Environment Agency

Evidence

Spatiotemporal modelling of nitrate and phosphorous for river catchments

Report – SC080041/R

The Environment Agency is the leading public body protecting and improving the environment in England and Wales.

It's our job to make sure that air, land and water are looked after by everyone in today's society, so that tomorrow's generations inherit a cleaner, healthier world.

Our work includes tackling flooding and pollution incidents, reducing industry's impacts on the environment, cleaning up rivers, coastal waters and contaminated land, and improving wildlife habitats.

This report is the result of research commissioned and funded by the Environment Agency.

Published by:
Environment Agency, Horizon House, Deanery Road, Bristol, BS1 5AH
www.environment-agency.gov.uk

ISBN: 978-1-84911-285-7

Further copies of this report are available from:
The Environment Agency's National Customer Contact Centre by emailing:
enquiries@environment-agency.gov.uk
or by telephoning 08708 506506.

Author(s):
Adrian Bowman, Claire Ferguson, Duncan Lee, Ana-Maria Magdalina, E. Marian Scott

Dissemination Status:
Publicly available

Keywords:
Spatiotemporal trends, seasonality, space/time interactions, catchment covariates

Research Contractor:
University of Glasgow

Environment Agency's Project Manager:
Dave Johnson (office)

Collaborator(s):
Chris Burgess, Linda Pope, Robert Willows

Project Number:
SC080041

Product Code:
LIT 7375

Evidence at the Environment Agency

Evidence underpins the work of the Environment Agency. It provides an up-to-date understanding of the world about us, helps us to develop tools and techniques to monitor and manage our environment as efficiently and effectively as possible. It also helps us to understand how the environment is changing and to identify what the future pressures may be.

The work of the Environment Agency's Evidence Directorate is a key ingredient in the partnership between research, guidance and operations that enables the Environment Agency to protect and restore our environment.

This report was produced by the Scientific and Evidence Services team within Evidence. The team focuses on four main areas of activity:

- **Setting the agenda**, by providing the evidence for decisions;
- **Maintaining scientific credibility**, by ensuring that our programmes and projects are fit for purpose and executed according to international standards;
- **Carrying out research**, either by contracting it out to research organisations and consultancies or by doing it ourselves;
- **Delivering information, advice, tools and techniques**, by making appropriate products available.

Miranda Kavanagh

Director of Evidence

Executive summary

The development of hypothesis-driven trend models for nutrients in hydrological catchments in England and Wales is required to provide better evidence for future nutrient management policy. Such models would enable policy decisions to be determined based on combined information from the catchment rather than on an individual site basis.

Statistical models are presented for both nitrate and phosphorus in river networks for 59 large hydrological areas (LHAs) in England and Wales to assess spatial and temporal trends along with seasonal patterns over the past 20–40 years. As these nutrients have similar sources (which include agriculture and population), a common modelling approach has been implemented. The models developed here are nonparametric, and enable flexible smooth functions to be fitted for trends and seasonality without constraining relationships to be linear.

Concentrations of orthophosphate-P (OP) have generally decreased over time for many of the LHAs with the exception of the combined LHA in Wales where the trend was fairly flat over recent years. In general, the seasonal pattern indicated low values at the beginning and end of the year, with one peak in the summer months. In a variety of LHAs, however, a trough was also evident in the spring along with the peak appearing later in the summer months.

For total nitrogen (TN) and total oxidised nitrogen (TON), trends over time and seasonal patterns are very similar across all LHAs in England and Wales. Over the past 10 years, total nitrogen and total oxidised nitrogen concentrations have generally been decreasing or been fairly constant. But prior to 2000, concentrations were generally on an upward trend. There are also several LHAs in, for example, the Environment Agency's Anglian and Southern Regions where very little trend is evident. Two LHAs in South West Region show that levels increased and then appear to have levelled off. The seasonal pattern generally highlights high values at the beginning and end of the year with one trough in the summer months.

There appeared to be more variation in trends for total oxidised nitrogen and total nitrogen across the LHAs than for orthophosphate-P. However, the seasonal patterns were more varied for orthophosphate-P.

Within each LHA, monitoring locations at the bottom of the catchment generally appear to have higher levels of total nitrogen and total oxidised nitrogen than those further up the catchments. The distribution of estimates across LHAs for total nitrogen and total oxidised nitrogen appear similar. However, the distribution of spatial estimates is different for total oxidised nitrogen and total nitrogen on the coast of the Lune LHA and in the south of the Severn LHA. The distribution of orthophosphate-P estimates is similar to that for total nitrogen and total oxidised nitrogen with the higher values for orthophosphate-P, total nitrogen and total oxidised nitrogen generally appearing to be clustered around the same locations. However, the range of values for the highest estimated orthophosphate-P concentrations, relative to the mean of each LHA, is larger. There appear to be more high values, relative to the mean, than for total oxidised nitrogen or total nitrogen indicated along the north-west coast of England with different spatial distributions evident for orthophosphate-P in, for example, the Severn, Hull, Ancholme, Great Ouse, Tone, Parrett and Frome/Bristol Avon LHAs.

Models that can help to explain possible catchment-scale influences on these trends such as different land uses, human and agricultural sources of nitrate and variations in rainfall have also been investigated. Data for these variables have been collated and aggregated to appropriate spatial and temporal scales. Models were fitted for LHAs with contrasting signals in terms of trends over time, that is, the Severn, Lune and

Trent LHAs (the Trent catchment is an area of high nitrate pollution and a Nitrate Vulnerable Zone).

The same nonparametric modelling framework was implemented as for the first analysis with relationships not constrained to be linear. The models highlighted that for 66, 75 and 54 per cent respectively of the variability for orthophosphate-P in the Severn, Lune and Trent LHAs could be explained using only the catchment covariate information; the corresponding figures were 52, 68 and 39 per cent respectively for total oxidised nitrogen and 64, 74 and 39 per cent respectively for total nitrogen.

Overall, the covariates are reasonably powerful and explain much of the patterns seen in the data. The relationships between each covariate and the nutrient levels are complex and different for each LHA and there does not seem to be one or two variables that explain the majority of the variability – many of the covariates contribute a small proportion each. Simplifying the models indicates that, in general and in no particular order, the variables baseflow index (BFI), flow, population, rainfall, land cover and the group made up of Agricultural Land Classification (ALC), slope and soil explain a moderate to large proportion of the overall variability that is explained.

Finally, the modelling approach used for the trends and covariates was extended to incorporate interactions between variables. This model is illustrated on one example LHA (Coquet, Wansbeck, Blyth) with interaction terms included in the model to investigate changes in spatial and seasonal patterns across time.

Acknowledgements

The project team at the University of Glasgow would like to thank Robert Willows, Chris Burgess, Linda Pope and David Johnson at the Environment Agency for their help and advice throughout this research project. The team is also extremely grateful to Chris Burgess for his time and assistance collating and aggregating covariate data.

Contents

1 Introduction

Previous modelling of nutrients in English and Welsh rivers has been carried out at individual monitoring locations in order to investigate trends over time and the effect of the contributing land at these locations. The development of hypothesis-driven trend models that incorporate information spatially within hydrological catchments is required to provide better evidence for future nutrient management policy. Such models would enable policy decisions to be determined based on combined information from the catchment rather than on an individual site basis.

This report presents models for both nitrate and phosphorus to assess spatial and temporal trends in river nutrient concentrations, along with seasonal patterns over the past 20–40 years. Since these nutrients have similar sources (including agriculture and population), a common modelling approach has been implemented. Because the models are nonparametric, flexible smooth functions can to be fitted without constraining relationships to be linear.

There is also a need to understand how the historical patterns identified for the water quality in a river are a consequence of catchment-scale influences such as different land uses, human and agricultural sources of nitrate, and variations in rainfall. This will again enable policy decisions to be determined based on information from catchments rather than on an individual site basis.

The overall objective of the project was to answer two questions:

- How has the overall pattern of nutrient concentrations in UK surface waters changed over the past 20–40 years?
- Can the spatiotemporal distribution of nutrient concentrations in surface waters within and between catchments be explained by catchment covariate information?

Specifically, it was of interest to develop spatiotemporal models within hydrological catchments that would allow analyses of nutrient trends on a catchment-wide basis (rather than individual sites) and within this to identify the effect of seasonal variation on observed concentrations. Co-variables and factors that can explain sources of spatial and temporal variation in nutrient concentrations have been collated and used to describe trends over time and space.

1.1 Report structure

Section 2 details the response data, sample types, river networks and spatial scales used. It describes the covariate data in terms of the physical characteristics of the land, land use, land cover, fertiliser, rainfall, flow and population that have been collated during the project. Full details are given on data aggregations within specific water bodies and to incorporate contributing land.

Section 3 describes the data transformations and manipulations required together with details of the three different models developed to investigate trends and seasonality, incorporate covariates and incorporate space/time interactions, respectively. The results of the statistical modelling are presented in section 4 and discussed in section 5 for each model. Finally, section 6 highlights the possibilities for future work.

Detailed technical information is presented in Appendix A and further results are given in Appendix B.

2 Data

This section describes the data used in this project. Details are given of:

- phosphate and nitrate determinand codes;
- types of samples;
- river network structure;
- data manipulations;
- covariate data.

2.1 Phosphate and nitrate codes

The Environment Agency provided the project team at the University of Glasgow with data for nitrate and phosphate from its Water Management Information System (WIMS) for its Anglian, Midlands, North East, North West, Southern, South West, Thames and Wales Regions. Tables 2.1 and 2.2 present the phosphate codes and nitrate codes, respectively, used within WIMS.

Table 2.1 Phosphate codes

Code	Description
0180[1]	Orthophosphate – as P in mg/l
0192	Phosphate as P in mg/l
0348	Phosphorous total – as P in mg/l
0346	Phosphorous dissolved as P in mg/l
9449	Phosphate – as P in mg/l
9856	Orthophosphate filtered – as P in mg/l

Notes: [1]Code used for phosphate

Table 2.2 Nitrate codes

Code	Description
0111[1]	Ammonia – as N in mg/l
0116[1]	Total oxidised nitrogen – as N in mg/l
0117[1]	Nitrate – as N in mg/l
0118[1]	Nitrite – as N in mg/l
0119	Ammonia un-ionised (calculated) as N in mg/l
9853	Nitrate filtered – as N as N in mg/l
9993	Ammonia filtered – as N in mg/l

Notes: [1]Codes used for nitrate

It was agreed that the determinand codes 0180 Orthophosphate-P (OP), 0111 Ammonia, 0116 Total oxidised nitrogen (TON), 0117 Nitrate and 0118 Nitrite would be used in this project with the nitrate codes combined as followings:

- Total oxidised nitrogen (TON) = 0116
- Total nitrogen (TN) = 0111 (Ammonia) +0116 (TON)

Determinand codes 0117 and 0118 were used when 0116 was not available; for full details of the data aggregations, see section A1 of Appendix A.

2.2 Sampling types

The original nutrient response data contained 12 sampling code types. These are listed along with a description of each type in Table 2.3.

It was agreed that the sampling types to be used were F1–F6, FC and GC.

Table 2.3 Sampling types

Sampling code	Description
F1[1]	FRESHWATER - RQO RE1
F2[1]	FRESHWATER - RQO RE2
F3[1]	FRESHWATER - RQO RE3
F4[1]	FRESHWATER - RQO RE4
F5[1]	FRESHWATER - RQO RE5
F6[1]	NON CLASSIFIED RIVER POINTS
FA	LAKES/PONDS/RESERVOIRS
FB	FRESHWATER - RIVER TRANSFER
FC[1]	COMPARATIVE INLET POINTS
FD	RIVER AUGMENTATION
GC[1]	WATER FOR POTABLE SUPPLY - RIVER ABSTRACTION
BB	GROUNDWATER - SPRING

Notes: [1]Sampling code used as surface water.

2.3 River network

It was agreed to use the General Quality Assessment (GQA) network as the source of the monitoring locations to be included in the analysis. These locations were associated with Water Framework Directive (WFD) water bodies within each of the 59 large hydrological areas (LHAs) across England and Wales. The LHAs are identified by name in Figure 2.1 and by number in Figure 2.2.

The Environment Agency also provided the tree structure for large hydrological areas which defines the contributing land for each monitoring location; see section A2 of Appendix A for details on small changes to the tree structure.

The monitoring period varies for each determinand by LHA (see Table 2.4), with every time series invariably ending in 2009. Many of the areas in Wales contain very few monitoring locations and hence have been combined together for the modelling. Full details are provided in section 3.

Figure 2.1 Large hydrological areas in England and Wales by name

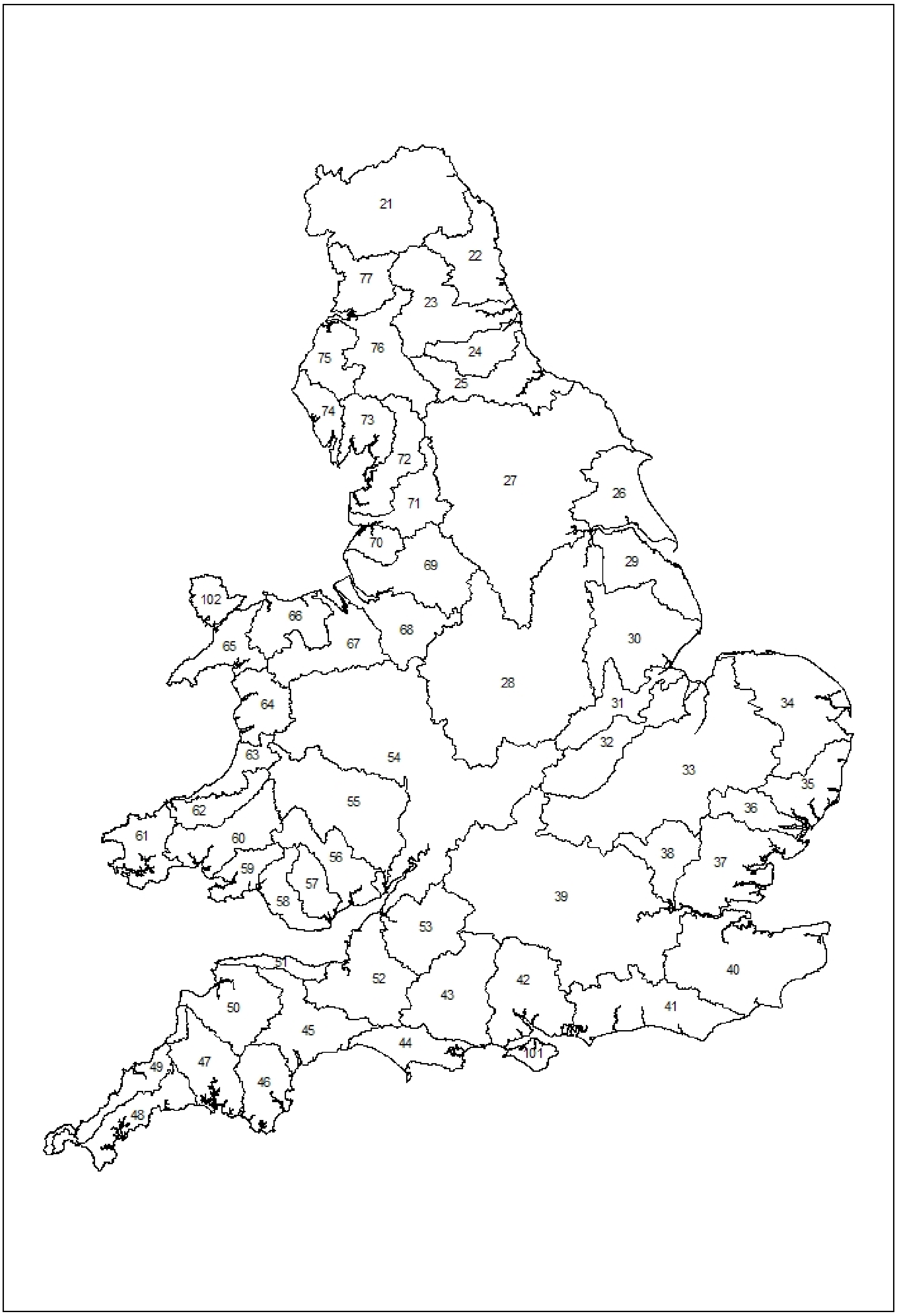


Figure 2.2 Large hydrological areas in England and Wales by number

Table 2.4 Time periods for each determinand in each large hydrological area

LHA number	OP	TON	TN
21	[1]	1994–2009	1994–2009
22	1997–2009	1994–2009	1994–2009
23	[1]	1994–2009	1994–2009
24	1995–2009	1994–2009	1994–2009
25	1994–2009	1994–2009	1994–2009
26	1974–2009	1974–2009	1974–2009
27	1973–2009	1973–2009	1973–2009
28	1986–2009	1986–2009	1986–2009
29	1988–2009	1981–2009	1981–2009
30	1983–2009	1981–2009	1981–2009
31	1981–2009	1981–2009	1981–2009
32	1981–2009	1981–2009	1981–2009
33	1981–2009	1981–2009	1981–2009
34	1981–2009	1981–2009	1981–2009
35	1981–2009	1981–2009	1981–2009
36	1981–2009	1981–2009	1981–2009
37	1978–2009	1978–2009	1978–2009
38	1974–2009	1974–2009	1974–2009
39	1972–2009	1972–2009	1972–2009
40	1974–2009	1974–2009	1974–2009
41	1971–2009	1971–2009	1971–2009
42	1977–2009	1977–2009	1977–2009
43	1967–2009	1964–2009	1964–2009
44	1974–2009	1974–2009	1974–2009
45	1974–2009	1974–2009	1974–2009
46	1974–2009	1974–2009	1974–2009
47	1973–2009	1973–2009	1973–2009
48	1974–2009	1974–2009	1974–2009
49	1975–2009	1974–2009	1974–2009
50	1974–2009	1974–2009	1974–2009
51	1974–2009	1974–2009	1974–2009
52	1985–2009	1985–2009	1985–2009
53	1966–2009	1966–2009	1966–2009
54	1971–2009	1967–2009	1967–2009
[2]55, 56, 57, 58, 59, 60, 61, 62, 63	[1]	1975–2009	1975–2009
64	[1]	1981–2009	1981–2009
[2]65,66,102	[1]	1980–2009	1980–2009
67	1986–2009	1983–2009	1983–2009
68	1974–2009	1974–2009	1974–2009
69	1974–2009	1953–2009	1953–2009
70	1974–2009	1971–2009	1971–2009
71	1971–2009	1969–2009	1969–2009
72	1971–2009	1963–2009	1971–2009
73	1967–2009	1954–2009	1956–2009
74	1971–2009	1965–2009	1965–2009
75	1975–2009	1975–2009	1975–2009
76	1972–2009	1972–2009	1972–2009
77	1975–2009	1975–2009	1975–2009
101	1979–2009	1979–2009	1979–2009

Notes: [1]Time periods missing because these areas had too many observations below the detection limit. They have therefore been omitted from the analysis.
[2]These areas were combined together to enable a spatial surface to be estimated.

2.4 Covariates

Tables 2.5 to 2.7 contain a summary of the covariate data collated, manipulated and aggregated as part of this project. These tables provide:

- a description of each variable;
- information on whether the variable is collected for an area (in km^2) or for a Water Framework Directive (WFD) water body;
- if there is information available over different years.

The covariates have been aggregated to a WFD water body. Details of these calculations are included in the tables, along with information on aggregations that include the contributing land for each water body.

Slope and ALC (Agricultural Land Classification) were originally compositional variables but have been converted here to categorical variables. Details of the aggregations are provided in Table 2.5; see section A3 of Appendix A for an explanation of the reasons for this.

For land cover, land use and population, the data have been standardised by:

- a geographical information system (GIS) based estimate of the local water body contributing area; or
- the size of the whole contributing area for variables that include the contributing land.

See Tables 2.6 and 2.7 for details of the standardisation.

An alternative approach would have been to standardise by the hydrologically effective rainfall. Some of the variables are hectares per water body for land cover and land use (crops) and these variables have been re-calculated to proportions per water body by dividing their values by 100 times the area of the corresponding water body (the water body area is given in km^2).

For a few of the variables (land cover, land use and population), data are available for several years but not for all years; see Table 2.6 for details. For these variables the data have been interpolated to complete the missing years. For example, if 1990 and 2000 are known, the years 1990–1995 use the 1990 data and the years 1996–2000 use the data from 2000.

For further details on covariates, see section A3 of Appendix A.

Table 2.5 Covariate files for physical characteristics of the contributing land

Variable	Description	Resolution	Aggregations	Contributed land
Soil	Soil group (km^2) (Heavy, Medium, Light-Medium,…,Unclassified)	SPATIAL	Soil group for the water body is taken to be the dominant soil group in the water body.	Soil group for the water body and contributing land is taken to be the dominant soil group in the whole area.
Slope	% of 1 km^2 grid Slope <3 (gentle) 3 <Slope <7 (moderate) 7 <Slope <12 (steep) Slope >12 (very steep)	SPATIAL	Sum each category across km^2 in each water body and take the category with the maximum value as representative of the water body	Sum each category across km^2 in each water body and contributing land and take the category with the maximum value as representative of the whole area.
ALC	% of 1 km^2 grid ALC1 ALC2 ALC3 ALC4 ALC5 Non-agricultural land	SPATIAL	Sum each category across km^2 in each water body and take the category with the maximum value as representative of the water body.	Sum each category across km^2 in each water body and contributing land and take the category with the maximum value as representative of the whole area.

Table 2.6 Covariate files for land cover, land use and fertiliser

Variable	Description	Resolution	Aggregations	Contributed land
Land cover Defra (WFD water body)	*Hectares of:* Arable Woodland Rough grazing Permanent grass	SPATIAL and TEMPORAL (yearly) 1969, 1981, 1987, 1990, 1995, 2000, 2005, 2007, 2008	Standardise by dividing by 100 times the water body size.	Standardise by dividing by 100 times the size of water body and contributing land (or divide by hydrologically effective rainfall).
Land use (WFD water body)	*Hectares of:* Cereals Potatoes Maize Sugar beet Stock feed Field vegetables Other arable Oilseed rape	SPATIAL and TEMPORAL (yearly) 1987, 1990, 1995, 2000, 2005, 2007, 2008 (for crops also 1969, 1981)	Standardise crops by dividing by 100 times the water body size.	Standardise crops by dividing by 100 times the size of water body and contributing land (or divide by 100 times the hydrologically effective rainfall).
	Counts of: Broad livestock (cows, poultry, pigs, sheep, other animals)		Standardise livestock by dividing by the water body size.	Standardise livestock by dividing by the size of the water body and contributing land (or divide by the hydrologically effective rainfall).
Fertiliser	Fertiliser application rates	TEMPORAL (yearly)		

Notes: Not all the land use covariates are present for each year. Maize is missing from the early Defra data (1969 and 1981) and oilseed rape is missing for 2000.
There are no livestock counts available for 1969 and 1981, with poultry missing for 1990 and 1995.

Table 2.7 Covariate files for rainfall, baseflow index, flow and population

Variable	Description	Resolution	Aggregations	Contributed land
Rainfall	Daily total rainfall at monitoring gauges	SPATIAL and TEMPORAL 1980–2009	Longest time series in water body used, aggregated to monthly total.	[1]
Baseflow index (BFI)	per km^2 (NSRI)[2] long-term average (1961–1990)	SPATIAL		Mean BFI from all contributing 1 km^2 grids.
Flow/ discharge	per WFD water body in m^3s^{-1} long-term average (1961–1990)	SPATIAL		
Population	Total annual population in WFD water body	SPATIAL and TEMPORAL 1981, 1985, 1986, 1990, 1991, 1995, 2000, 2001, 2005, 2008	Standardise by dividing counts per water body by the water body size.	Standardise by dividing counts per water body and contributing land by the size of the whole area (or divide by hydrologically effective rainfall).

Notes: [1]The total rainfall from the entire upstream area is aggregated using the gauged site data and long-term average rainfall. For further details, see section A3 in Appendix A for a full description of the process.
[2] National Soil Resources Institute

3 Statistical modelling

This section describes the statistical modelling approaches used to investigate and describe the trends in the nutrient data.

3.1 Transformations

A transformation was required for TN, TON and OP to stabilise the variability in the data. In all cases a natural log transformation (ln) was applied because:

- this transform appeared the most appropriate in stabilising the variance;
- the data are assumed to come from a lognormal distribution or to be normally distributed after applying the transformation.

For consistency the same transformation was applied in each case.

3.2 Detection limits

The proportion of non-detects (that is, measurements that are flagged as being below the limit of detection) was assessed for each determinand in each LHA.

- If the proportion of non-detects was small (that is, less than 3 per cent), the less-than values were substituted by the recorded value.
- If the proportion of non-detects was large (that is, greater than 50 per cent), then the LHA was not analysed for that determinand.
- For sites with a proportion of non-detects between 3 and 50 per cent, imputation is a more appropriate method than simple substitution (Helsel 2005). For determinands in LHAs with this proportion of non-detects, the values were imputed (replaced by a simulated value) using the methods detailed in section A4 of Appendix A.

3.3 Spatiotemporal modelling

Three models were investigated.

- Model 1 was structured around each of the 59 principal hydrological catchments (LHAs) to examine the temporal and spatial trends and seasonality within each catchment using data at a calendar date.
- In Model 2, the covariates listed in Tables 2.5 to 2.7 were used in an attempt to explain the trends in TON, TN and OP by working with WFD water bodies as the units of measurement and aggregating response data to monthly.
- Model 3 develops the methodology implemented in Models 1 and 2 to incorporate space/time interactions and an appropriate covariance structure. Because incorporating space/time interactions and spatiotemporal covariance into these models is difficult (especially with the data dimensions involved), this model was developed for one example LHA.

The same, underlying statistical modelling framework is employed for each model. Regression models are used in which the explanatory variables are incorporated as smooth functions instead of linear relationships. The benefit of such an approach is the flexibility to model smooth trends in space and time along with flexible non-restricted seasonal patterns. Data can be incorporated for a number of monitoring locations to estimate the trends and seasonal patterns that appear, on average, across the area of the monitoring locations. The approach makes fewer assumptions about the nature of trends and seasonal patterns within the data compared with a parametric approach and the methods are more robust to the presence of outliers in the data. However, it can become computationally intensive for data that are highly dimensional. These techniques are data driven and additional smooth or linear covariates for environmental factors can be incorporated, though there are still choices to be made about the amount of smoothing for each function. See sections A6.1 and A6.2 in Appendix A for further details.

3.3.1 Model 1

The model estimates:

- an overall mean for each area (LHA) of interest;
- a smooth surface for space (monitoring location);
- a smooth trend for time;
- a smooth seasonal pattern throughout the year.

Further details are provided in section A6.1 of Appendix A.

For many of the LHAs, all the data can be used for the model For the larger LHAs, however, the data dimensionality is a complication.

The models used here can be applied for data in time and space up to a dimension of around 30,000 observations. While the models can be applied for slightly higher dimensions than this, the computational time increases to several hours per model. For individual models this is manageable. However, alternative approaches are required in order to fit a substantial number of models.

In situations where the data dimension is >30,000 and has to be reduced, both the temporal and spatial dimensions have been reduced by constructing grids across both time and space. As a result of the irregular nature of the data, this grid construction has to be considered for each LHA individually and an effort has been made to reduce the data to as close to 30,000 observations (across time and space collectively) as possible in each case.

Typical grids involve taking 500 sites across space and 1,500 time points, with these values modified, if required, to ensure the maximum possible amount of data is used. See section A5 of Appendix A for more details, examples and a list of the LHAs for which gridding was performed.

The modelling approach adopted was to use all available data, wherever possible. Therefore, for LHAs with a data dimension of approximately 30,000 observations or less all data were used and gridding was not performed.

To apply the model in each LHA, the following criteria were used:

- Extract sampling types F1–F6, FC and GC (see Table 2.3).
- Eliminate rows with a zero in the OP, TON or TN response.

- Create TON and TN (determinand code 0111+TON) and extract OP (determinand code 0180).
- Select sites to give ~30,000 observations.
- Impute values for the detection limits (if required).
- Log transform data.

For each LHA, the model produces estimates for the trend over time, the trend over space and the seasonal pattern.

The spatial estimates can be presented as a surface plot or overlaid on ArcGIS maps at the relevant monitoring locations; the latter is presented here.

For the trend over time and the seasonal pattern, plots are produced with estimates and standard errors. These plots provide an indication of the trend over the time periods of interest and the seasonal pattern, with the standard errors helping to identify where the estimates are significantly different from zero.

As well as graphical displays, the percentage of variability in the response which is explained by the fitted model can also be produced. This is referred to as R^2 and provides an indication of how much of the variability in each large hydrological area can be explained by the trends over space and time and the seasonal patterns in combination. R^2 is a statistic that provides information about the fitted model and is not available for individual model components here since the same decomposition of the total sum of squares is not available here as with a linear model.

3.3.2 Model 2

In the second model, it is of interest to investigate how much of the variation in the responses of TN, TON and OP can be explained by the covariates collated in Tables 2.5 to 2.7. In order to incorporate the covariate data into a model, both the responses and the covariates have been aggregated to WFD water bodies. In addition, the response data used in Model 1 have been aggregated to monthly mean values.

In the situation where there is more than one monitoring location in a water body, the response data have been averaged across the monitoring locations. The aggregations for the covariates are fully described in Tables 2.5 to 2.7 for covariates in a specific water body and to aggregate covariates to include the contributing land. The latter aggregations were mainly undertaken by the Environment Agency using the tree structure for the river networks.

Three different forms of the covariate model can be investigated here. To enable information from all contributing land for each water body to be incorporated into the model (Model 2A), monitoring locations in a LHA were selected that have no upstream water bodies. This selection ensures that contributing areas do not overlap and hence there is little correlation spatially between measurements for a particular covariate. Hence, in this model, the response would be nutrient concentrations within headwater streams and rivers, and the covariates provide information on the local land area draining to that water body.

In order to explain as much of the variability in the nutrients as possible, initial models contain the covariates from Tables 2.5 to 2.7 along with the year, month within the year (seasonality) and the spatial trend. Data from all contributing land to a particular water body was included for covariates, where appropriate, as described in the final column of Tables 2.5 to 2.7. Further details on the modelling are provided in section A6.2 of Appendix A.

Most of the covariate information is spatial, and since many of the covariates do not vary with time, there is very little additional temporal information. To benchmark the effect of the covariates, a model with only covariates was also fitted. Subsequent models highlight the effect of particular groups of covariates on the percentage of the variability explained (R^2); this has been adjusted here for the number of covariates within each model.

Model 2A was fitted for three example areas:

- Severn (54);
- Lune (72);
- Trent (28).

The first two areas were chosen as ones in which the trends from Model 1 were contrasting. The final LHA was chosen as an example of a Nitrate Vulnerable Zone (NVZ).

Two alternative forms for the covariate model could include:

- monitoring locations selected for which only the covariate information in the related water body should be used (that is, the contributing area information is not relevant (Model 2B);
- all possible monitoring locations and covariates used within each LHA to develop a spatiotemporal model for each LHA (Model 2C).

In Model 2C, the response would be the nutrient levels within a specific water body and the covariates the covariate information for the same water body. But in order to incorporate data from all land contributing to a specific water body, it would be necessary to incorporate a river network covariance structure (for further details, see O'Donnell 2010) into the model. This was outside the scope of this project.

3.3.3 Model 3

The additive models fitted for Model 1 and 2 can in principle be extended to allow interactions between the covariates involved; this was carried out by Bowman et al. (2009) on a large sulphur dioxide (SO_2) dataset. However, very large sample sizes create significant computational difficulties. A research project at the Department of Statistics at the University of Glasgow is developing spatiotemporal models through P-splines, which offer an alternative technical mechanism for the implementation of smoothing techniques. Eilers and Marx (1996) provide a general introduction to this well-established approach, while Lee and Durban (2011) show how this can be implemented very efficiently in a spatiotemporal setting.

This methodology has been implemented for one example determinand and LHA here, TON in LHA 22 (Coquet, Wansbeck, Blyth), incorporating smooth functions for year, day of the year, monitoring location, an interaction between year and day of the year, and an interaction between year and monitoring location. This enables investigation of changes in the spatial and seasonal patterns across time.

The interesting feature of this model is that it can demonstrate, for example, how the amplitude of seasonal variation and distribution of nutrient concentrations across space have changed throughout time.

4 Results

The results are described below for each of the models in turn. LHAs are referred to by their name (and number). The names and numbers for each LHA are shown in Figure 2.1 and Figure 2.2, respectively.

For Model 1, results are described for trends over time and seasonal patterns for OP, TON and TN in each LHA in England and Wales grouped by Environment Agency region. Plots for one region are provided in section 4.1 with the plots for the other regions given in section B1 of Appendix B. This is followed by plots displaying the estimates for the spatial trends over the LHAs and a summary of the percentage of the variability in each LHA explained by Model 1.

For Model 2, the percentage of variation explained is displayed and discussed for models which either contain all covariates or a selection of possible covariates (models A1–A12). The results for the percentage of the variability explained for all 12 models are shown for OP, TON and TN in three selected LHAs – the Severn (54), the Lune (72) and the Trent (28). The plots for time trend and seasonality for OP in these three LHAs are shown in section 4.2 and the results for TON and TN are given in section B2 of Appendix B. The relationships between the covariates and responses are displayed in section 4.2 for OP in the Trent (28) LHA and the remaining plots are given in section B3 of Appendix B.

For Model 3, TON in LHA 22 (Coquet, Wansbeck, Blyth) is used as an example (section 4.3). A series of plots is displayed to:

- highlight the initial results from this LHA using a similar model to that for Model 1;
- display the results for incorporating interactions in the model.

The covariance structure is also discussed.

The time periods for each LHA and each determinand are summarised in Table 2.4. For each of the LHAs, the full time period available was used except in cases where there were very few samples in the early years.

4.1 Model 1: trends over time and space and seasonality

Model 1 was fitted to all 59 LHAs across England and Wales to explore trends over time, trends over space and seasonal patterns for natural log transformations of OP, TON and TN. The only exceptions are a few LHAs in North East Region and all those in Wales, where there is a very high percentage of OP data that are below the limit of detection (Table 4.1). These LHAs were not analysed for OP because more than 50 per cent of their data were below the limit of detection.

Table 4.1 Excluded LHAs

Region	LHA[1]	Percentage OP data below limit of detection
North East	Tweed (21)	60.18
	Tyne (23)	51.47
Wales	Anglesey (102)	51.72
	Cleddau (61)	50.14
	Clwyd Conwy (66)	65.40
	Dovey (64)	68.67
	Glaslyn (65)	72.98
	Rheidol Ystwyth (63)	75.67
	Taff (57)	72.27
	Teifi (62)	42.00
	Tywi (60)	57.34
	Usk (56)	73.70
	Wye (55)	67.20

Notes: [1] Also Loughor (59) and Tawe Neath (58).

Similarly, the small number of monitoring locations in Tywi (60) and Loughor (59) for OP and the small number of monitoring locations in all the LHAs in Wales – with the exception of the Dee (67) for TON and TN – made it impossible to estimate a spatial surface.

Therefore, in these cases the following LHAs were grouped together to fit the initial model:

- Wye (55), Usk (56), Taff (57), Tawe Neath (58), Loughor (59), Tywi (60), Cleddau (61), Teifi (62) and Rheidol Ystwyth (63);
- Glaslyn (65), Clwyd Conwy (66) and Anglesey (102).

The LHAs of the Severn (54) and Dovey (64) are taken to be in Midlands Region.

For Model 1, the results for each LHA are grouped by Environment Agency region. The plots obtained for the LHAs in Southern Region are displayed in Figures 4.1, 4.2 and 4.3 for trend over time (left) and seasonality (right) for each of the three determinands. The plots for the other regions are given in section B1 of Appendix B.

In these plots, the estimates are on a common scale for trend and a common scale for seasonality across all LHAs. However, the scales are different for trend and seasonality within each LHA. The plots for trend over time and the seasonal pattern display the estimates along with dashed lines at a distance of two standard errors from the estimates. For the trend over time, there is an edge effect which results in the standard errors at the beginning and end of the time periods being wider; wider standard errors also indicate a reduction in data availability.

In each case, the model extracts an overall mean initially and hence all trends are comparable across regions; the overall mean for each LHA and each determinand is displayed on maps in section B4 of Appendix B. The fitted spatial estimates have been extracted from the model for each determinand in each LHA and are displayed in Figures 4.4, 4.5 and 4.6 for OP, TON and TN respectively.

An overall summary for each determinand of interest within each LHA is provided in terms of the percentage of variation explained by the fitted model (R^2) for Model 1, that is, by the trend over time, the trend over space and the seasonal pattern in combination. The R^2 value provides an indication of how much of the variability is explained overall by a model that includes trends over time and space and seasonality in each LHA. The R^2 values are provided on top of their corresponding LHAs in Figures 4.7, 4.8 and 4.9.

Figures 4.1 to 4.9 are displayed at the end of section 4.1.

4.1.1 Time trends and seasonality

North West Region (LHAs 68–77; Appendix B, Figures B.4 to B.6)

In general OP has a decreasing trend in this region, with a seasonal pattern highlighting one peak in OP in the summer months. The exceptions to this are the Lune (72) and Lyne, Esk (77) LHAs where OP levels appear to have remained more constant over the time period. Many LHAs contain slight increases around the late 1970s/early 1980s and around 1995. The seasonal pattern highlighted is also slightly different for the Eden (76) and Lyne, Esk (77) LHAs where OP levels in the winter months appear much higher than in the spring.

The trends over time and seasonal patterns for TON and TN are quite similar. In general for the Leven, Kent, Duddon, Derwent, Eden and Lyne, Esk (73–77) LHAs, an increasing trend is evident until around the mid 1980s where it levels off, with a small decrease evident again in later years. There is evidence of a slight decline in both TN and TON for the Ribble (71) LHA, while the trend for TON in the Douglas (70) LHA follows the trend above for LHAs 73–77 and TN appears to have remained fairly flat with evidence of only a slight decline in later years. A similar pattern is evident for TON and TN in the Weaver (68) LHA. TON in the Mersey (69) LHA appears to have been relatively constant since around the late 1970s, although TN saw a decreasing trend from this time.

Both TON and TN highlight a seasonal pattern that has higher levels in the winter months and low levels in the summer. This is consistent across all LHAs in North West Region with stronger seasonal patterns evident in the Lune (72) and Lyne, Esk (77) LHAs.

North East Region (LHAs 21–27; Appendix B, Figures B.10 to B.12)

The Hull (26) and Ouse, Humber Estuary (27) LHAs have sufficient data to enable modelling from the mid-1970s. However, data for LHAs 21–25 are only available over approximately the last 15 years.

For OP, the Hull (26) and Ouse, Humber Estuary (27) LHAs highlight a generally decreasing trend with lower levels evident in the early 1990s. In the Wear (24) LHA, the levels appear to have been decreasing over the past five years. The wide standard errors in the results for the Wear (24) LHA highlight the small amount of data available at the beginning of this time series. In the Coquet, Wansbeck, Blyth (22) and Tees (25) LHAs, there has been a general decrease from around 2000 with a slight levelling off for the values is highlighted around 2005–2007. There are no results for the Tweed (21) and Tyne (23) LHAs as a result of the high percentage of values below the limit of detection in these LHAs.

Results are available for all LHAs for TN and TON. With the exception of the Ouse, Humber Estuary (27) LHA, the trends over time are similar for both determinands. The Ouse, Humber Estuary (27) LHA shows increases around 1975–1980 and the late 1980s to the late 1990s for TON, but with stronger evidence for a declining trend over the past 10 years. While small increases are evident in the trend around the same times for TN, the general pattern here is of a decrease in levels since the early 1970s. In the remaining sites, data are only available for the past 15 years, with the Coquet, Wansbeck, Blyth (22), Wear (24) and Tees (25) LHAs highlighting features in the trend with levels increasing in the late 1990s and around 2005 but declining levels suggested in recent years. However, the trends for TON and TN in the Tyne (23) LHA are fairly flat over the time period with suggestions of a decrease in levels in recent years.

The seasonal patterns for OP in the Wear (24), Tees (25) and Ouse, Humber Estuary (27) LHAs are consistent with those elsewhere, with generally low values in the winter months rising to a peak in summer. The peak for the Wear (24) and Tees (25) LHAs is broader, probably as a result of less data. While the pattern for the Coquet, Wansbeck, Blyth (22) LHA is similar, there is more evidence of a dip in the values between winter and spring. This is much more prominent in the Hull (26) LHA where the peak is also shifted later in the year.

For TON and TN again the seasonal patterns are very similar for both variables and across all LHAs in the region. The only differences appear to be for LHAs where there were less data available to estimate the patterns and the troughs for these patterns are slightly broader. In all cases the lowest levels are in summer.

Midlands Region (LHAs 28, 54 and 64; Appendix B, Figures B.1 to B.3)

The three LHAs included as the Midlands Region were the Trent (28), Severn (54) and Dovey (64). Of these LHAs there are no results for OP in the Dovey (64) LHA as a result of the large percentage of values below the limit of detection. In the other two LHAs, the trends over time for OP are generally decreasing. Both LHAs display a fairly flat trend in the late 1980s with a very similar decline thereafter.

In the late 1960s to early 1980s, there appears to have been an increase in TON and TN values in the Severn (54) LHA with values being fairly stable and showing just a slight decline since the early 1980s. A similar slightly declining trend is highlighted for the Trent (28) LHA for both variables. An increase in levels for the Dovey (64) LHA is evident until the late 1980s, with the values remaining fairly constant from this time apart from a slight decrease in recent years. However, the width of the standard error bands indicates the small amount of data here in comparison with the other two LHAs.

The seasonal patterns for OP in the Trent (28) and Severn (54) LHAs are very similar and of the same magnitude, with the peak in OP occurring in late summer. In both these LHAs, the seasonal patterns for TN and TON are also very similar with higher values evident in the winter months and lower levels in summer. For the Dovey (64) LHA, the seasonal pattern is again similar for TON and TN but lower values are evident throughout late spring and summer, presenting a broader trough in the shape. Again, this is likely to be an artefact of the relatively small amount of data.

Wales Region (LHAs (55–63), 67 and (65, 66, 102); Appendix B, Figures B.13 to B.15)

There are two complications with the data from Wales. First, the data for OP were not analysed for any of the LHAs except the Dee (67) since more than 50 per cent of the data are marked as being below the limit of detection. Imputation is not appropriate in such situations as a result of more than half of the data being estimated.

Secondly, the number of monitoring locations is far fewer for many of the LHAs in Wales than in the LHAs in England, and as a result a spatial surface cannot be estimated. Therefore, in these cases the LHAs have been combined together to allow estimation of a spatial surface. Although this approach is not ideal as independent river networks have been combined together and analysed as though they are one network, it does provide an indication of the patterns in Wales.

In general for all LHAs in Wales the trends in the later years appear fairly flat; the wide standard errors at the beginning of the time series indicate the small amount of data available here in order to estimate the trends. In the combined area for the Glaslyn (65), Clwyd, Conwy (66) and Anglesey (102) LHAs, there is a suggestion of a decrease in values in the last 10 years for TON and TN. Although TN increases slightly around 2010, too much emphasis cannot be put on this effect since it is near the end points of the data. In the other combined LHAs [Wye, Usk, Taff, Tawe, Neath, Loughor, Tywi, Cleddau, Teifi and Rheidol, Ystwyth (55–63)], there is evidence of a reduction in levels for TON and TN until around 1985 with an increase evident in the early 1990s.

Both TON and TN highlight a seasonal pattern that has higher levels in the winter months and a dip in the summer. This is consistent across all LHAs in Wales. However, the seasonal pattern appears stronger in the combined area of the Glaslyn (65), Clwyd, Conwy (66) and Anglesey (102) LHAs. For the one LHA analysed for OP, the Dee (67), the levels are higher in the summer with lower values evident in winter.

Anglian Region (LHAs 29–37; Appendix B, Figures B.7 to B.9)

In all LHAs, there has been a general decrease in OP levels since the early 1990s, the only exceptions being a feature in many of the plots where there is a slight increase or levelling off around 1995, and the Gipping (35) LHA where the trend is fairly flat in recent years. In the Witham, Welland, Nene, Great Ouse (30–33), Gipping (35) and Stour, East Anglia (36) LHAs, there is evidence of an increase in values up until 1990. In the Blackwater, Chelmer (37) LHA, however, the values have generally decreased from the early 1980s.

In the Ancholme (29), Welland (31), Great Ouse (33) and Bure, Waveney (34) LHAs, there is very little trend over time for TON and TN. However, there is a generally decreasing trend over time for the Nene (32) and Blackwater, Chelmer (37) LHAs. For the Blackwater, Chelmer (37) LHA, the trend is fairly flat in the 1990s. For the Witham (30) and Gipping (35) LHAs, there is little trend evident early on in the time series but there is evidence of a decreasing trend for both TN and TON in the last 10 years. In the Stour, East Anglia (36) LHA, a decreasing trend is evident in the late 1980s with the trend fairly flat thereafter.

