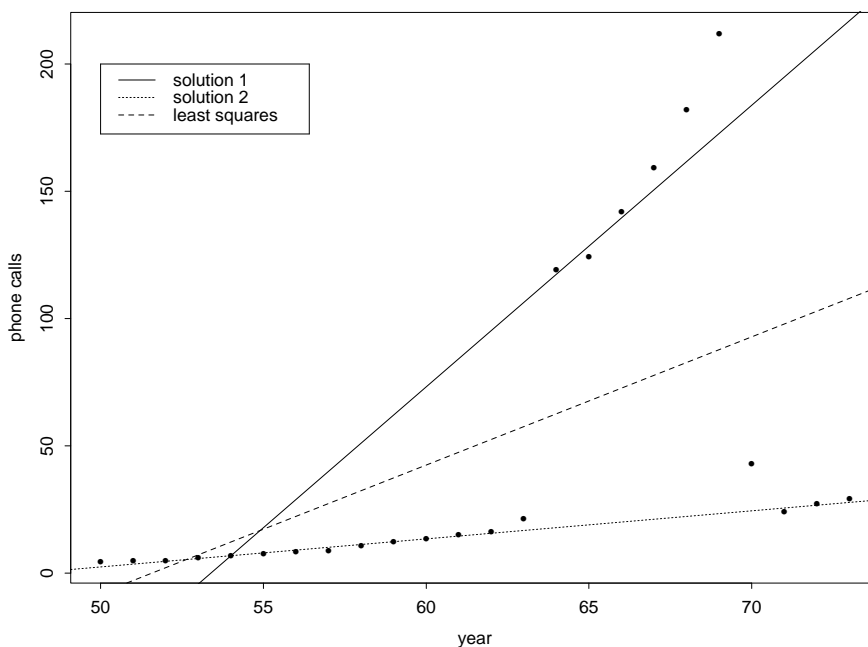


Figure B.1: *Three regression lines fitted to the Belgian telephone data.*

# Bibliography

- Campbell, N. A., Lopuhaá, H. P., and Rousseeuw, P. J. (1998). On the calculation of a robust s-estimator of a covariance matrix. *Statistics in Medicine*, 17:2685–2695.
- Friedman, J. H. (1991a). Adaptive spline networks. In Lippmann, R. P., Moody, J. E., and Touretzky, D. S., editors, *Advances in neural information processing systems 3*. Morgan Kaufmann.
- Friedman, J. H. (1991b). Multivariate adaptive regression splines (with discussion). *Annals of Statistics*, 19:1–141.
- Hastie, T. J. (1992). Generalized additive models. In Chambers, J. M. and Hastie, T. J., editors, *Statistical Models in S*, chapter 7. Wadsworth and Brooks, Pacific Grove, CA.
- Hodgson, G. (2000). A comprehensive analysis of borehole data from the SS2020 project area. Technical Report XX/00, CSIRO Land and Water.
- Rousseeuw, P. J. and Leroy, A. M. (1987). *Robust Regression and Outlier Detection*. Wiley, New York.
- Salama, R., Hutton, T., Shamir, R., Dyce, P., Dawes, W., and Owens, W. (1996). Dry-land salinity risk assessment in the Loddon and Campase catchments. Management and analysis of GIS data. Technical Memorandum 96.29, Division of Water Resources, CSIRO.
- Salama, R. B., Bartle, G. A., Ye, L., Williamson, R. R., Watson, G. D., and Knapton, A. (1997). Hydrogeomorphology and hydrogeology of the Upper Kent River catchment and its controls on salt distribution and patterns of groundwater discharge. Technical Report Technical Report 27/97, Division of Water Resources, CSIRO, Australia.
- Venables, W. N. and Ripley, B. D. (1994a). *Modern Applied Statistics with S-Plus*. Springer, New York, first edition.
- Venables, W. N. and Ripley, B. D. (1994b). *Modern Applied Statistics with S-Plus*. Springer, New York, second edition.