E

TWC/29/22
ORIGINAL: English
DATE: May 24, 2011

UPOV

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS
GENEVA

TECHNICAL WORKING PARTY ON AUTOMATION AND COMPUTER PROGRAMS

Twenty-Ninth Session
Geneva, June 7 to 10, 2011

ANALYSIS OF THE RELATION BETWEEN LOG SD AND MEAN OF VARIETIES

Document prepared by experts from Denmark and the United Kingdom

Introduction

1. At the twenty-eighth session on the Technical Working Party on Automation and Computer Programs, held in Angers, France, from June 29 to July 2, 2010, the bias in the present COYU method was discussed (document TWC/28/27; also previously documents TWC/26/17 and TWC/27/15). One of the possible approaches to overcome the bias in the present method was to use a linear and quadratic adjustment instead of the moving average method. However, it was questioned whether such an adjustment would be appropriate in all cases. It was decided to carry out a survey on the relationship between log SD and the mean of varieties in order to see if this could be modelled sufficiently well using a linear or quadratic regression. In such case it would be appropriate to introduce a new method for COYU based on a linear and quadratic effect adjustment of the log SD.

2. *The following data were received:*

- Data on *Lolium perenne* (perennial ryegrass) from DE, NL and UK through the years 1993-2002. Part of the data has previous been used in TWC/28/31, A study on Grass Reference Collections in Different Locations.
- Data on *Brassica napus* L. *oleifera* (spring oil seed rape) from DK through the years 1997-2005

- Data on *Pisum sativum* (field pea) from DK through the years 1997-2005

Data

Lolium perenne from Germany, Netherlands and United Kingdom

3. Data from two groups of oil seed rape grown in DUS trials in Germany, the Netherlands and United Kingdom was examined. Data were available for two subgroups of varieties (Amenity and Forage) for the years 1993 to 2002.

Table 1: Minimum, mean and maximum number of varieties over the combinations of year and country

Crop	Statistics	UPOV Character number								
		02	07	08	09	10	11	12	13	14
Lolium perenne, Amenity	Min	44	44	44	44	44	44	44	44	52
	Mean	64	64	64	64	64	64	64	64	65
	Max	77	80	74	74	74	74	74	74	73
Lolium perenne, Forage	Min	35		34	40	34	34	35	35	35
	Mean	57		57	58	57	57	56	57	57
	Max	73		73	73	73	73	73	73	73

4. For Amenity, the UPOV characteristics 02, 07, 08, 09, 10, 11, 12, 13 and 14 were available although some characteristics were not present from all countries. For Forage, the UPOV characteristics 02, 08, 09, 10, 11, 12, 13 and 14 were available although one characteristic was not present in one country and two characteristics were missing in some years in one country. The number of varieties for each combination of country and year varied between 34 and 74 (see Table 1).

Table 2: Description of the characteristics for *Lolium perenne*

UPOV	Description (TG/4/7)	Units
02	Plant: growth habit in autumn of year of sowing[a]	Note
07	Plant: natural height in spring	cm
08	Time of inflorescence emergence in 2nd year	days
09	Plant: natural height at inflorescence emergence	cm
10	Flag leaf: length at inflorescence emergence	mm
11	Flag leaf: width	mm
12	Stem : length of longest stem (inflorescence included when fully expanded)	cm
13	Inflorescence: length	cm
14	Inflorescence: number of spikelets	no

[a)] Recorded as degrees in UK and from this transformed to note using the formula: Note=degree×8/90 +1

Brassica napus (spring oil seed rape) and *Pisum sativum* (field pea) from Denmark

5. Data on reference varieties from two species grown in DUS trials in Denmark was examined. Data were available for the years 1997 to 2005. For *Brassica napus*, the UPOV characteristics 02, 03, 6, 13, 14, 17, 18, 19 and 20 were used. For *Pisum sativum*, the UPOV

characteristics 12, 13, 23, 24, 25, 48 and 49 were used. The number of varieties for each combination of species and year varied between 64 and 154 (see Table 3).

Table 3: Minimum, mean and maximum number of varieties over the years

Species	Statistics	UPOV Character number								
		02	03	06	13	14	17	18	19	20
Brassica napus	Min	76	76	65	65	64	65	65	65	65
	Mean	97	97	94	94	94	95	94	94	94
	Max	113	113	112	110	110	111	111	111	111
		UPOV Character number								
		12	13	31	32	48	49			
Pisum sativum	Min	82	82	82	82	82	82			
	Mean	120	120	120	120	120	120			
	Max	154	154	154	154	154	154			

Table 4: Description of the characteristics for *Brassica napus* and *Pisum sativum*

UPOV	Description (TG/36/6 corr. and TG/7/9)	Units
***Brassica napus* L. *oleifera* (spring oil seed rape)**		
02	Cotyledon: length	mm
03	Cotyledon: width	mm
06	Leaf: number of lobes	no
13	Flower: length of petals	mm
14	Flower: width of petals	mm
17	Plant: total length including side branches	cm
18	Siliqua: length	mm
19	Siliqua: length of beak	mm
20	Siliqua: length of peduncle	mm
***Pisum sativum* (field pea)**		
12	Stem: length.	cm
13	Stem: number of nodes	no
31	Stipule. length	mm
32	Stipule. width	mm
48	Pod length	mm
49	Pod: width	mm

Method

6. For each combination of crop (or crop subgroup) country, year and characteristic, the data from all varieties was analysed in the following linear regression model:

$$Y_v = \mu + \alpha_1 x_v + \alpha_2 x_v^2 + \alpha_3 x_v^3 + E_v$$

where

Y_v is the log of the pooled standard deviation for variety *v*

x_v is the pooled mean of the observations for variety *v*

E_v is the residual (random) variation not explained by the model. It was assumed that the residual term was normally distibuted with mean zero and constant variance.

μ is the intercept

α_1, α_2 and α_3 is the linear, quadratic and cubic effect of the mean

After fitting the model the following three tests were performed:

1. Is α_1 significant different from zero given that μ already was included in the model
2. Is α_2 significant different from zero given that μ and α_1 already were included in the model
3. Is α_3 significant different from zero given that μ, α_1 and α_2 already were included in the model

7. The cubic term is included to check whether a linear or quadratic model is sufficient. The results are summarised by the number of times that each of these tests was significant for each characteristic. In addition, the number of times that at least one of these tests was significant was tabulated.

8. For the characteristics where the overall test for the cubic term was significant all combinations of country and year that had a significant term were plotted. In those plots the fit using a straight line, a 2nd degree polynomial and a 3rd degree polynomial are shown together with the fit using a cubic spline with 4 degrees of freedom.