While the peak in OP levels is still evident in the LHAs for this region, it appears to be slightly later in many LHAs than previously seen and many LHAs show a dip in levels earlier in the year. In contrast, the seasonal patterns for TON and TN are very similar to each other and also across all LHAs in the region, with high values in winter decreasing to lowest levels in summer.

Thames Region (LHAs 38, 39; Appendix B, Figures B.16 to B.18)

There are only two LHAs in the Thames region – the Lee (38) and Thames (39) LHAs.

In the Lee (38) LHA, the OP values declined in the mid-1970s before remaining fairly constant until the late 1990s where the level started to decline again. This pattern is also evident in the Thames (39) LHA, with the exception of the early 1980s where levels increased slightly. For TON and TN in the Lee (38) LHA, the levels have generally decreased over the time period. However, there is evidence of levels increasing slightly throughout the 2000s, with this level possible remaining fairly constant or even decreasing slightly in recent years. For the Thames (39) LHA, the trend is fairly flat for TON but a slight decreasing trend is evident for TN.

The seasonal pattern for OP in both LHAs indicates low levels in early spring and a peak in levels around late summer. The patterns for TON and TN are very similar to other LHAs with high levels in the winter months and lowest levels in late summer.

Southern Region (LHAs 40–42, 101)

In all LHAs in Southern Region, the trends for OP are generally decreasing (Figure 4.1); over the time periods with the flattest trend are evident for the Isle of Wight (101) LHA. OP in the Medway, Stour (40) LHA has a slightly different trend with an increase in levels evident until 1985 followed by a decreasing trend thereafter. For TON and TN, there is very little trend evident in the Arun, Ouse, Cuckmere (41), Test (42) and Isle of Wight (101) LHAs (Figures 4.2 and 4.3). However, in the Medway, Stour (40) LHA, the trend is mainly decreasing from around the mid-1980s onwards and there is slight evidence of an increase before this.

In the Medway, Stour (40) and Arun, Ouse, Cuckmere (41) LHAs, the seasonal pattern mainly displays a peak around mid-to-late summer for OP. In the Isle of Wight (101) LHA, however, lower levels are evident in spring along with a peak in summer (Figure 4.1). The pattern in the Test (42) LHA is quite different with no peak evident later in the year and a trough at 100 days. The seasonal patterns for TN and TON are very similar to each other, and in each LHA of this region, with low values around mid-to-late summer. However, the seasonal pattern in the Test (42) LHA appears weaker than the other LHAs (Figures 4.2 and 4.3).

South West Region (LHAs 43–53; Appendix B, Figures B.19 to B.21)

After a slight indication of an increase in the late 1960s to early 1970s, for OP a decline is evident until the early 1980s from when the trend decreases more gradually for the Avon, Hants (43), Piddle, Frome (44) and Frome, Bristol, Avon (53) LHAs. This general decline in levels over these time periods continues for the other LHAs with the exception of a slight increase in levels for OP in the early 2000s for the Tamar (47) and Tone, Parrett (52) LHAs. The Camel (49) LHA also highlights a slight increase between 1985 and 1990, and there are indications of both these features in the Dart (46) and Fal (48) LHAs. The increase in the 1980s is slightly earlier in time for the Exe (45) LHA.

For TON and TN in the Tamar (47) and Frome, Bristol, Avon (53) LHAs, the levels increase from the late 1960s until the early 1980s and remain fairly constant in later years. A similar pattern is highlighted in the Exe (45), Dart (46) and Torridge, Taw (50) LHAs, but there is also an indication of a decreasing trend from around the mid-1980s. The trend for the Camel (49) LHA is similar, though an initial decrease is evident in the late 1970s. In contrast, in the Tone, Parrett (52) LHA there is little trend evident except for a reduction in levels in the past five years for TON. Levels for TN appear to have decreased in the late 1980s before being reasonably constant until the decrease of the last five years. The trends in the Avon, Hants (43), Piddle, Frome (44), Fal (48) and East and West Lyns (51) LHAs are quite different with a generally increasing trend evident for both TN and TON; the wide standard error bands early on indicate the small amount of data available until the late 1980s. However, there is slight evidence of a decrease for TON in the Piddle, Frome (44) LHA over the past 10 years.

The seasonal patterns for OP, TON and TN are very similar to what has been highlighted for the other regions in terms of a peak in summer for OP and a dip for TON and TN. The peak in summer for OP generally occurs in mid-to-late summer. In the

Avon, Hants (43) and East and West Lyns (51) LHAs, a dip is evident in early spring with a peak that is slightly later in the year. The pattern for the Piddle, Frome (44) LHA is quite different with the prominent feature being a dip at 100 days. The lowest values for TON and TN occur in mid-to-late summer, with the dip for the Dart (46), Fal (48) and Camel (49) LHAs occurring slightly later.

4.1.2 Spatial trends

Figures 4.4, 4.5 and 4.6 display the spatial estimates from each of the models for OP, TON and TN respectively at each of the monitoring points used in the modelling in each LHA. In LHAs with a large amount of data, gridding has been applied in both space and time (see section 3 and section A5 of Appendix A for further details) and so not all the monitoring points are contained in these maps.

The circles on the figures represent the monitoring points and are shaded to indicate the difference between the spatial estimate for that location and the overall mean for the particular LHA (see section A4 of Appendix A for overall mean ln concentrations in each LHA) containing the monitoring point. Dark red colours on the maps indicate where the estimated concentrations at a particular monitoring location are much higher than the mean for the LHA, with the colour scale going to blue indicating locations where the estimated concentrations are lower than the mean for the LHA. The colour bands have been chosen arbitrarily and therefore monitoring points that have high estimated concentrations relative to the mean of the LHA are indicated in red.

The figures illustrate how the estimated concentrations change across the LHAs and indicate, for example, monitoring points with larger concentrations in each of the LHAs. Estimated concentrations from LHAs, such as in Wales, will have a higher variability associated with them as a result of the sparser monitoring network.

The distribution of spatial estimates across the LHAs for TON and TN (see Figures 4.5 and 4.6 respectively) is similar. In each LHA, monitoring points at the bottom of the catchment tend, in general, to have higher levels of TN and TON than those further up the catchments. There are only a couple of LHAs where a different distribution of the spatial estimates is evident and these are on the coast of the Lune (72) LHA and in the south of the Severn (54) LHA, where in both cases the estimated concentrations for TON are indicated as being smaller than the mean for these LHAs, whereas TN appears to be higher than the LHA means.

For TN and TON, the estimated concentrations are indicated as red and therefore are higher than the mean for the respective LHA in the south-east of the Tyne (23), the south of the Dart (46), the north and west of the Derwent (75), the south of the Duddon (74) and Leven, Kent (73) LHAs and on the coast of the Douglas (70) and Mersey (69) LHAs.

The distribution of OP estimates (Figure 4.4) is also quite similar to that for TN and TON. However, the range of values for the highest estimated concentrations (indicated in red) relative to the mean of each LHA is larger. The highest values are generally clustered around the same locations as for TON and TN. However, there appears to be many more high values, relative to the mean, indicated along the north-west coast of the Mersey (69), Douglas (70), Ribble (71) and Lune (72) LHAs. In the Severn (54) LHA, the higher OP concentrations appear to be concentrated towards the middle of the LHA in contrast to TON, where the higher concentrations are clustered towards the south-east, and TN where the concentrations in this LHA are not much higher than the mean. Other differences between the spatial distribution of OP and TON/TN estimates are evident in:

- the Hull (26) and Ancholme (29) LHAs, where estimated concentrations on the coast appear higher than the mean;
- the Great Ouse (33) LHA where estimates appear larger in the middle of the catchment;
- the Tone, Parrett (52) and Frome, Bristol, Avon (53) LHAs where higher concentrations than the mean are also evident.

4.1.3 R^2

In Figures 4.7 to 4.9, the percentage of variability explained by the fitted models including spatial and temporal trends and seasonality is:

- 12–68 per cent for OP;
- 19–67 per cent for TON;
- 19–70 per cent for TN.

For OP, the smallest amount of variability explained is for the Test (42) LHA with the largest amount explained in the Dee (67) LHA.

For TON, the Thames (39) LHA has the smallest percentage explained with the largest percentage explained in the East and West Lyns (51) LHA.

The Thames also has the smallest percentage of variability explained for TN with the largest percentage variability explained in the Dee (67) LHA. However, the East and West Lyns (51) LHA is only a few per cent less than this.

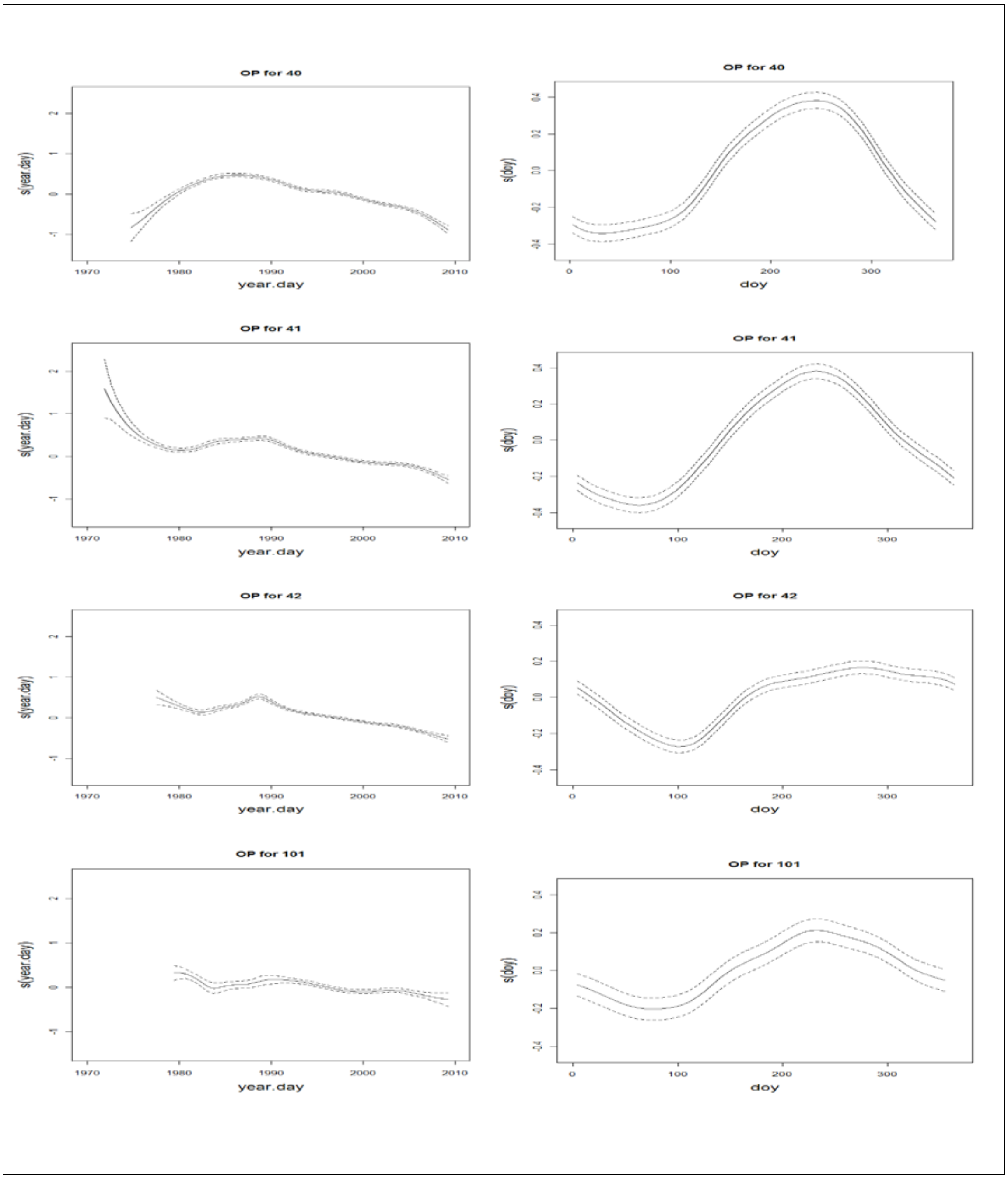


Figure 4.1 In OP for the four LHAs in Southern Region

Notes: For each LHA, panels on the left highlight the year trend with the seasonal effects on the right.
The dashed lines lie at a distance of two standard errors from the estimates.
See Figures 2.1 and 2.2 for the LHA names and codes respectively.

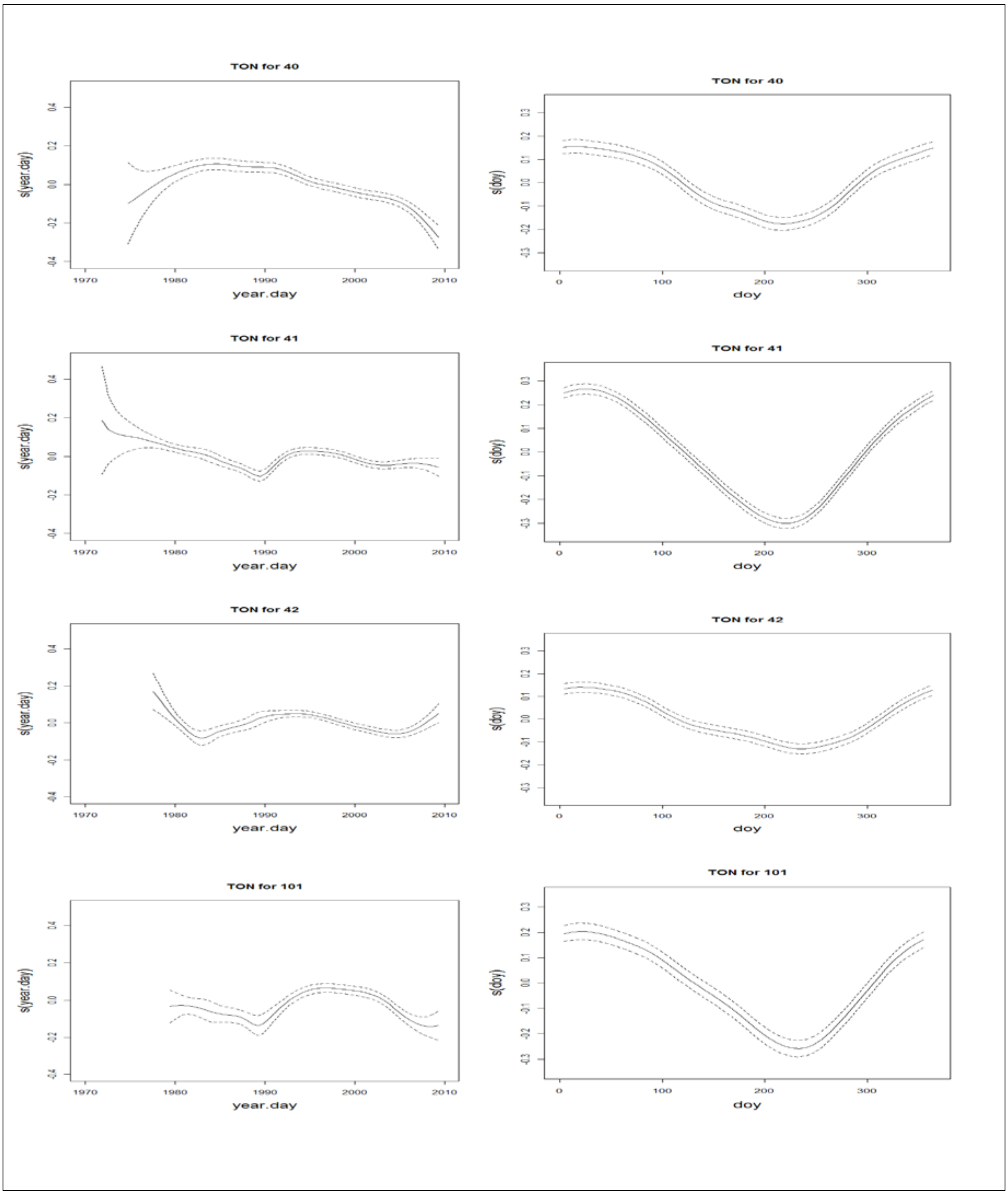


Figure 4.2 In TON for the four LHAs in Southern Region

Notes: For each LHA, panels on the left highlight the year trend with the seasonal effects on the right.
The dashed lines lie at a distance of two standard errors from the estimates.
See Figures 2.1 and 2.2 for the LHA names and codes respectively.

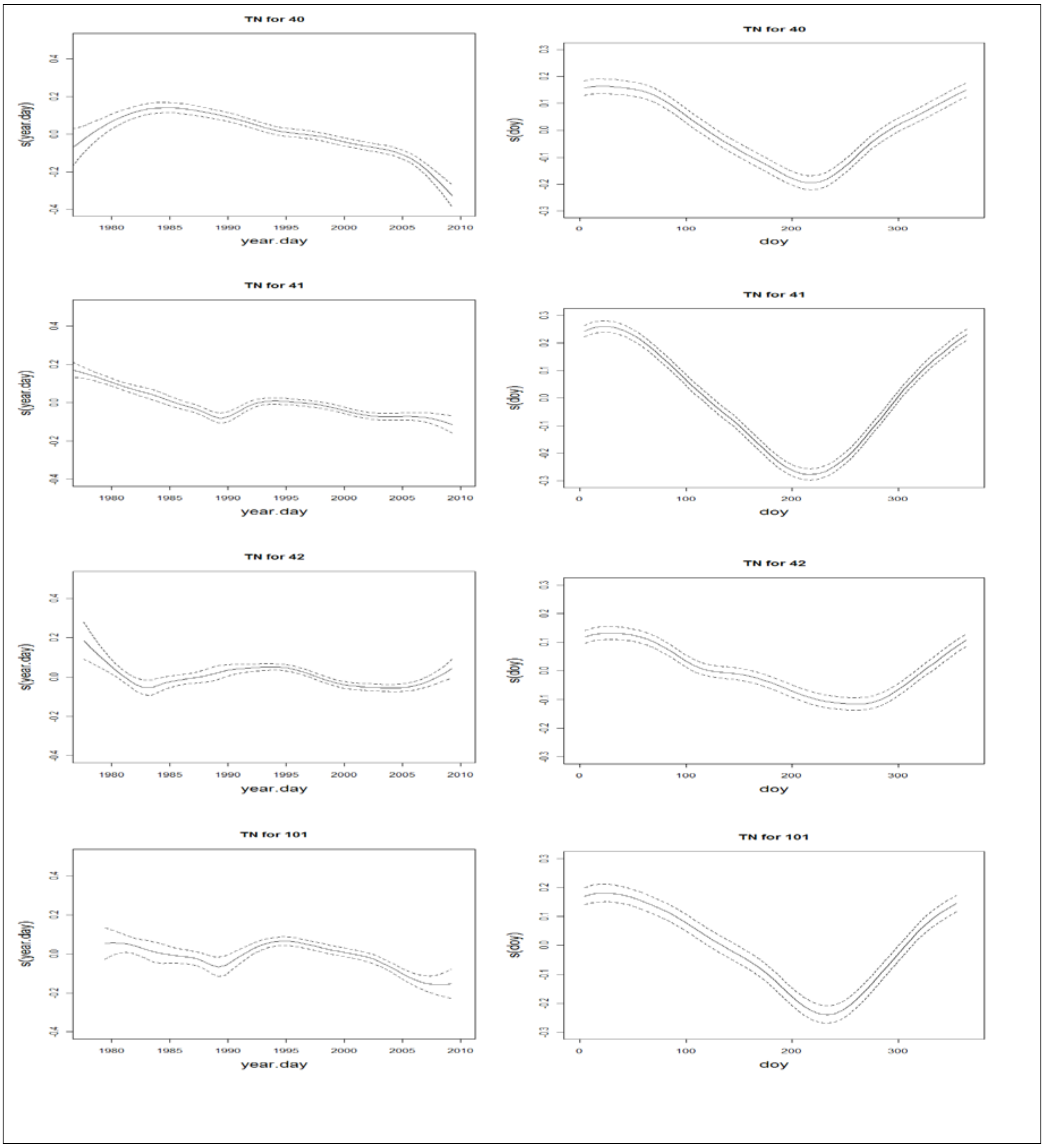


Figure 4.3 In TN for the four LHAs in Southern Region

Notes For each LHA, panels on the left highlight the year trend with the seasonal effects on the right.
The dashed lines lie at a distance of two standard errors from the estimates.
See Figures 2.1 and 2.2 for the LHA names and codes respectively.

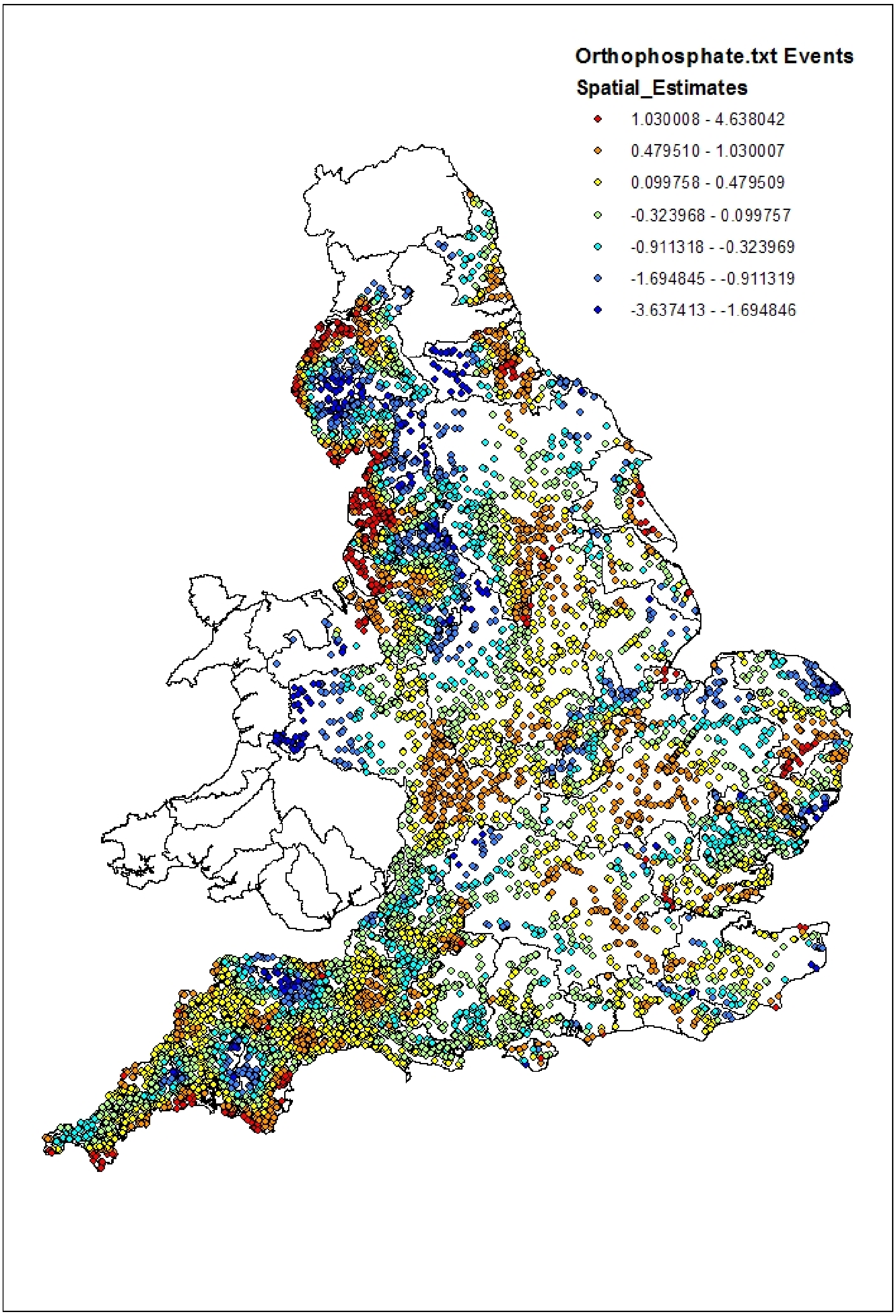


Figure 4.4 Spatial estimates (relative to mean of each LHA) for ln TN in each LHA

Note Data have been gridded in time and space, where necessary, and as a result not all monitoring points are included in the map.

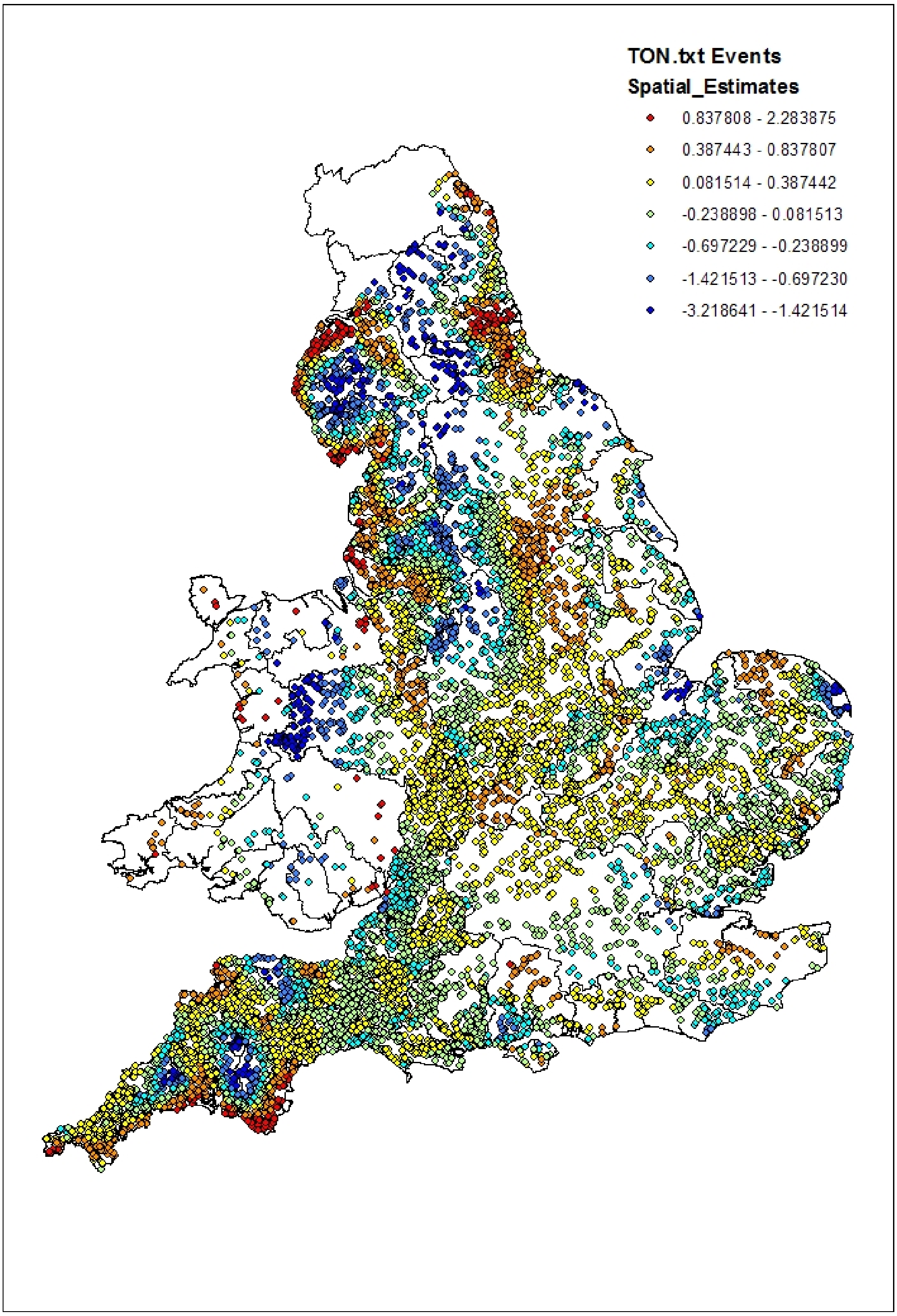


Figure 4.5 Spatial estimates (relative to mean of each LHA) for ln TON in each LHA

Note Data have been gridded in time and space, where necessary, and as a result not all monitoring points are included in the map.

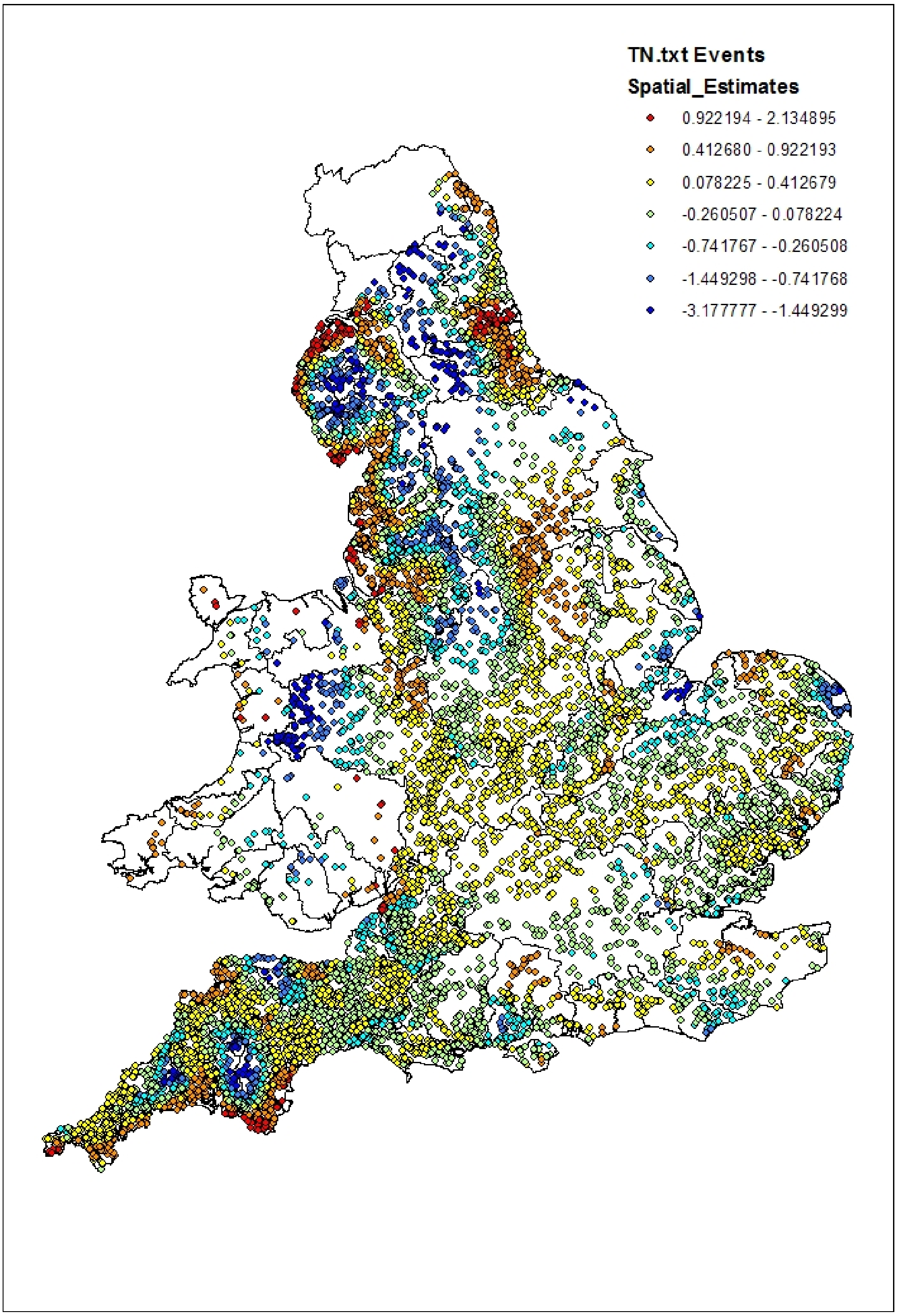


Figure 4.6 Spatial estimates (relative to mean of each LHA) for ln TN in each LHA

Note Data have been gridded in time and space, where necessary, and as a result not all monitoring points are included in the map.

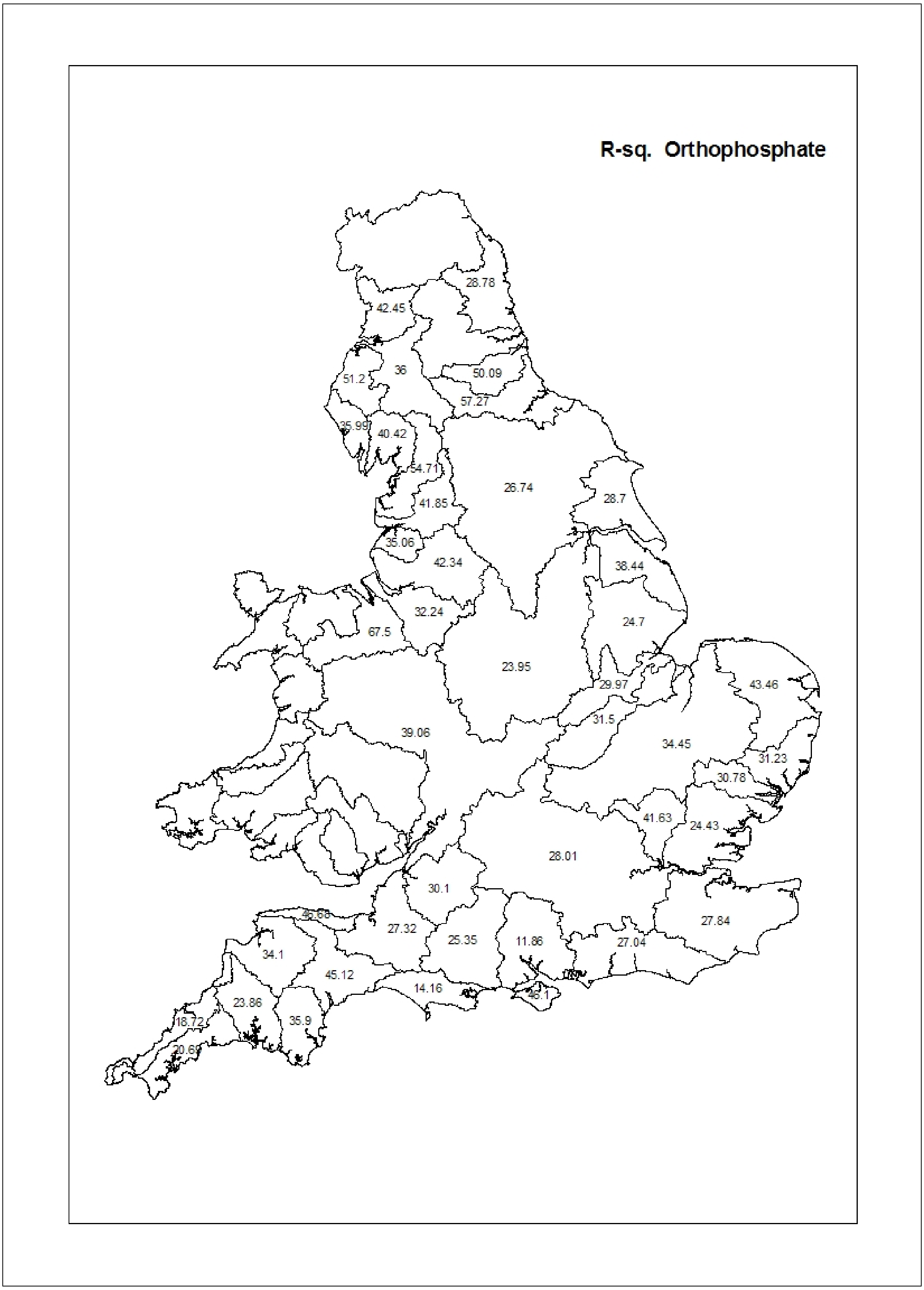


Figure 4.7 R^2 values (%) for ln OP in each LHA from fitting Model 1

Notes: There are no results for the LHAs in Wales and the Tweed (21) or Tyne (23) LHAs as a result of the high proportion of non-detects for OP.

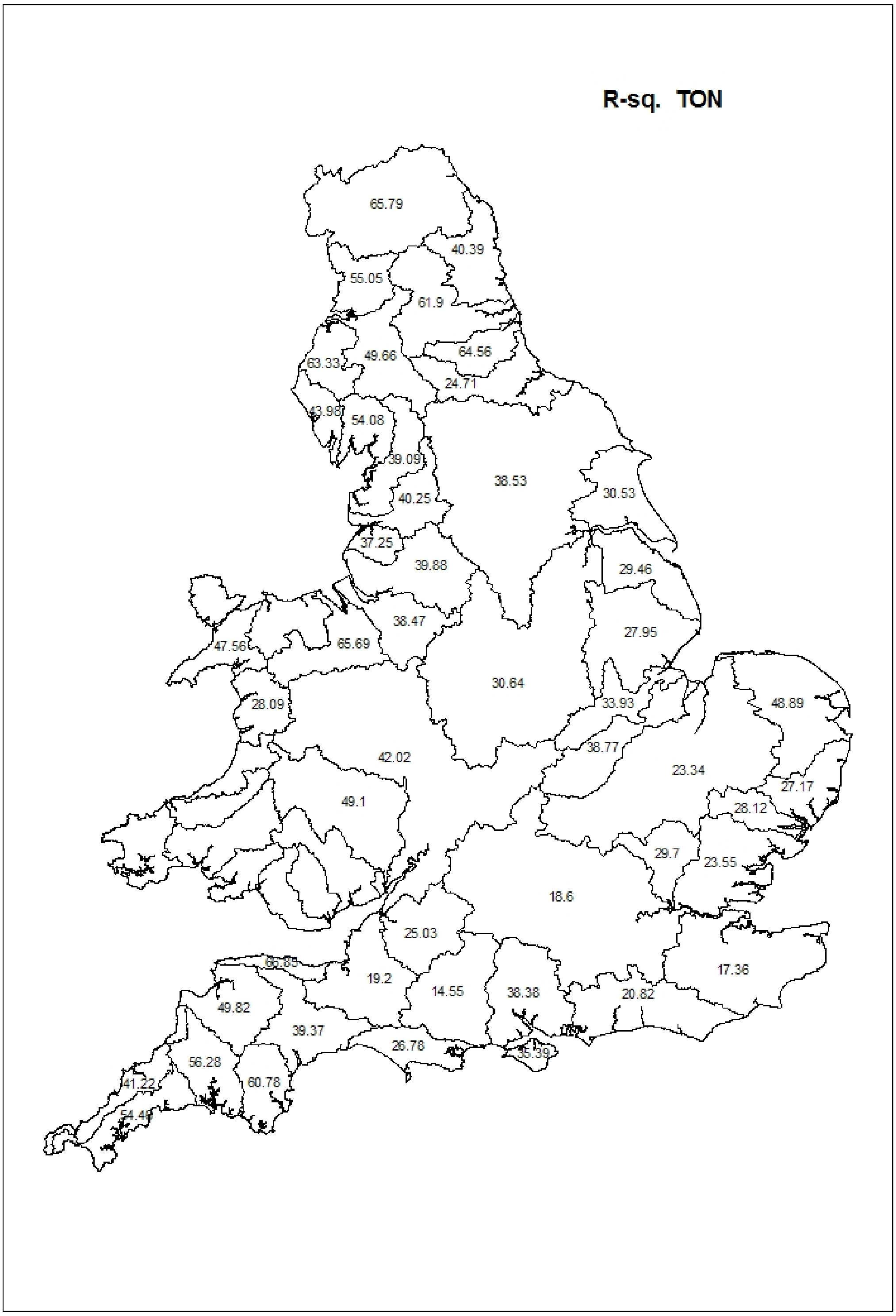


Figure 4.8 R^2 values (%) for In TON in each LHA from fitting Model 1

Notes: The LHAs in Wales have been combined in two clusters. The R^2 value is 47.6 per cent for the North Wales cluster (LHAs 65, 66, 102) and 49.1 per cent for the South Wales cluster (LHAs 55–63).

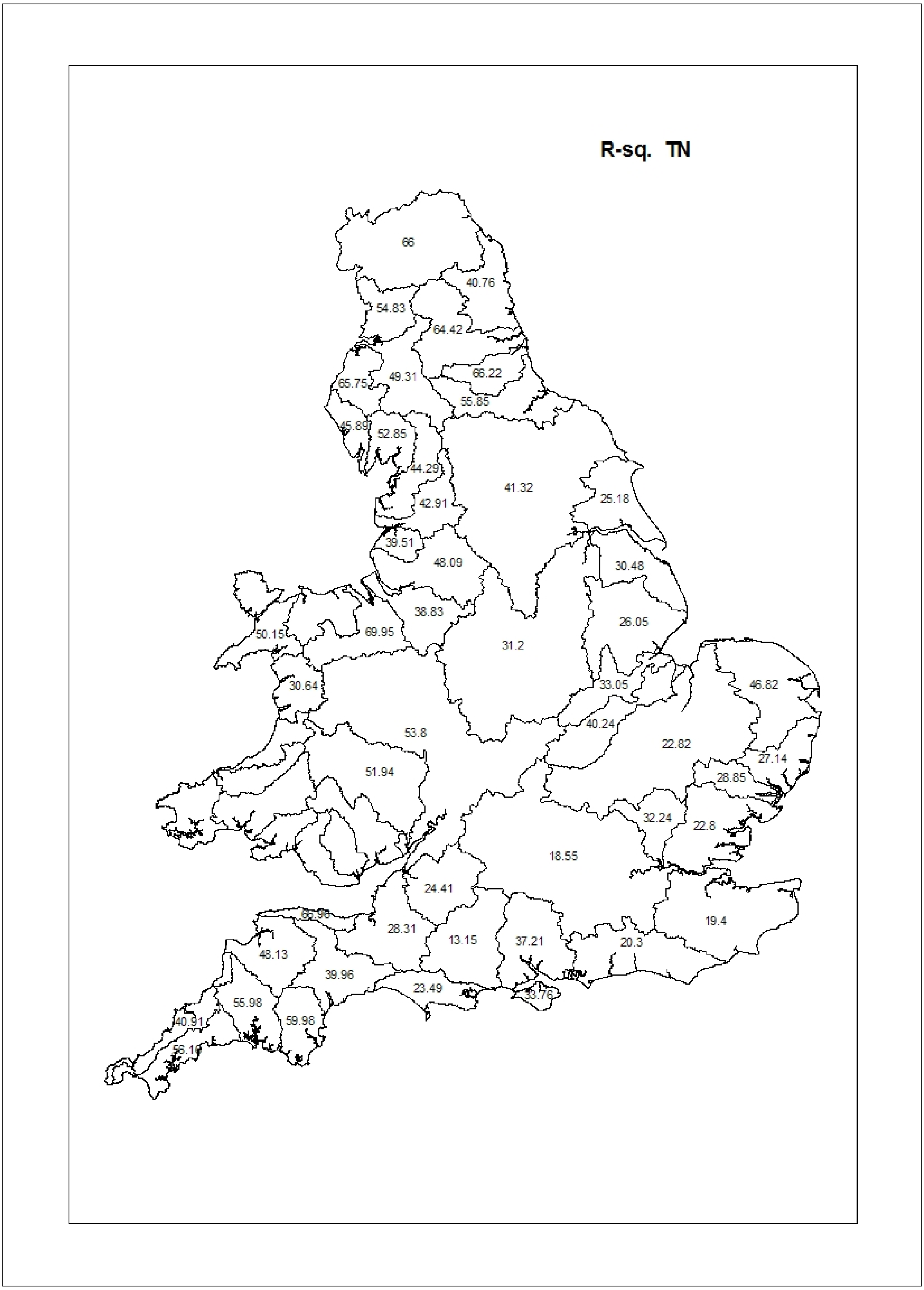


Figure 4.9 R^2 values (%) for ln TN in each LHA from fitting Model 1

Notes: The LHAs in Wales have been combined in two clusters. The R^2 value is 50.2 per cent for the North Wales cluster (LHAs 65, 66, 102) and 51.9 per cent for the South Wales cluster (LHAs 55–63).

4.2 Model 2: describing trends

To investigate the percentage of variability in the responses of OP, TON and TN that could be explained by the covariates listed in Tables 2.5–2.7, Model 2A was fitted to the Trent (28), Severn (54) and Lune (72) LHAs. The data used to fit Model 1 after transformation of the responses, data imputations and gridding were used to fit the covariate models.

The LHAs were chosen to be contrasting in terms of the signals in the response data as highlighted by Model 1 and with the Trent (28) LHA also representing a Nitrate Vulnerable Zone of interest.

Plots for the time trend and seasonality for OP are shown for all LHAs in Figure 4.10, with the remaining plots displayed in section B2 of Appendix B. Figure 4.10 illustrates the contrasting trends for OP in the Severn (54) and Lune (72) LHAs.

