rivm
National Institute
for Public Health
and the Environment

Report 680717016/2010

J. Ferreira

Estimation of net decreases in nitrate concentrations

Sample size required to demonstrate future decrease

rivm

RIVM Report 680717016/2010

Estimation of net decreases in nitrate concentrations

Sample size required to demonstrate future decreases

J. A. Ferreira (Consultant statistician)

Contact:
José Ferreira
Expertisecentrum voor Methoden en Informatie
Jose.Ferreira@rivm.nl

This investigation has been performed by order and for the account of B. Fraters, within the framework of Landelijk Meetnet effecten Mestbeleid (LMM).

rivm

Abstract

Estimation of net decreases in nitrate concentrations
Sample size required to demonstrate future decreases

Measurements of nitrate concentrations are regularly carried out in Dutch farms to assess the impact of governmental policies aiming to reduce the levels of nitrate. The RIVM has developed a method for determining the ‘sampling characteristics’, namely the number of farms and the number of measurements per farm, that are required in order to exhibit a future reduction in the levels of nitrate due to the introduction of governmental policies. The method is based on a statistical model that compares nitrate concentrations and related measurements from two different years. Using this method, sampling characteristics have been determined based on data from 1992 to 2008.

The underlying statistical model should be realistic and at the same time simple enough to allow the consideration of ‘scenarios’, or hypothetical situations. Suppose that nitrate levels decrease by 20% as a consequence of a new governmental policy. How many farms and how many measurements per farm are required in order to be able to exhibit such a decrease? Besides the percent of decrease and the sampling characteristics, a scenario includes the specification of certain parameters of the statistical model. One such parameter is the amount of variability in the measurements carried out within a farm.

A statistical model describing the changes in nitrate concentrations has recently been proposed by L.J.M. Boumans en B. Fraters in a preliminary RIVM report where it is suggested that the nitrate concentrations in agricultural farms have decreased between 1992 and 2006 as a result of the introduction of governmental policies. However, the model of Boumans and Fraters seems too complicated to allow the determination of sampling characteristics. Thus, although its objective is to provide estimates of sampling characteristics, the best part of the present report is devoted to the problem of setting up, estimating and validating a model to describe changes in nitrate concentrations and which can be used to determine sampling characteristics. As a by-product of this effort, this report supports the overall conclusion of Boumans en Fraters to the effect that nitrate concentrations have decreased during the 1992-2006 period as a result of governmental policies and/or changes in farm management.

Key words:
nitrate concentrations, detection of changes, sample size calculations, statistical models

Rapport in het kort

Schatten van de gerealiseerde daling van nitraatconcentraties
Benodigde steekproefomvang om een toekomstige daling aan te tonen

Op landbouwbedrijven worden regelmatig de concentraties nitraat gemeten om te kijken of die door beleidsmaatregelen afnemen. Het RIVM heeft een methode ontwikkeld om de 'steekproefomvang' te bepalen, die aangeeft aan op hoeveel landbouwbedrijven moet worden gemeten en hoeveel metingen per bedrijf moet worden verricht om zo'n afname te kunnen aantonen. De methode is gebaseerd op een statistisch model dat de nitraatconcentraties en bijbehorende variabelen van twee jaren vergelijkt. Met behulp van deze methode zijn schattingen van de steekproefomvang gepresenteerd gebaseerd op data uit de periode 1992 tot 2008.

Het statistische model achter de methode moet zowel realistisch als simpel zijn om aan te kunnen geven wat de kans is om het effect van bepaalde beleidsdoelen waar te nemen, bijvoorbeeld een afname van de nitraatconcentratie met twintig procent. Behalve het percentage afname en de steekproefomvang moeten voor dergelijke scenarioanalyses nog andere berekende parameters in het statistische model worden gespecificeerd. Een voorbeeld daarvan is de mate waarin de meetresultaten op een bedrijf variëren.

In 2009 heeft het RIVM al een statistisch model in concept ontwikkeld (Boumans en Fraters). Hierin is aannemelijk gemaakt dat de nitraatconcentraties op landbouwbedrijven tussen 1992 en 2006 zijn gedaald als gevolg van beleidsmaatregelen. Het model is echter te complex om de benodigde steekproefomvang te bepalen om een toekomstige daling aan te tonen. Daarom is een eenvoudiger model ontworpen en gevalideerd, waarmee vervolgens de benodigde steekproefomvang is vastgesteld. Dit model bevestigt overigens de eerdere bevinding van het RIVM dat de nitraatconcentraties gedurende de periode 1992-2006 zijn gedaald door beleidsmaatregels en/of door verandering in de bedrijfsvoering.