Results

9. The numbers of significant linear, quadratic and cubic relations between the logarithms of the pooled standard deviation and the means for each characteristic are shown in Tables 5 and 6 for the two groups of *Lolium perenne*, Table 7 for *Brassica napus* and in Table 8 for *Pisum sativum*.

Lolium perenne, Amenity

10. For the Amenity group, about 65% (140 out of 214 cases) of the examined combinations showed a significant relationship at the 5 % level for at least one of the 3 terms in the model (recorded as “≥1” in the table). In most of these cases the linear term was also significant. The quadratic and cubic terms were also significant in several cases. However some of these may be expected to be type I errors. Looking at the tests for cubic relation in each characteristic one finds that the number of tests for characteristic with UPOV number, 02, 08, 11 and 13 exceed the number expected with a significance level of 5 % and given the total number of tests for the characteristics. For the same tests at the 1% level only characteristics 08 and 11 exceed the limit for the number of tests.

Table 5: Number of significant effects for Lolium perenne: Amenity

UPOV no	Significant at the 5% level				Significant at the 1% level				Tests total
	Linear	Quadratic	Cubic	≥1[a]	Linear	Quadratic	Cubic	≥1[a]	
02	9*	3	6*	14	9*	2	0	11	20
07	4*	1	1	5	3*	1	1	4	10
08	21*	13*	11*	26	16*	10*	6*	22	29
09	11*	3	2	13	10*	2	1	13	20
10	26*	4*	1	26	23*	2	0	23	29
11	17*	6*	6*	20	13*	4*	4*	16	29
12	6*	2	1	8	4*	0	0	4	29
13	17*	1	5*	18	10*	0	1	10	29
14	8*	3	0	10	6*	0	0	6	19
Sum	119	36	33	140	94	21	13	109	214

a) number of cases where at least one of the three terms were significant
b) the number attached with an asterix exceeds the threshold that can be accepted given the level of significance and total number of tests

11. For UPOV characteristics 02, 08, 11 and 13 all combinations of country and years that were significant at the 5% level are shown graphically in Appendix 1. The figures show that even when the 3rd degree term of the polynomial was significant, the 2nd degree polynomial described the relationship almost as well in many cases. The exceptions seemed to be the relations between Log SD and the mean for characteristic 08 in DE 1997 and 1999 where the top around 50 to 55 days was clearly better described by the 3rd degree polynomial (blue line) than the 2nd degree polynomial (green line). In many cases the 3rd degree polynomial seemed to be rather sensitive to a few extreme observations with low or high means. The cubic spline was in most cases located between the 2nd and the 3rd degree polynomial. In some cases both the spline and the 2nd degree polynomial seemed to describe the relationship almost equally well. However, for some characteristics, such as characteristic 08 in DE 1999 the spline described the relationship better than both the 2nd and 3rd degree polynomial (better than the 2nd degree polynomial at means between 45 and 55 and better that the 3rd degree polynomial at means above 65).

Lolium perenne, Forage

12. For the Forage group, about 55% (159 out of 213 cases) of the examined combinations showed a significant relationship at the 5 % level for at least one of the 3 terms in the model. In most of these cases the linear term was also significant. The quadratic and cubic terms were also significant in several cases. However some of these may be expected to be type I errors. Looking at the tests for cubic relation in each characteristic, one finds that the number of tests for characteristic with UPOV number, 08 and 14 exceed the number needed for being overall significant at the 5 % level and 1% level given the total number of tests for the characteristics.

Table 6: Number of significant effects for Lolium perenne: Forage

UPOV no	Significant at the 5% level				Significant at the 1% level				Tests total
	Linear	Quadratic	Cubic	≥1[a]	Linear	Quadratic	Cubic	≥1[a]	
02	8*	4*	1	11	8*	2	1	10	20
08	20*	13*	12*	26	18*	10*	8*	24	30
09	10*	2	0	12	9*	2	0	11	23
10	28*	4*	3	29	28*	1	0	28	30
11	24*	2	1	25	20*	0	1	20	30
12	17*	1	2	19	9*	0	0	9	30
13	23*	1	3	25	17*	1	0	17	30
14	9*	1	5*	12	5*	1	3*	8	20
Sum	139	28	27	159	114	17	13	127	213

a) number of cases where at least one of the three terms were significant
b) the number attached with an asterix exceeds the threshold that can be accepted given the level of significance and total number of tests

13. For UPOV characteristics 08 and 14, all combinations of country and years that were significant at the 5% level are shown graphically in Appendix 2. The figures show that, even when the 3rd degree term of the polynomial was significant, the 2nd degree polynomial described the relationship almost as well as the 3rd degree polynomial in most cases. The exceptions seemed to be the relations between Log SD and the mean for characteristic 08 in the NL (and DE) in 1995 and 1999 where the top around 40 to 50 days was clearly better described by the 3rd degree polynomial (blue line) than the 2nd degree polynomial (green line). Likewise, the 3rd degree polynomial also described the top around 35-45 days better for the same characteristic in UK 1997. For characteristic 14, the 3rd degree polynomial described one variety with very few spikelets better than the 2nd degree polynomial (two different varieties). However, overall the 3rd degree polynomial did not seem to describe the relationship for characteristic 14 much better than the 2nd degree polynomial. Also here the cubic spline was most often located between the 2nd and the 3rd degree polynomial. In some cases the spline was clearly better than both polynomials, .e.g. characteristic 11 in UK 1997.

Brassica napus

14. For *Brassica napus,* about 80% (65 out of 81 cases) of the examined combinations showed a significant relationship at the 5 % level for at least one of the 3 terms in the model (Table 7). In most of these cases the linear term was also significant. The quadratic and cubic terms were also significant in several cases. However some of these may be expected to be type I errors. Looking at the tests for cubic relation in each characteristic, one finds that the number of tests for characteristic with UPOV number 02 and 03 exceed the number needed for being overall significant at the 5 % level or at the 1 % level given the total number of tests for the characteristics.