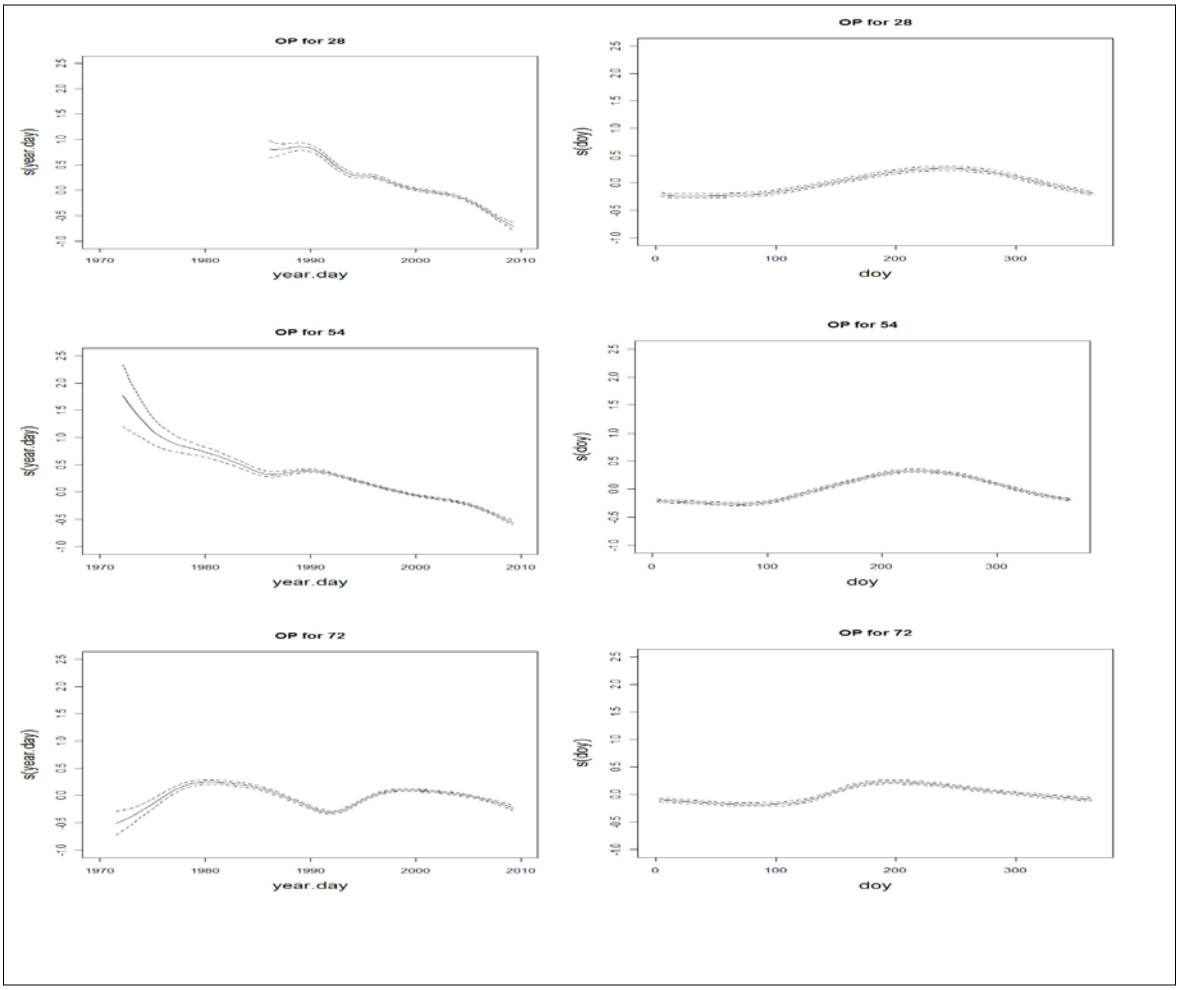


Figure 4.10 Time trend (left) and seasonality (right) for ln OP in the Trent (28), Severn (54) and Lune (72) LHAs with all plots on a common scale

Notes: The dashed lines are two standard errors from the estimates.

The models fitted for each determinand in each of the three LHAs are listed in Table 4.2.

Table 4.2 Models fitted for each determinand in the Trent (28), Severn (54) and Lune (72) LHAs

Model number	Variables
A1	Space, time and seasonality plus all the covariates
A2	Only covariates
A3	BFI, flow/discharge, population, rainfall, land cover, livestock, crops
A4	BFI, flow/discharge, population, rainfall, crops, ALC, slope, soil, livestock
A5	BFI, flow/discharge, population, rainfall, livestock
A6	BFI, flow/discharge, population, rainfall, ALC, slope, soil
A7	BFI, flow/discharge, population, rainfall, crops
A8	BFI, flow/discharge, population, rainfall, land cover (ALC, slope, soil)
A9	BFI, flow/discharge, population, rainfall, land cover
A10	BFI, flow/discharge, population, rainfall
A11	Flow/discharge, population, rainfall
A12	BFI, population, rainfall

In each of the models listed in Table 4.2, a natural log transform (ln) was applied to all the covariates from Tables 2.5–2.7 with the exception of the categorical variables of slope, soil and ALC (to reduce skewness in the data distributions). A small constant of 0.01 was added to covariates that contained values of zero before applying the log transformation. The relationships were fitted initially as smooth functions for all continuous covariates to enable them to be data driven and not constrained to be linear.

Contributing land was incorporated for all catchment covariates except flow and fertiliser. A couple of variables displayed strong relationships with one another such as:

- cereals and total arable (land cover);
- managed grass (calculated as the sum of temporary and permanent grass) (land use) and grass (land cover).

Hence only the total arable and grass variables were included in the models.

The percentage of the variability explained (R^2), adjusted for the number of covariates in each model, is presented for each of the 12 models in Tables 4.1 to 4.3 for the three LHAs respectively.

Models A1 and A2 are of most interest initially, with the remaining models indicating the percentage of variability explained by subsets of the models.

Fertiliser explained a very small percentage of the variability and hence was not included in the reduced models (A3–A12). The small contribution by fertiliser is likely to be a consequence of the data available for use in this study. The data available were annual application rates for the whole of England and Wales; data were not available for individual LHAs. Hence there is no information to explain spatial trends and little information to explain temporal trends within each LHA.

Each of the 12 models (A1–A12) was fitted for the three contrasting LHAs. The resulting relationships between the responses and covariates are specific to these LHAs and thus can indicate the sensitivity of a river to changes in particular covariates. Strong relationships are indicated by the covariate values where the interval, provided graphically, for the estimates and standard errors does not include zero.

The plots obtained for Model A2 are displayed for OP in the Trent (28) LHA in Figure 4.11 (at the end of section 4.2) and the plots for the other models in section B3 of Appendix B. Each plot illustrates the fitted relationship between a covariate (on the *x*-axis) and the response (indicated by the smooth function scale on the *y*-axis), with dashed lines to indicate a distance of two standard errors from the estimates.

4.2.1 Trent (28) LHA

For the Trent (28) LHA, for which data span 1986–2009, there are 137 water bodies for which OP is modelled, 155 for TON and 154 for TN.

The 12 models listed in Table 4.2 were fitted for the three determinands and the corresponding R^2 values are recorded for each model in Table 4.3.

The trends for time and space and the seasonal pattern explain the largest percentage of the variation at 58, 52 and 52 per cent for OP, TON and TN respectively (A1 in Table 4.3). In each case the percentage of variability explained by only the covariates from Tables 2.5–2.7 are 54, 39 and 39 per cent respectively (A2 in Table 4.3).

The covariates are therefore doing quite well in explaining the patterns in the nutrients. Each of the covariates or covariate groups contributes a small percentage to the variability explained and it is not clear from the models fitted here (A1–A12) that one set of covariates explains a large percentage of the variability.

Models that incorporate either the land cover variables or the group of variables, that is, ALC, slope and soil (A3 and A4) explain similar amounts of variability (43 and 44 per cent for OP, 32 and 37 per cent for TON, and 33 and 37 per cent for TN respectively).

For all determinands, approximately 30 per cent of the variability is explained – in no particular order – by BFI, flow, population, rainfall, land cover and the group ALC/slope/soil.

Table 4.3 R^2 values for Trent (28) LHA (%)

Model number	OP	TON	TN
A1	57.8	51.9	51.5
A2	53.7	39.2	38.8
A3	42.8	32.3	32.5
A4	44.0	37.3	37.1
A5	24.8	24.5	26.6
A6	24.5	28.5	28.2
A7	23.6	24.9	25.7
A8	30.3	30.8	30.2
A9	18.7	23.0	23.6
A10	12.5	20.0	21.0
A11	11.3	15.9	17.2
A12	6.17	10.6	11.4

All the variables for OP and most of the variables for TON and TN included in the models are statistically significant. In general this is to be expected due to the large amount of data being used. However, there are a few variables that appear as not significant and hence further refinement of the models is required. For variables that are not significant, it could be that there is truly no relationship. Alternatively, it may be that there is a relationship between this variable and one of the other covariates and hence both are not required in the model.

In the Trent (28) LHA, it was observed that potatoes and sugar beet, stock feed and field vegetables, cows and sheep, and BFI and total cumulative rainfall are related.

Variables that are related in any way can cause problems in the model; the relationship does not have to be linear and in these cases both variables may not be required. Therefore, the models were re-fitted after eliminating the sheep, field vegetables and sugar beet variables.

If sheep, field vegetables and sugar beet are removed from A1 and A2 for TON, in which field vegetables are not statistically significant, the R^2 values are reduced by approximately 1 per cent for A2 and 0.5 per cent for A1. Sometimes it is sufficient to remove only these factors and all the variables that were initially not statistically significant become significant. For example, the rainfall variable is not significant in A4 for TN and the rainfall and arable land variables are not significant in A2 for TN, but eliminating the sheep, field vegetables and sugar beet makes them significant and the reduction produced in R^2 is small, that is, 2 per cent for A4 and 1 per cent for A2.

While there is a relationship between BFI and total cumulative rainfall, both these variables are significant in most of the models for the Trent (28) LHA, with the exception of A2 and A4 for TN where rain is not significant. If rainfall is removed from the models, it reduces R^2 by approximately 10 per cent suggesting that both BFI and total cumulative rainfall should be retained in these models.

4.2.2 Severn (54) LHA

For the Severn (54) LHA, for which data span 1971–2009 for OP and 1967–2009 for TON and TN, there are 173 water bodies for which OP is modelled, 183 for TON and 179 for TN.

The 12 models listed in Table 4.2 were fitted for the three determinands and the corresponding R^2 values are recorded for each model in Table 4.4.

For each of the determinands, the model including all possible covariates and the trends for time and space and the seasonal pattern explain the largest percentage of the variation at 70, 58 and 71 per cent for OP, TON and TN respectively (A1 in Table 4.4). In each case the percentage of variability explained by using only the covariates from Tables 2.5–2.7 are 66, 52 and 64 per cent respectively (A2 in Table 4.4).

For all three determinands the covariates are doing very well at explaining the patterns in the nutrients. Each of the covariates or covariate groups contributes a small percentage to the variability explained and it is not clear from the models fitted here (A1–A12) that one set of covariates explains a large percentage of the variability.

Models that have either the land cover variables or the group of variables, that is, ALC, slope and soil (A3 and A4) explain similar amounts of variability (62 and 61 per cent for OP, 51 per cent for both for TON, and 60 and 62 per cent for TN respectively).

It was found that 51 per cent for OP, 49 per cent for TON and 58 per cent for TN of the variability is explained – in no particular order – by BFI, flow, population, rainfall, land cover and the group ALC/slope/soil.

Table 4.4 R^2 values for Severn (%)

Model number	OP	TON	TN
A1	70.1	57.7	70.6
A2	65.7	52.4	64.3
A3	61.9	51.0	60.0
A4	61.4	51.0	61.9
A5	52.7	43.2	50.9
A6	48.1	54.6	56.2
A7	46.0	44.5	52.7
A8	51.3	49.1	58.3
A9	46.5	39.5	48.2
A10	41.1	48.1	47.0
A11	36.4	47.8	45.5
A12	18.6	13.5	14.6

Most of the variables included in each of the models are statistically significant. However, those models with a few variables that are not significant were investigated further.

In the Severn (54) LHA, there are less covariates that appear to be related and, in comparison with the Trent (28) LHA, only potatoes and sugar beet of the crop variables indicate a strong relationship and none of the livestock variables appear to be related.

Rainfall is one of the variables most often identified as being not statistically significant. However, removing rainfall from the models invariably produces a drop in R^2 of 5–10 per cent. The non-significant result for rainfall is likely to be a consequence of the fact that there is a relationship evident between BFI and rainfall. It therefore appears from the model that rainfall is not significant though the contribution to the R^2 value suggests that it should be retained in the model.

The other non-significant variables are:

- fertiliser in A1 for both OP and TN;
- rough grazing in A3 for TN;
- arable in A1 and A2 for TON;
- BFI in A8 for TON.

Eliminating these reduces R^2 by less than 1 per cent (typically 0.01 per cent).

4.2.3 Lune (72) LHA

For the Lune (72) LHA, data for which span 1971–2009, there are only 38 water bodies for which OP, TON and TN can be modelled. This LHA contains much less data than the other two LHAs investigated, and because of the reduced amount of data, many of the variables are not significant – especially population and the ones describing the various crops.

The 12 models listed in Table 4.2 were fitted for the three determinands and the corresponding R^2 values are recorded for each model in Table 4.5.

For each of the determinands, the model including all possible covariates and the trends for time and space and the seasonal pattern explains 79, 74 and 79 per cent for OP, TON and TN respectively (A1 in Table 4.5). In each case the percentage of variability explained by only the covariates from Tables 2.5–2.7 are 75, 68 and 74 per cent respectively (A2 in Table 4.5).

Hence again it appears that the covariates are doing well in explaining the patterns in the nutrients. Each of the covariates or covariate groups contributes a small percentage to the variability explained and it is not clear from Table 4.3 that one set of covariates explains a large percentage of the variability.

It was found that 69 per cent for OP, 65 per cent for TON and 71 per cent for TN of the variability is explained – in no particular order – by BFI, flow, population, rainfall, land cover and the group ALC/slope/soil.

Table 4.5 R^2 values for Lune (72) LHA (%)

Model number	OP	TON	TN
A1	77.8	74.3	78.7
A2	74.7	68.3	74.1
A3	66.3	67.7	68.6
A4	72.7	57.3	73.2
A5	57.8	58.6	61.8
A6	65.7	62.3	68.8
A7	54.6	55.0	60.5
A8	68.5	65.0	71.4
A9	53.0	55.8	60.8
A10	41.0	47.9	52.7
A11	41.0	41.0	46.4
A12	19.5	19.4	21.8

Repeating the models A1–A12 but without the non-significant variables makes the R^2 values decrease with changes of less than 1 per cent.

Total cumulative rainfall and BFI are related. The models comparing the contribution of flow, BFI, population and rainfall (A10, A11 and A12) have all the components significant for all determinands. Total cumulative rainfall only appears to be non-significant for OP, and when removed from the models, it produces an R^2 value lower than when the BFI is removed. This implies that the total cumulative rainfall contains more information than the BFI in this LHA and should be retained over BFI. The fact that rainfall is not significant for OP suggests that the source of OP is sewage treatment works rather than agricultural runoff.

Figure 4.11 contains plots from Model A2 for OP in the Trent (28) LHA to highlight the relationships identified between the covariates and OP. The plots for other determinands are shown in section B3 of Appendix B. On these plots, the solid line highlights the fitted values with the dashed lines indicating ±2 standard errors. Small dashes on the *x*-axis indicate the distribution of the data points. The covariate is on the *x*-axis with a smooth function illustrating the fitted smooth relationship with the response on the *y*-axis.

4.2.4 Examples of complex relationships in the three LHAs

The plots highlight the complex relationships detected between the nutrients and many of the covariates. For example, for pigs and sheep there are interesting features in Figure 4.11 around mid-to-high values of the covariate which require further investigation in a more detailed study of the shape of relationships for particular LHAs. Examples of some of the relationships are discussed below for the three LHAs investigated.

OP

In Figure 4.11 for the Trent (28) LHA, relationships with sugar beet and cattle appear strongly positive with a generally negative relationship indicated for grass. The relationship for arable land appears mainly negative but becomes positive for water bodies with high quantities of contributing arable land. For ‘other animals’, a positive relationship is evident with mid-to-high counts and there is a slight indication of a positive relationship with fertiliser. The relationship with population is mainly flat for smaller population sizes but there is an indication of a positive relationship with mid-to-high population but a negative relationship with very high populations. While clear relationships are evident for BFI and flow/discharge, they are driven by potentially influential observations or sparse data distributions.

The results in the Severn (54) LHA (see section B3 of Appendix B, Figure B.26), are generally contrasting with an indication of a positive relationship with population at higher population counts. The relationship with grass is generally positive, except for at high values, and the relationship with cattle is more complex.

A few of the covariates in the Trent (28) LHA that appeared to be related to each other and hence the relationships were investigated after removing field vegetables, sugar beet, sheep and cumulative rainfall from the model. The only relationship that appeared to change was for BFI, where the positive relationship towards higher values became negative. However, since this is driven by a very small amount of data it cannot be considered a true effect.

TON

For the Trent (28) LHA (Figure B.24 in Appendix B), positive relationships are highlighted for high population counts with a strong generally positive relationship with grass, sheep, poultry and ‘other animals’. Rough grazing has a negative relationship.

In the Severn (54) LHA (Figure B.27 in Appendix B), potatoes have a generally negative relationship with generally positive relationships for maize and oilseed and indications of positive relationships at higher values of rainfall and for fertiliser.

A positive relationship with rainfall is the most apparent in the Lune (72) LHA (Figure B.30 in Appendix B).

TN

The relationship with BFI is mainly positive with a positive relationship with flow/discharge indicated at medium to high flows in the Trent (28) LHA (Figure B.25 in Appendix B). Sheep, poultry and ‘other animals’ have generally positive relationships.

In contrast the relationship with flow/discharge is mainly negative in the Severn (54) LHA (Figure B.28 in Appendix B), with indications of positive relationships for sugar beet and arable.

Again for the Lune (72) LHA (Figure B.31 in Appendix B), the suggestion of a positive relationship at higher values of rainfall is highlighted along with a strong positive relationship with grass.

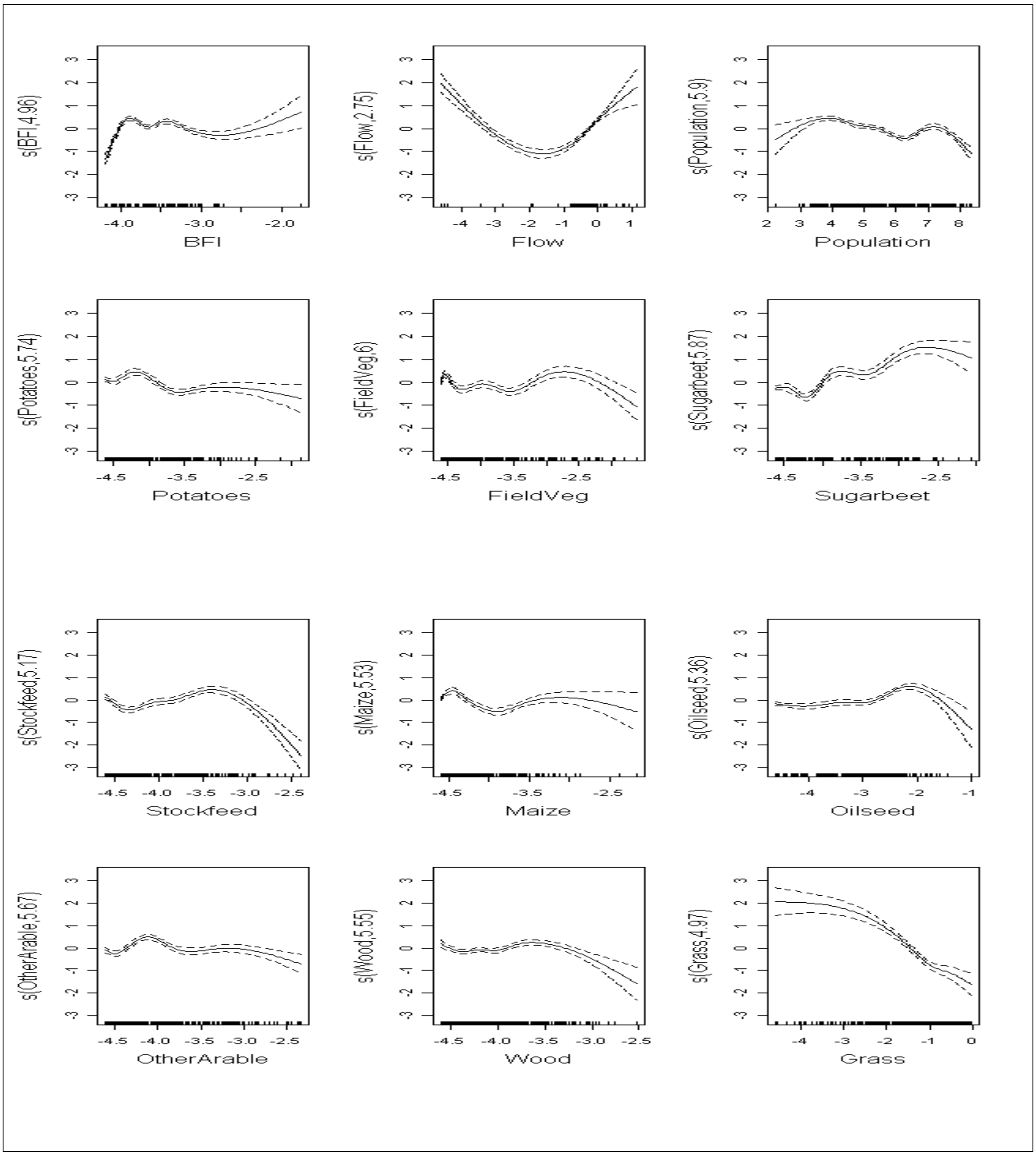


Figure 4.11 Trent In OP – Model A2

Notes: For each covariate, panels highlight the relationship with the response. The dashed lines lie at a distance of two standard errors from the estimates and the vertical dashes on the *x*-axis display the data distribution.

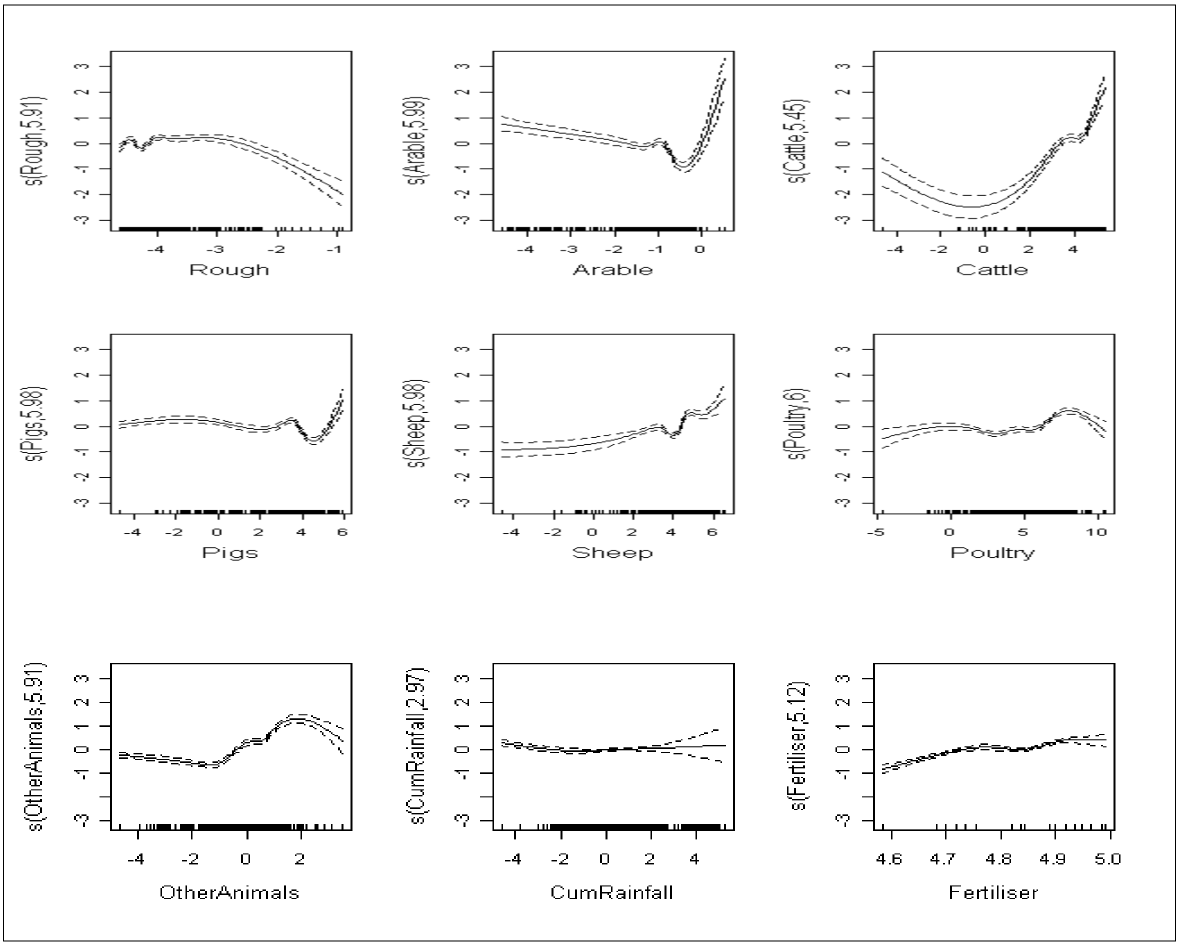


Figure 4.11 Trent ln OP – Model A2 (continued)

Notes: For each covariate, panels highlight the relationship with the response. The dashed lines lie at a distance of two standard errors from the estimates and the vertical dashes on the *x*-axis display the data distribution.

4.3 Model 3: incorporating interactions

The data from the Coquet, Wansbeck, Blyth (22) LHA was used to illustrate the modelling potential. This involves 13,517 observations over space and time. Figure 4.12 shows the results of fitting an additive model with terms for space, time (in years) and seasonal effects for a natural log transform of TON.

These estimates agree well with those listed earlier in this report and the fitted model explains 39 per cent of the variability in ln TON. The correspondence is not exact because a different method of smoothing was employed. However, it is of interest to extend this simple model to allow interactions, for example, permitting the spatial or seasonal effects to change with time. To incorporate more complex terms such as interactions, a different method of smoothing using *p*-splines was required due to the large data dimensions involved in this work.

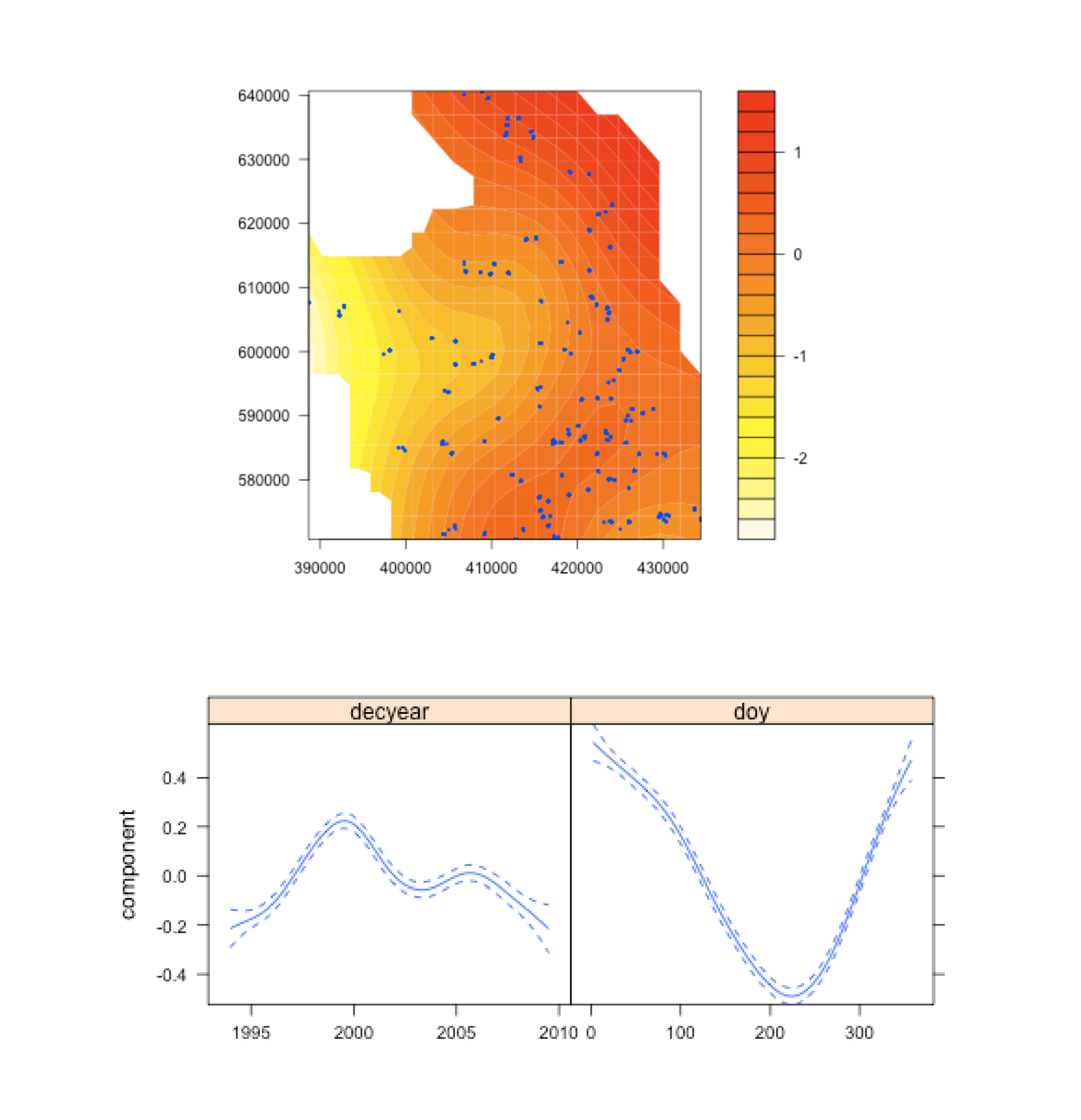


Figure 4.12 Spatial component (top panel) from a simple additive model for In TON in the Coquet, Wansbeck, Blyth (22) LHA, Easting (horizontal axis) and Northing (vertical axis). Lower panel: year trend (left) and seasonal (right) effects

Notes: The dotted lines lie at a distance of two standard errors from the estimates.

In Figure 4.13, the top left and lower panels correspond to those shown for the additive model. Although it is clear that these main effects are very similar to the previous Model 1 for this LHA, the top right panel shows one of the interaction terms – in this case between day of the year (on the vertical axis) and time (on the horizontal axis). For each position on the year axis, the values on the vertical axis show the size of the adjustment which should be added to the main effect for season. This indicates that the seasonal effect was greater in the early years (with a negative adjustment around day 200 where the main effect is smallest) and smaller in the later years (with a positive adjustment around day 200). The contours on this plot indicate the distance from zero (no adjustment) in units of standard errors. The interpretation of this result is that the amplitude in seasonal variation has decreased over this period.

The interaction between location (space) and time cannot be displayed in a single two-dimensional plot. Figure 4.14 illustrates this interaction through a series of plots which indicate the spatial adjustments that should be made at particular time points. Again, the contours on this plot indicate the distance from zero (no adjustment) in units of standard errors. Only 1994, 2001 and 2009 are displayed here.

The interpretation of this result is that there have been small changes in the distribution of the spatial estimates over the time period. Concentrations are generally lower in the north-west of the LHA but increase moving south-east through the LHA in the early years of the time period. By the end of 2009, however, concentrations appear higher in the north-east but decrease moving south-east.

The model including the interaction terms has an R^2 of 42 per cent, indicating that the more complex model only explains a small percentage more of the variability than the simple additive one. This suggests that the changes identified seasonally and spatially over time are small.

The models fitted above assume that the errors are independent. For spatial and temporal data, however, correlated errors should be considered to enable accurate inference to be performed. If correlated errors are not accounted for appropriately then standard errors may be underestimated.

This was carried out by calculating the residuals from the interaction model and fitting a simple (separable) covariance structure which decays exponentially as spatial distance and temporal distance increases in each case. It is clear from these calculation that, after the removal of spatial and temporal trends, spatial and temporal covariance in the error term is weak. It is possible, in principle, to adjust the standard errors shown in the earlier figures to account for spatial and temporal covariance. However, this is a major computational task. In view of the weak nature of the covariances involved, and because this is likely to weaken further with the incorporation of additional covariates, no adjustments have been made here.

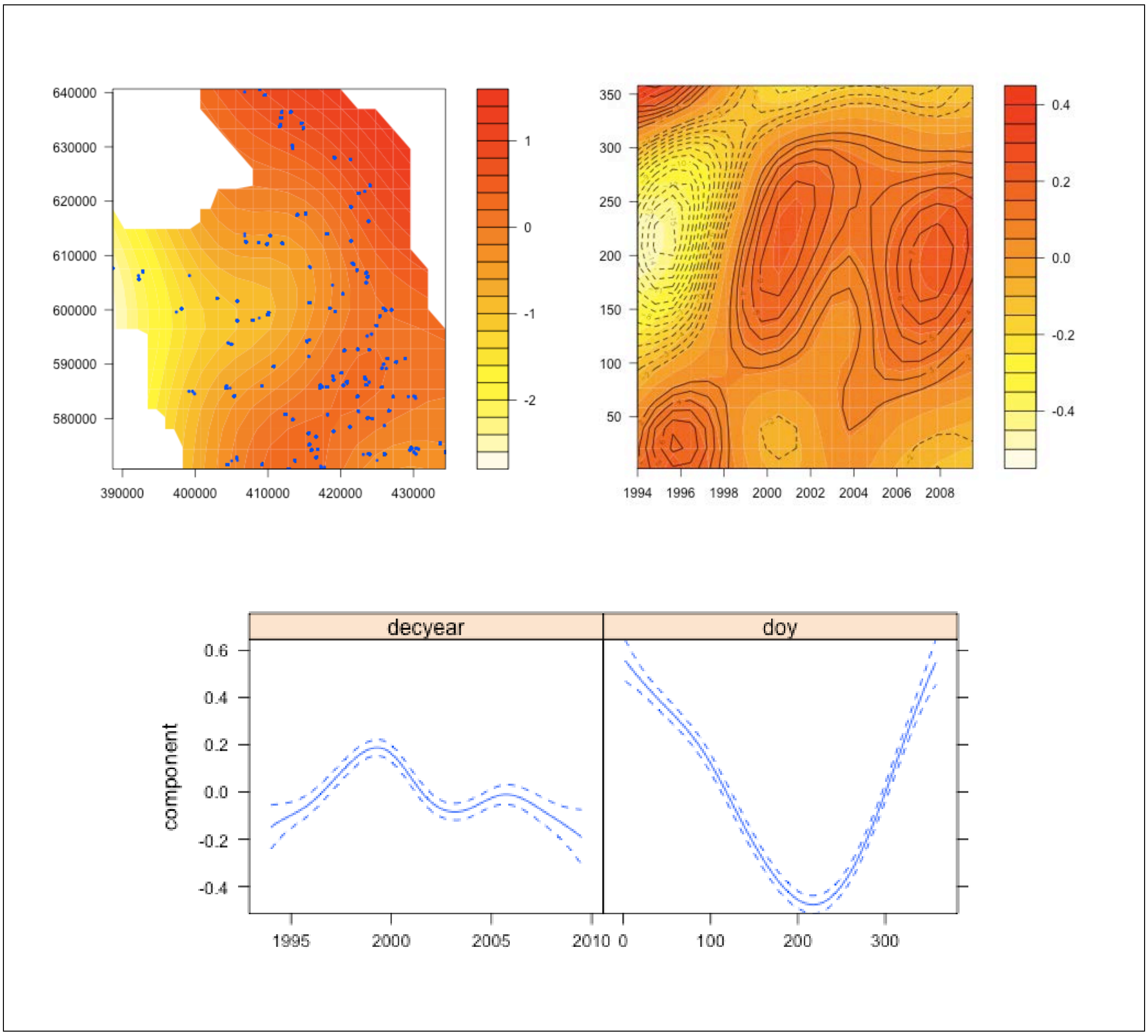


Figure 4.13 Top left and lower panels: overall spatial trend (Easting, horizontal axis and Northing, vertical axis), year trend (left) and seasonal (right) effects from the interaction model for ln TON in the Coquet, Wansbeck, Blyth (22) LHA. Top right panel: estimated interaction between the seasonal (vertical axis) and time (horizontal axis) effects

Notes: Contours indicate the distance from zero in units of standard errors.

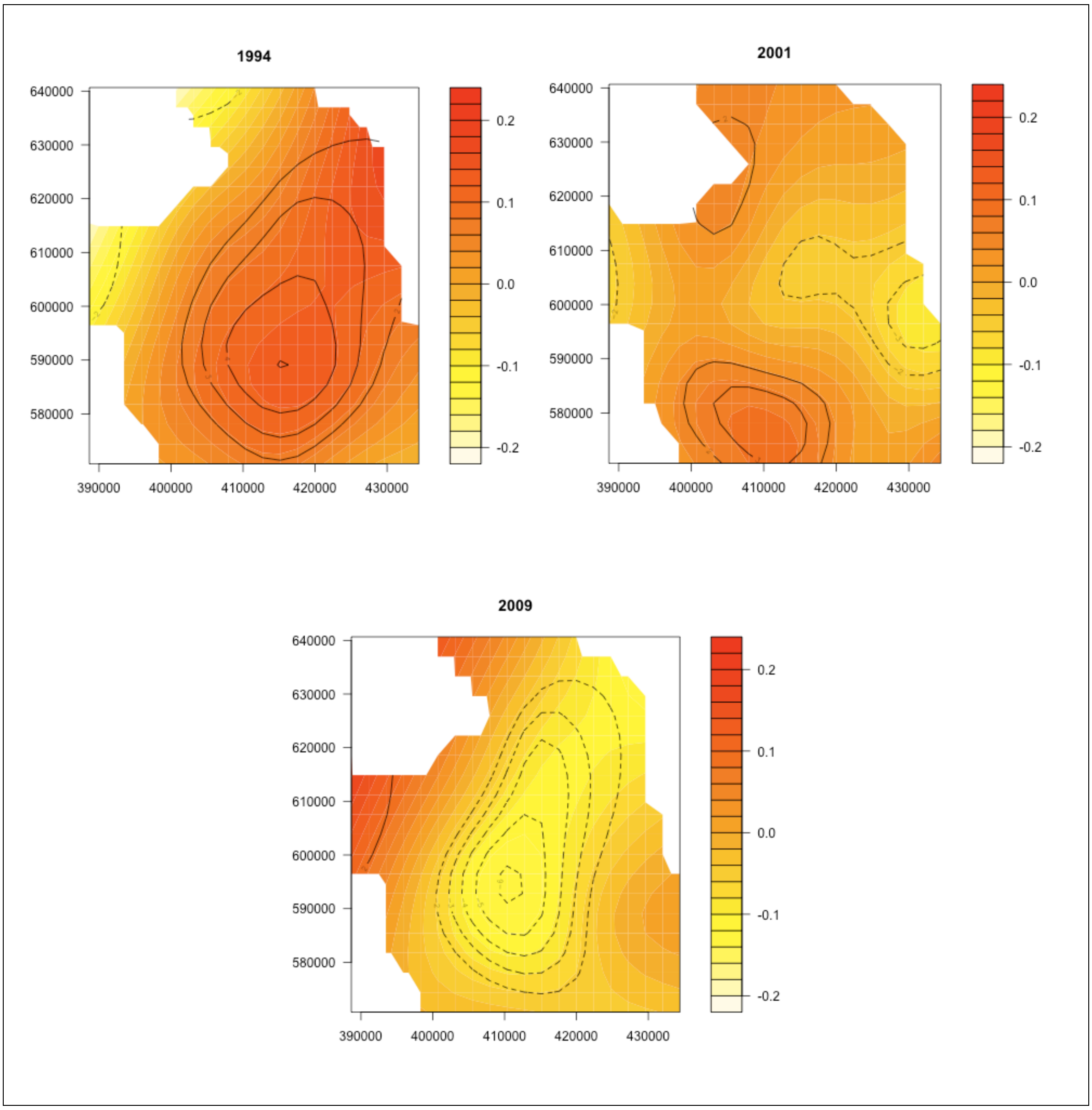


Figure 4.14 Interaction effect between location (Easting, horizontal axis and Northing, vertical axis) and time with plots for 1994, 2001 and 2009 for ln TON in the Coquet, Wansbeck, Blyth (22) LHA

Notes: Each panel shows the adjustment which should be added to the main effects.
The contours quantify the size of the effects as distance from zero (corresponding to no adjustment) in units of standard errors.

5 Discussion

This work used nonparametric models, enabling flexible smooth functions to be fitted for trends, seasonality and relationships with covariates without constraining relationships with responses to be linear. This is important since, over the period of these data, there have been large changes in the loading of nutrients to catchments. Important influences included changes in land use and land management practice such as:

- changes in the total amounts of manure and inorganic fertiliser used (N);
- changes in population;
- the use of secondary and tertiary treatments leading to the removal of phosphates from the discharges from sewage treatment works (Bowes et al. 2010; Neal et al. 2010; Young et al. 1999).

To provide greater power for the detection of trends, models have been fitted to data from monitoring locations within defined LHAs with results grouped by Environment Agency region.

5.1 Model 1

Initial models were developed for each LHA in England and Wales (with the exception of a few that were combined in Wales) to investigate trends over time, trends over space and seasonal patterns. These models highlighted the fact that OP has generally decreased over all the time periods studied with the exception of the LHA in Wales where the trend was fairly flat over recent years. In general, the seasonal pattern indicated low values at the beginning and end of the year with one peak in the summer months. This indicates the significance of high winter rainfall and river flows diluting the overall catchment phosphorus load. However, in a variety of LHAs, a trough was also evident in the seasonal pattern in spring along with a peak appearing later in the summer months.

For TON and TN, trends over time and seasonal patterns are very similar for both determinands and across all LHAs. For both TN and TON, levels are generally decreasing or fairly constant from around the year 2000 although the levels have often increased before that. There are also several LHAs in Anglian and Southern Region where very little trend is evident and a couple of LHAs in South West Region where levels increased and then appeared to have levelled off. Where there are differences in trends between TON and TN, these differences can be attributed to the contribution of ammonia.

In contrast to the seasonal pattern for OP, the data for TON and TN generally highlight high values at the beginning and end of the year with a single trough in the summer months. This pattern is believed to be characteristic of the mobilisation of mineralised nitrogen from land following autumn and winter rainfall.

Trends were more variable over LHAs for TON and TN than for OP. However, the seasonal patterns were more varied between LHAs for OP.

Within each large hydrological area, monitoring locations at the bottom of the catchment generally appear to have higher levels of TON and TN than those further up the catchments. Estimates for TON and TN appear similar. However, the distribution of spatial estimates is different for TON and TN on the coast of the Lune (72) LHA and in the south of the Severn (54) LHA. The distribution of OP estimates is also similar to

TON and TN with the higher values for OP, TON and TN appearing to be clustered around the same locations. However, the range of values for the highest estimated OP concentrations, relative to the mean of each LHA, is larger. There appears to be many more high OP values relative to the mean indicated along the north-west coast compared to TON and TN, with different spatial distributions evident in the Severn (54), Hull (26), Ancholme (29), Great Ouse (33), Tone, Parrett (52) and Frome, Bristol, Avon (53) LHAs.

For Wales, it was necessary to combine LHAs in order to estimate a spatial surface as a result of the small number of monitoring points within each LHA. In general the number of monitoring locations required to fit statistical models depends on the power to detect trends required and the spatial variability in the determinand of interest. However, an approximate guide to fit models such as those used in this project would be that more than 50 monitoring points within each LHA, with a time series of monthly data over at least 10 years, would be required in order to provide a reasonable estimate of the spatial surface, trends and seasonality. With LHAs that contain less than approximately 30 monitoring locations, spatial surfaces will become difficult to estimate accurately and cannot be estimated in the case of only a handful of monitoring locations.

For each LHA, an R^2 value was also given, which provided a measure with which to compare LHAs in terms of the percentage of the variability explained by the temporal and spatial trend and seasonal patterns in combination. This indicated that the overall prominence of the trends and seasonality are quite different in each of the LHAs.

5.2 Model 2

The second model extended the modelling framework of Model 1 to incorporate other covariates in a series of 12 models (A1–A12) in an attempt to describe the nutrient levels. In this project, water bodies were selected that had no further upstream monitoring points and contributing land LHAs that do not overlap. The covariates contained information on all land contributing to a particular water body (where appropriate). These models were fitted for LHAs that had contrasting signals in terms of trends over time – the Severn (54), the Lune (72) and Trent (28) LHAs (the Trent LHA is a NVZ). The models highlighted the fact that 66, 75 and 54 per cent respectively of the variability for OP for the three LHAs could be explained using only the catchment covariate information; the values were 52, 68 and 39 per cent respectively for TON and 64, 74 and 39 per cent respectively for TN. Models that include the temporal and spatial trends and seasonality do not explain much more of the variability. The covariates are reasonably powerful and explain much of the patterns seen in the data.