Table 7: Number of significant effects for *Brassica napus*

UPOV no	Significant at the 5% level				Significant at the 1% level				Tests total
	Linear	Quadratic	Cubic	≥1[a]	Linear	Quadratic	Cubic	≥1[a]	
02	5*	2	2	7	4*	1	2*	5	9
03	7*	2	3*	9	5*	0	1	6	9
06	5*	2	1	5	3*	2*	0	3	9
13	2	1	1	2	2*	1	1	2	9
14	5*	2	1	7	4*	0	0	4	9
17	5*	1	2	8	5*	1	0	6	9
18	9*	0	1	9	6*	0	0	6	9
19	9*	1	0	9	9*	1	0	9	9
20	9*	0	0	9	9*	0	0	9	9
Sum	56	11	11	65	47	6	4	50	81

[a)] number of cases where at least one of the three terms were significant
[b)] the number attached with an asterix exceeds the threshold that can be accepted given the level of significance and total number of tests

15. For UPOV characteristics 02 and 03, all combinations of country and years that were significant at the 5% or 1% level are shown in Appendix 3. The figures show that, even when the 3rd degree term of the polynomial was significant, the 2nd degree polynomial described the relationship almost as well as the 3rd degree polynomial in 2001, but in 2004 the 3rd degree polynomial described the relationship better than the 2nd degree polynomial. Again the cubic spline was in most cases located between the 2nd and the 3rd degree polynomial

Pisum sativum

16. For *Pisum sativum,* about 70% (38 out of 54 cases) of the examined combinations showed a significant relationship at the 5 % level for at least one of the 3 terms in the model. In most of these cases the linear term was also significant. The quadratic and cubic term was also significant in some cases. However, some of these may be expected to be type I errors. Looking at the tests for cubic relation in each characteristic, the number of significant cases did not exceed the number needed for being overall significant at the 5 % level given the total number of tests for any of the characteristics.

Table 8: Number of significant effects for *Pisum sativum*

UPOV no	Significant at the 5% level				Significant at the 1% level				Tests total
	Linear	Quadratic	Cubic	≥1[a]	Linear	Quadratic	Cubic	≥1[a]	
12	8*	0	0	8	7*	0	0	7	9
13	7*	2	0	8	5*	0	0	5	9
31	3*	1	2	5	2*	0	0	2	9
32	9*	0	1	9	9*	0	0	9	9
48	0	1	0	1	0	0	0	0	9
49	4*	5*	1	7	3*	1	1	4	9
Sum	31	9	4	38	26	1	1	27	54

a) number of cases where at least one of the three terms were significant
b) the number attached with an asterix exceeds the threshold that can be accepted given the level of significance and total number of tests

17. For the characteristics with UPOV numbers 31, 32 and 49 all combinations of country and years that were significant at the 5% level are shown in Appendix 4. The figures show that, even when the 3rd degree term of the polynomial was significant, the 2nd degree polynomial described the relationship almost as well as the 3rd degree polynomial except for one or two very extreme values at the end of the scale. It should be considered whether such extreme varieties should be considered as comparable reference varieties for the candidate in question. Again the cubic spline was most often located between the 2nd and the 3rd degree polynomial.

Discussion and conclusions

18. For most cases a model with a linear and quadratic effect described the relation between Log SD and the mean sufficiently well. In many of the cases where the preferred model included a significant cubic term, this seemed to be caused by a few unusual varieties. The 3rd degree polynomial fit seemed to be strongly influenced by extreme standard deviations particularly for varieties with either low or high means. In such cases it might be questioned whether a model that displaying such sensitivity should be used or whether such extreme varieties should be left out of the model fitting. The cubic spline applied seemed to be less influenced by extreme observations and seemed to describe the relationship at least as well as the 2nd and 3rd degree polynomial. In most cases the cubic spline (with 4 degrees of freedom set) was in most cases located between the 2nd and 3rd degree polynomial. However the cubic spline would be a little more difficult to implement than the polynomial regressions and some technical challenges remain.

APPENDIX 1

APPENDICES:

Figures showing the relations between log SD and mean for combinations of country, years where the 3rd degree polynomial described the relation significant better than the 2nd degree polynomial. There is a figure for each characteristic where the threshold for the number of significant tests was passed.

Legends:

Black circles:	Plot of individual Log SD versus mean for each variety in the given country and year
Purple line:	Fit for the relation between log SD and mean using a straight line
Green line:	Fit for the relation between log SD and mean using a 2nd degree polynomial
Blue line:	Fit for the relation between log SD and mean using a 3rd degree polynomial
Red line:	Fit for the relation between log SD and mean using a cubic spline with 2 degrees of freedom

APPENDIX 1

<u>Appendix I, Amenity</u>

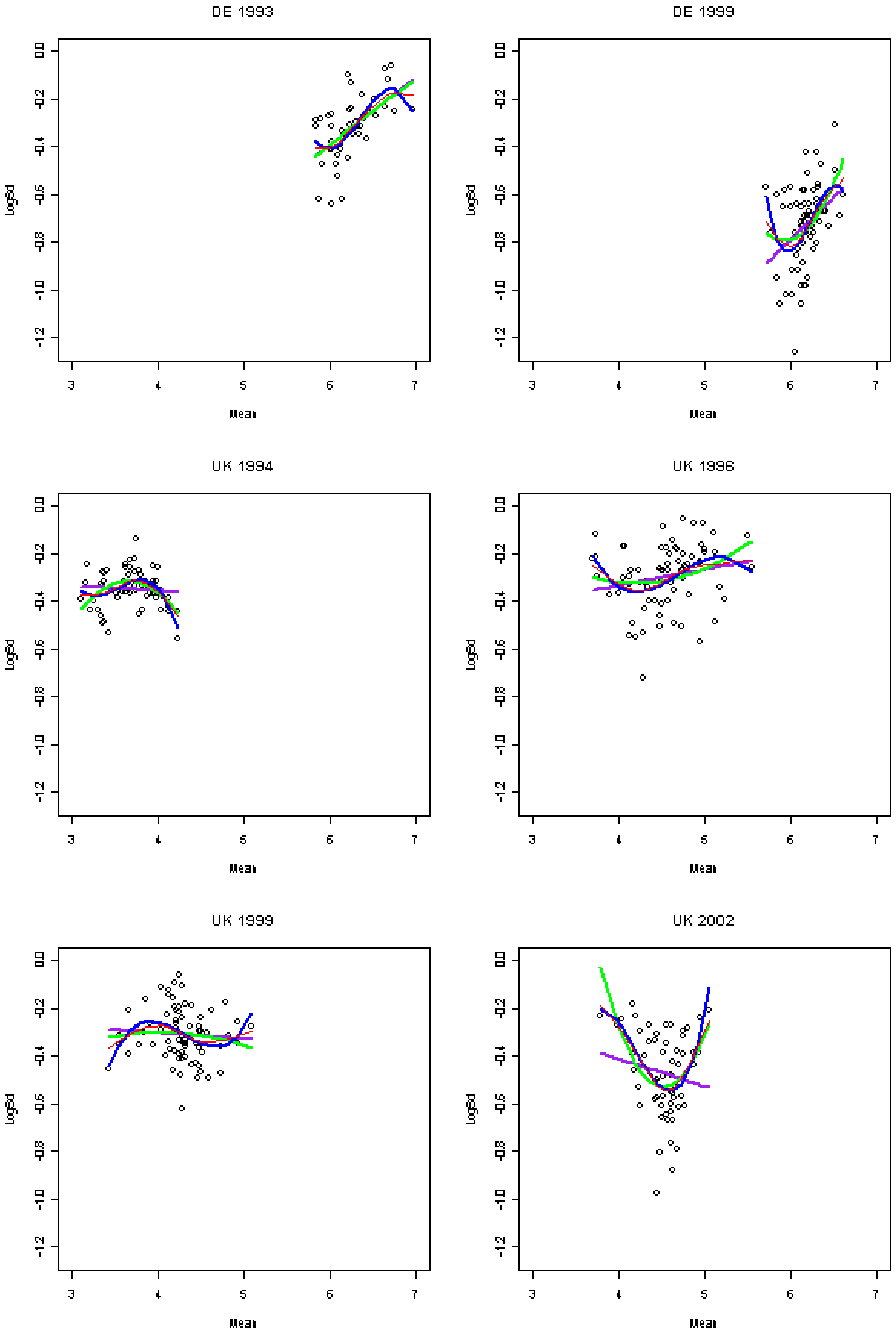

Fig 1.1 Amenity characteristic 02

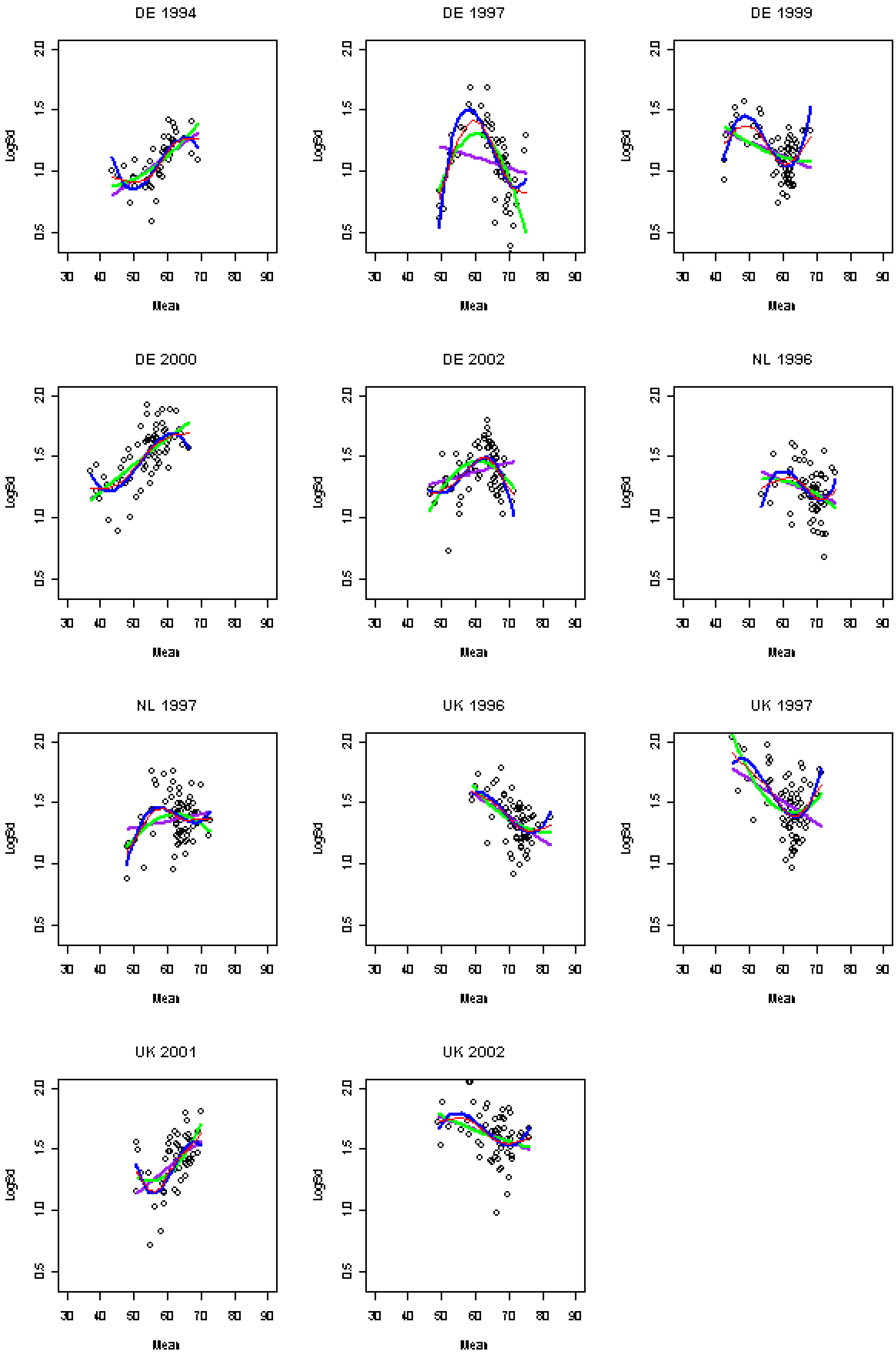

Fig 1.2 Amenity characteristic 08