The relationships between each covariate and the nutrient levels are complex and different for each LHA, and there does not appear to be one or two variables that explain the majority of the variability; many of the covariates contribute a small proportion each. Simplifying the models indicates that BFI, flow, population, rainfall, land cover and the group ALC/slope/soil explains:

- 51, 69 and 30 per cent for respectively OP in the three LHAs studied;
- 49, 65 and 31 per cent respectively for TON;
- 58, 71 and 30 per cent respectively for TN.

Twelve models were fitted for three contrasting LHAs. The resulting relationships between the responses and covariates are specific to these LHAs, indicating the sensitivity of a river to changes in particular covariates. Strong relationships are indicated by covariate values where the interval, provided graphically, for the estimates and standard errors does not include zero.

The covariate model developed could also be fitted to monitoring points that are only tributaries/headwaters, with covariate data based only on the information from the respective water body and not containing any contributed information. Model 2C (see section 3) is the natural model to develop into a complete spatiotemporal model. But to incorporate the contributing land information appropriately, a river network covariance structure is required which was out with the scope of this project.

In this project variables such as land use, population and rainfall were standardised by using the size of the LHA over which the contributing variables had been calculated. An alternative would have been to standardise by hydrologically effective rainfall.

5.3 Model 3

The initial models here assume an independent error structure. But since the emphasis here is on investigating patterns over time, the estimates are the feature of interest and these are not affected by a correlated error structure. The form of the initial models was extended to incorporate space/time interactions and to investigate a covariance structure for one example LHA – Coquet, Wansbeck, Blyth (22). The results highlighted a small change in both the seasonal and spatial patterns over time for this particular catchment. There was little structure remaining in the errors and hence incorporating a more complex covariance structure was not justified. The effects of the interaction terms were small, as indicated by the small change in the percentage of the variability explained. Therefore, a simpler additive model may be adequate to capture the temporal, spatial and seasonal patterns in this particular example LHA.

6 Future work

The models presented in this report to investigate trends over time, trends over space and seasonal patterns for all LHAs in England and Wales do not incorporate interaction terms or spatial covariance structures. The methodology developed for this project was illustrated by an example LHA. However, incorporating these features is not trivial and becomes extremely complex as the data dimensionality increases. Future work could extend this model to all LHAs by developing techniques to deal with the data dimensionality. This could potentially be useful in investigating changes in seasonal and spatial patterns over time. The example LHA suggests small changes in spatial patterns and seasonality over time, but this is a small LHA with a time series from 1994. The development of methods for fitting interactions for highly dimensional data could enable models to be fitted for LHAs with many monitoring points and long time series to investigate how changes in the loading of nutrients over time has impacted changes in spatial and seasonal estimates.

The modelling in this report has indicated several difficulties when data become sparse in space or time. The sparseness of the monitoring points in the LHAs for Wales resulted in these LHAs being grouped into two clusters. Where there were less data available, seasonal patterns and trends were more difficult to identify – as highlighted by large standard errors on the plots. However, it was necessary to grid data for those LHAs that had an overall data dimension of >30,000 observations, and hence for these LHAs, not all the available data were used in the estimation. Continued methodological development is required to enable higher dimensional data to be analysed. A full simulation study would be required to estimate an optimal number of monitoring locations and temporal monitoring frequency, which would depend on the specification of objectives for the monitoring. Recent work at the University of Glasgow has investigated optimal monitoring strategies over time for linear, non-linear and nonparametric trends and varying seasonal components, but this would require to be extended spatially.

The percentages of variation (R^2) explained by each of the fitted models for Model 1 indicated the percentage of the variability that can be explained by the temporal and spatial trends and seasonal patterns in combination for each LHA. These percentages provide a measure with which to compare the signals for each of the LHAs across England and Wales. The associated plots containing estimates and standard errors for trend over time and seasonal patterns illustrate where the estimates are significantly different from zero. However, future work could involve formal model testing of the statistical significance of trends and seasonal patterns and comparison between LHAs. Such methods would enable approximate inference to be carried out to assess whether nonparametric trends and seasonality are significant, are changing throughout time, and if such terms are possibly linear.

To incorporate catchment information into the models, information from the contributing land was incorporated into the covariates and the water bodies restricted to an appropriate subset. This along with terms for time trend, spatial trend and seasonality describes a reasonable amount of the variability here, assuming independent errors. The most relevant covariance structure for these models would take account of the river network structure. However, this was outside the scope of this project. All the models described in this report could be developed to incorporate such a structure. A river network covariance structure would be required for the spatiotemporal Model 2C. This would enable a spatiotemporal model to be constructed using data from all water bodies (or from points within the river network) in a particular LHA as the response and to incorporate information on the influence of different land uses and other catchment variables might have on observed water quality temporal and spatial trends. Such covariance structures use both river distance and Euclidean distance to model the flow direction and network structure appropriately. Additional catchment information could be incorporated within such models with the aim of describing even more of the variability in observed water quality.

It would be beneficial to develop a hierarchical model for these data that could model all LHAs in the country simultaneously while accounting for the different spatial structures, that is, water bodies within LHAs and LHAs within England and Wales. This would enable a full assessment of differences within and between catchments.

Spatiotemporal modelling remains a complex area in statistics with many challenges and the possibilities for many future developments as mentioned above. However, the models and results from this project could provide the basis for all future work having successfully identified patterns within all LHAs and explained a substantial percentage of the variability in specific LHAs using catchment covariates.

References

Bowes, M.J., Neal, C., Jarvie, H.P., Smith, J.T., Davies, H.N. (2010). Predicting phosphorus concentrations in British rivers resulting from the introduction of improved phosphorus removal from sewage effluent. *Science of the Total Environment*, 408, 4239-4250

BOWMAN, A.W., GIANNITRAPANI, M. and SCOTT, E.M., 2009. Spatiotemporal smoothing and sulphur dioxide trends over Europe. *Journal of the Royal Statistical Society*, *Series C*, 58 (5), 737-752 .

EILERS, P.H.C. and MARX, B.D., 1996. Flexible smoothing with B-splines and penalties. *Statistical Science*, 11(2), 89-121.

HELSEL, D.R., 2005. More than obvious: better methods for interpreting nondetect data. *Environmental Science and Technology*, 39(20), 419A-423A.

Neal, C., Jarvie, H.P., Withers, P.J.A., Whitton, B.A., Neal, M. (2010). The strategic significance of wastewater sources to pollutant phosphorus levels in English rivers and to environmental management for rural, agricultural and urban catchments. *Science of the Total Environment* 408, 1485-1500

LEE, D.-J. and DURBAN, M., 2011. P-spline ANOVA-type interaction models for spatio-temporal smoothing. *Statistical Modelling*, 11(1), 49-69.

O'DONNELL, D., BOWMAN, A.W., SCOTT, E.M. and HALLARD, M., 2009. Stream distance based prediction on river networks, submitted to *Applied Statistics*.

WOOD, S.N., 2006. *Generalized Additive Models: An Introduction with R*. Chapman & Hall, Boca Raton, FL.

Young, K., Morese, G.K., Scrimshaw, M.D., Kinniburgh, J.H., MacLeod, C.L., Lester, J.N. (1999) The relation between phosphorus and eutrophication in the Thames catchment, UK. Science of the Total Environment, 228: 157-183

List of abbreviations

ALC	Agricultural Land Classification
BFI	baseflow index
GIS	geographical information system
GQA	General Quality Assessment
LHA	large hydrological area
ln	natural log transform
NSRI	National Soil Resources Institute
NVZ	Nitrate Vulnerable Zone
OP	orthophosphate-P
TN	total nitrogen
TON	total oxidised nitrogen
wb	water body
WFD	Water Framework Directive

Appendix A Technical appendix

Spatiotemporal Modelling of Nitrate and Phosphorous for River Catchments Technical Appendix

This document contains further details on the data aggregations performed and models developed and implemented for the project above.

1 Response Data

It was agreed that where determinand code 0116 (Total Oxidised Nitrogen, see Table 2.2 of the main report) was missing the sum of 0117 (Nitrate as N) and 0118 (Nitrite as N) would be used. If both 0118 and 0116 were missing at a particular monitoring point then 0117 would be used. There are a few cases where 0117 and 0118 are both marked as having values that are below the limit of detection (non-detects). The measurements are larger for 0117 than 0118 and hence, in cases where only 0118 is marked as a non-detect, then the sum of 0117 and 0118 is not marked as a non-detect. When 0117 is a non-detect the sum of 0117 and 0118 is also marked as a non-detect and imputed (see section 4 below for more details on non-detects).

In general for the response data, there were also occasional values that were marked as being above the limit of detection. However, the recorded values were small relative to the larger values in the respective determinands. These values were also imputed, therefore, as described in section 3 of the main report and in section 4 below. In a small number of cases an addition sign was observed next to the measurement and in these cases the recorded value was taken as the measurement.

A few zero measurements were found in the response data or in the coordinates of the monitoring locations and in these instances these entries were excluded from any analysis.

2 River Network

There were a couple of problems with items in the tree structure that defines the contributing land for monitoring points. Two separate tree structure numbers were duplicated such that

two different waterbodies with different id numbers had the same tree structure number. Therefore, the tree structure numbers for the two waterbodies were re-coded:
Waterbody ID GB107041011970 - changed the tree-structure id to 601/12 and
Waterbody ID GB107041013130 - changed the tree-structure id to 601/13.

3 Covariate Data

Slope and ALC (Agricultural Land Classification) were originally compositional variables, see Table 2.5 in the main report. For each of these variables, there were separate columns of data corresponding to each category, with the column entries in percentages of the km^2 grid. Variables of this type have a couple of complications for statistical modelling. The categories for each variable will be highly related and will sum to 100% across the area being considered. The result of this is that variables cannot be included in their present form in the statistical models. To overcome this, slope and ALC have been coded as categorical variables (see Table 2.5 in the main report for details). An alternative approach, statistically, would have been to create new variables by computing log ratios of each level of the compositional variable with respect to one of the other levels, see Aitchison (1986) for details. However, this would have resulted in variables that were more difficult to interpret.

For rainfall, the following procedure was used to extrapolate the data to all waterbodies of interest. Many of the waterbodies had multiple rainfall sites within them. A single rainfall site was selected to represent precipitation within the waterbody based on identifying the point with the longest time series of data. The data for this time series were then aggregated to obtain monthly total rainfall in mm/month. Waterbodies without rainfall sites within them were linked to rainfall sites that were within 5km of the waterbody; it was decided that it was reasonable to assume that the rainfall recorded at these sites could be considered representative of more sites than those within the waterbody.

In order to estimate the total contributing rainfall for a particular waterbody, the relationship between the total rainfall for an individual waterbody and the total rainfall from the contributing area, in rainfall data sources with national coverage, has been used. For this the Met. Office annual average rainfall for the period 1961-1990 has been used, which is a 1km^2 dataset and:

- the total annual average rainfall for each waterbody has been calculated;
- the upstream accumulation technique used with the other covariates has been used to calculate the total annual average rainfall from the upstream area;
- a factor which represents how much on average during the period 1961-1990 the total from the entire contributing area was greater than that of the individual waterbody has been calculated;
- each monthly total has been multiplied by this factor.

There were a couple of other minor modifications to the covariate data. In cases where dates were recorded wrongly, the corresponding data were removed (e.g. rainfall had a year of 6006). In cases where two different values were recorded for the same waterbody for flow/discharge and BFI the higher of the two values has been used for this analysis. For landuse and landcover, data were available from a variety of sources: DEFRA, CEH, ADAS and CSF. For all analysis here the DEFRA files have been used.

4 Non-Detects

If the proportion of non-detects in a specific LHA is small for a specific determinand, i.e. less than 3%, the less-than values are substituted by the recorded value, and if the proportion of non-detects is large, i.e. greater than 50%, then the LHA has not been analysed for that determinand.

For LHAs with a proportion of non-detects between 3% and 50%, imputation is a more appropriate method than simple substitution (Helsel, 2005). A series of parametric, non-parametric and robust statistical approaches were implemented to explore the effect of non-detects on the distribution of the data. In these methods the measurements that are recorded as being below the limit of detection are treated as censored observations. Here it is assumed that the measurements are less than or equal to the stated value. The following methods: Maximum Likelihood Estimation (MLE), Regression on Order Statistics (ROS) and Kaplan-Meier (KM) were compared across determinands (see Helsel (2005) for full details on these approaches). Each method estimates the summary statistics for the response of interest i.e. the mean, median and standard deviation of the distribution of the data, where the measurements below the limit of detection are incorporated as censored observations.

It was decided that the nonparametric methods (KM and ROS) would be applied since they require less assumptions about the data and in trials provided superior results. The summary statistics obtained with the KM and ROS methods were quite similar. Therefore, the ROS method (Helsel, 2005) was applied using the `NADA` package in `R`. The method was applied to all of the sites for the specific determinand of interest in the relevant LHA in order to estimate the mean and standard deviation of the distribution of the data for the specific determinand in the specific LHA, incorporating the censored observations and assuming the data are from a lognormal distribution. The estimated means and variances were then used to simulate data from a lognormal distribution with these parameters and the resulting simulated values were then used to impute measurements for the non-detects i.e. the original values, which were marked as non-detects, were replaced with these simulated values, which were constrained to be less than or equal to the original value.

5 Gridding

In LHAs with an overall data dimension of >30,000 observations gridding has been performed in both time and space to reduce the data dimensionality. An example of this is given in Figure 1 for the LHA of the Great Ouse. In this case the data are gridded by taking every second monitoring location in space and every second time point. Figure 1 highlights that even when the data are gridded the data are still representative of the entire region.

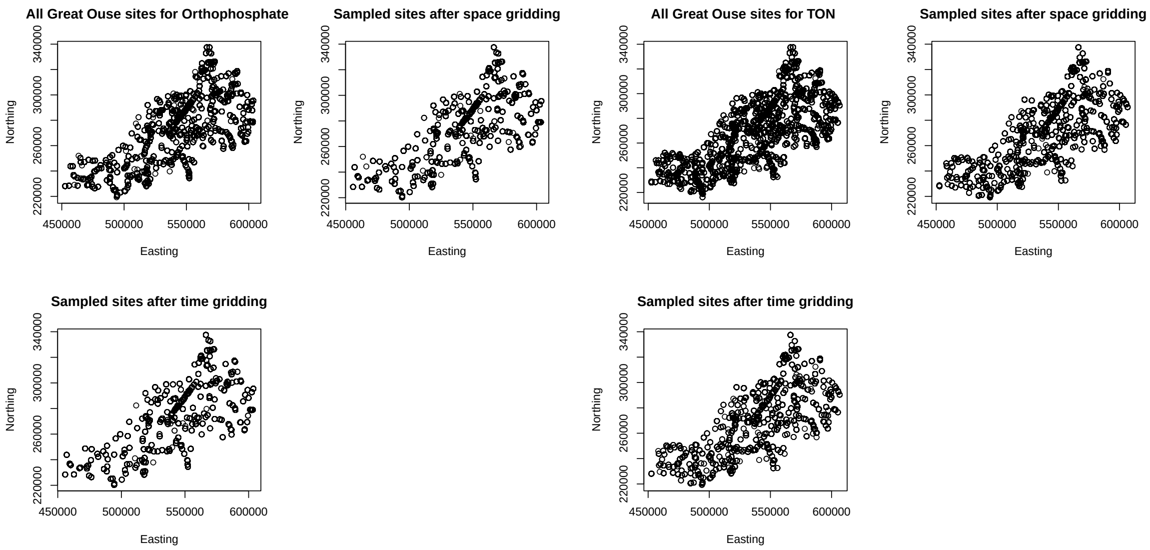


Figure 1: Monitoring locations for the Great Ouse (LHA 33) for Orthophosphate-P (left) and Total Oxidised Nitrogen (right). In both cases the plot at the top left shows all monitoring locations, top right: the monitoring locations after gridding spatially and bottom left: the monitoring locations after gridding in space and time.

Gridding was applied to fit the models in the LHAs contained in Table 1.

6 Spatiotemporal Modelling

6.1 Trends over time, space and seasonality

An additive model has been fitted for each determinand in each LHA. The response variable is expressed as a sum of smooth functions of the space and time components:

$$y = \alpha + s(\text{Easting}, \text{Northing}) + s(\text{year.day}) + s(\text{doy}) + \epsilon \tag{1}$$

where doy is the day of the year and year.day = year + doy/366.

This model assumes additive effects and fits a smooth spatial surface, a smooth temporal trend and a smooth seasonal pattern and the errors are assumed to be independent,

LHA number	LHA name
27	Ouse, Humber Estuary
28	Trent
33	Great Ouse
34	Bure, Waveney
37	Blackwater, Chelmer
39	Thames
40	Medway, Stour
41	Arun, Ouse, Cuckmere
42	Test
43	Avon(Hants)
44	Piddle, Frome
45	Exe
46	Dart
47	Tamar
48	Fal
50	Torridge, Taw
52	Tone, Parrett
53	Frome, Bristol, Avon
54	Severn
69	Mersey
71	Ribble

Table 1: Gridding was applied to the data in each of these LHAs before statistical analysis was performed.

$\epsilon \sim N(0, \sigma^2)$. For each response here a natural log transform has been applied to stabilise the variance and hence $y = \ln(\text{Orthophosphate-P})$, $\ln(\text{Total Oxidised Nitrogen})$ or $\ln(\text{Total Nitrogen})$.

An `R` function, developed at Glasgow, called `sm.additive` (Bowman et al., 2009) has been used in order to fit each of these models.

In order to fit this model the amount of smoothing for each of the terms in the model has to be chosen. In all cases here the smoothing parameters have been chosen by declaring a required degrees of freedom for each of the smooth functions. These have been set to default values of 6 for a univariate component such as year.day and 12 for a bivariate component such as the spatial surface to provide a reasonable degree of smoothness. The smooth term for day of the year is computed using a cyclical smoother since day of the year is a cyclic component and the estimates on the 31st December should smoothly lead into the 1st January.

For full details of the fitting procedures and estimation of the smooth functions see Bowman et al. (2009). The benefit of fitting the models through this approach is that the model can be extended to incorporate temporal and spatial correlation in the errors and interaction terms. While not trivial, temporal and spatial correlation could be incorporated into the fitting using the `sm.additive` function. The methodology has been developed to fit interaction terms using a different method of smoothing using p-splines (Eilers & Marx, 1996). This method has been employed here to enable the fitting of interaction terms. However, these methods are still under development and the results in the report are provided as an example of the potential of the techniques.

In order to produce a simple measure to compare LHAs, the proportion of the variability explained by the fitted model (R^2) for Model 1 has been computed. This indicates the proportion of the variability in each LHA than can be explained by the smooth spatial surface, the smooth temporal trend and the smooth seasonal pattern in combination. An R^2 value is computed by comparing Model (1) above to a model that only includes a mean:

$$y = \alpha + \epsilon$$

6.2 Describing trends

The modelling approach to incorporate the covariates used a different `R` function of `gam` in the `mgcv` library of `R`. This is the same type of model as that fitted in Model (1). However, the fitting procedures are slightly different, see Wood (2006) for full details. This methodology is well estabilished in `R` and has the advantage of being able to handle large datasets with many covariates efficiently. However, these techniques cannot be extended as naturally, to account for correlation in the errors or interaction terms, as the methods used above. For the models that incorporate many covariates there is unlikely to be much correlation remaining in the residuals and hence the assumption of independence here appears appropriate.

Covariate models are of the form:

$$\begin{aligned} y &= \alpha + s(\text{Easting}, \text{Northing}) + s(\text{year.month}) + s(\text{month}) + s(\text{flow}) + s(\text{BFI}) \\ &\quad + s(\text{landuse}) + s(\text{landcover}) + \text{ALC} + \text{Slope} + \text{Soil} \\ &\quad + s(\text{rainfall}) + s(\text{population}) + s(\text{fertiliser}) + \epsilon \end{aligned}$$

where for landuse and landcover a series of different covariates are included individually such as potatoes, field vegetables, cows, pigs etc. The smooth terms in the model above are simply to indicate that these groups of variables are included. Categorical variables are incorporated into the model for ALC, slope and soil.

This model assumes additive effects and the errors are assumed to be independent, $\epsilon \sim N(0, \sigma^2)$.

In this R function the degree of smoothing is automatically selected by the model fitting procedure. However, it has been constrained here to allow a maximum of 6 degrees of freedom for each univariate component in a similar way to Model (1). Since the main emphasis here was on the shape of relationships and contribution of a covariate to describing the nutrient responses, for some covariates, where the data distribution was slightly sparse, the smoothing was constrained further to prevent the relationships detected from becoming too 'wiggly'.

7 Software

All of the statistical analysis was performed in R (version 2.9.2). The data were provided in Microsoft Access 2003 database files and small manipulations of the data were performed in this. ArcGIS (version 9.3.1) was used to map the results of the analysis.

8 References

Aitchison, J. (1986). *The Statistical Analysis of Compositional Data.* Chapman & Hall.

Bowman, A.W., Giannitrapani, M. and Scott, E.M. (2009). *Spatiotemporal smoothing and sulphur dioxide trends over Europe.* Journal of the Royal Statistical Society, Series C, 58, 5, 737-752.

Eilers, P.H.C and Marx, B.D. (1996). *Flexible Smoothing with B-Splines and Penalties.* Statistical Science, 11(2), 89-121.

Helsel, D.R. (2005). *Nondetects and data analysis: statistics for censored environmental data.* New York: Wiley-Interscience.

Wood, S.N. (2006). *Generalized Additive Models, An Introduction with R.* Chapman & Hall, Boca Raton.

Appendix B Results appendix

Spatiotemporal Modelling of Nitrate and Phosphorous for River Catchments

Results Appendix

This document contains further plots that are referred to in the report but could not be included due to the large number of results.

Section 1

Section 1 contains plots to illustrate the results from fitting Model 1, see section 3 of the main report, for each LHA grouped by region for the responses of a natural log transform (ln) of Orthophosphate-P (OP), Total Oxidised Nitrogen (TON) and Total Nitrogen (TN), with the exception of the Southern region which is presented in the main report.

The solid lines indicate the estimates for the trend (left plot) and the seasonality (right plot), after an overall mean ln concentration for each LHA has been removed, and the dashed lines indicate a distance of two standard errors from the estimates. Wider standard errors at the beginning and end of the trend plots indicate edge effects and less data in the early years.

The smoothing parameters are chosen for most of the models using 6 degrees of freedom but where the trends or seasonal patterns appeared to be too “wiggly”, which is likely to be a result of sparser data, further smoothing was applied by setting the degrees of freedom to 3 (e.g. TON and TN in the LHA of Dovey(64)).

Figure A.1: Trend over time (left) and seasonality (right) for ln Orthophosphate-P in the Midlands region (LHAs Trent(28), Severn(54); Dovey(64) has insufficient Orthophosphate-P data for Model 1)

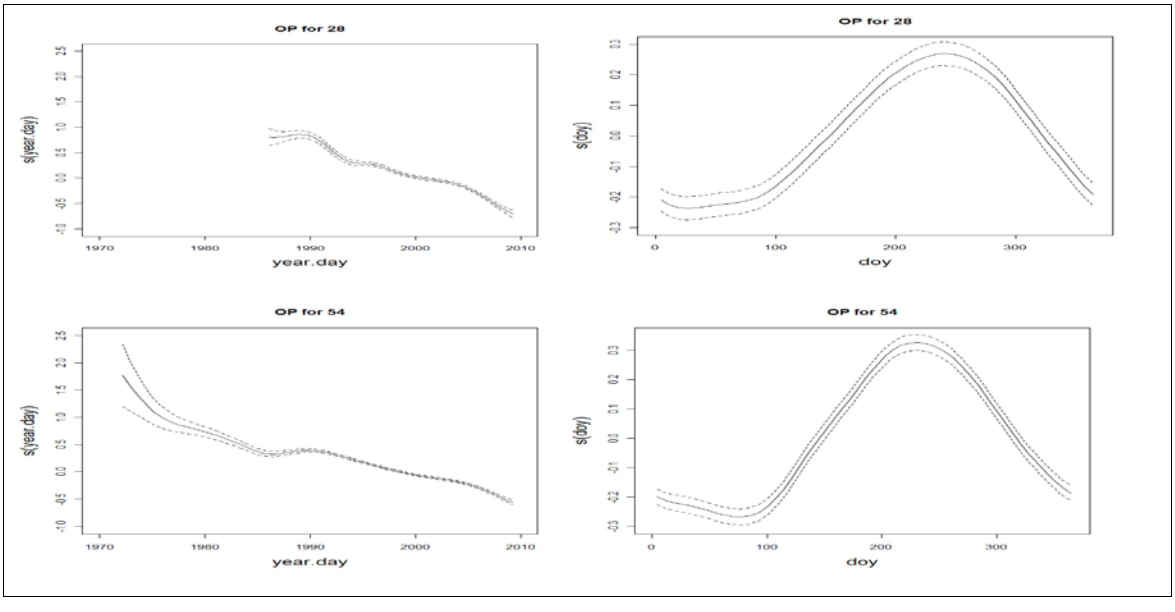


Figure A.2: Trend over time (left) and seasonality (right) for ln Total Oxidised Nitrogen in the Midlands region (Trent(28), Severn(54), Dovey(64))

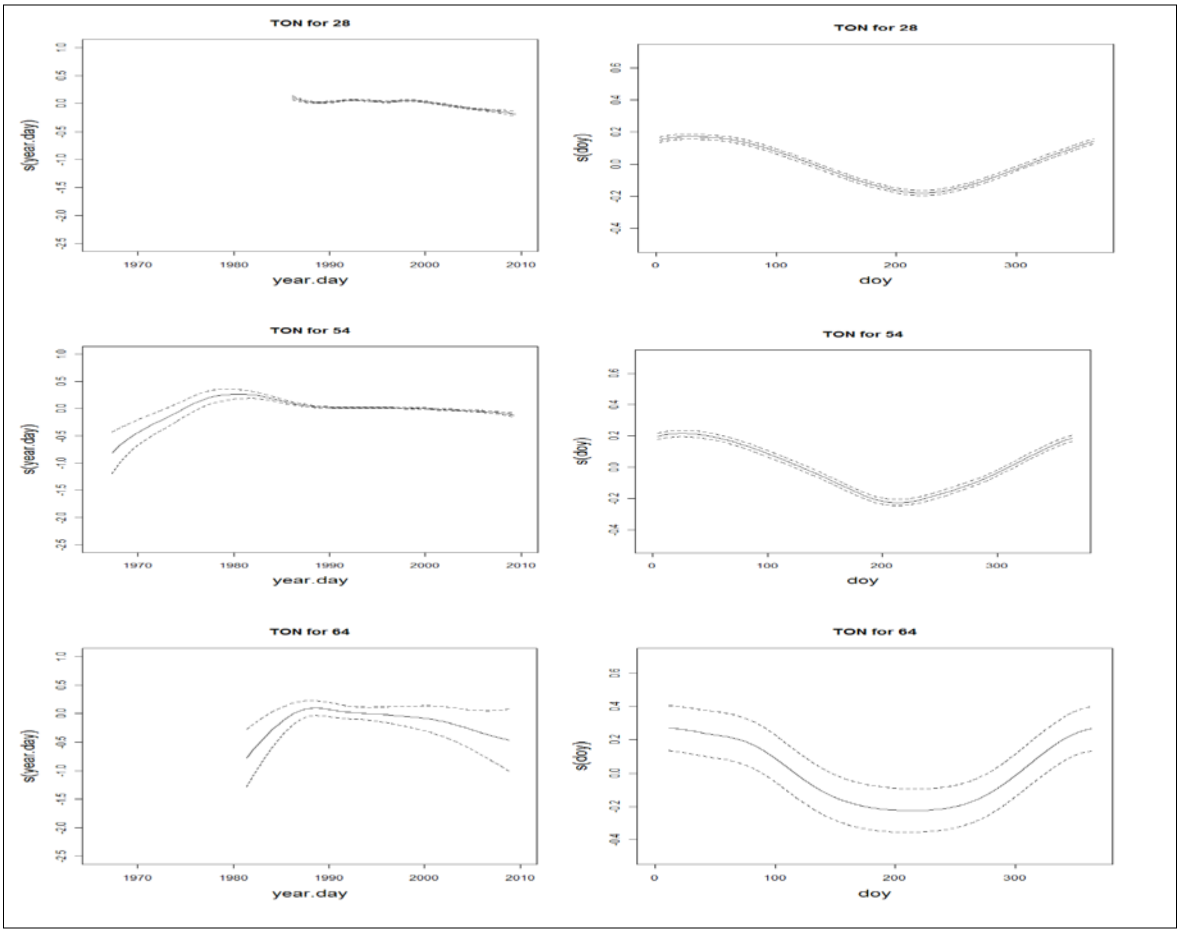
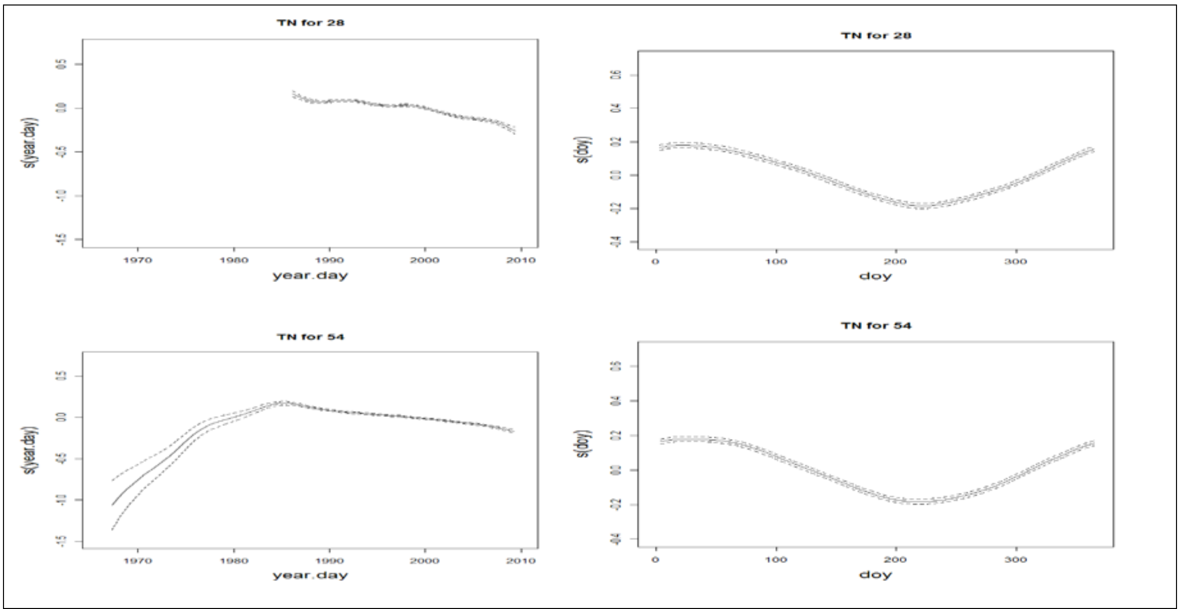


Figure A.3: Trend over time (left) and seasonality (right) for ln Total Nitrogen in the Midlands region (Trent(28), Severn(54), Dovey(64))



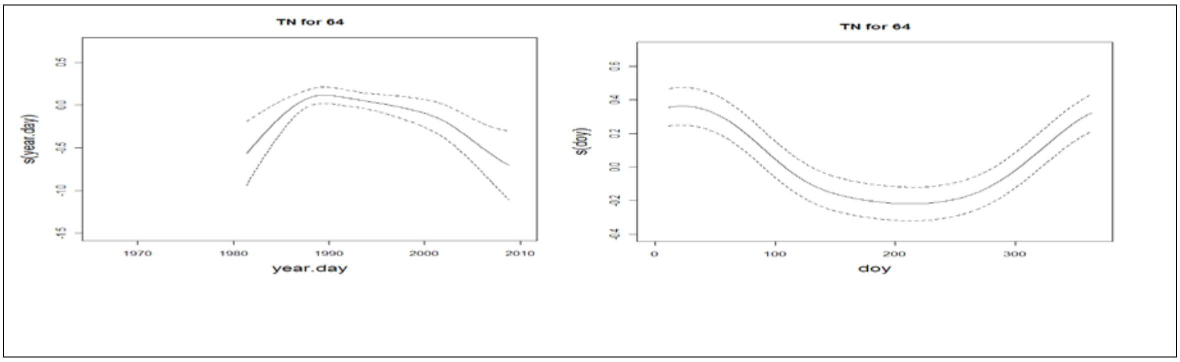
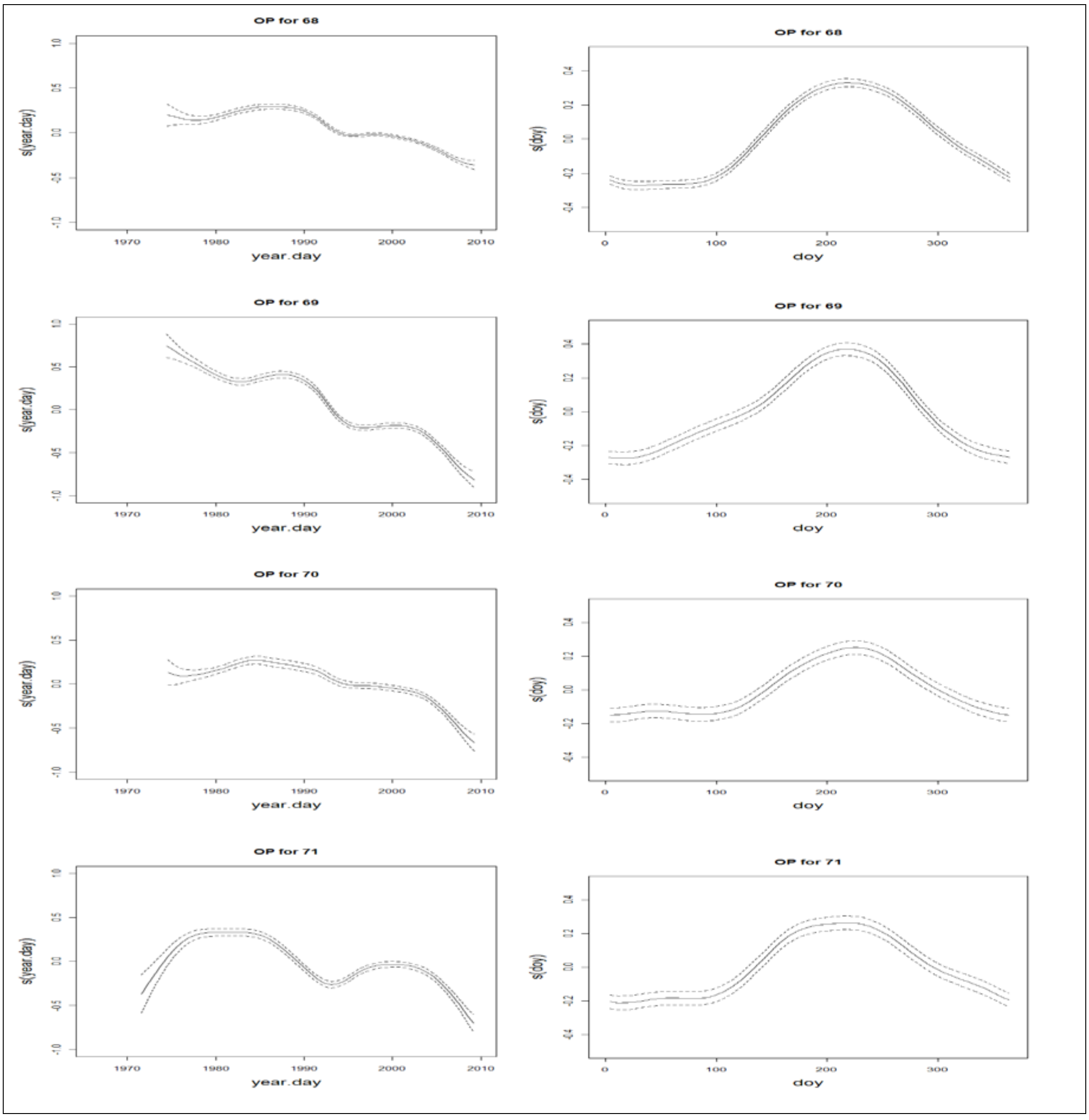


Figure A.4: Trend over time (left) and seasonality (right) for ln Orthophosphate-P in the North West Region (Weaver(68), Mersey(69), Douglas(70), Ribble(71), Lune(72), Leven,Kent(73), Duddon(74), Derwent(75), Eden(76), Lyne,Esk(77) LHAs)



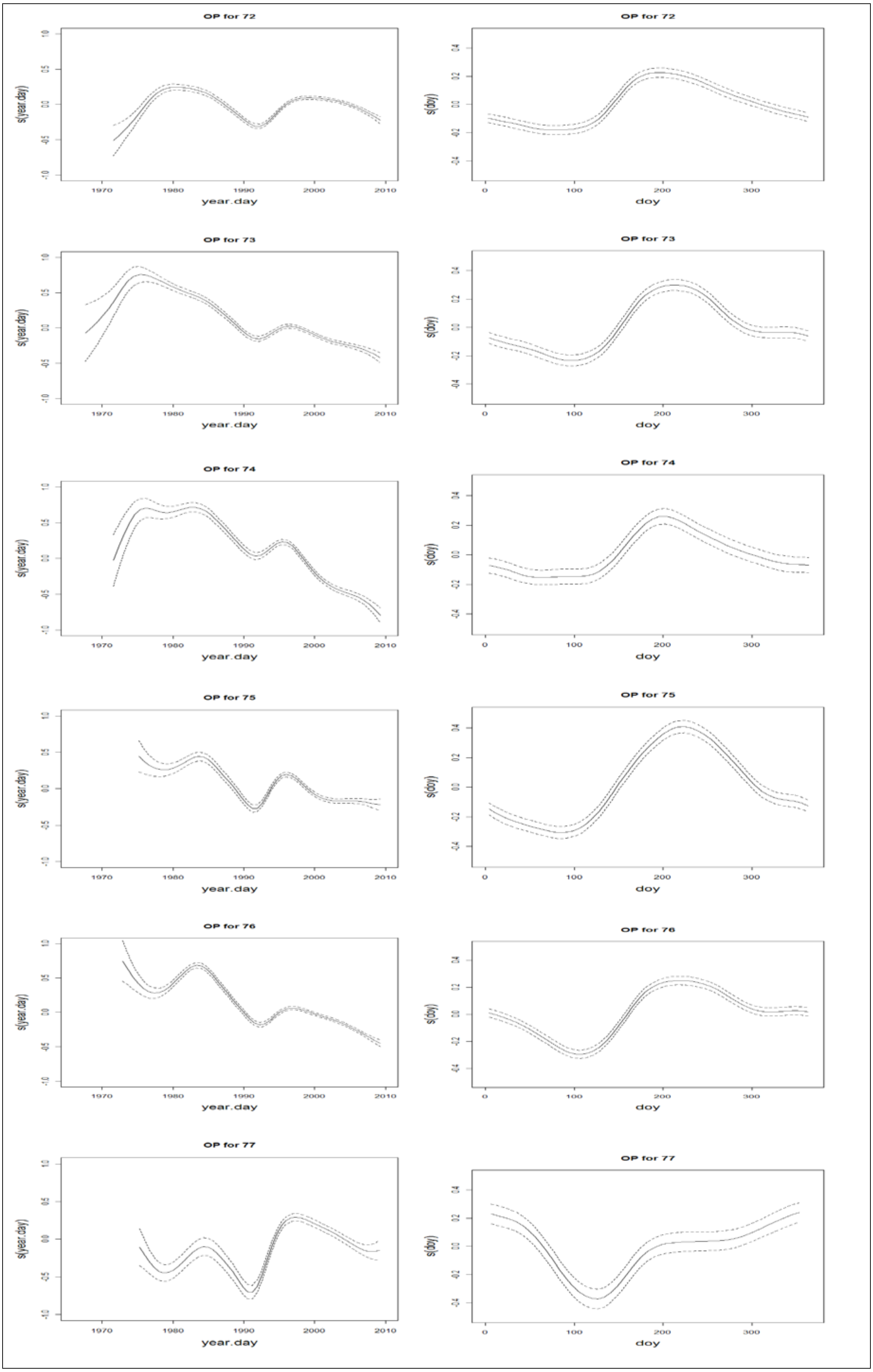
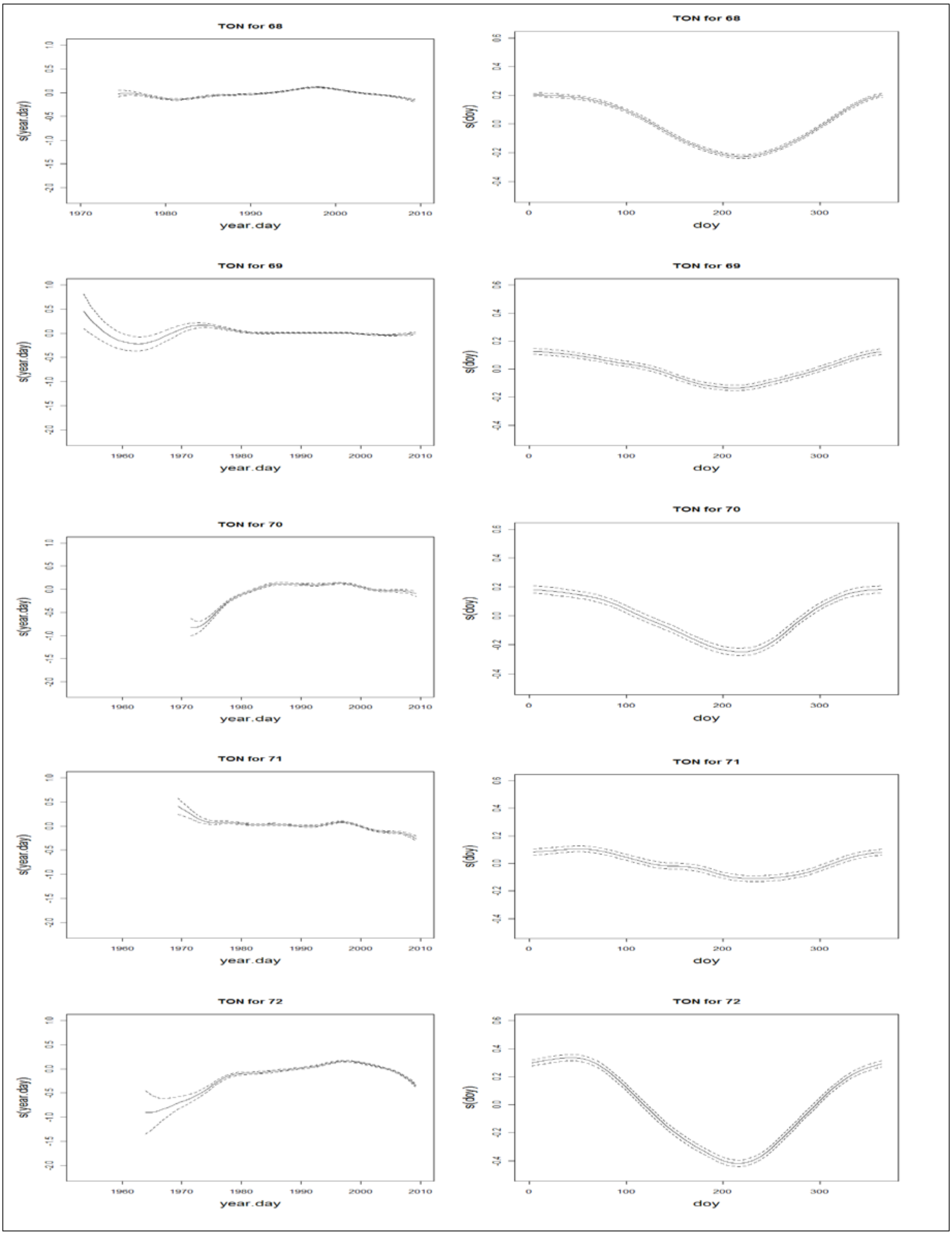

OP for 72
s(year.day)
year.day
OP for 72
s(doy)
doy
OP for 73
s(year.day)
year.day
OP for 73
s(doy)
doy
OP for 74
s(year.day)
year.day
OP for 74
s(doy)
doy
OP for 75
s(year.day)
year.day
OP for 75
s(doy)
doy
OP for 76
s(year.day)
year.day
OP for 76
s(doy)
doy
OP for 77
s(year.day)
year.day
OP for 77
s(doy)
doy

Figure A.5: Trend over time (left) and seasonality (right) for ln Total Oxidised Nitrogen in the North West Region (Weaver(68), Mersey(69), Douglas(70), Ribble(71), Lune(72), Leven,Kent(73), Duddon(74), Derwent(75), Eden(76), Lyne,Esk(77) LHAs)