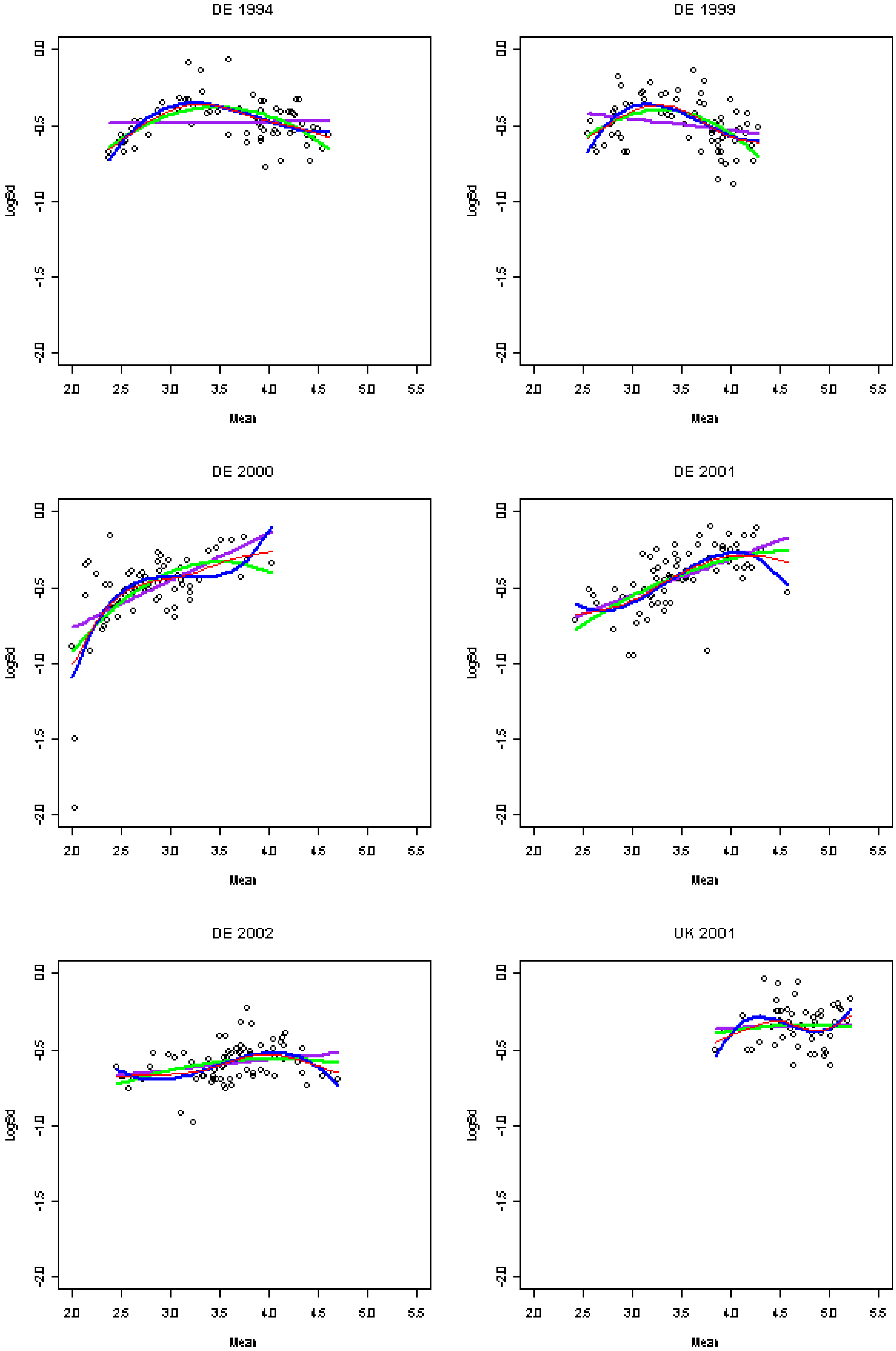

Fig 1.3 Amenity characteristic 11

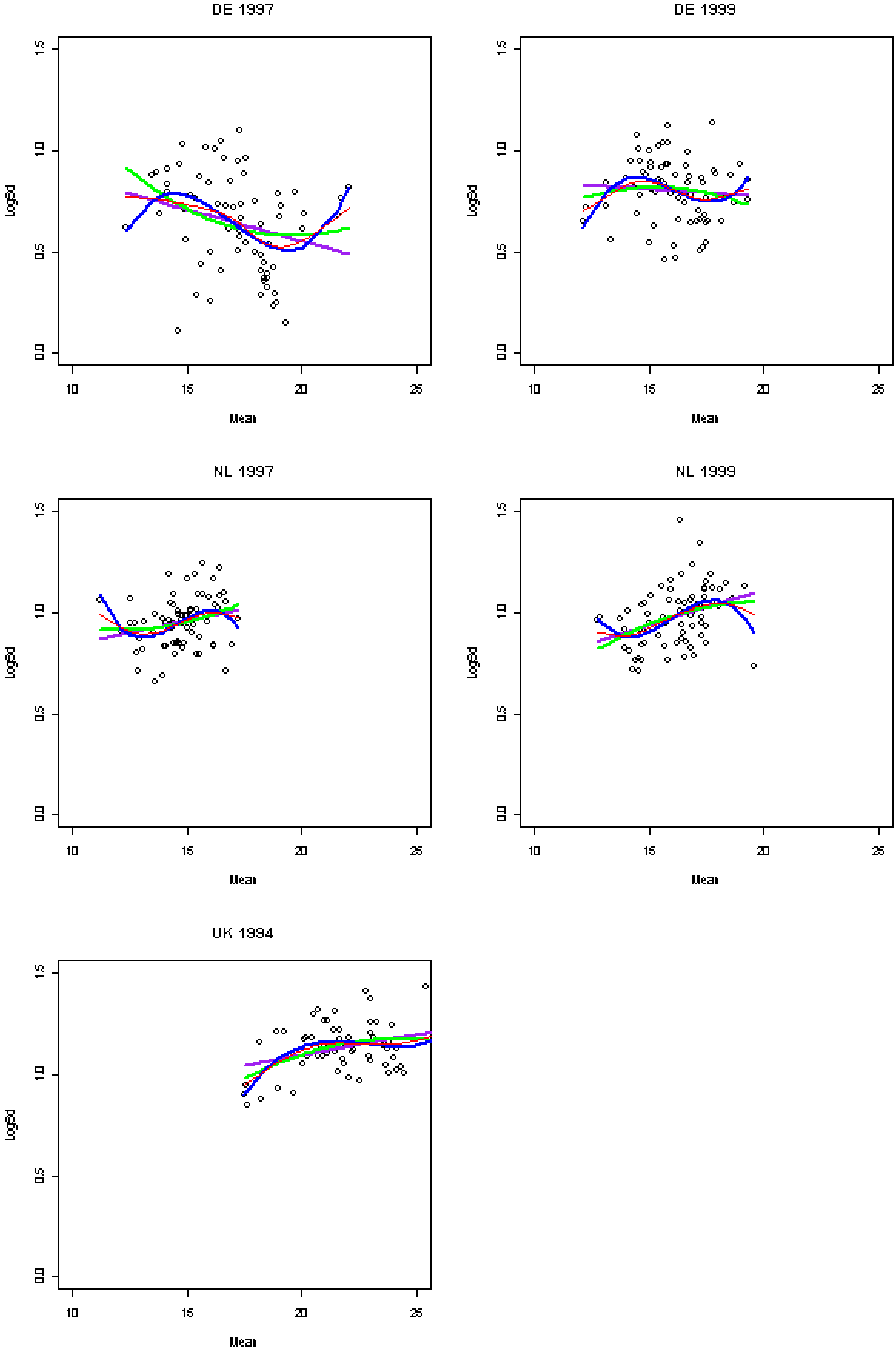

Fig 1.3 Amenity characteristic 11

APPENDIX 2

Appendix 2, Forage

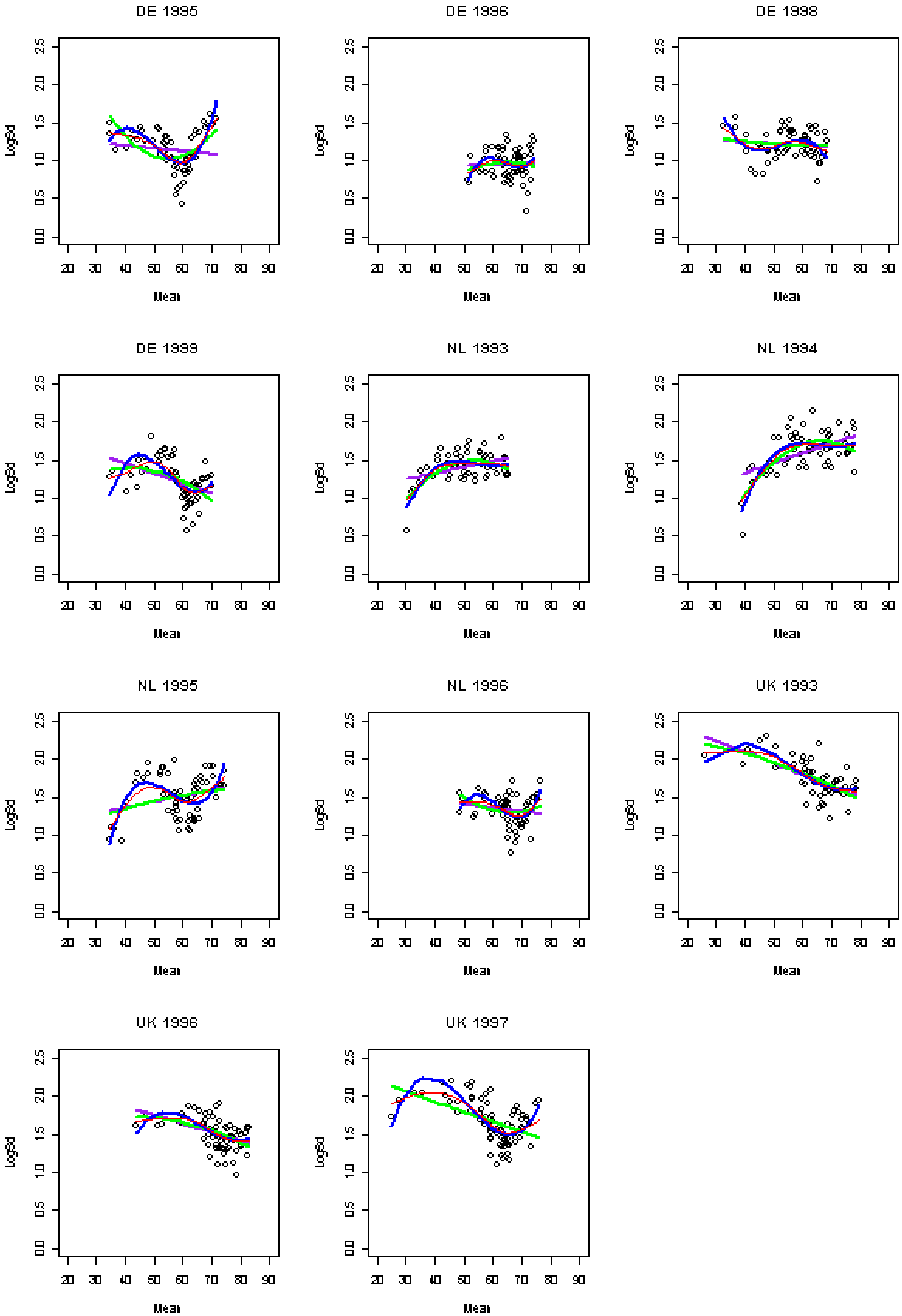

Fig 2.1 Forage characteristic 08

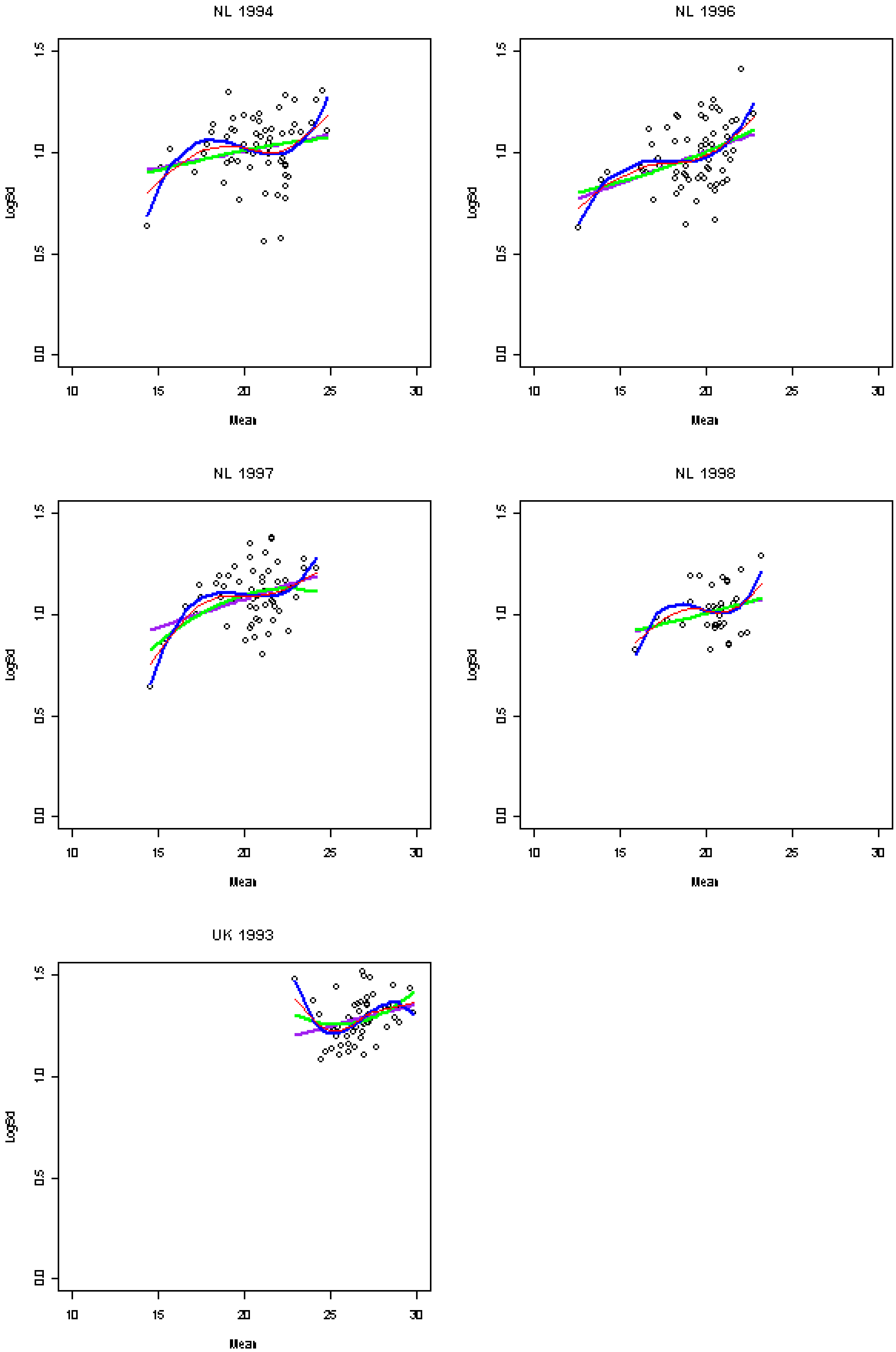

Fig 2.2 Forage characteristic 14

Appendix 3: Brassica napus

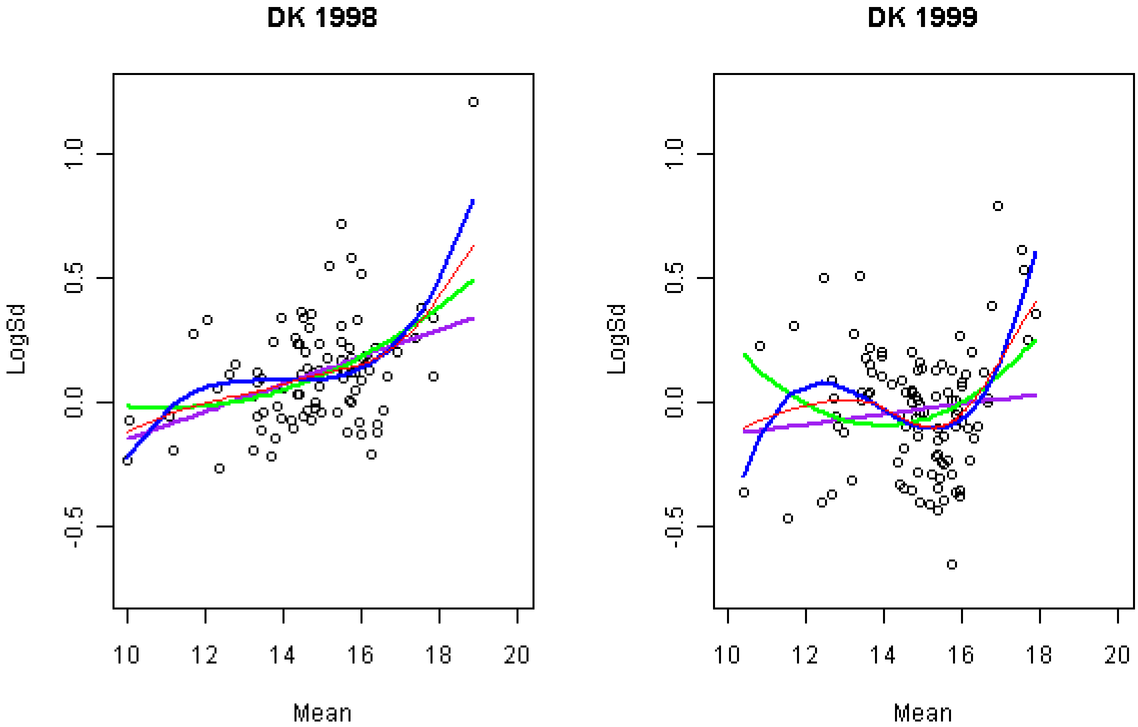

Fig 3.1 Brassica characteristic 02