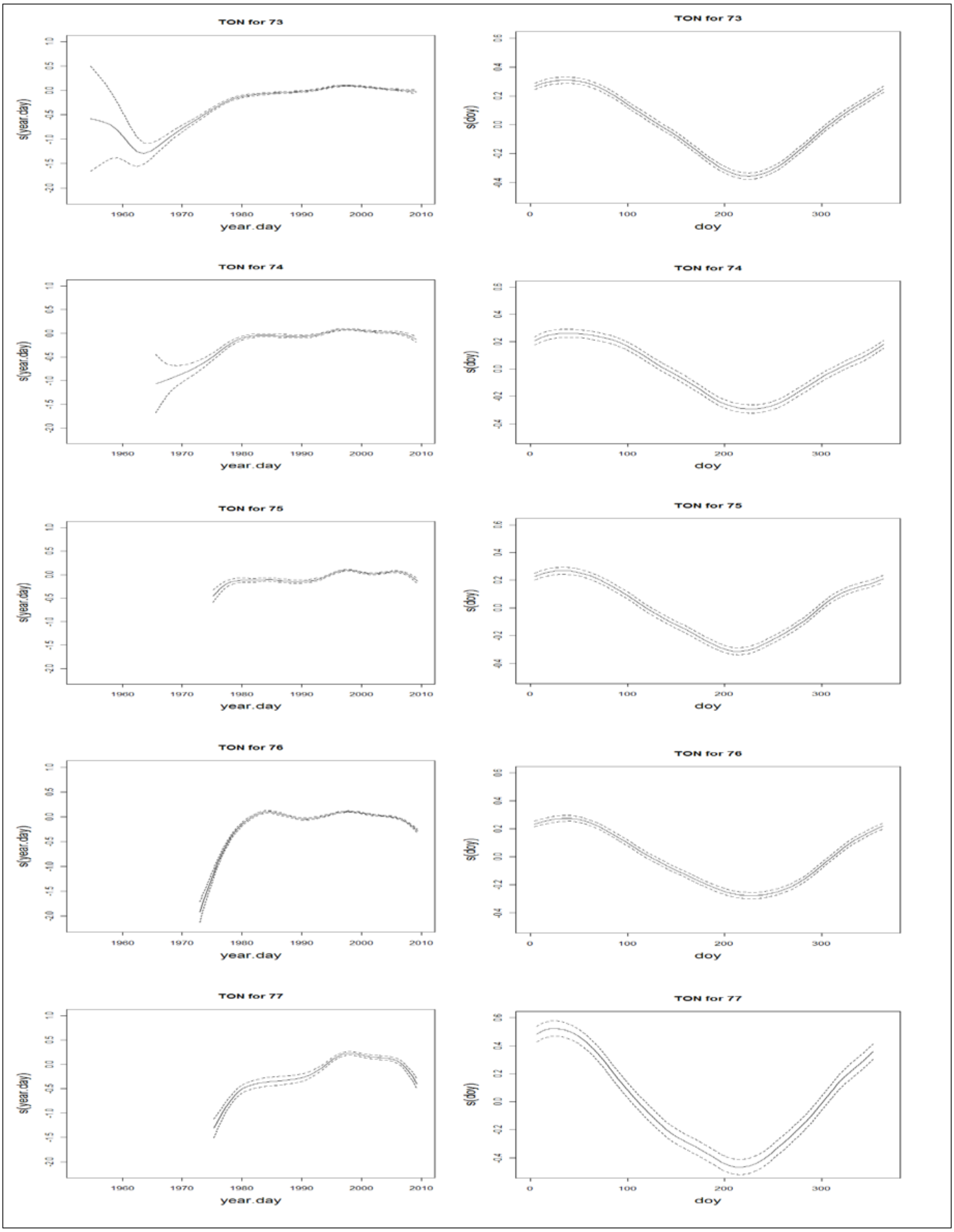
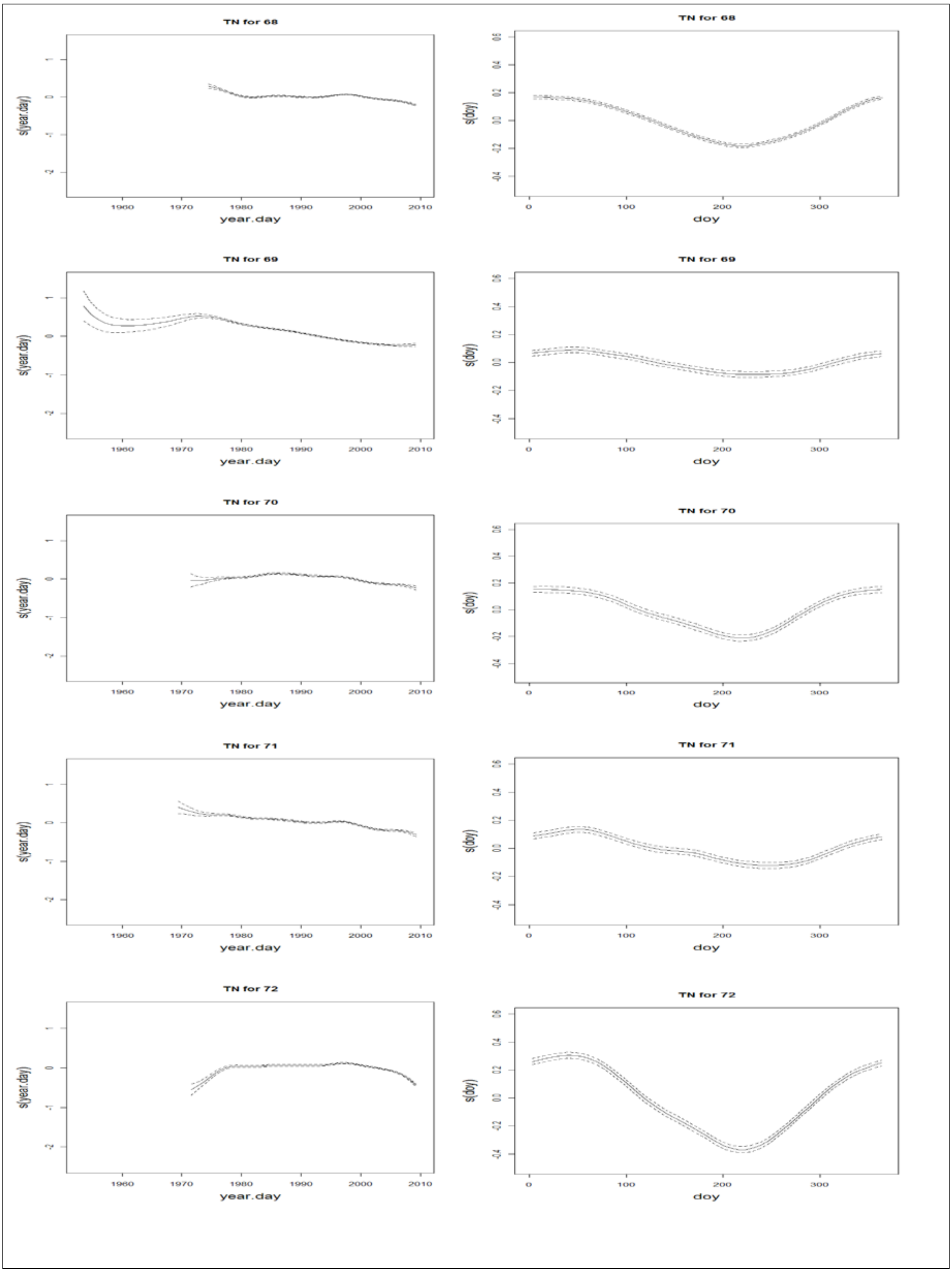

TON for 73
s(year.day)
year.day
TON for 73
s(doy)
doy
TON for 74
TON for 74
TON for 75
TON for 75
TON for 76
TON for 76
TON for 77
TON for 77

Figure A.6: Trend over time (left) and seasonality (right) for ln Total Nitrogen in the North West Region (Weaver(68), Mersey(69), Douglas(70), Ribble(71), Lune(72), Leven,Kent(73), Duddon(74), Derwent(75), Eden(76), Lyne,Esk(77) LHAs)



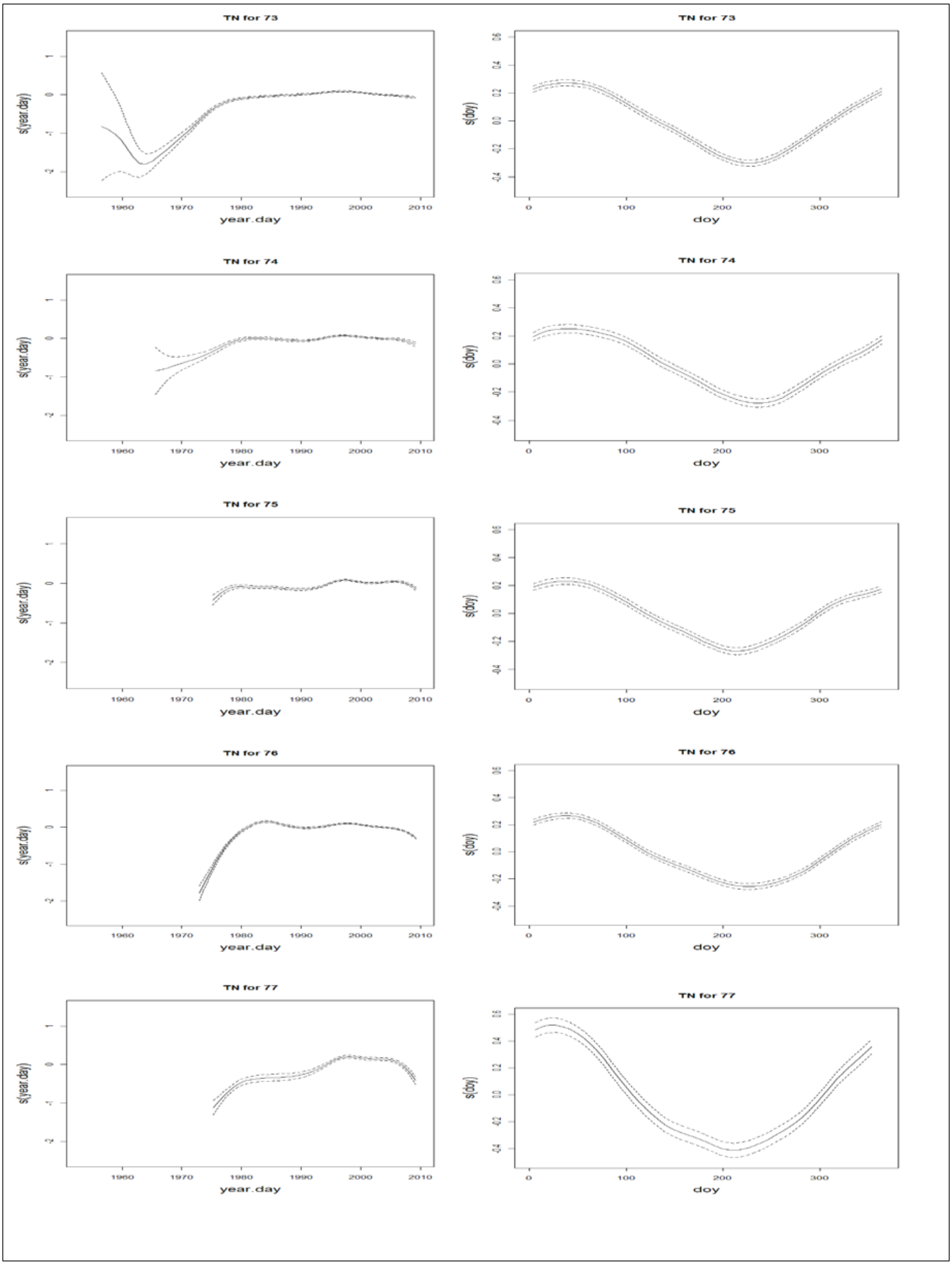
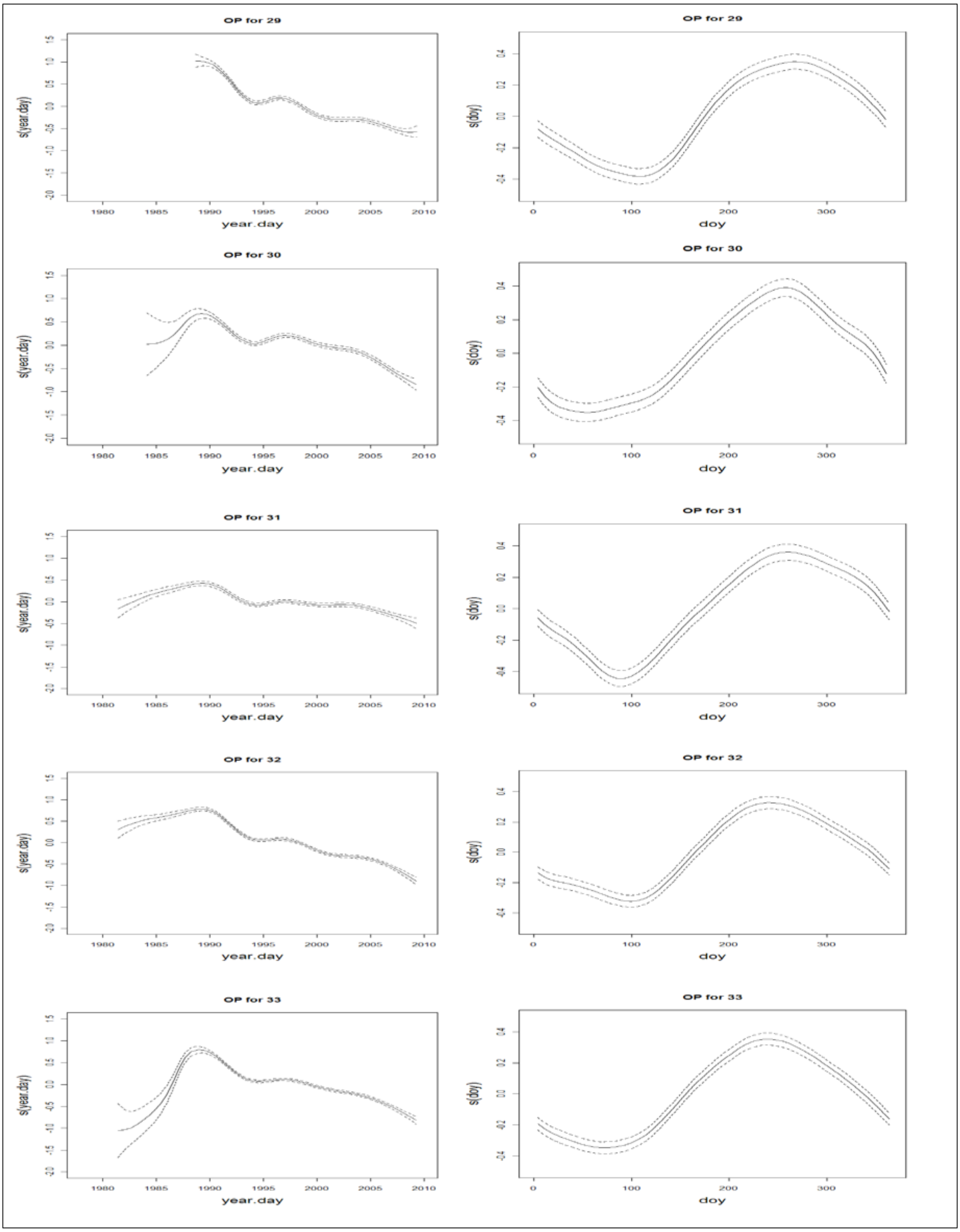

TN for 73
s(year.day)
year.day
TN for 73
s(doy)
doy
TN for 74
s(year.day)
year.day
TN for 74
s(doy)
doy
TN for 75
s(year.day)
year.day
TN for 75
s(doy)
doy
TN for 76
s(year.day)
year.day
TN for 76
s(doy)
doy
TN for 77
s(year.day)
year.day
TN for 77
s(doy)
doy

Figure A.7: Trend over time (left) and seasonality (right) for ln Orthophosphate-P in the Anglian Region (Ancholme(29), Witham(30), Welland(31), Nene(32), Great Ouse(33), Bure, Waveney(34), Gipping(35), Stour, E.Anglia(36), Blackwater,Chelmer(37) LHAs)



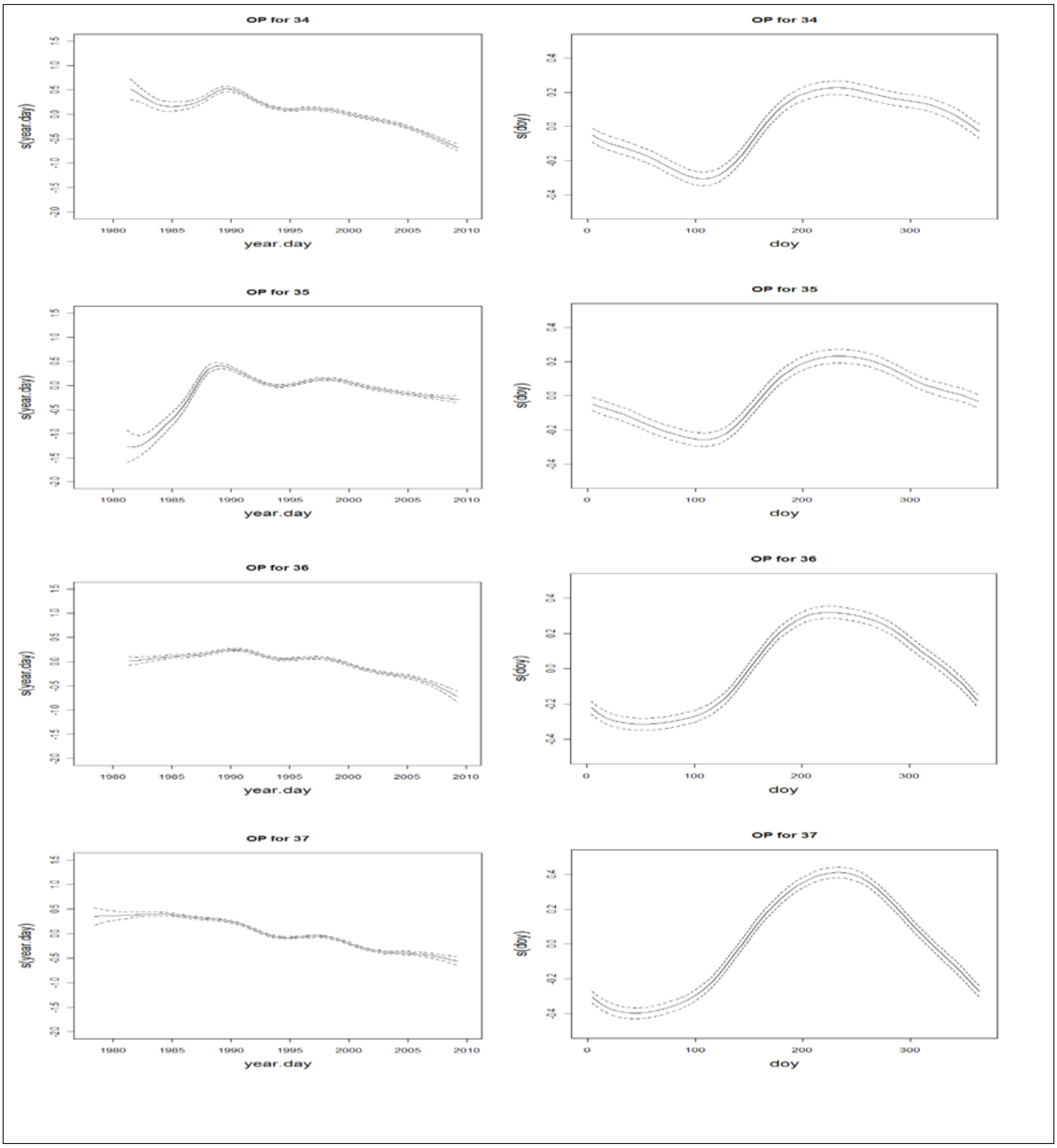
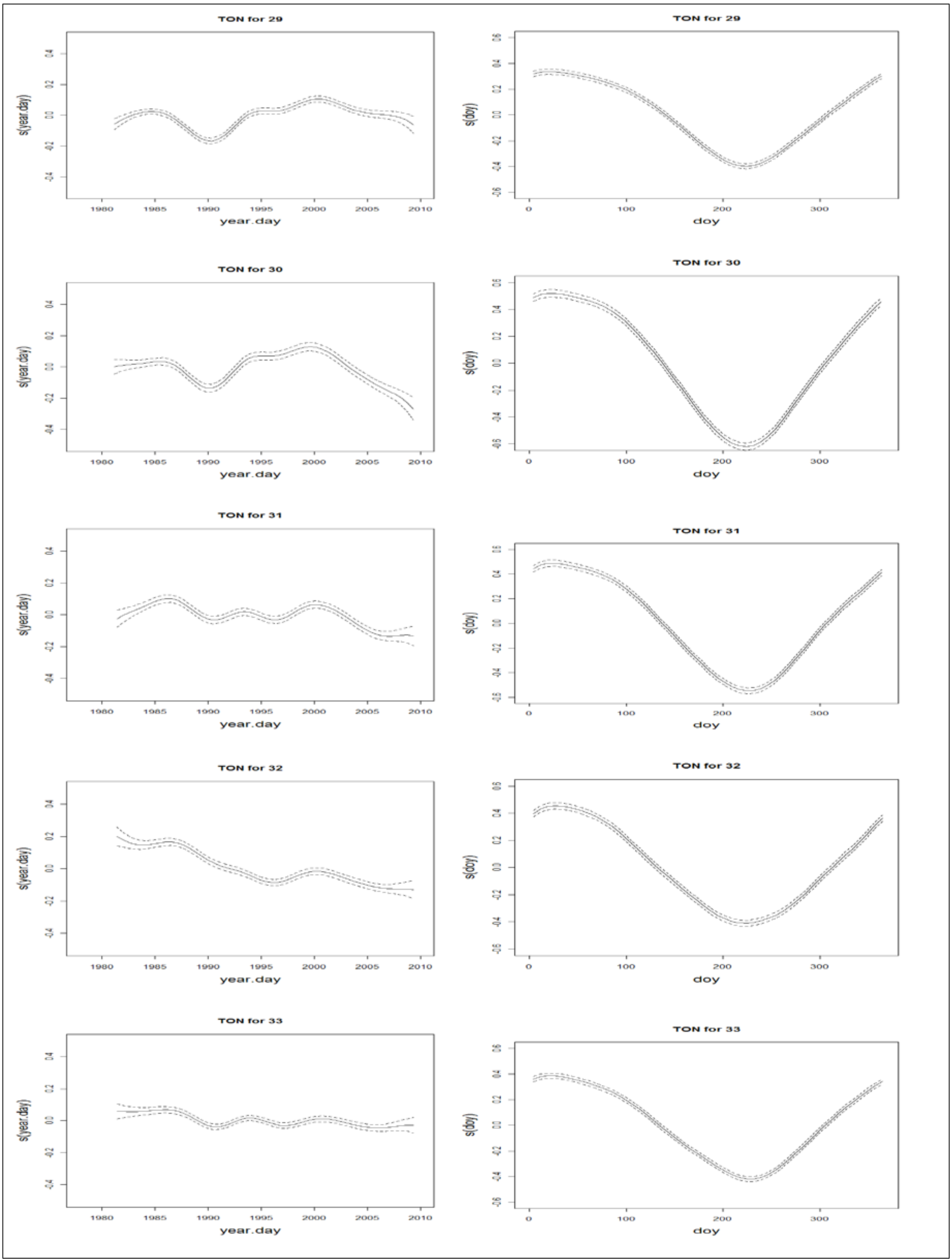

OP for 34
s(year.day)
year.day
OP for 34
s(doy)
doy
OP for 35
s(year.day)
year.day
OP for 35
s(doy)
doy
OP for 36
s(year.day)
year.day
OP for 36
s(doy)
doy
OP for 37
s(year.day)
year.day
OP for 37
s(doy)
doy

Figure A.8: Trend over time (left) and seasonality (right) for ln Total Oxidised Nitrogen in the Anglian Region (Ancholme(29), Witham(30), Welland(31), Nene(32), Great Ouse(33), Bure, Waveney(34), Gipping(35), Stour, E.Anglia(36), Blackwater,Chelmer(37) LHAs)



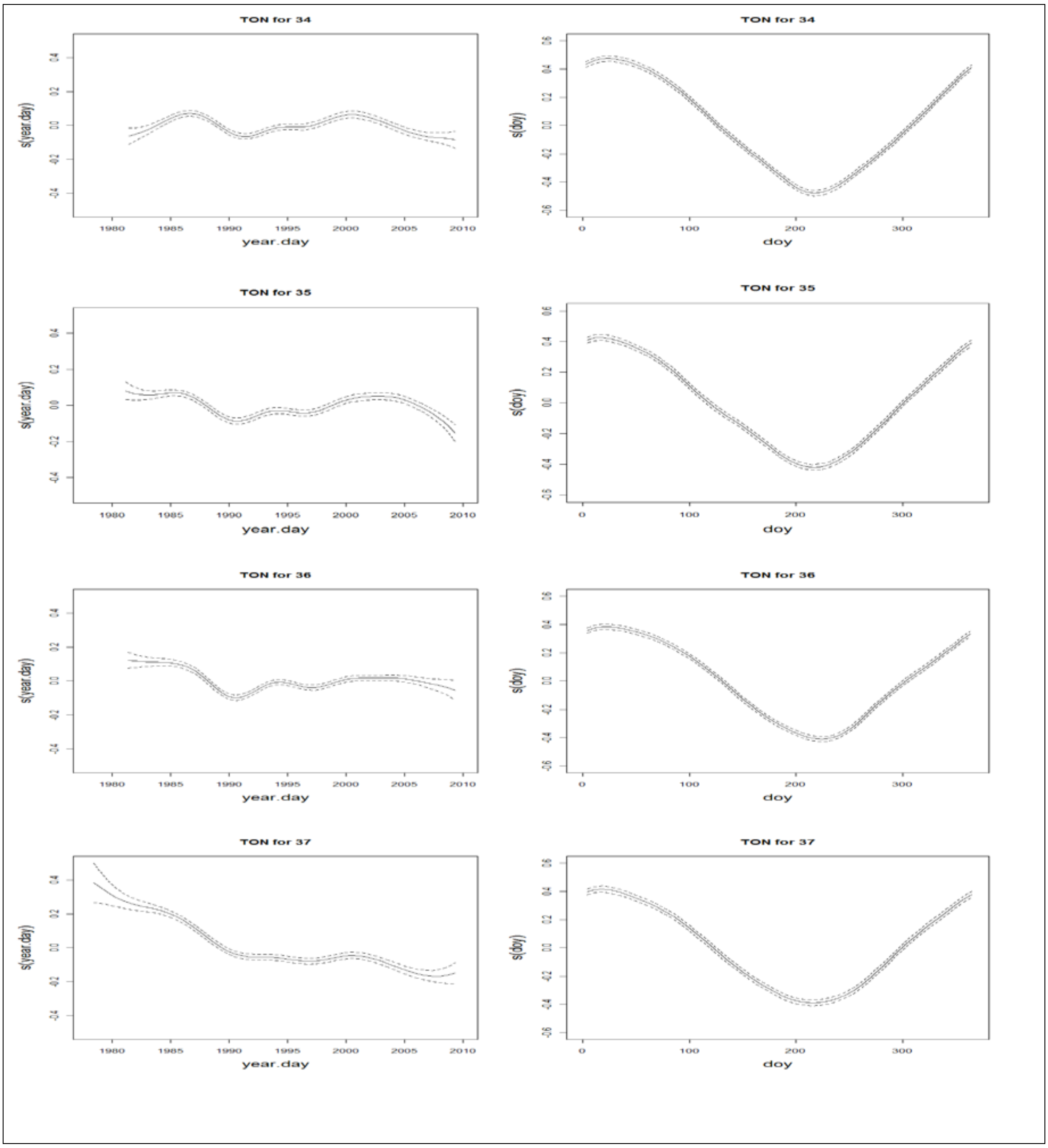
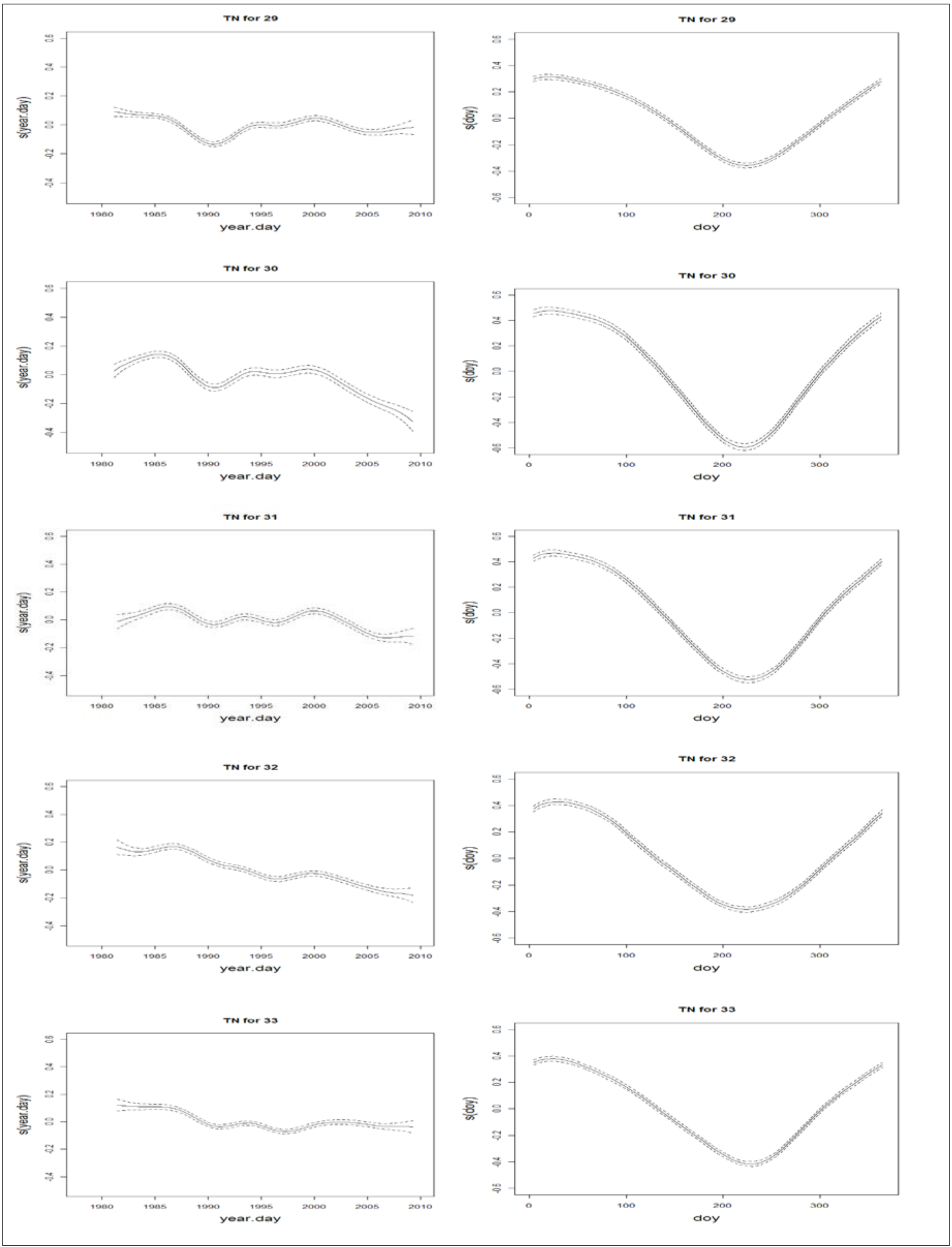

TON for 34
TON for 34
TON for 35
TON for 35
TON for 36
TON for 36
TON for 37
TON for 37
s(year.day)
year.day
s(doy)
doy

Figure A.9: Trend over time (left) and seasonality (right) for ln Total Nitrogen in the Anglian Region (Ancholme(29), Witham(30), Welland(31), Nene(32), Great Ouse(33), Bure, Waveney(34), Gipping(35), Stour, E.Anglia(36), Blackwater,Chelmer(37) LHAs)



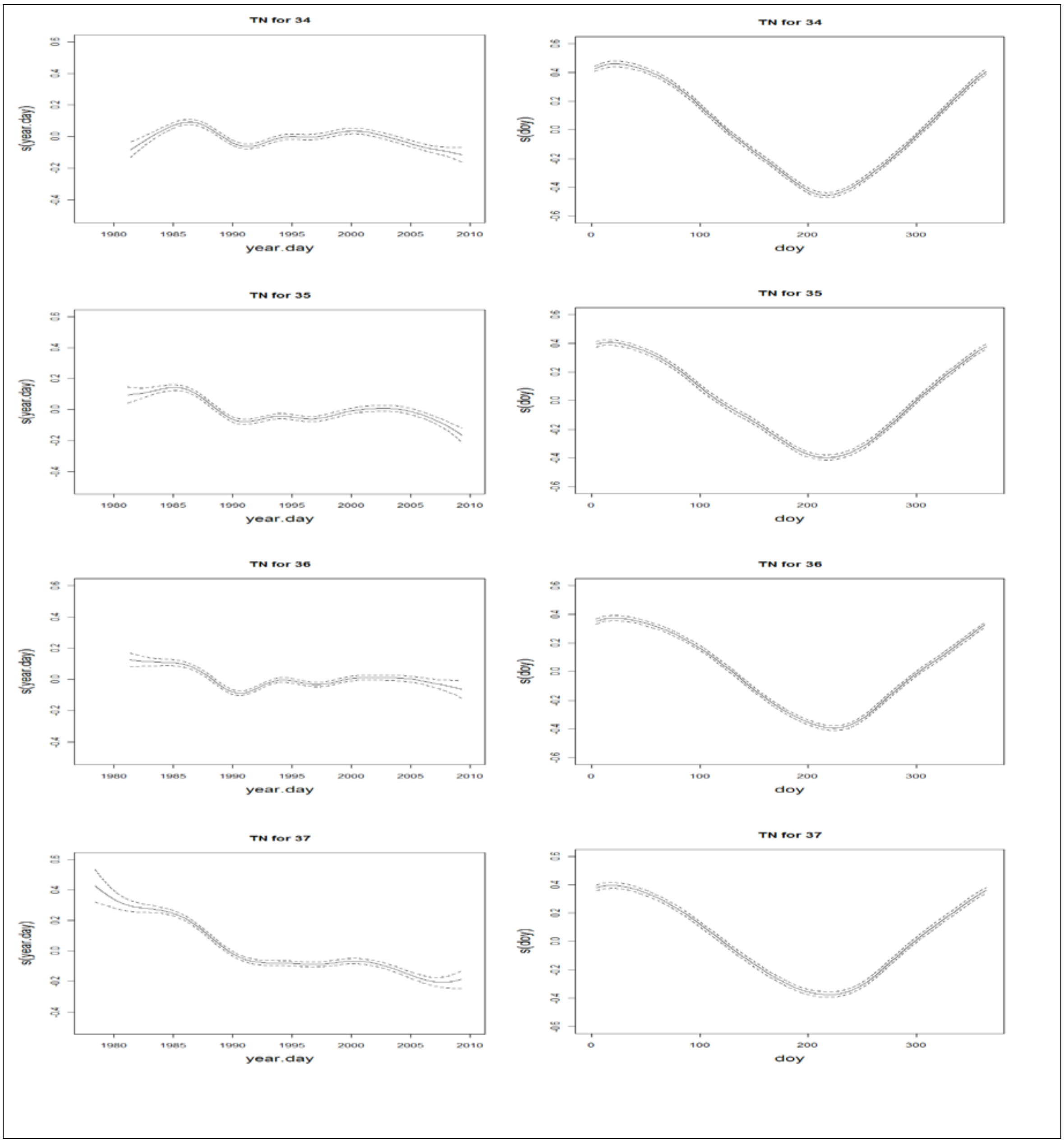

TN for 34
s(year.day)
year.day
TN for 34
s(doy)
doy
TN for 35
s(year.day)
year.day
TN for 35
s(doy)
doy
TN for 36
s(year.day)
year.day
TN for 36
s(doy)
doy
TN for 37
s(year.day)
year.day
TN for 37
s(doy)
doy

Figure A.10: Trend over time (left) and seasonality (right) for ln Orthophosphate-P in the North East Region (Tweed(21), Coquet,Wansbeck,Blyth(22), Tyne(23), Wear(24), Tees(25), Hull(26), Ouse, Humber Estuary(27) LHAs; Tweed(21) and Tyne(23) have insufficient data for modelling Orthophosphate-P)

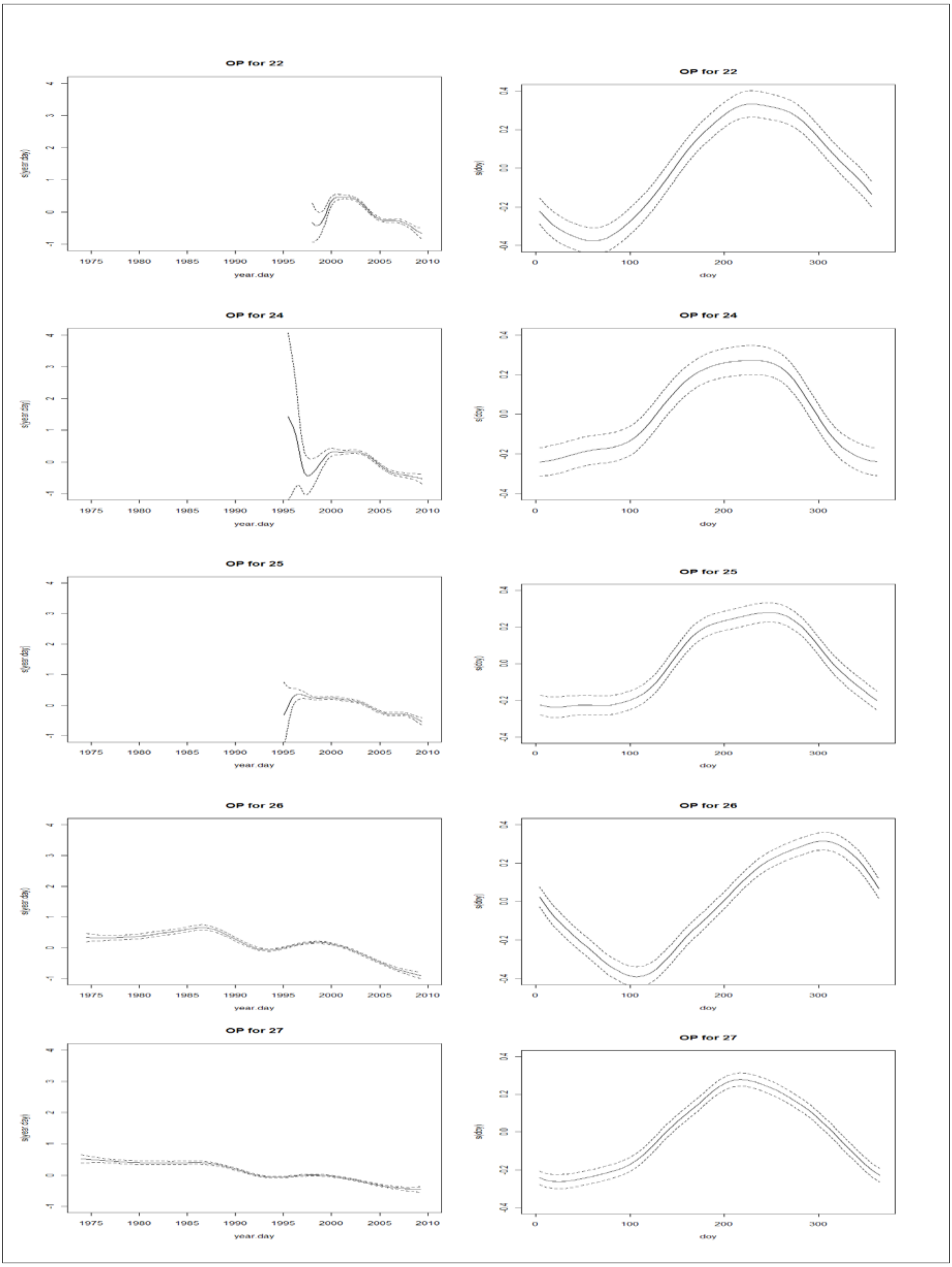
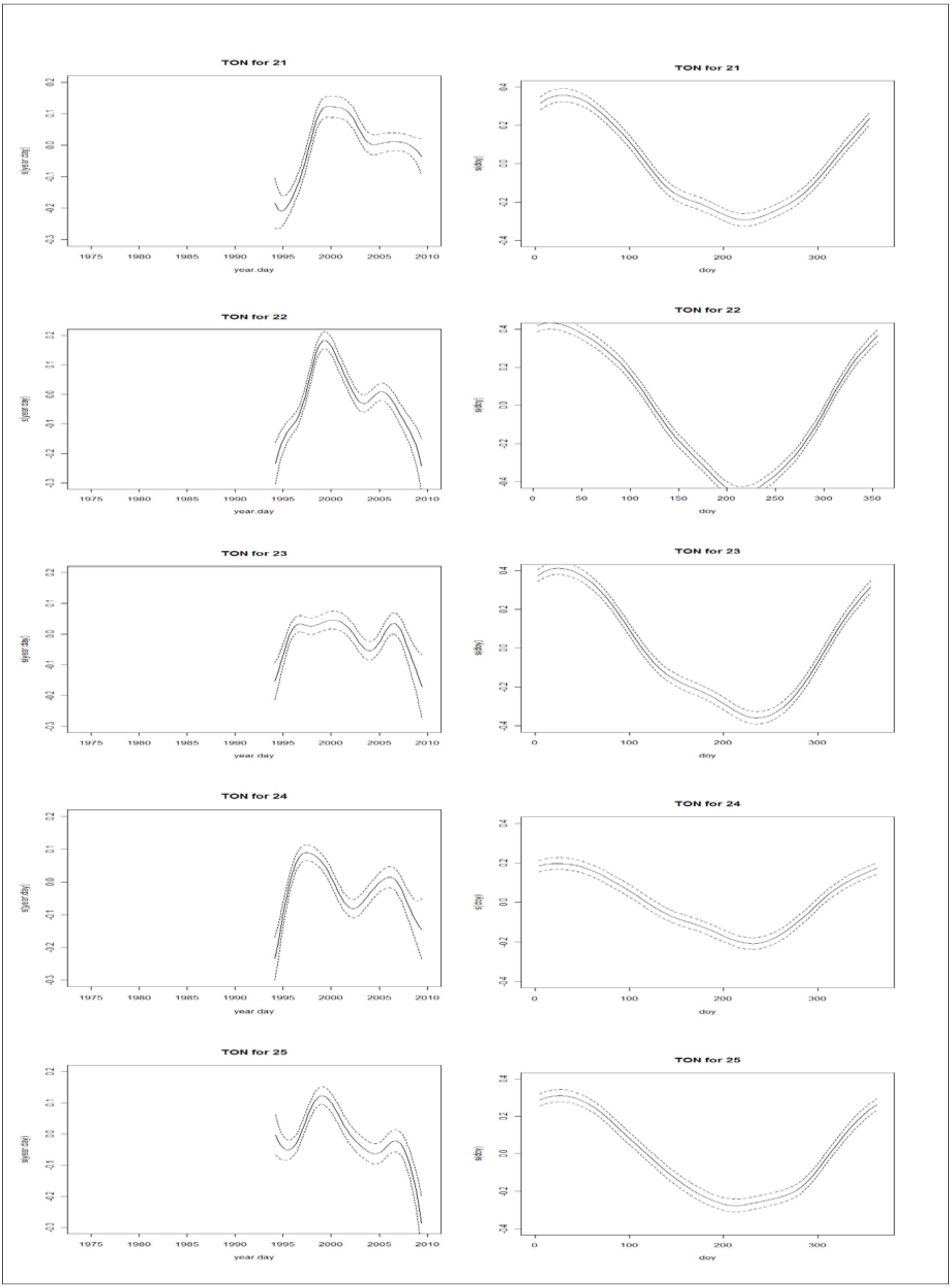


Figure A.11: Trend over time (left) and seasonality (right) for ln Total Oxidised Nitrogen in the North East Region (Tweed(21), Coquet,Wansbeck,Blyth(22), Tyne(23), Wear(24), Tees(25), Hull(26), Ouse, Humber Estuary(27))



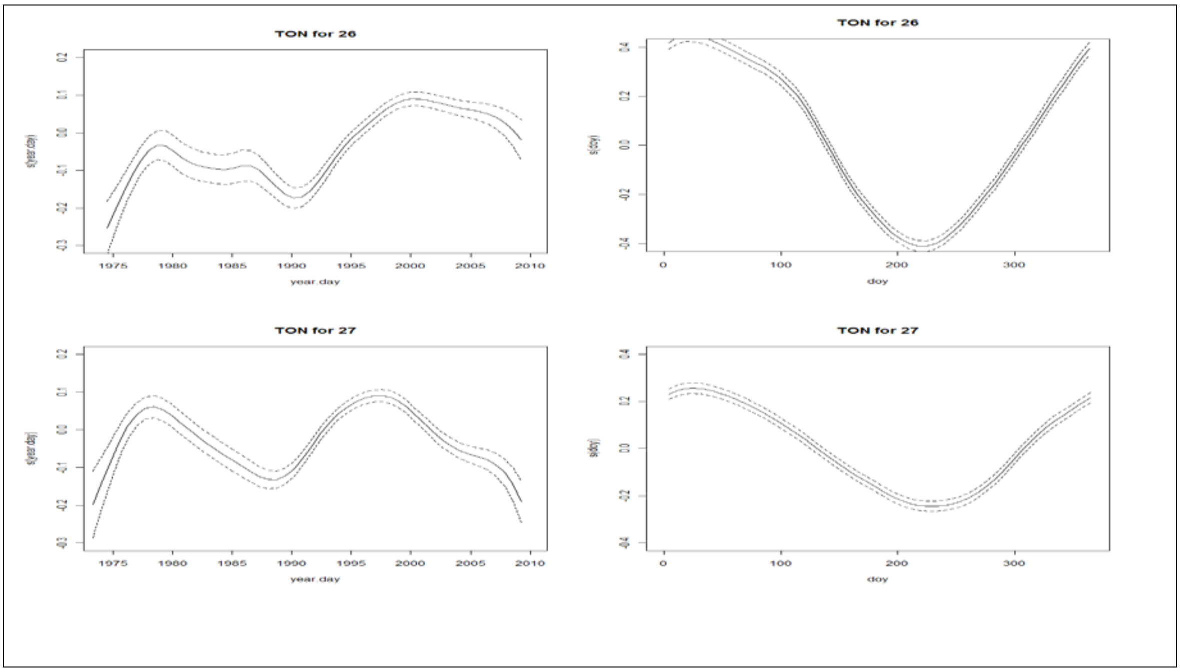
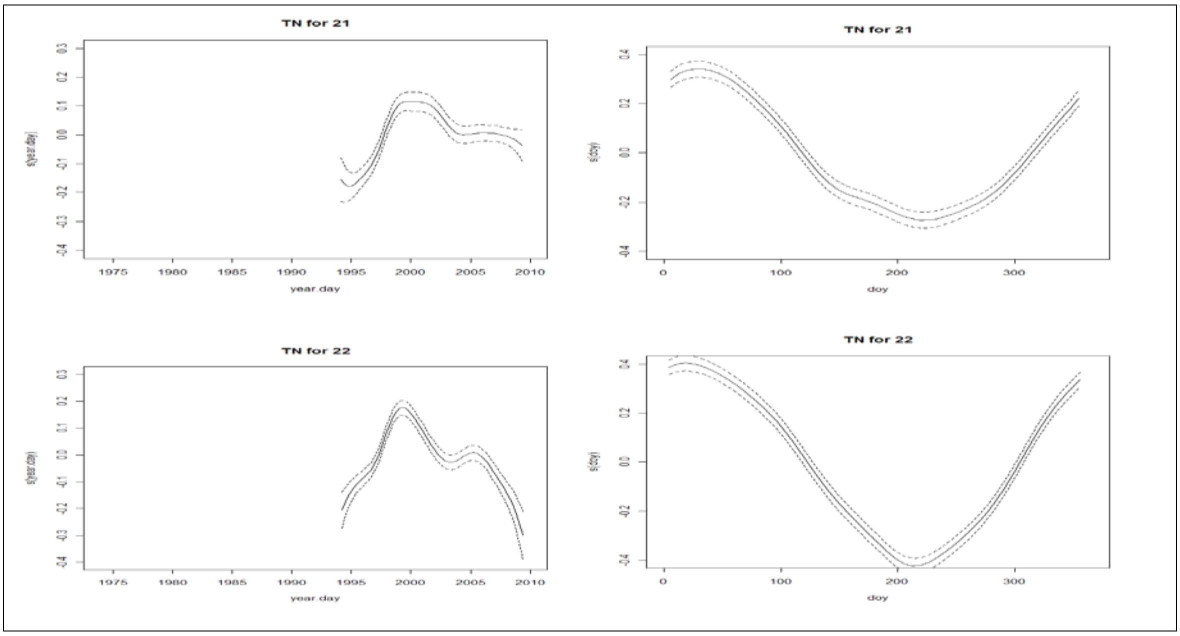


Figure A.12: Trend over time (left) and seasonality (right) for ln Total Nitrogen in the North East Region (Tweed(21), Coquet,Wansbeck,Blyth(22), Tyne(23), Wear(24), Tees(25), Hull(26), Ouse, Humber Estuary(27))



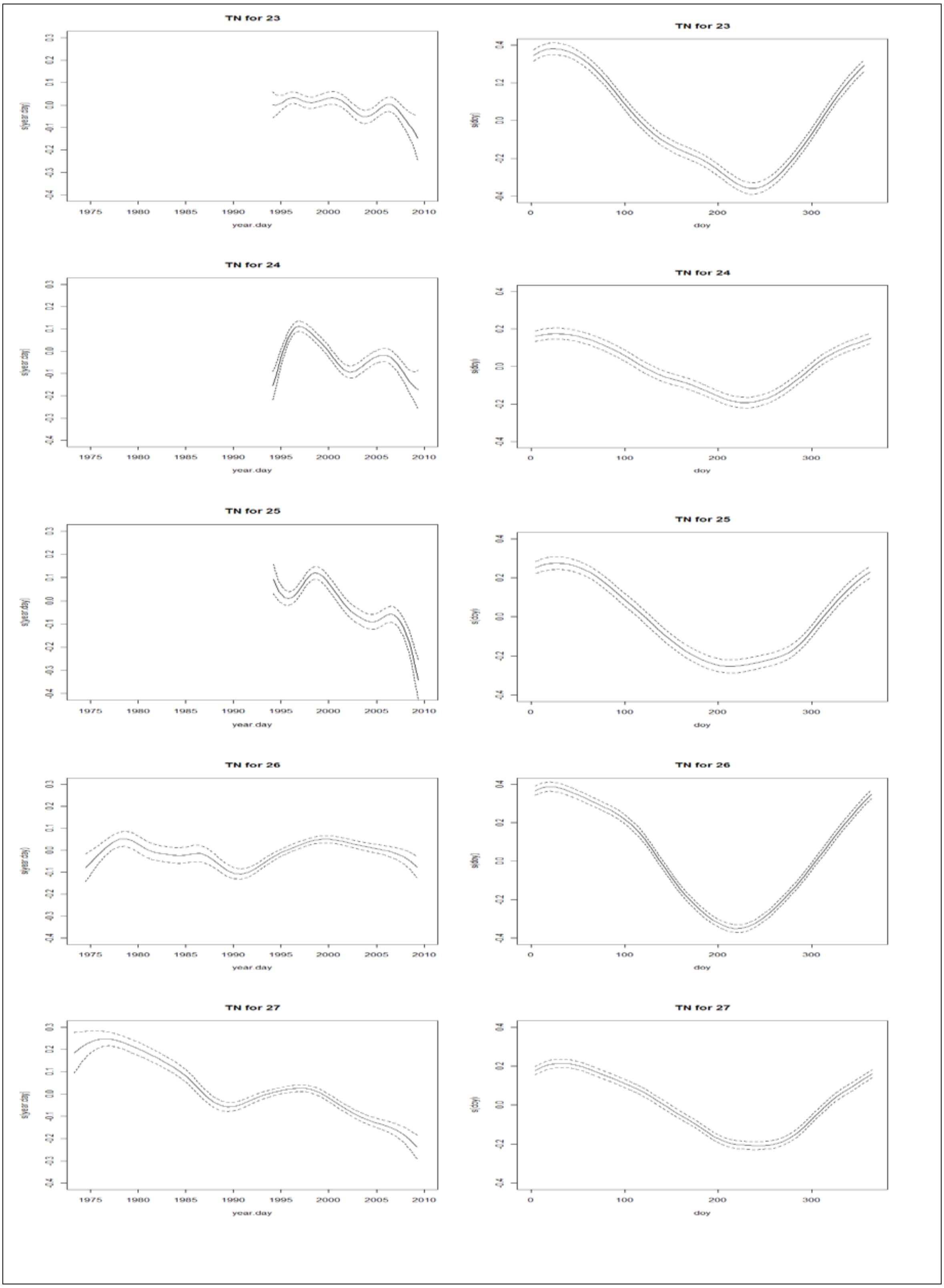

TN for 23
s(year.day)
year.day
TN for 23
s(doy)
doy
TN for 24
s(year.day)
year.day
TN for 24
s(doy)
doy
TN for 25
s(year.day)
year.day
TN for 25
s(doy)
doy
TN for 26
s(year.day)
year.day
TN for 26
s(doy)
doy
TN for 27
s(year.day)
year.day
TN for 27
s(doy)
doy

Figure A.13: Trend over time (left) and seasonality(right) for ln Orthophosphate-P in the Wales Region (only the LHA of the Dee(67) has sufficient data for modelling Orthophosphate-P)

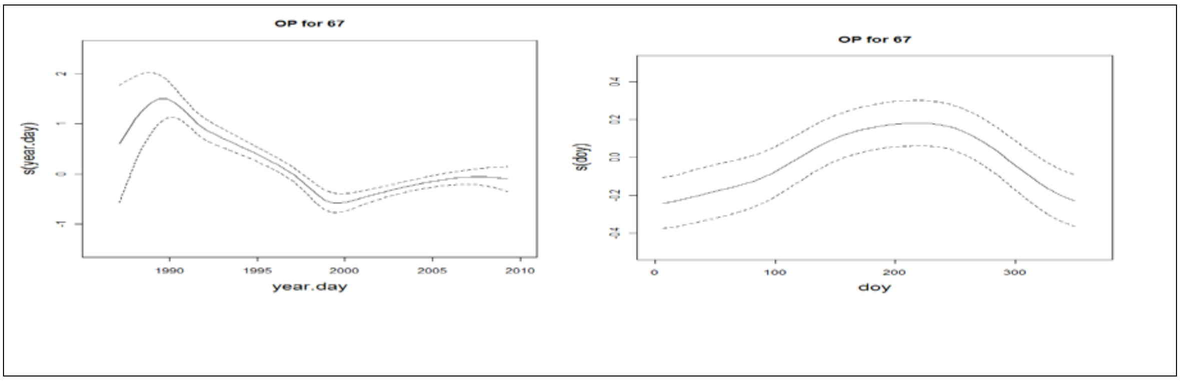


Figure A.14: Trend over time (left) and seasonality (right) for ln Total Oxidised Nitrogen in the Wales Region (Dee(67) and the clusters Glaslyn(65), Clwyd, Conwy(66),Anglesey(102) and Wye(55), Usk(56), Taff(57), Tawe,Neath(58), Loughor(59), Tywi(60), Cleddau(61), Teifi(62), Rheidol,Ystwyth(63))

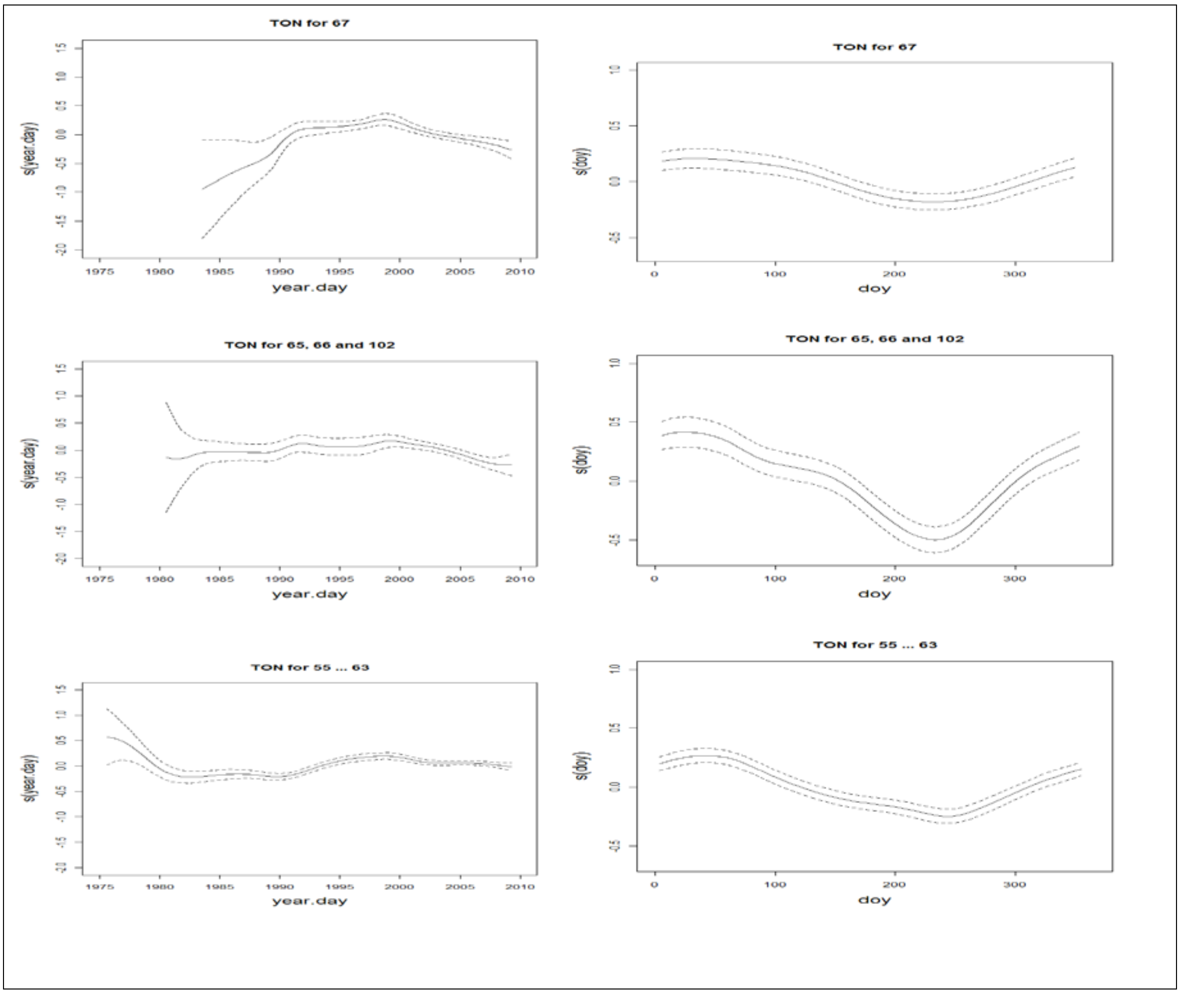


Figure A.15: Trend over time (left) and seasonality (right) for ln Total Nitrogen in the Wales Region (Dee(67) and the clusters Glaslyn(65), Clwyd, Conwy(66),Anglesey(102) and Wye(55), Usk(56), Taff(57), Tawe,Neath(58), Loughor(59), Tywi(60), Cleddau(61), Teifi(62), Rheidol,Ystwyth(63))

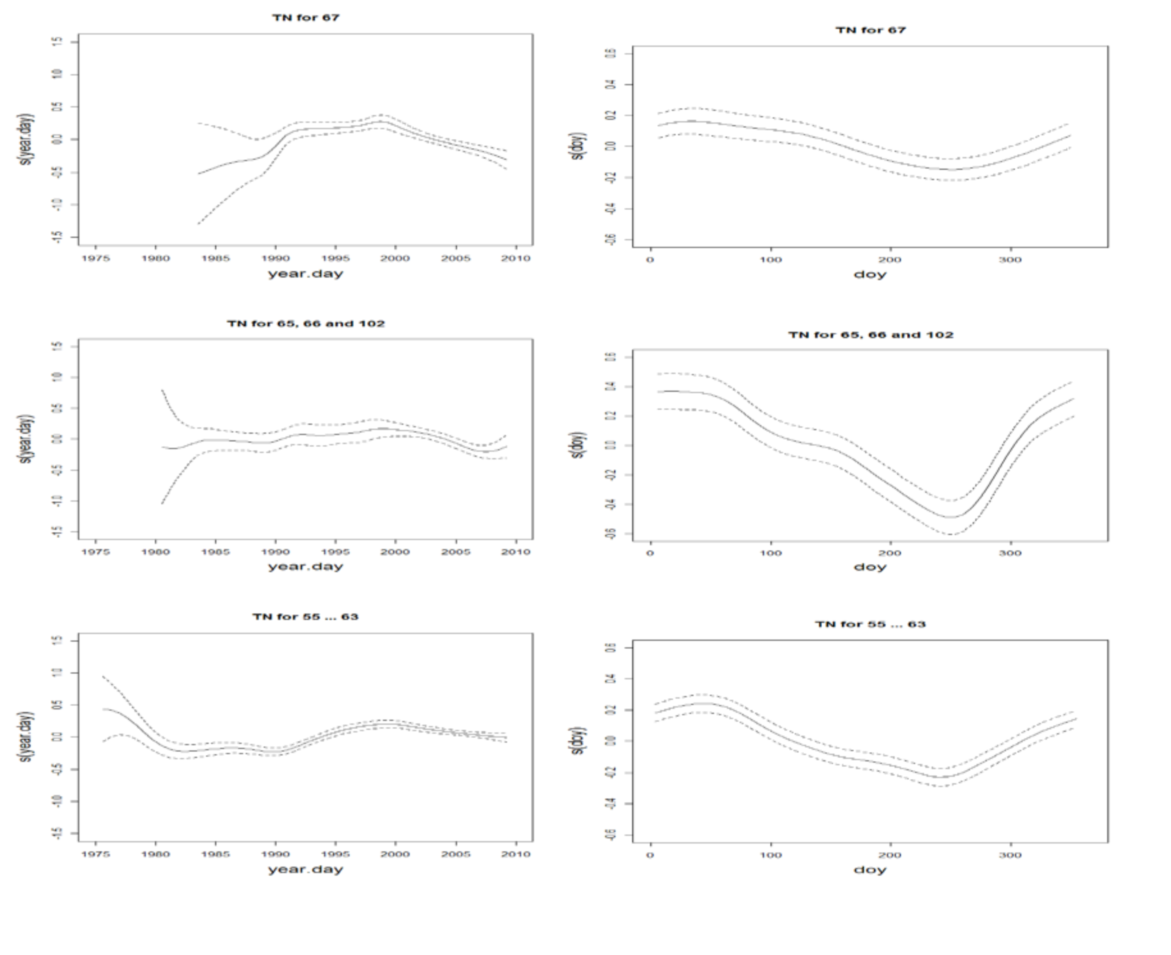


Figure A.16: Trend over time (left) and seasonality (right) for ln Orthophosphate-P in the Thames Region (Lee(38), Thames(39) LHAs)

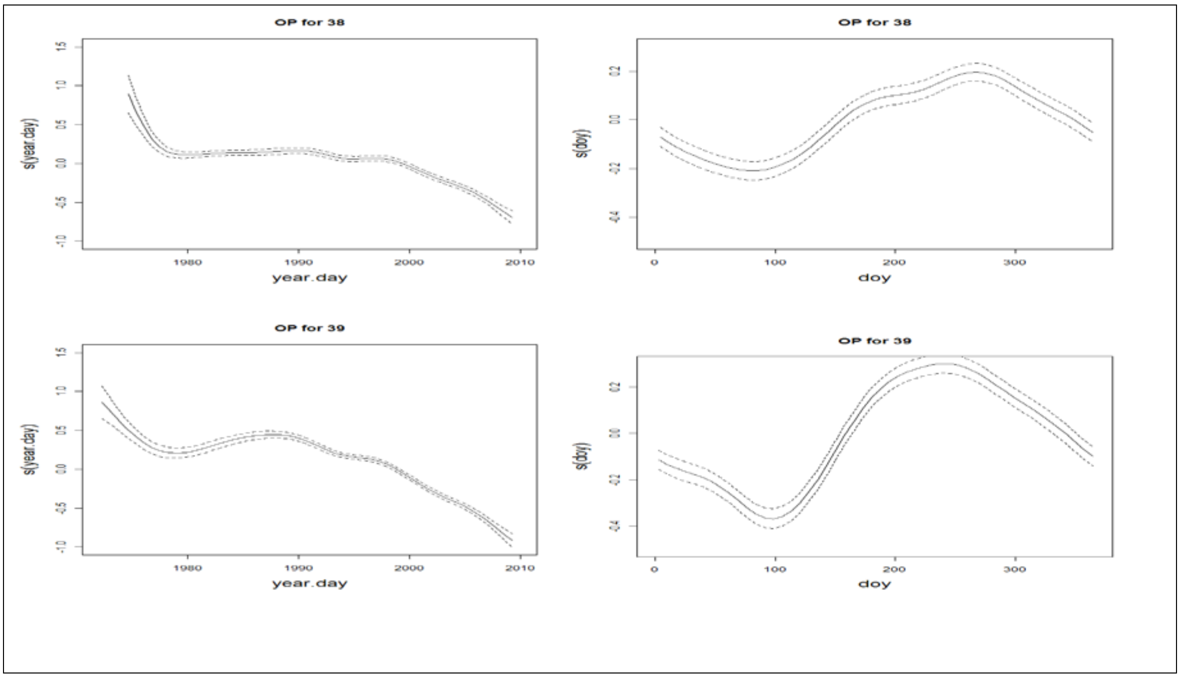


Figure A.17: Trend over time (left) and seasonality (right) for ln Total Oxidised Nitrogen in the Thames Region (Lee(38), Thames(39) LHAs)

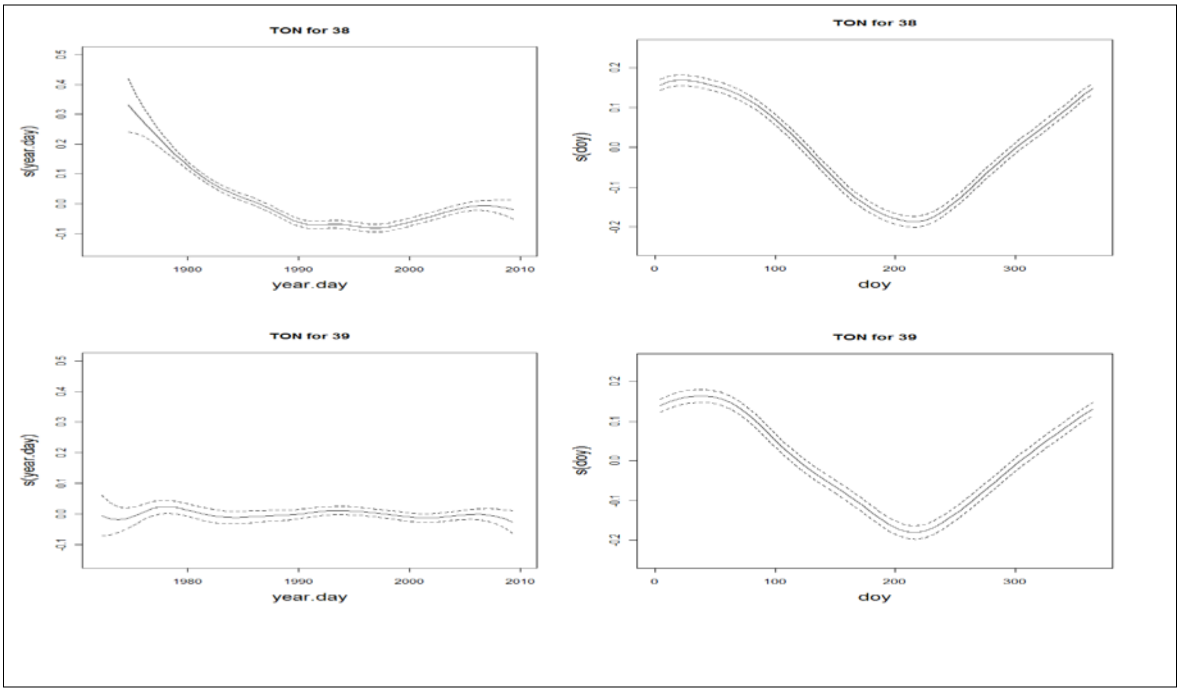


Figure A.18: Trend over time (left) and seasonality (right) for ln Total Nitrogen in the Thames Region (Lee(38), Thames(39) LHAs)

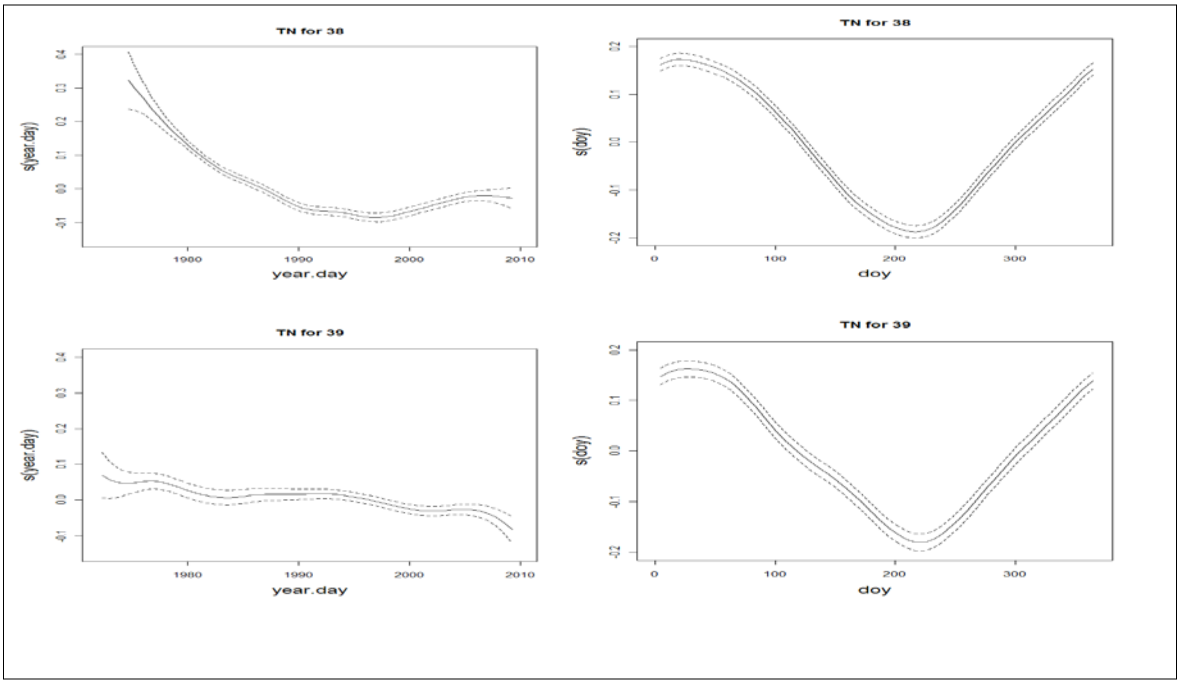
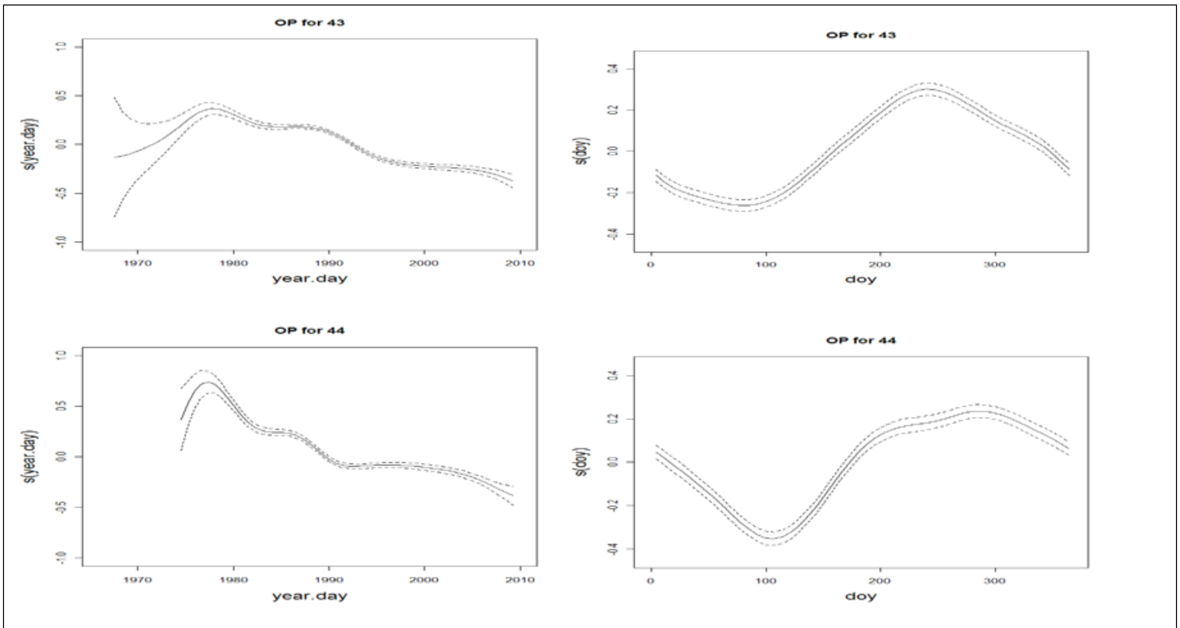
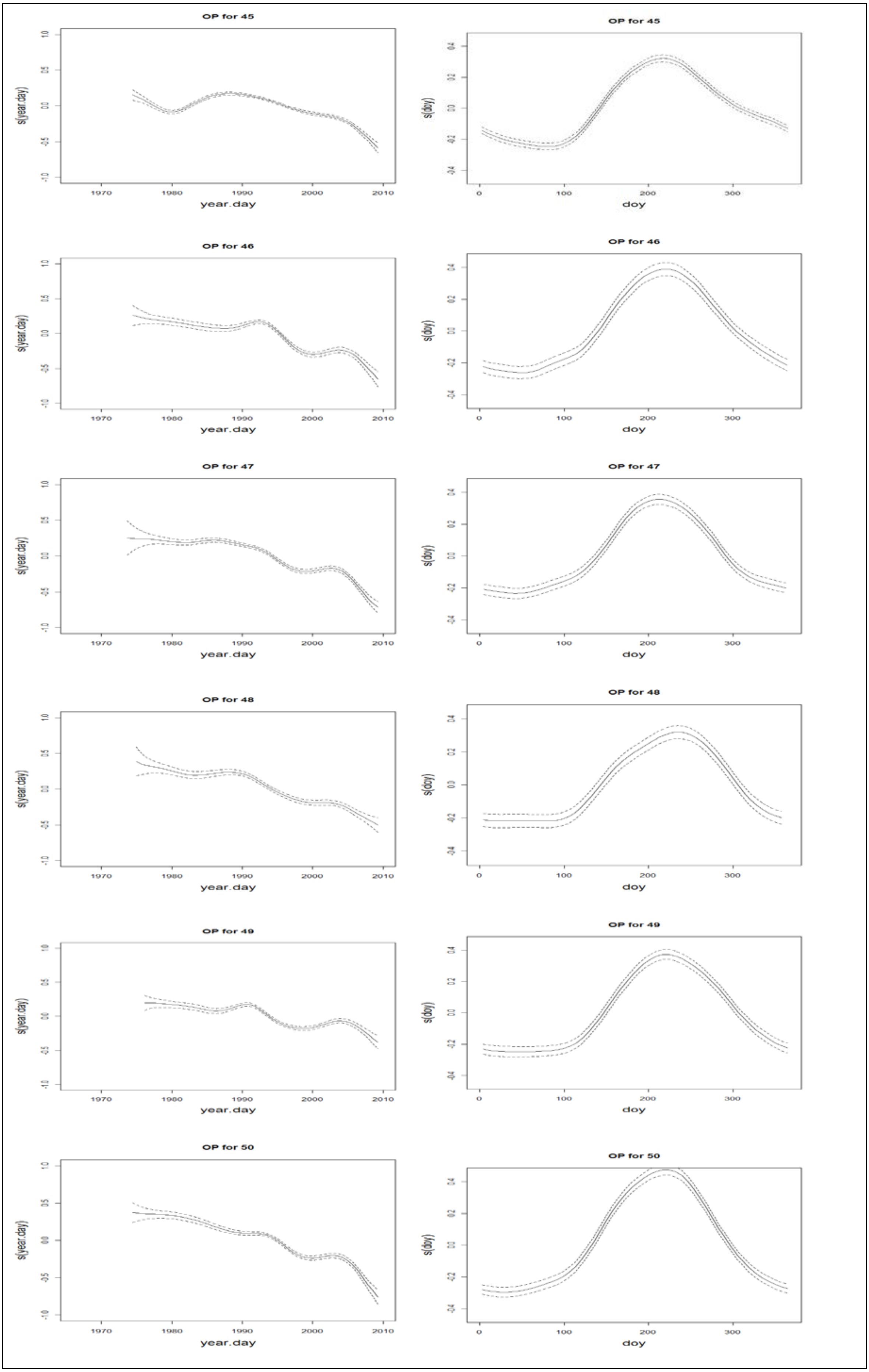


Figure A.19: Trend over time (left) and seasonality (right) for ln Orthophosphate-P in the South West Region (Avon,Hants(43), Piddle,Frome(44), Exe(45), Dart(46), Tamar(47), Fal(48), Camel(49), Torridge,Taw(50), East and West Lyns(51), Tone, Parrett(52), Frome, Bristol, Avon(53) LHAs)




OP for 45
s(year.day)
year.day
OP for 45
s(doy)
doy
OP for 46
s(year.day)
year.day
OP for 46
s(doy)
doy
OP for 47
s(year.day)
year.day
OP for 47
s(doy)
doy
OP for 48
s(year.day)
year.day
OP for 48
s(doy)
doy
OP for 49
s(year.day)
year.day
OP for 49
s(doy)
doy
OP for 50
s(year.day)
year.day
OP for 50
s(doy)
doy

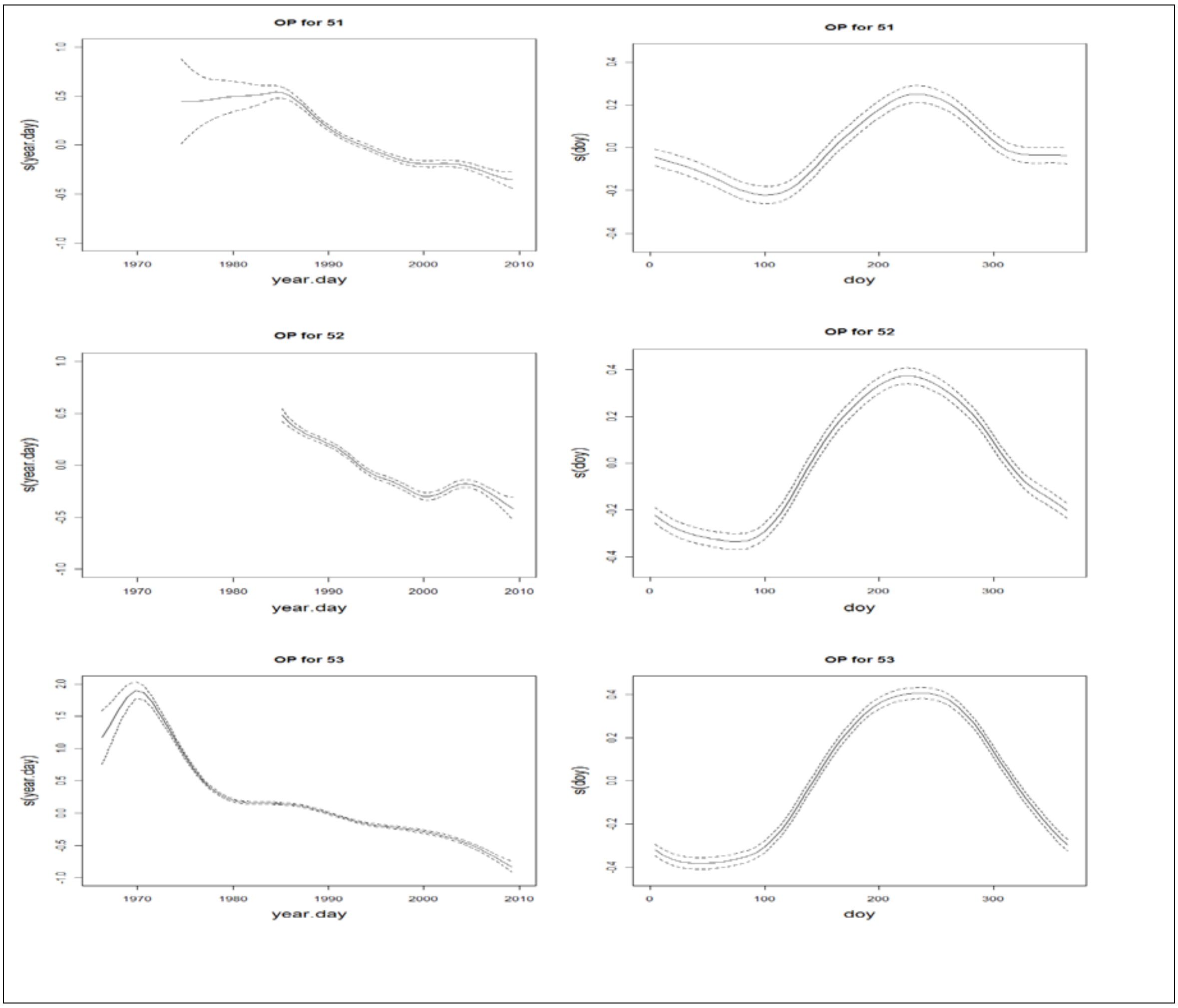
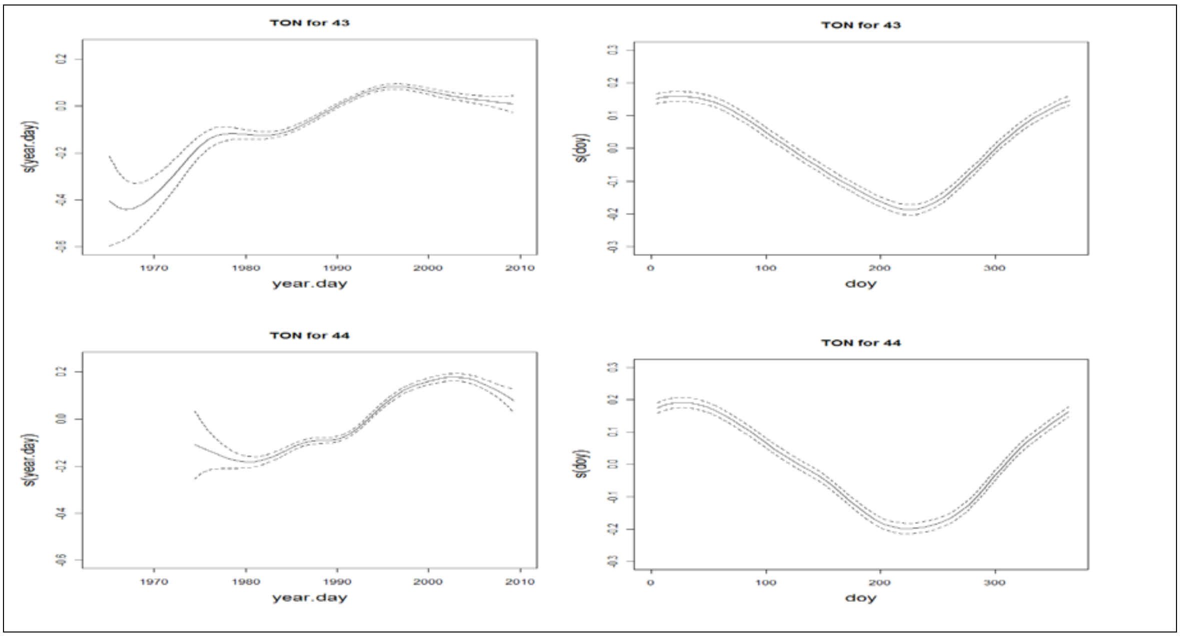
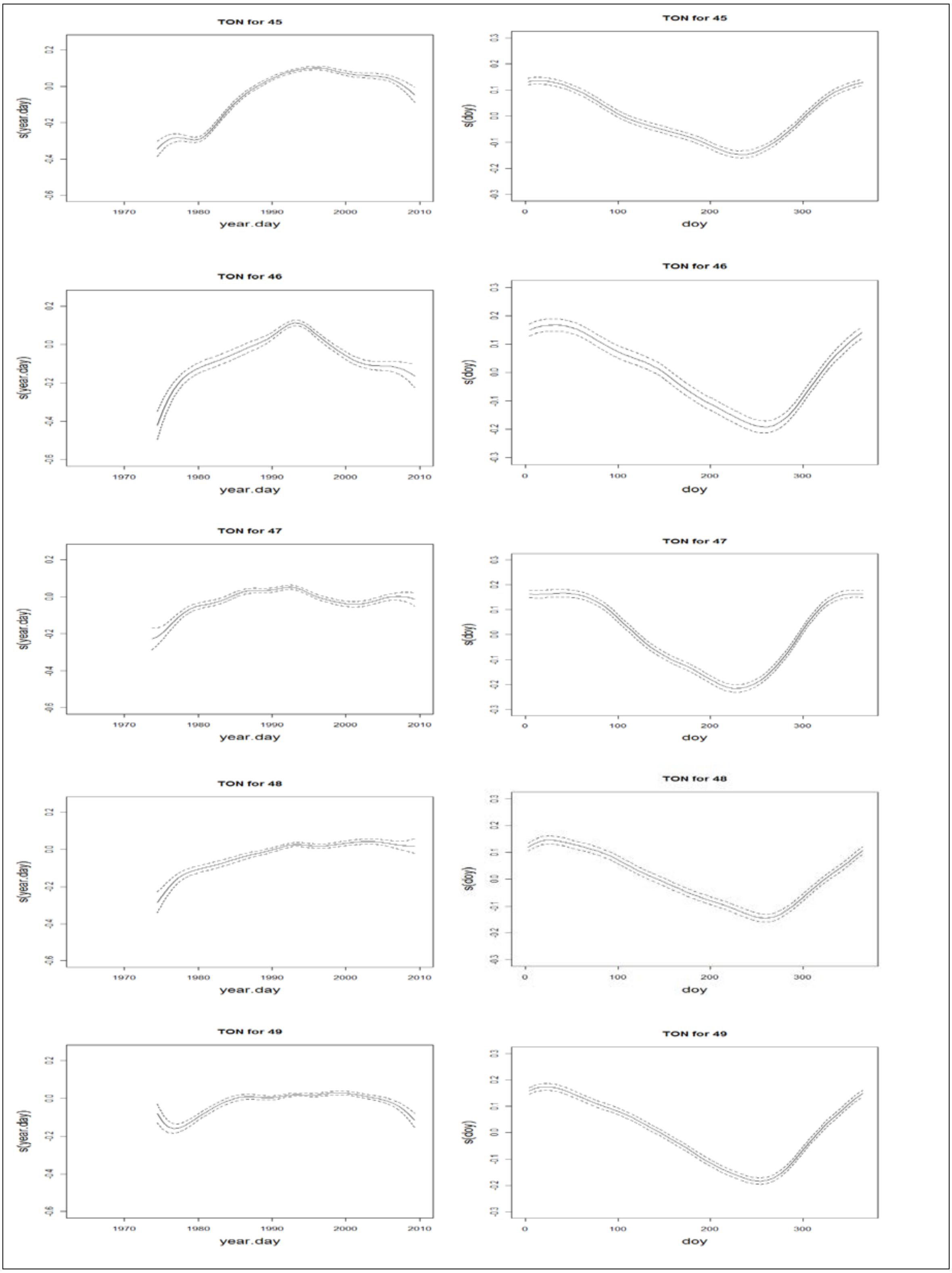


Figure A.20: Trend over time (left) and seasonality (right) for ln Total Oxidised Nitrogen in the South West Region (Avon,Hants(43), Piddle,Frome(44), Exe(45), Dart(46), Tamar(47), Fal(48), Camel(49), Torridge,Taw(50), East and West Lyns(51), Tone, Parrett(52), Frome, Bristol, Avon(53) LHAs)




TON for 45
s(year.day)
year.day
TON for 45
s(doy)
doy
TON for 46
s(year.day)
year.day
TON for 46
s(doy)
doy
TON for 47
s(year.day)
year.day
TON for 47
s(doy)
doy
TON for 48
s(year.day)
year.day
TON for 48
s(doy)
doy
TON for 49
s(year.day)
year.day
TON for 49
s(doy)
doy