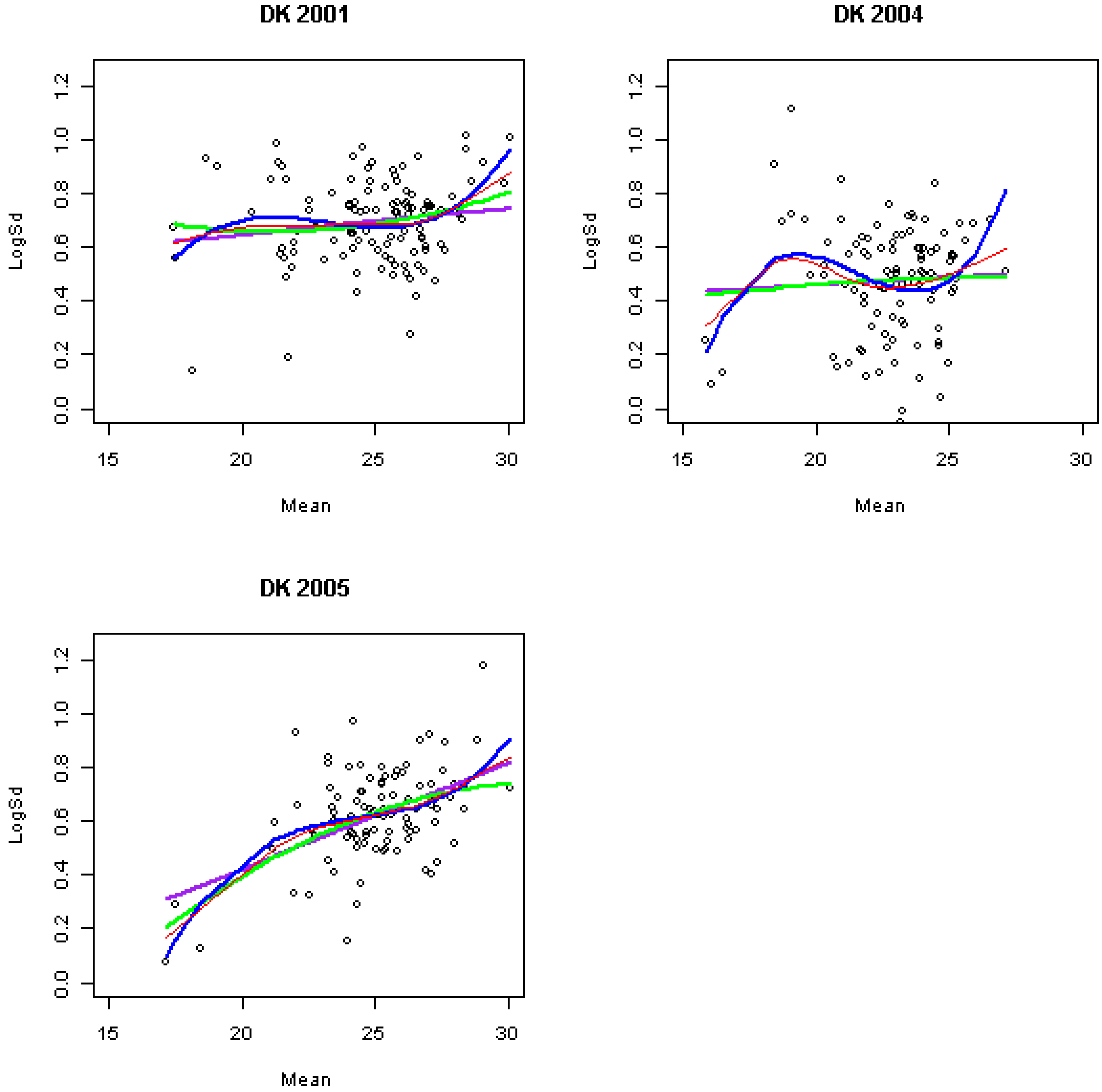

Fig 3.2 Brassica characteristic 03

APPENDIX 4

Appendix 4, Pisum sativum

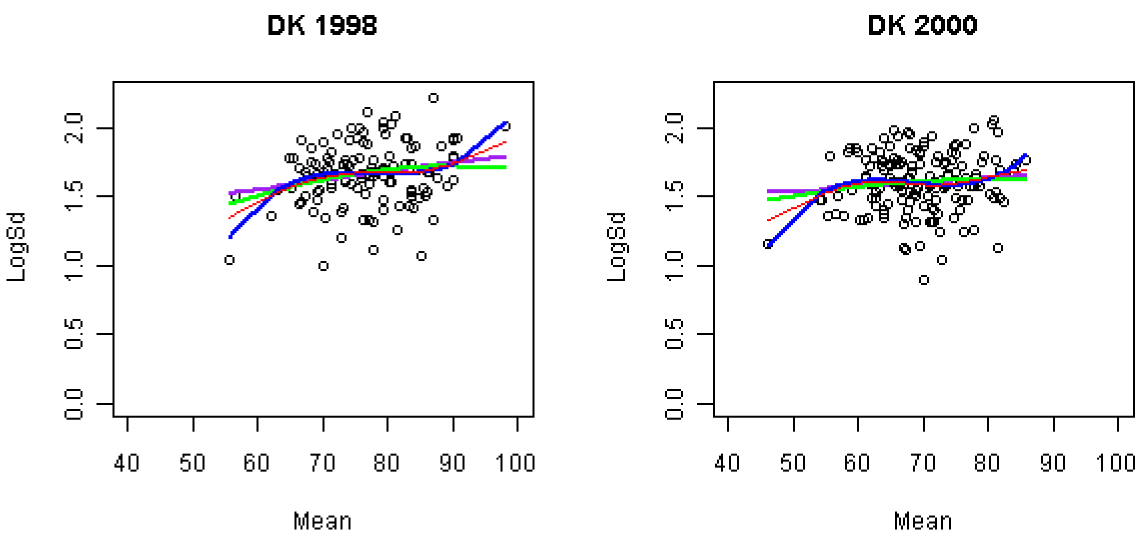

Fig 4.1 Pisum characteristic 31