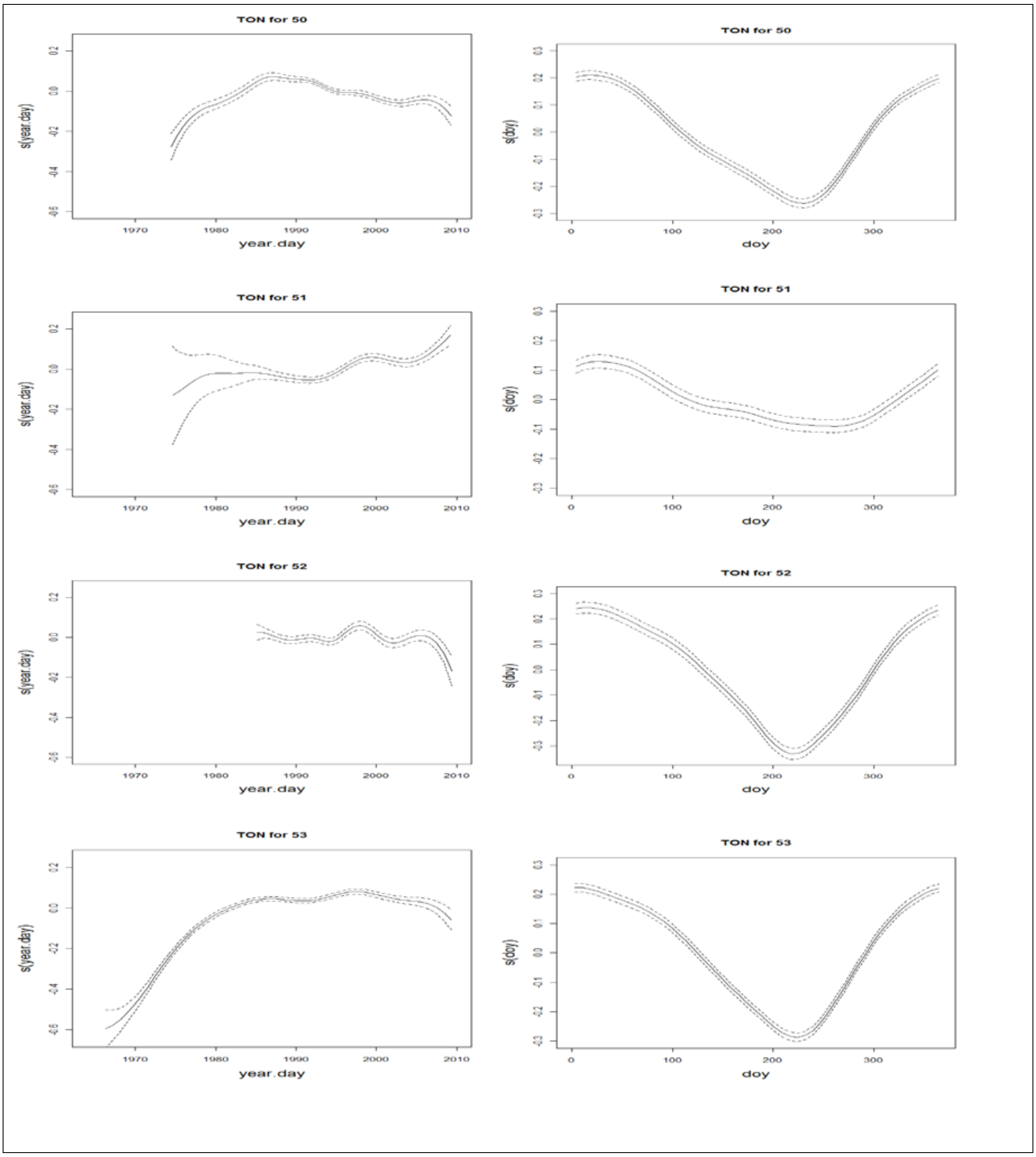
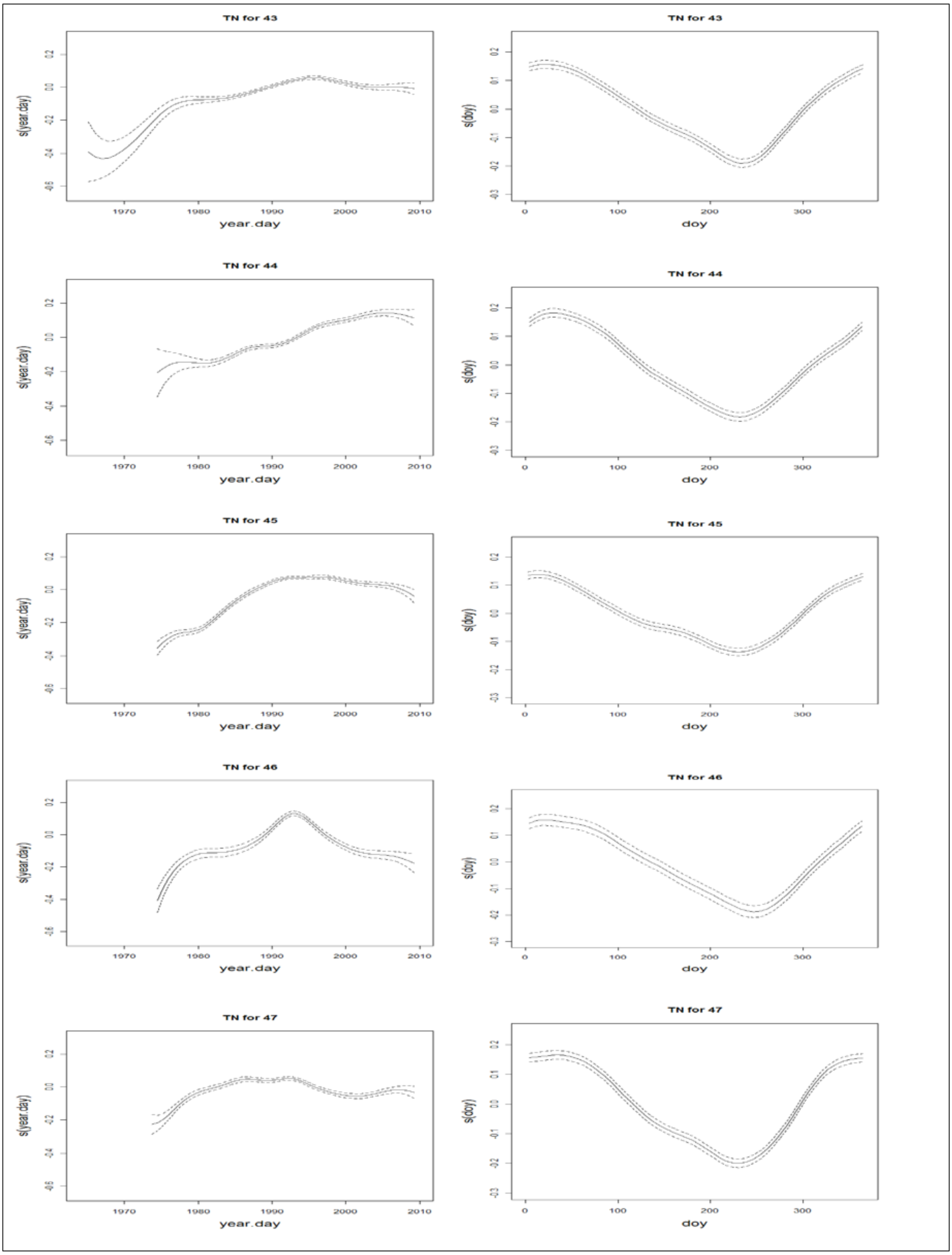
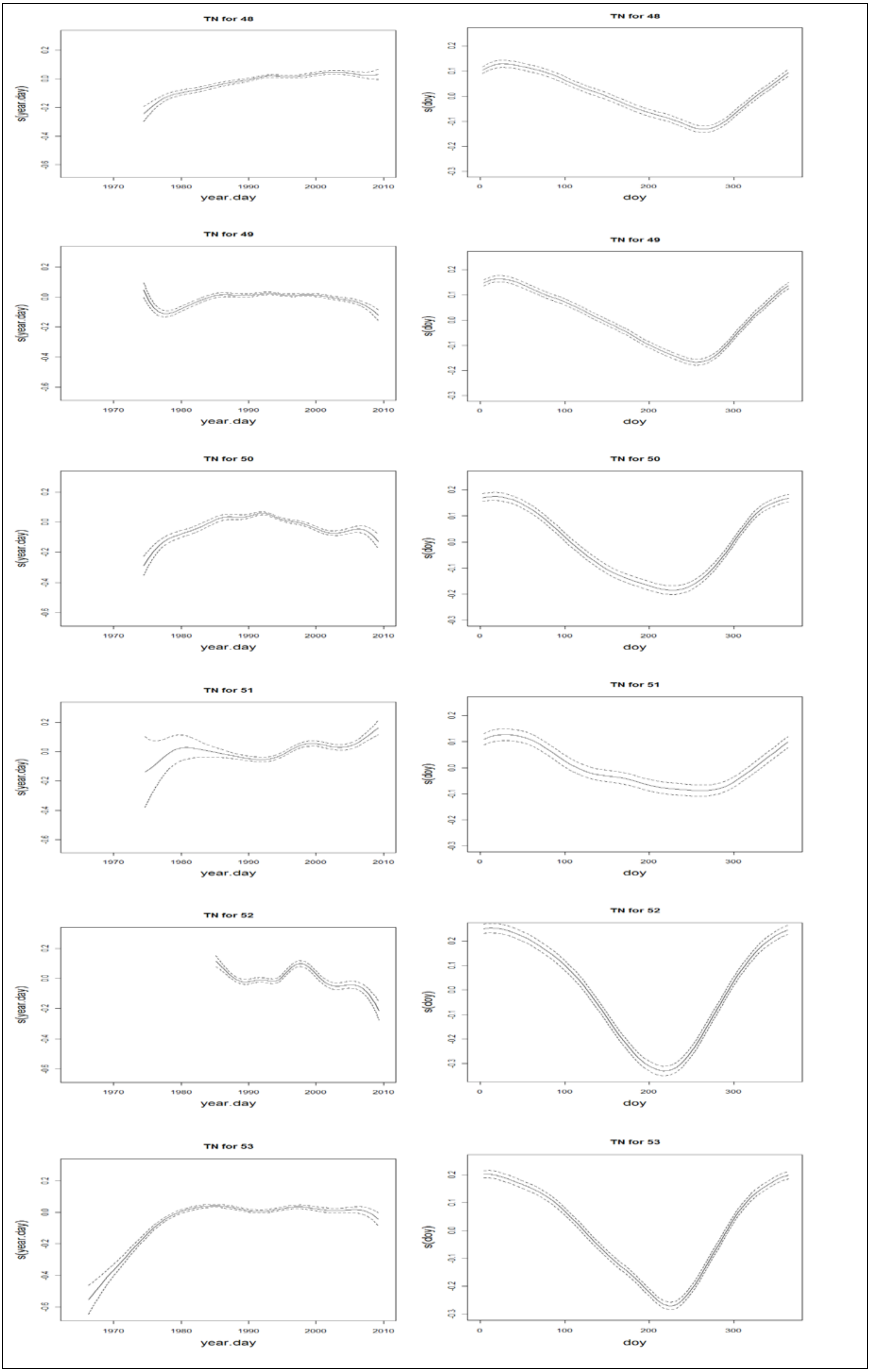

TON for 50
s(year.day)
year.day
TON for 50
s(doy)
doy
TON for 51
s(year.day)
year.day
TON for 51
s(doy)
doy
TON for 52
s(year.day)
year.day
TON for 52
s(doy)
doy
TON for 53
s(year.day)
year.day
TON for 53
s(doy)
doy

Figure A.21: Trend over time (left) and seasonality (right) for ln Total Oxidised Nitrogen in the South West Region (Avon,Hants(43), Piddle,Frome(44), Exe(45), Dart(46), Tamar(47), Fal(48), Camel(49), Torridge,Taw(50), East and West Lyns(51), Tone, Parrett(52), Frome, Bristol, Avon(53) LHAs)




TN for 48
s(year.day)
year.day
TN for 48
s(doy)
doy
TN for 49
s(year.day)
year.day
TN for 49
s(doy)
doy
TN for 50
s(year.day)
year.day
TN for 50
s(doy)
doy
TN for 51
s(year.day)
year.day
TN for 51
s(doy)
doy
TN for 52
s(year.day)
year.day
TN for 52
s(doy)
doy
TN for 53
s(year.day)
year.day
TN for 53
s(doy)
doy

Section 2

Section 2 contains plots for Model 1, see section 3 of the main report, for ln Total Oxidised Nitrogen and ln Total Nitrogen in the Trent(28), Severn(54) and the Lune(72) LHAs. Both the trend over time and seasonality are plotted on the same scale within each LHA. Similar plots for Orthophosphate-P are included in Figure 4.10 of the main report.

Figure A.22: Trend over time (left) and seasonality (right) for ln Total Oxidised Nitrogen in the Trent(28), Severn(54) and Lune(72) LHAs

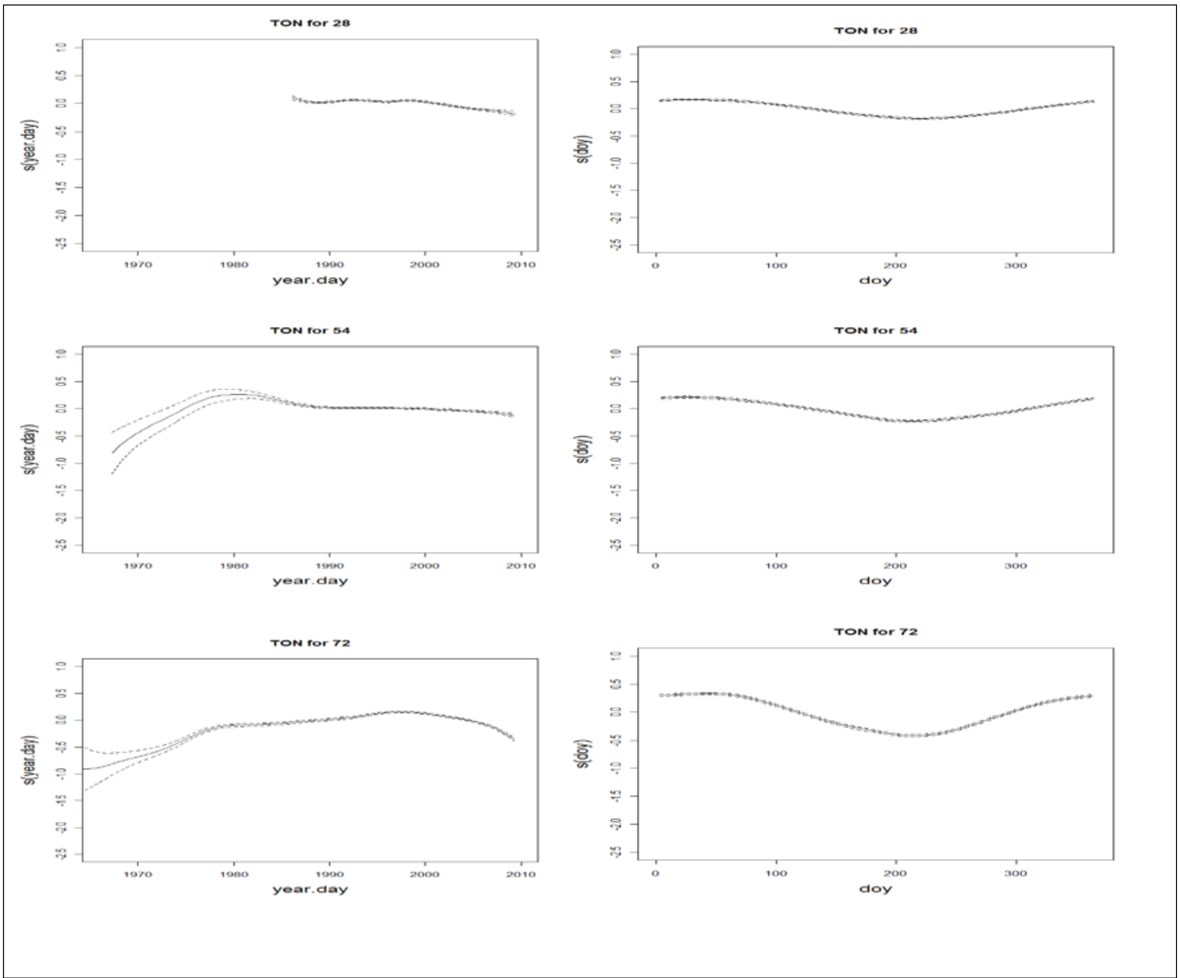
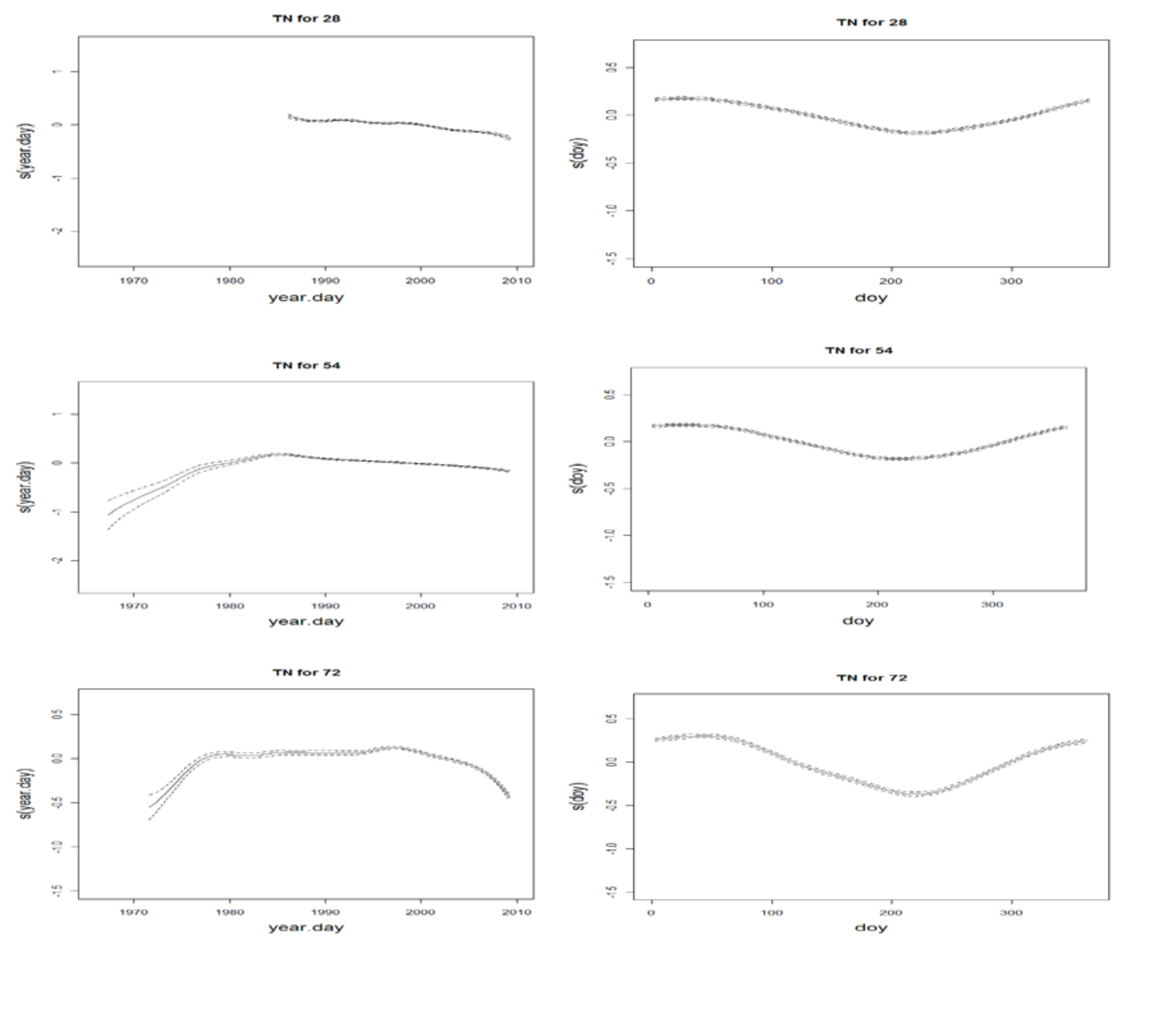


Figure A.23: Trend over time (left) and seasonality (right) for ln Total Nitrogen in the Trent(28), Severn(54) and Lune(72) LHAs.



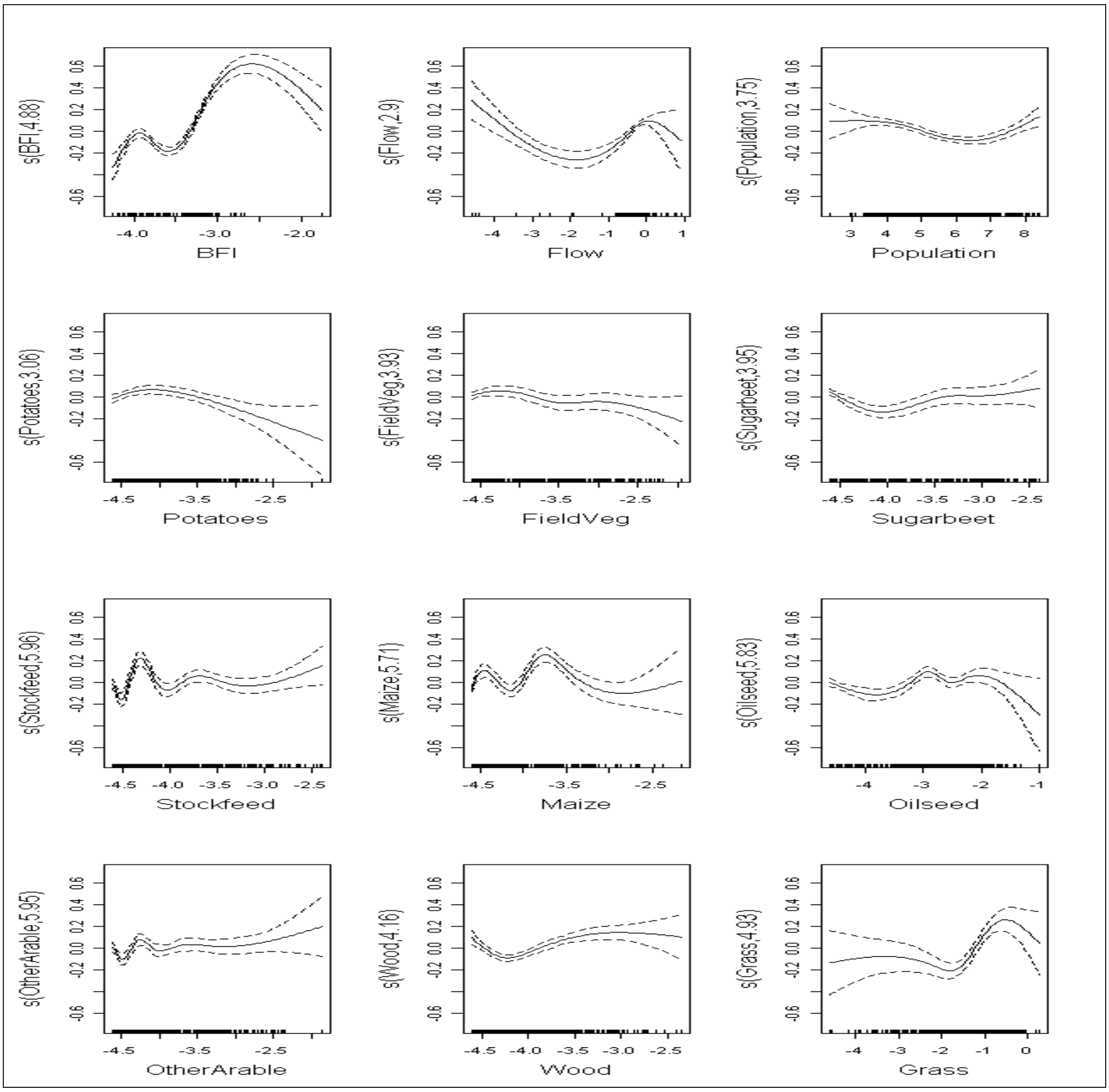
Section 3

Section 3 contains plots to illustrate the results from fitting Model A2, see Model 2 in section 3 of the main report, for the responses of ln Orthophospate-P, ln Total Oxidised Nitrogen and ln Total Nitrogen in each of the Trent (28), Severn (54) and the Lune (72) LHAs. For each covariate (*x* axis), plots highlight the relationship with the response (*y* axis) with the estimates depicted by a solid line and standard errors as dashed lines.

Some of the relationships are driven by sparse data distributions which is reflected by wider intervals between the standard error lines. The tick marks on the horizontal axis highlight the data distribution for each covariate. A natural log transform (ln) has been applied to all continuous covariates to reduce skewness in the data distributions.

The degree of smoothing is constrained to allow a maximum of 6 degrees of freedom for each component but where sparse data leads to “wiggly” relationships, further constraining is applied by limiting the degrees of freedom to 4. The degree of smoothing is specified on the *y* axis.

Figure A.24: Relationships between ln continuous covariates and ln Total Oxidised Nitrogen – Model A2 in the Trent(28) LHA.



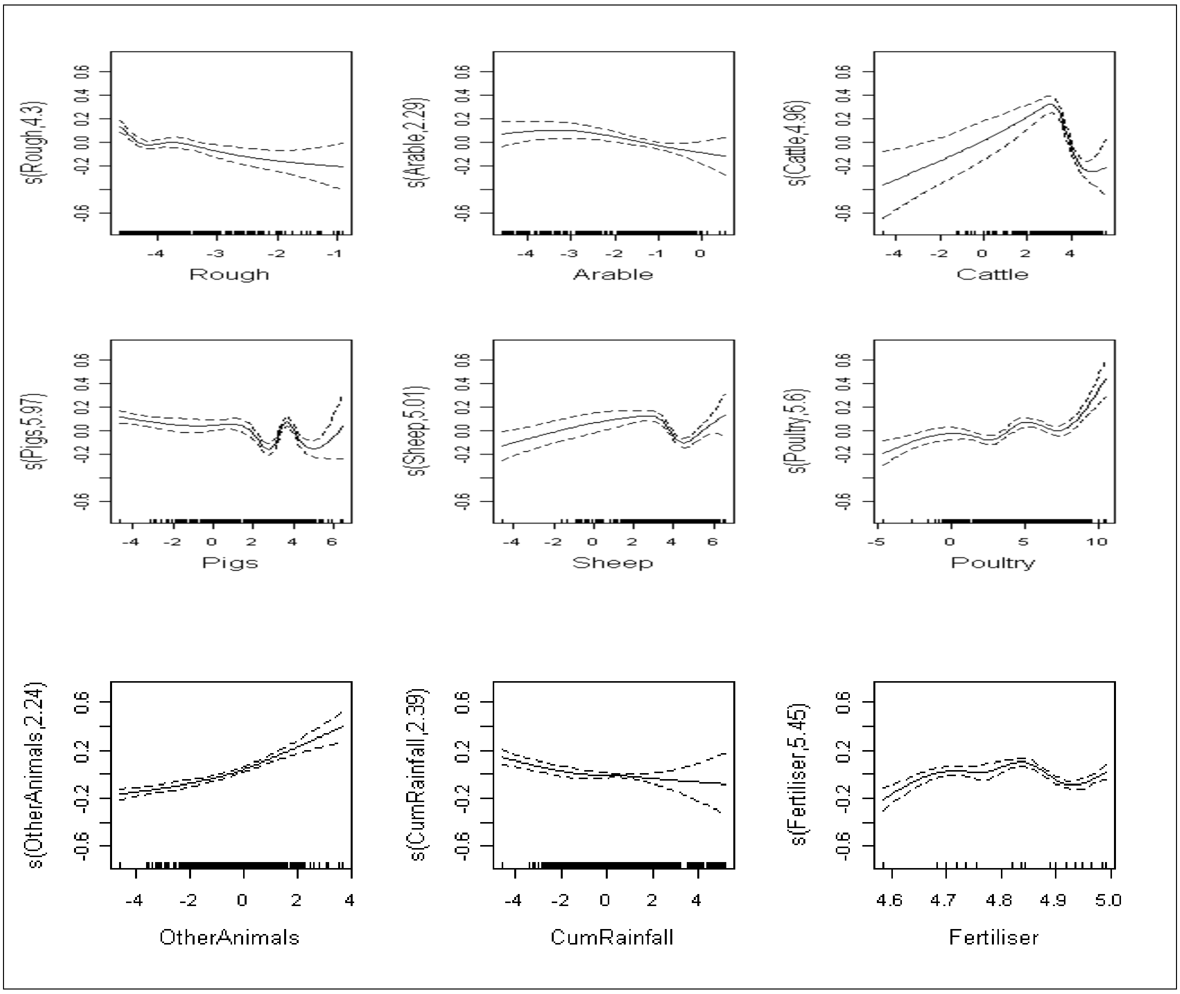
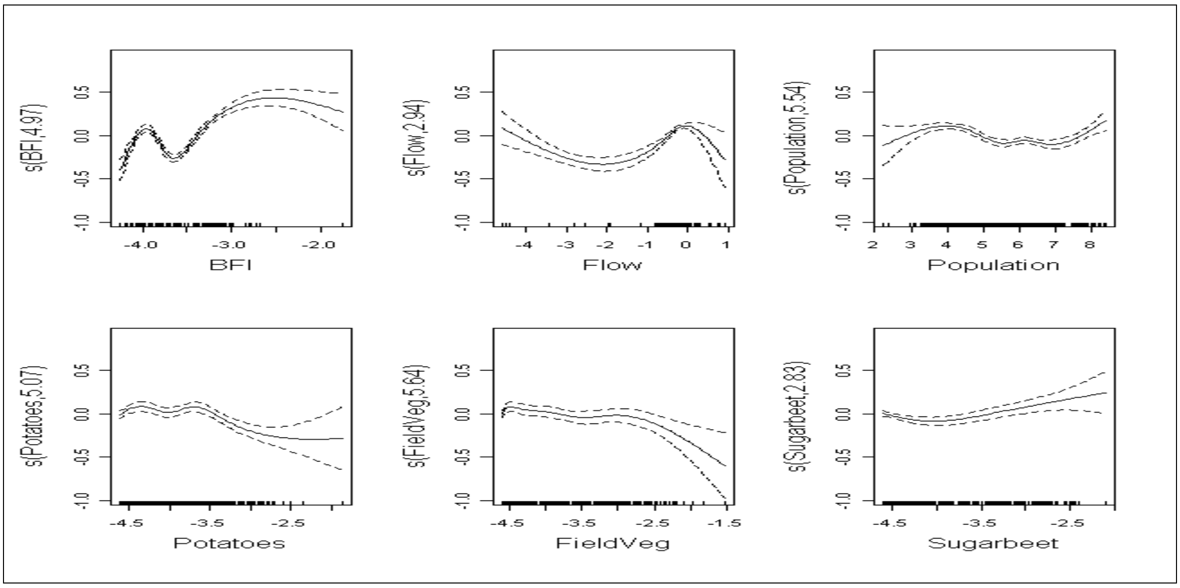


Figure A.25: Relationships between ln continuous covariates and ln Total Nitrogen – Model A2 in the Trent(28) LHA



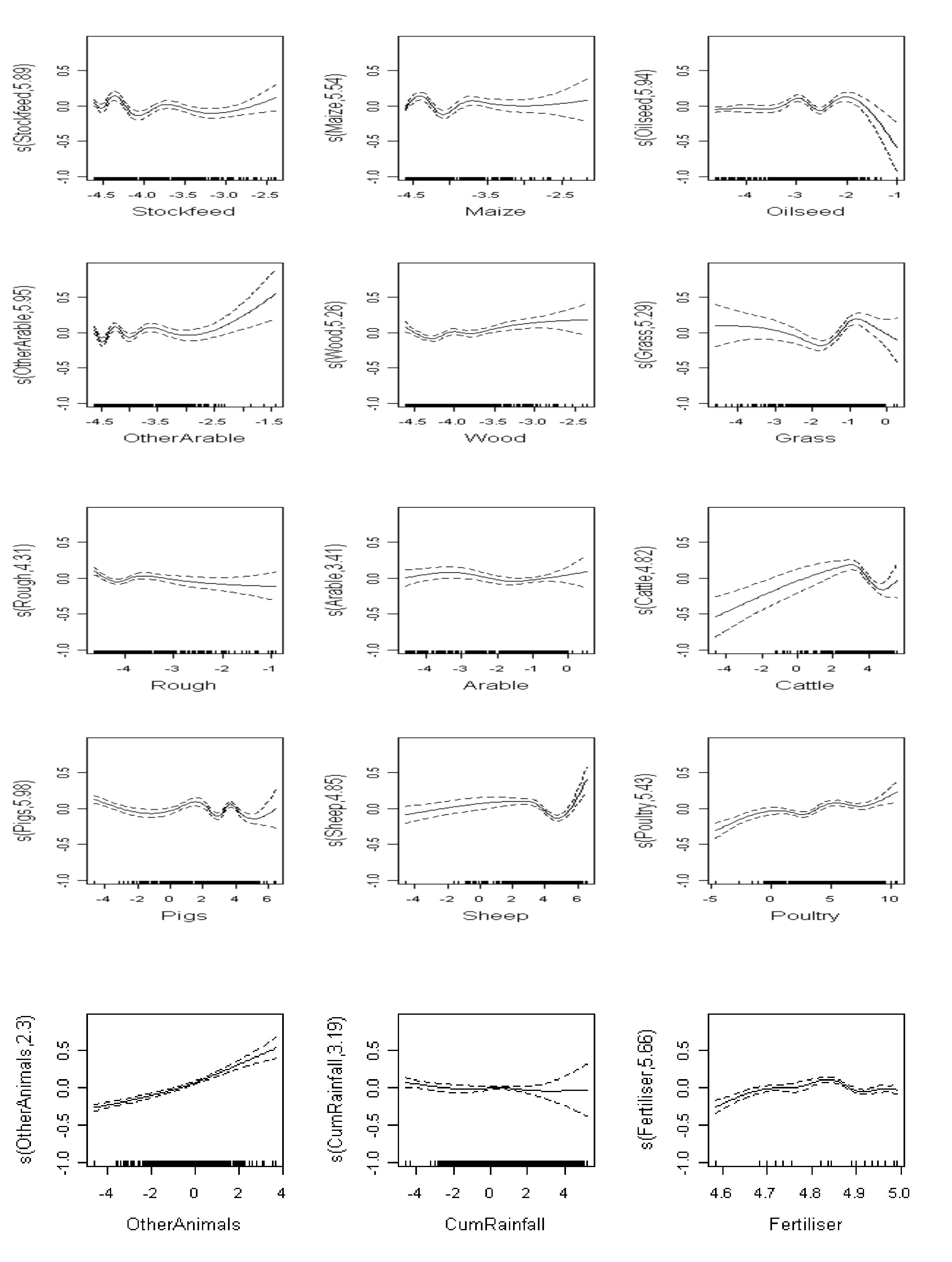
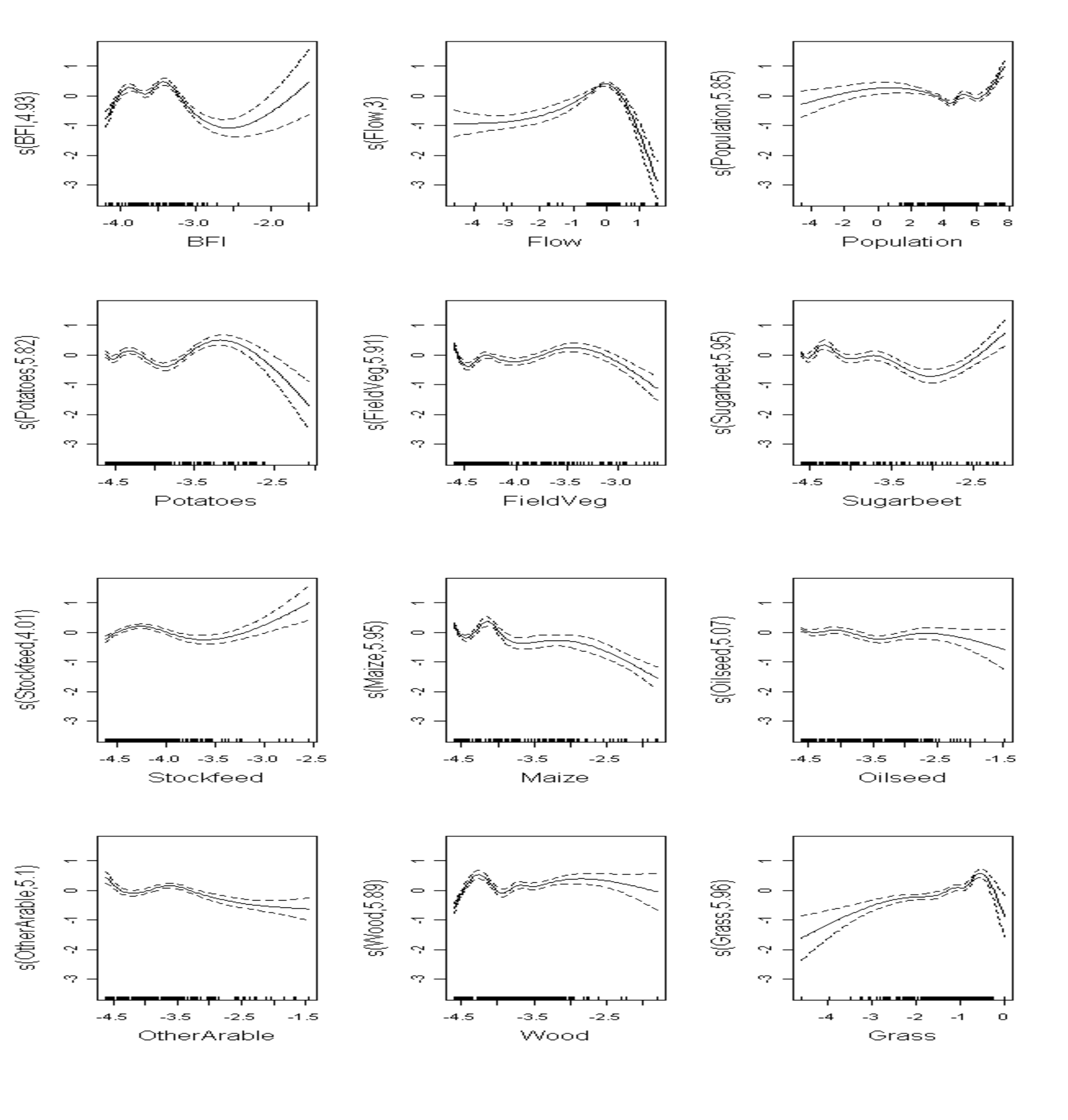

s(Stockfeed,5.89)
Stockfeed
s(Maize,5.54)
Maize
s(Oilseed,5.94)
Oilseed
s(OtherArable,5.95)
OtherArable
s(Wood,5.26)
Wood
s(Grass,5.29)
Grass
s(Rough,4.31)
Rough
s(Arable,3.41)
Arable
s(Cattle,4.82)
Cattle
s(Pigs,5.98)
Pigs
s(Sheep,4.85)
Sheep
s(Poultry,5.43)
Poultry
s(OtherAnimals,2.3)
OtherAnimals
s(CumRainfall,3.19)
CumRainfall
s(Fertiliser,5.66)
Fertiliser

Figure A.26: Relationships between ln continuous covariates and ln Orthophosphate-P – Model A2 in the Severn(54) LHA



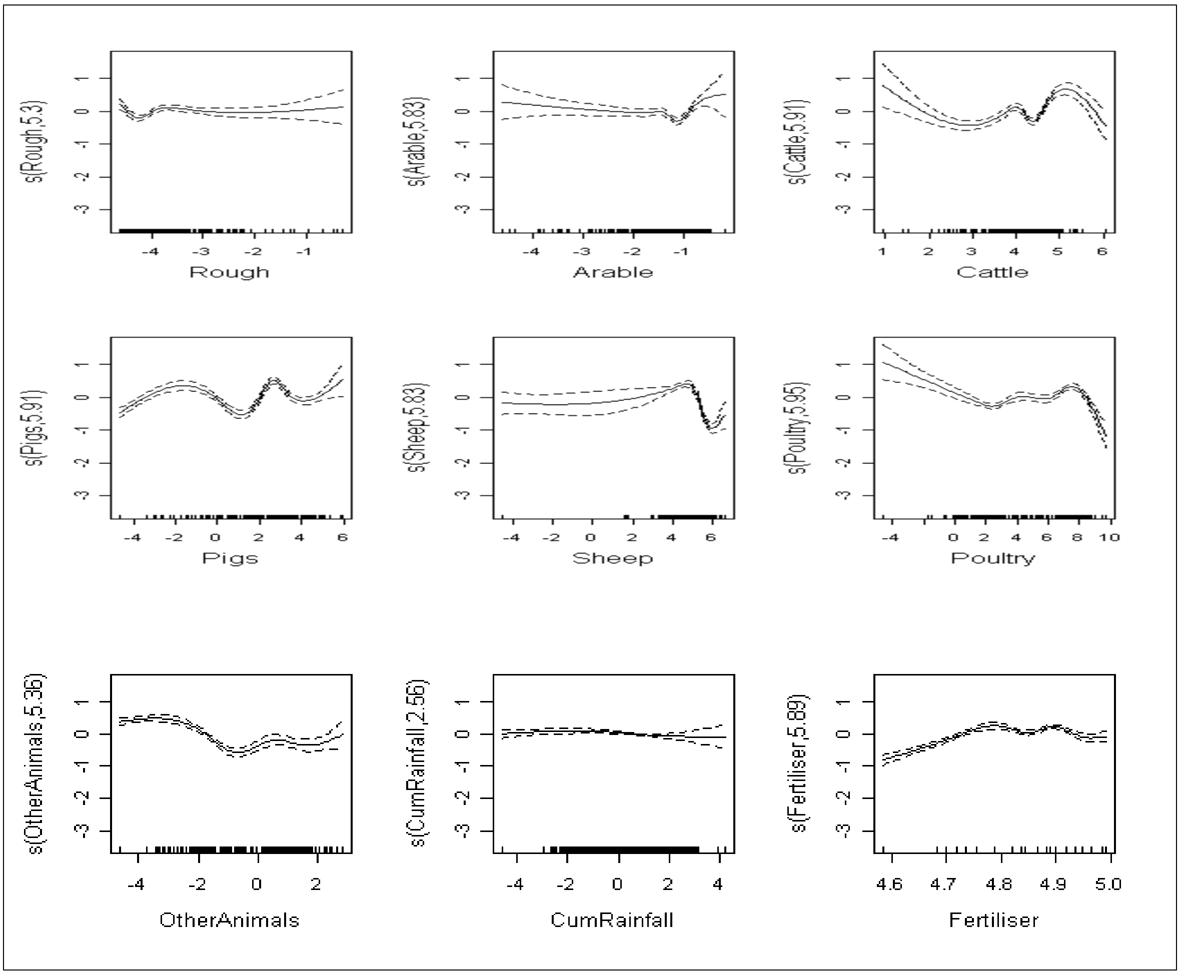
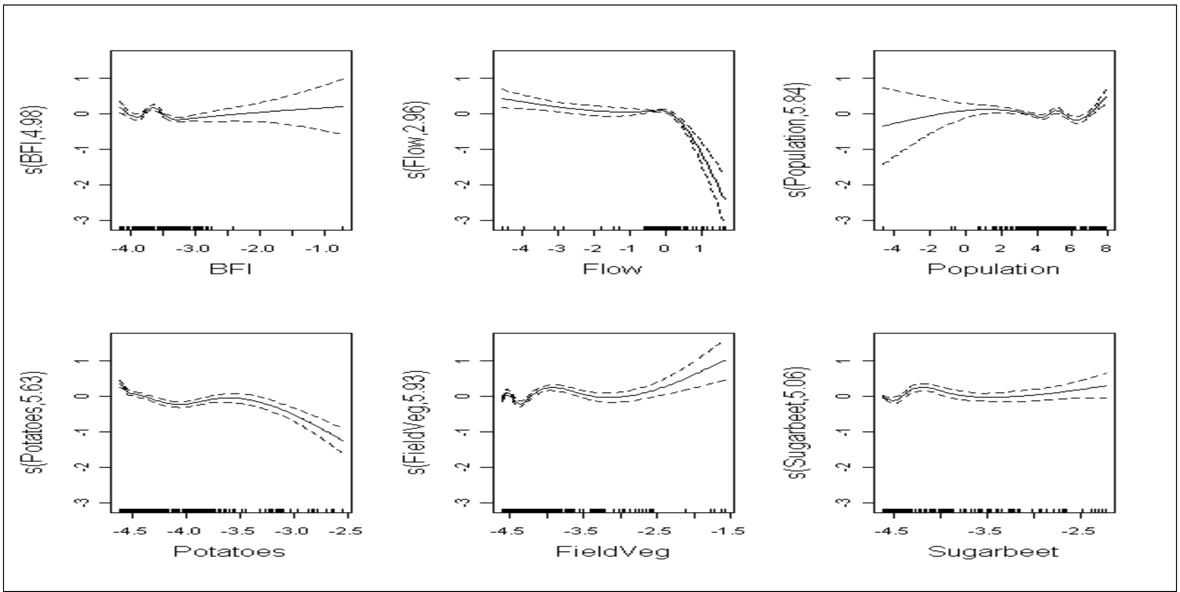