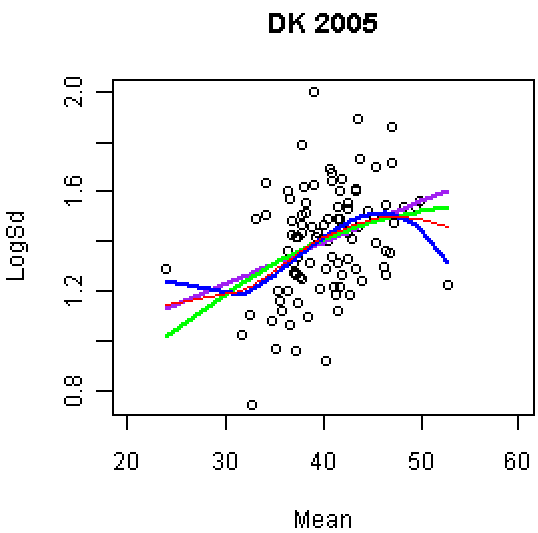

Fig 4.2 Pisum characteristic 32

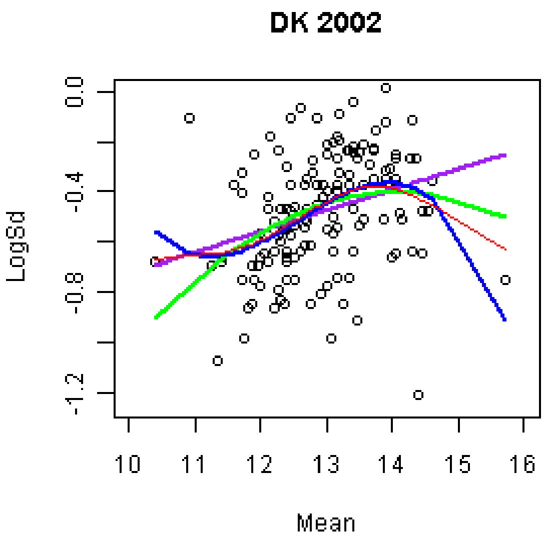

Fig 4.3 Pisum characteristic 49

Authors: Kristian Kristensen and Adrian Roberts

[End of document]