Figure A.27: Relationships between ln continuous covariates and ln Total Oxidised Nitrogen – Model A2 in the Severn(54) LHA



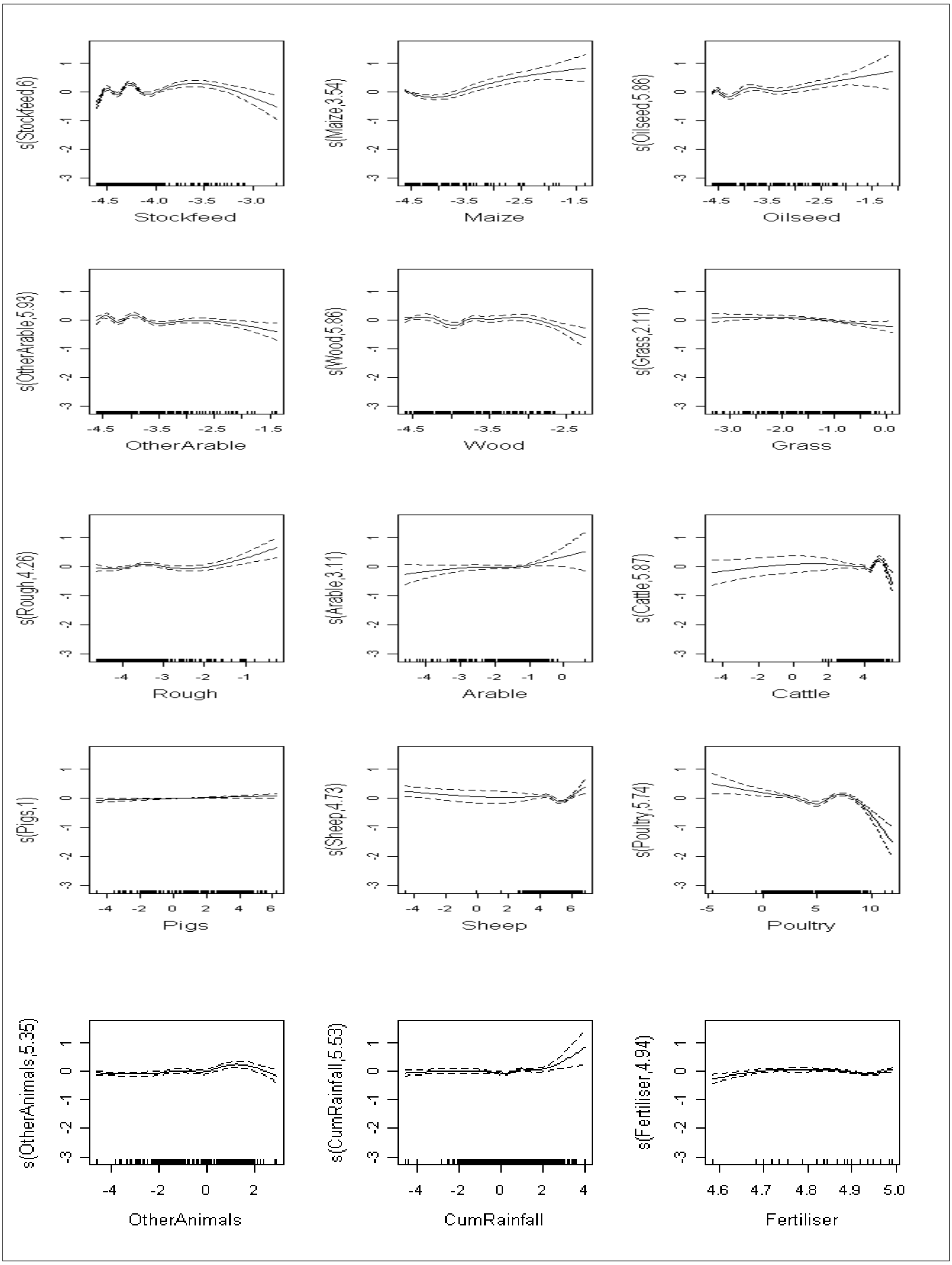
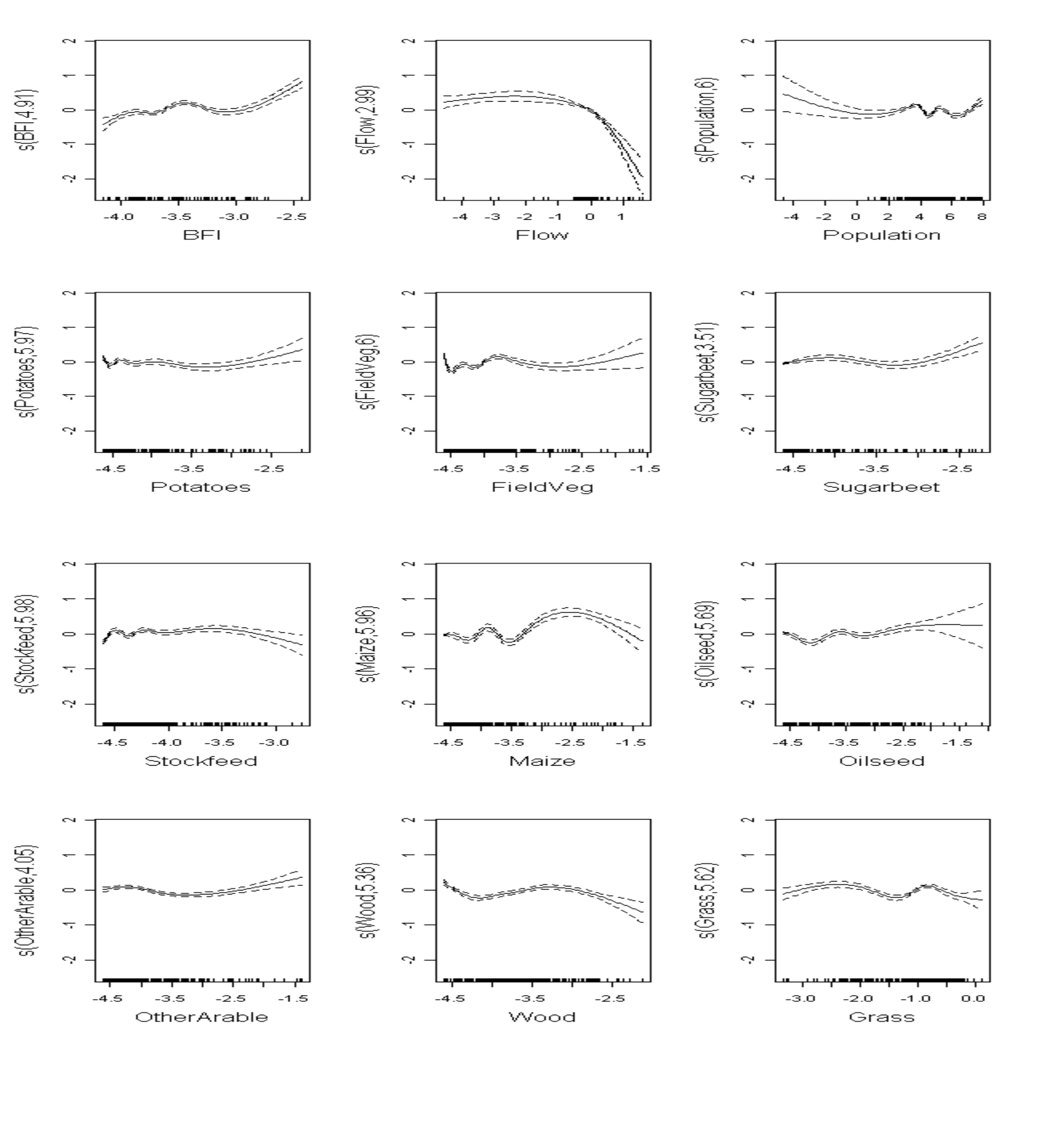

s(Stockfeed,6)
Stockfeed
s(Maize,3.54)
Maize
s(Oilseed,5.86)
Oilseed
s(OtherArable,5.93)
OtherArable
s(Wood,5.86)
Wood
s(Grass,2.11)
Grass
s(Rough,4.26)
Rough
s(Arable,3.11)
Arable
s(Cattle,5.87)
Cattle
s(Pigs,1)
Pigs
s(Sheep,4.73)
Sheep
s(Poultry,5.74)
Poultry
s(OtherAnimals,5.35)
OtherAnimals
s(CumRainfall,5.53)
CumRainfall
s(Fertiliser,4.94)
Fertiliser

Figure A.28: Relationships between ln continuous covariates and ln Total Nitrogen – Model A2 in the Severn(54) LHA



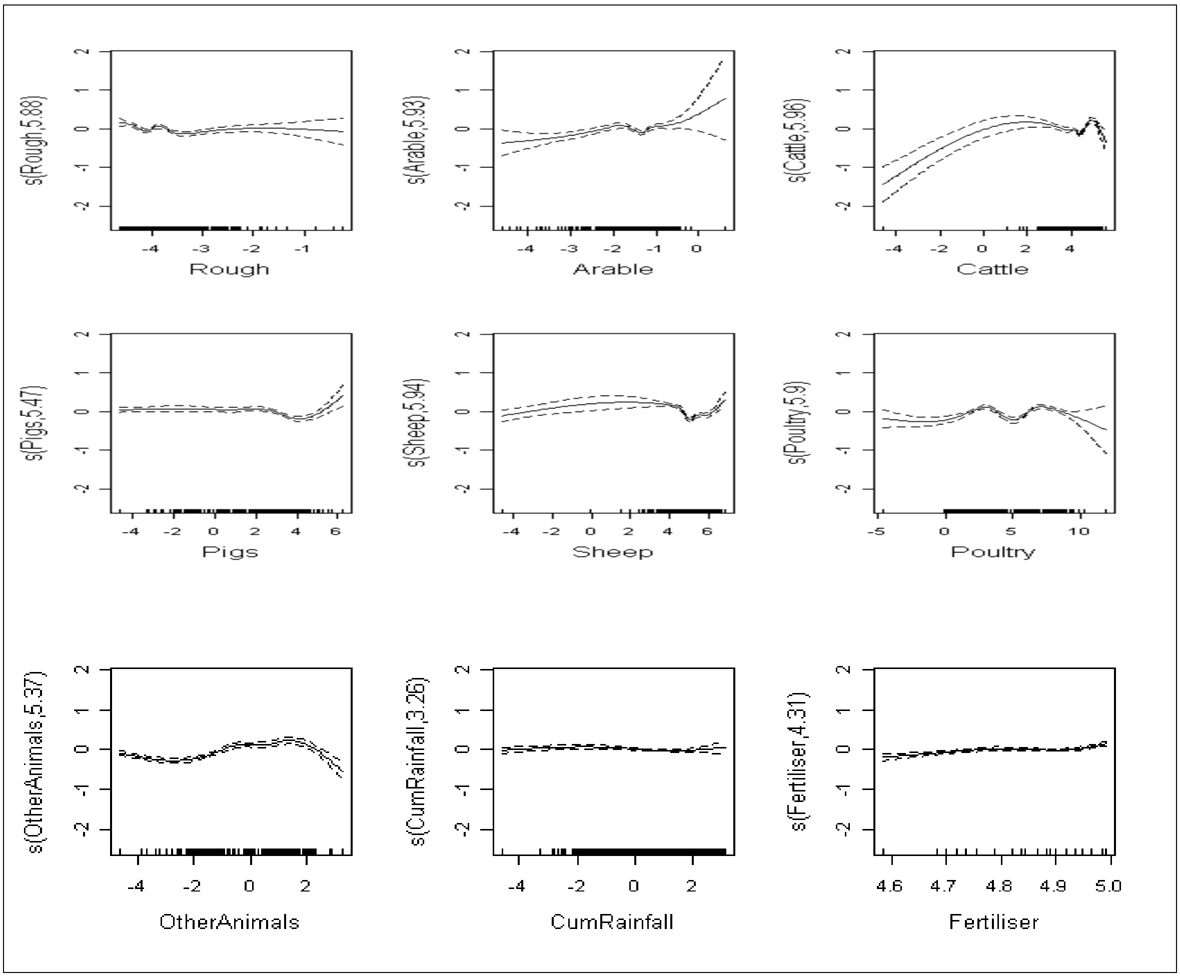
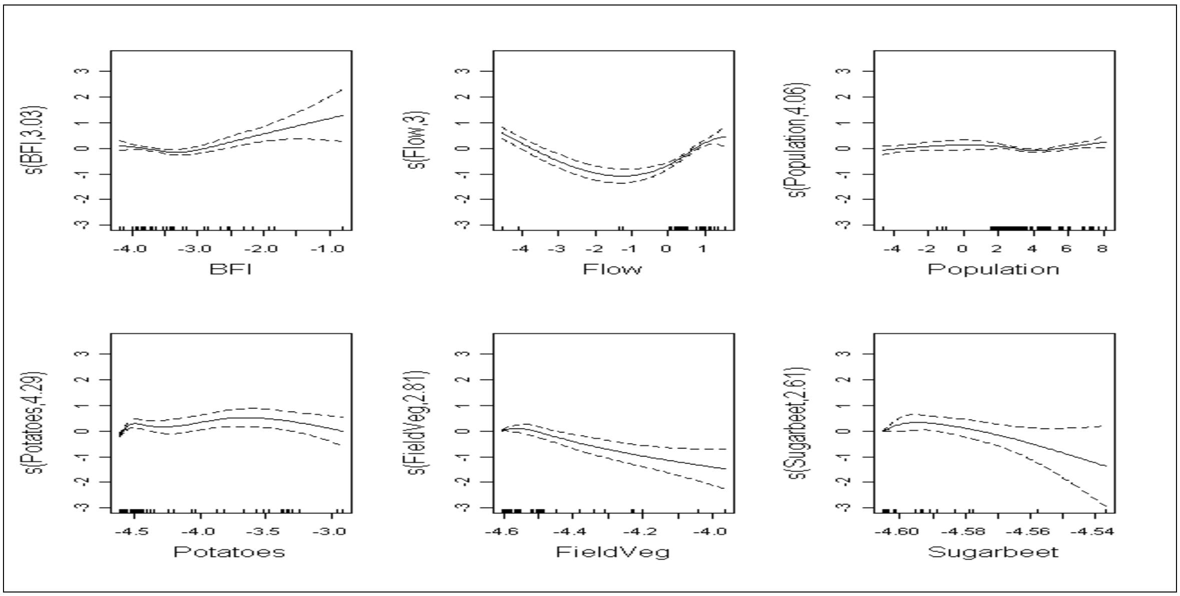


Figure A.29: Relationships between ln continuous covariates and ln Orthophosphate-P – Model A2 in the Lune(72) LHA



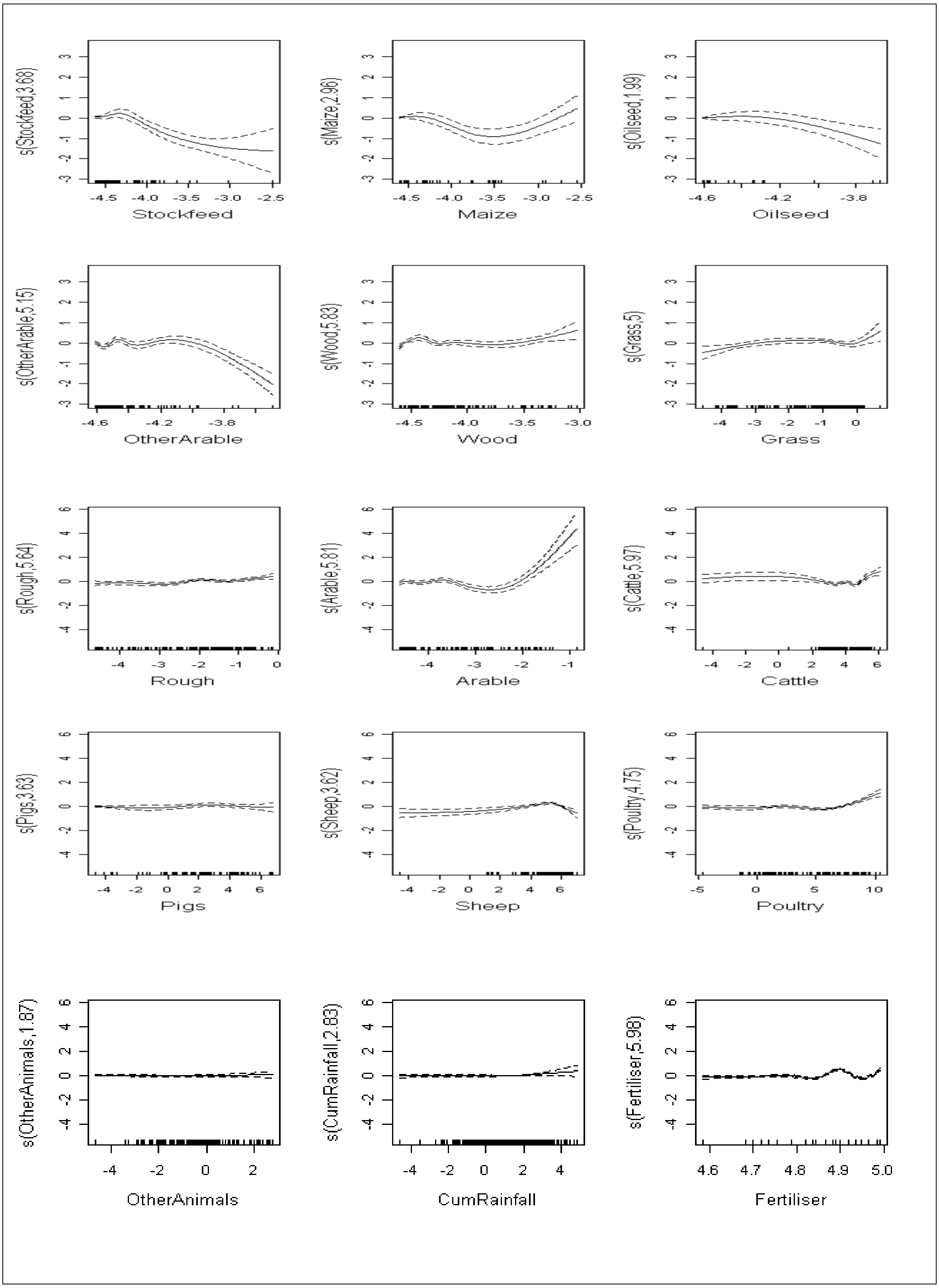
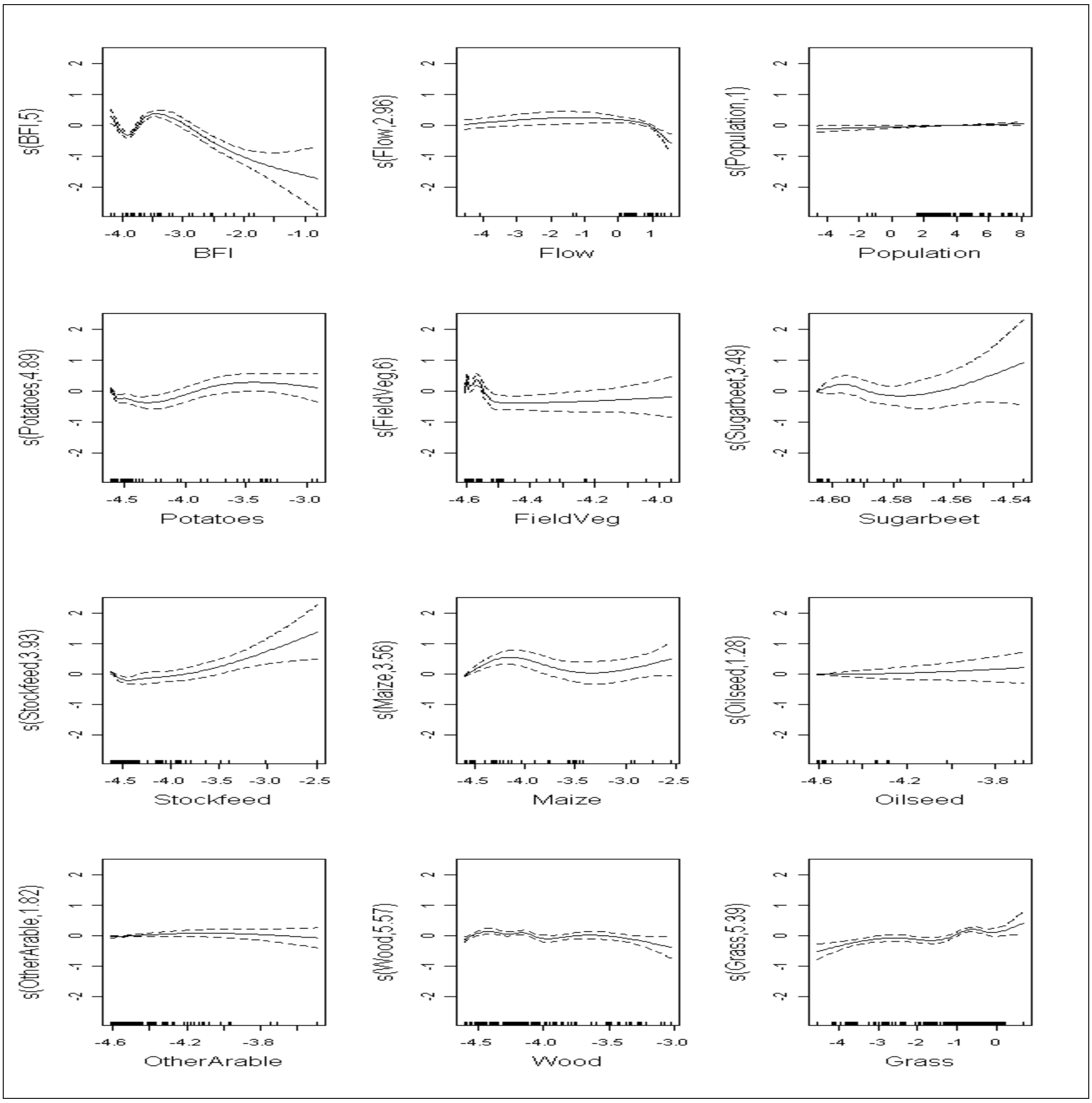

s(Stockfeed,3.68)
Stockfeed
s(Maize,2.96)
Maize
s(Oilseed,1.99)
Oilseed
s(OtherArable,5.15)
OtherArable
s(Wood,5.83)
Wood
s(Grass,5)
Grass
s(Rough,5.64)
Rough
s(Arable,5.81)
Arable
s(Cattle,5.97)
Cattle
s(Pigs,3.63)
Pigs
s(Sheep,3.62)
Sheep
s(Poultry,4.75)
Poultry
s(OtherAnimals,1.87)
OtherAnimals
s(CumRainfall,2.83)
CumRainfall
s(Fertiliser,5.98)
Fertiliser

Figure A.30: Relationships between ln continuous covariates and ln Total Oxidised Nitrogen – Model A2 in the Lune(72) LHA



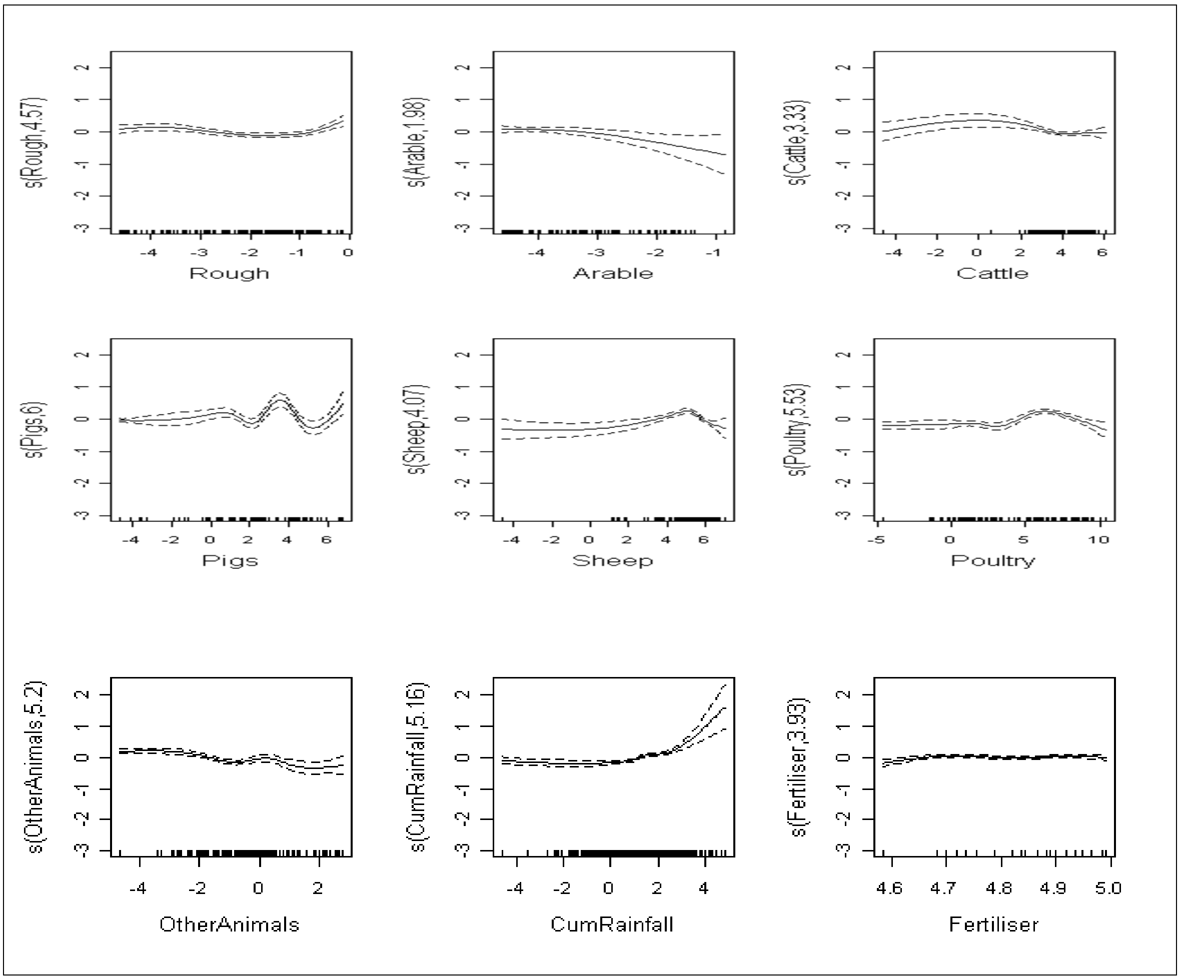
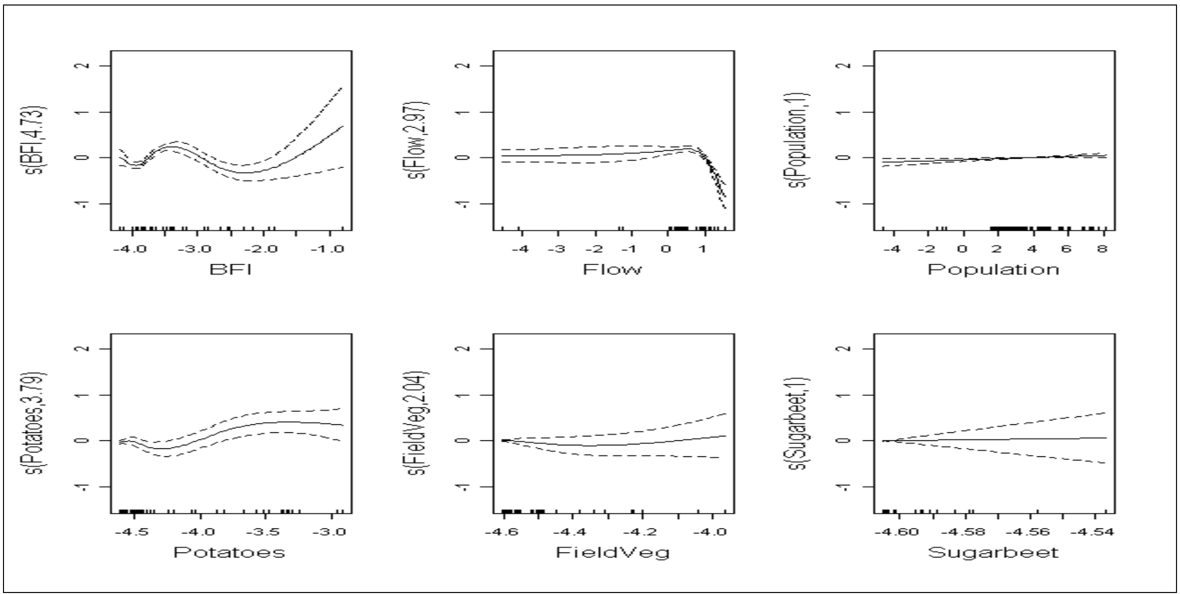
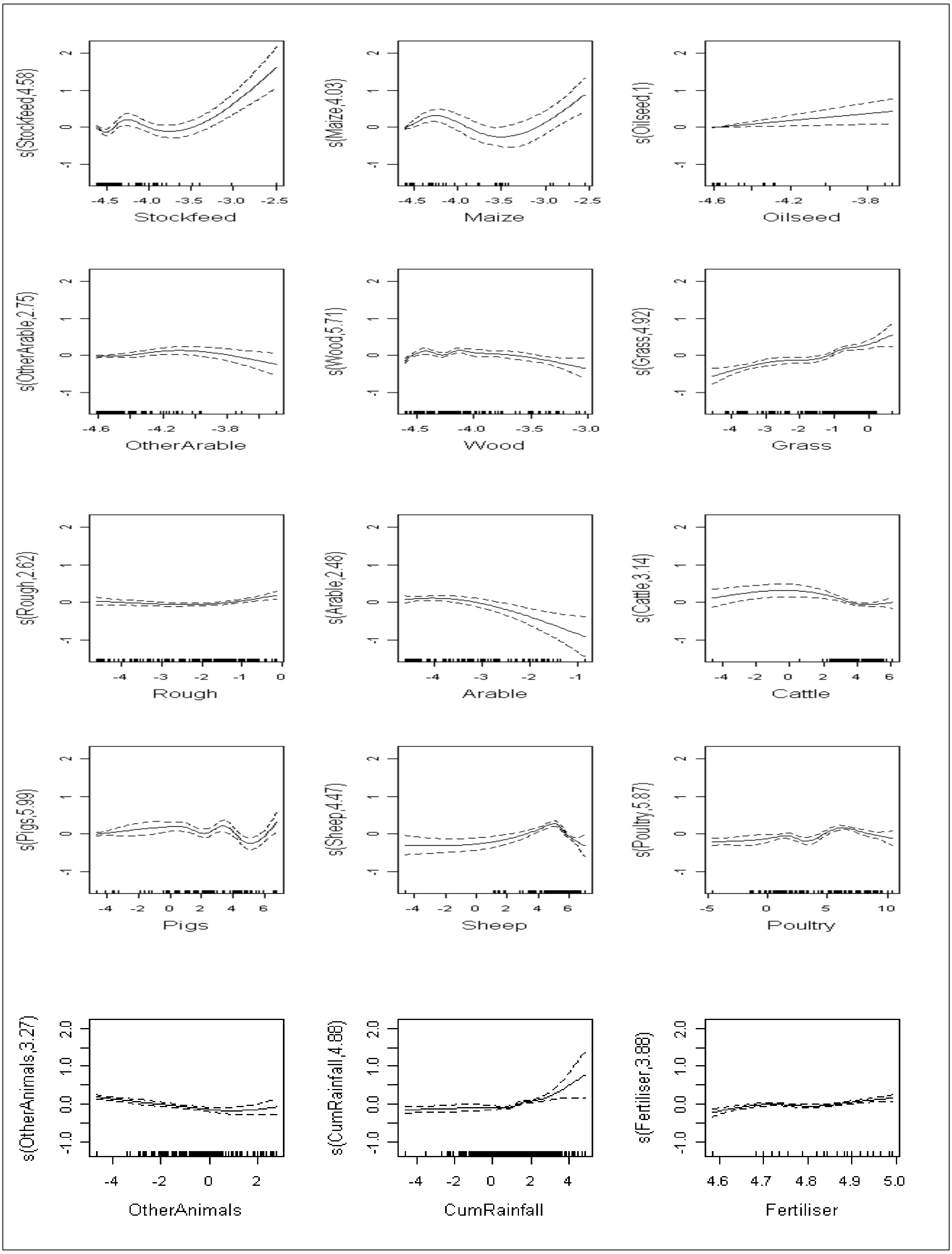


Figure A.31: Relationships between ln continuous covariates and ln Total Nitrogen – Model A2 in the Lune(72) LHA





s(Stockfeed,4.58)
Stockfeed
s(Maize,4.03)
Maize
s(Oilseed,1)
Oilseed
s(OtherArable,2.75)
OtherArable
s(Wood,5.71)
Wood
s(Grass,4.92)
Grass
s(Rough,2.62)
Rough
s(Arable,2.48)
Arable
s(Cattle,3.14)
Cattle
s(Pigs,5.99)
Pigs
s(Sheep,4.47)
Sheep
s(Poultry,5.87)
Poultry
s(OtherAnimals,3.27)
OtherAnimals
s(CumRainfall,4.88)
CumRainfall
s(Fertiliser,3.88)
Fertiliser

Section 4

Section 4 contains maps of the overall means for each of the LHAs for Model 1, see section 3 of the main report, for responses of ln OP, ln TON and ln TN. Model 1 estimates an overall mean for each LHA and all the trend and seasonal estimates are relative to this. All results are on the natural log (ln) scale.

Figure A.32: Overall means for ln Orthophosphate-P - Model 1.

Notes: There are no results for the LHAs in Wales or the Tweed(21) and the Tyne(23) as a result of the high proportion of non-detects for Orthophosphate-P.

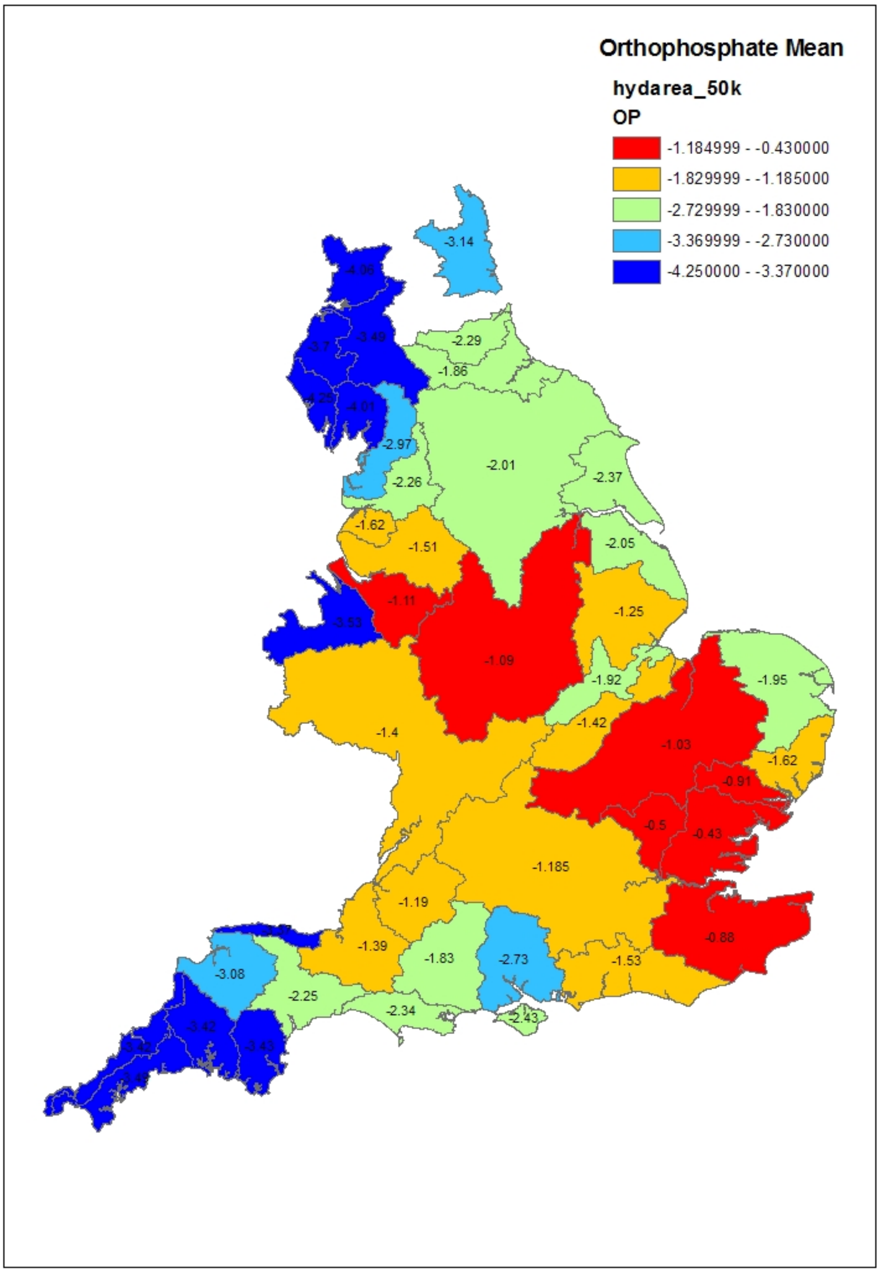


Figure A.33: Overall means for ln Total Oxidised Nitrogen - Model 1.

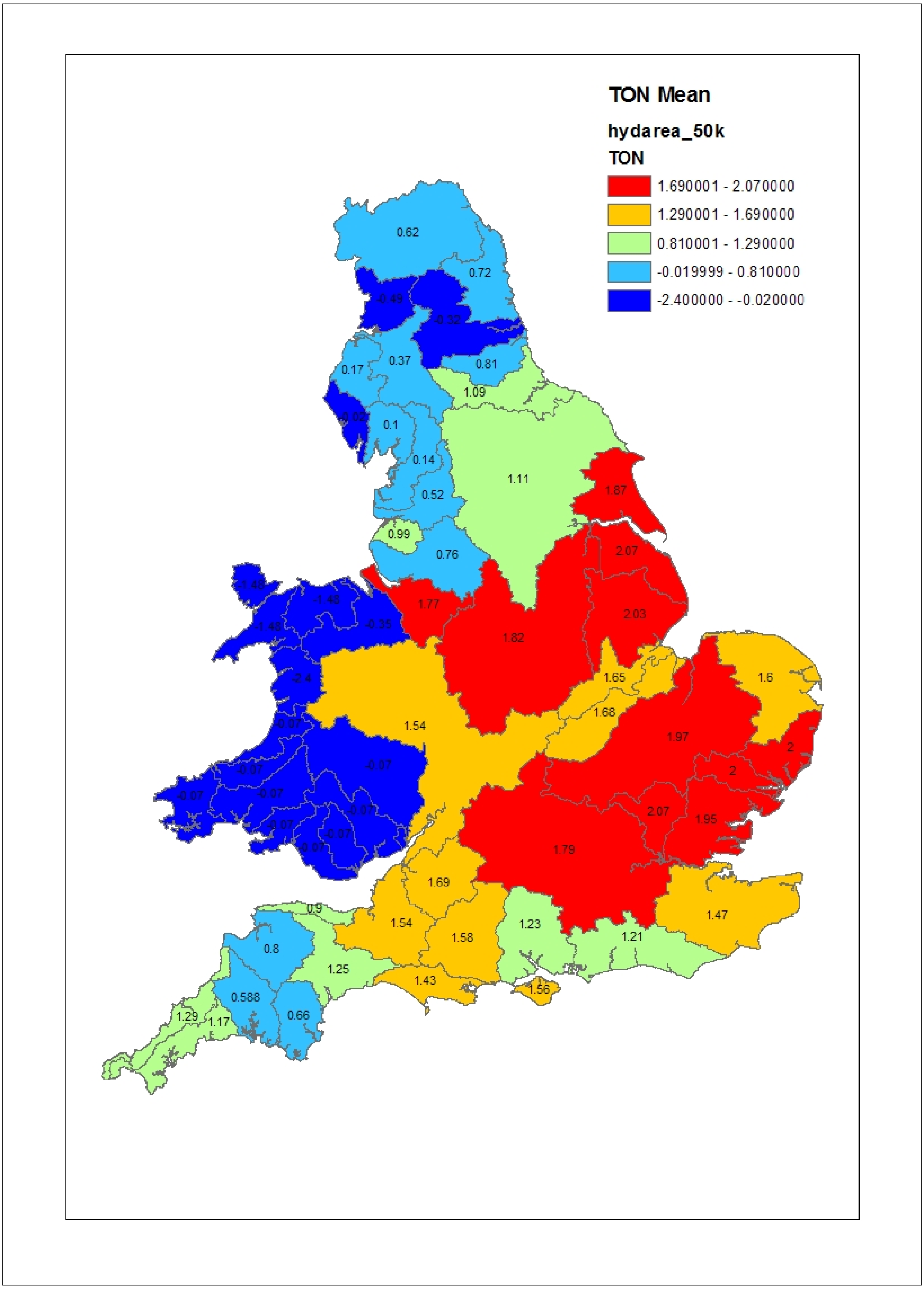
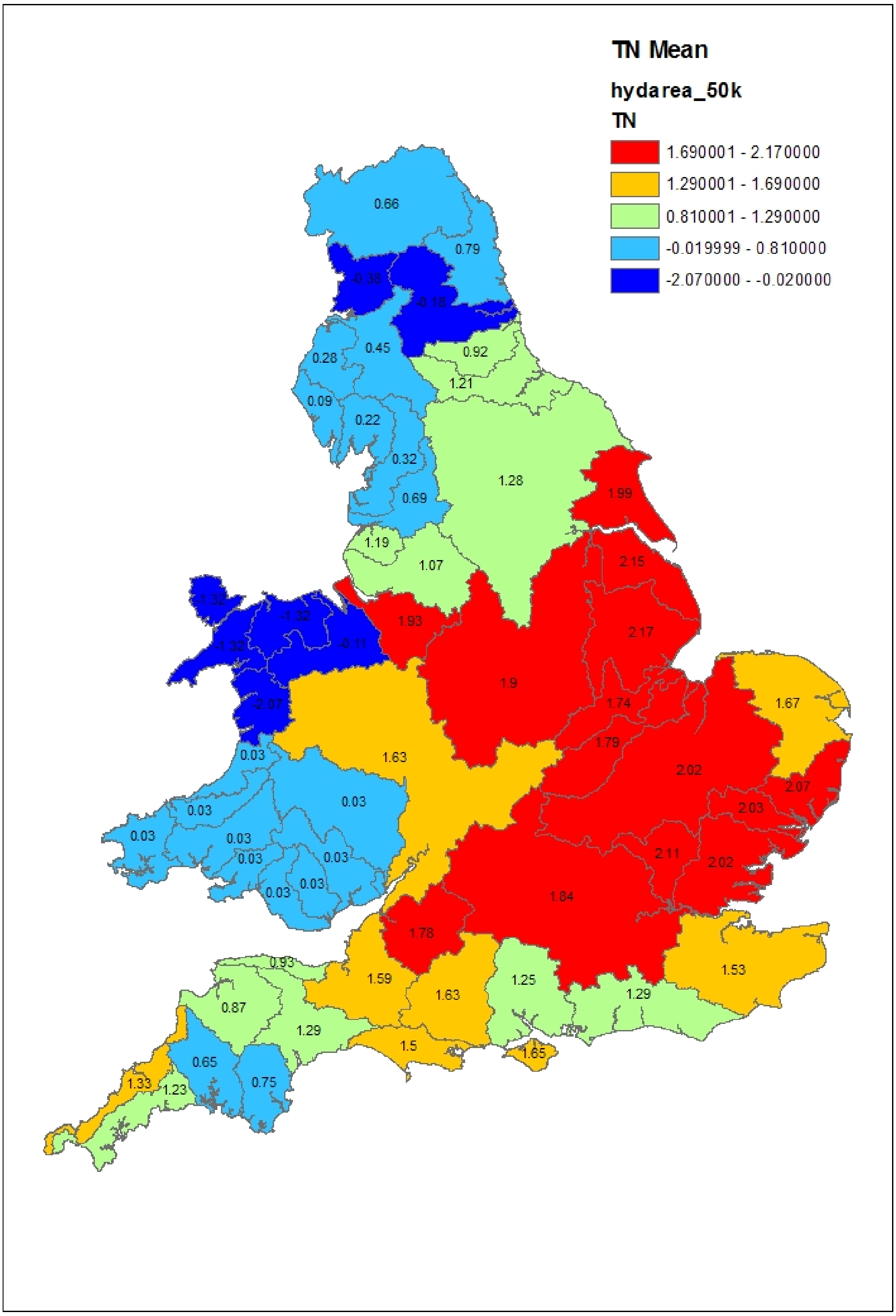


Figure A.34: Overall means for ln Total Nitrogen - Model 1.



Appendix C Supplementary appendix