Trefwoorden:
nitraat concentraties, detectie van verandering, steekproef karakteristieken, statistische modellen

rivm

Contents

Summary

Measurements of nitrate concentrations are regularly carried out in Dutch farms in order to assess the impact of governmental policies aiming to reduce the levels of nitrate. A question of interest in this context is whether a decrease in nitrate levels observed during a given period can be attributed to a governmental policy introduced prior to the observation period rather than to climatic variations and/or other factors known to influence the levels of nitrate. This question can be addressed by means of statistical methods. A statistical model that describes the changes in nitrate occurring between two given years is set up and used to test, on the basis of nitrate measurements and other relevant data, the hypothesis that a *net decrease*—i.e. a decrease attributable to governmental policies—has occurred, and eventually to estimate it. However, the ability of the model to actually exhibit a net decrease depends on the amount of data collected and on the magnitude of the decrease: the more data are available, the easier it is to provide evidence of the decrease; and the bigger the decrease, the easier it is to bring it to light. Thus it can happen that, due to a scarcity of data, one fails to exhibit a relatively small but practically relevant net decrease. In order to prevent this, and also to avoid wasting unnecessary resources, it is useful to determine a *sampling plan*, consisting of estimates of the number of farms that need to be sampled and of the number of measurements that need to be made per farm, which will (with high probability) allow one to detect a decrease of a given magnitude.

The determination of a sampling plan requires the setting up of a statistical model that is realistic and at the same time simple enough to allow the specification of hypothetical situations and of parameters governing the amount of variability in the data. This report proposes such a model, develops a method for determining sampling plans, and provides numerical examples of sampling plans based on existing data. Analyses of data from 1992 tot 2008 indicate that the model is accurate enough for the purposes of detecting net changes in nitrate levels and determining sampling plans. The estimates of the number of farms and number of measurements per farm provided, though admittedly rough, are deemed useful in assisting decision-makers.

As a by-product of the effort of developing and validating a statistical model to describe changes in nitrate concentrations, this report supports the overall conclusion of a preliminary RIVM report by L.J.M. Boumans en B. Fraters to the effect that nitrate concentrations have decreased during the 1992-2006 period as a result of governmental policies and/or changes in farm management.

1 Introduction

The main problem considered in this report concerns the estimation of the net decrease in mean nitrate concentrations in ground water that is supposed to have taken place in the Netherlands from 1992 to 2008, as a result of policies governing the application of nitrate in farms. The estimation is based on a data set consisting of nitrate concentration measurements and measurements of other relevant variables obtained from around 500 Dutch farms during the 1992–2008 period. The problem has already been considered by Boumans and Fraters (2009) on the basis of the same data but using a different approach. The reason for considering it anew is motivated by the need to determine the **sample size ─** namely the number of farms and the number of measurements per farm ─ required in order to exhibit a net decrease of mean nitrate concentrations in the future, which is the second problem we consider here. Indeed, the model used by Boumans and Fraters (2009) seems to be too complicated to allow such a sample size determination. On the other hand, Boumans and Fraters (2009) do not provide confidence intervals for net decreases, nor do they provide separate estimates of net decreases for the different farm types; consequently, the analysis, estimates and confidence intervals presented here should be of interest in their own right (and not simply as means of providing a platform for sample size determinations) and regarded at least as a complement to the work of Boumans and Fraters (2009).

A more detailed explanation as to why we require a reanalysis of the data used by Boumans and Fraters (2009) in order to devise a method of sample size determination is provided in **Appendix A**. A preliminary version of the report of Boumans and Fraters (2009) can be found in **Appendix D**.

The amount of nitrate in ground water found at a given time in a given farm is not only a function of the amount of nitrate applied by the farmer in the form of fertilizer and manure, but also a function of precipitation patterns and of the rate of *denitrification*, the process by which nitrate is decomposed into molecular nitrate. Furthermore, the impact of precipitation and denitrification on the leaching of nitrate into ground water depends in turn on the type of cultures/activities in the farm and on the type of soil. Consequently, in order to be able to estimate a **net** or **compensated decrease** in nitrate levels during a given period ─ i.e., a decrease that is due solely to a decrease in the application of nitrate and/or to changes in farm management ─ on the basis of nitrate concentration measurements one has, in principle, to account (or 'correct') for the relevant precipitation, farm and soil type characteristics, coupled with the nitrate concentrations. The data set used by Boumans and Fraters (2009) has been prepared with this in mind: together with the nitrate concentration measurements obtained in a given year at a given farm, it contains the farm's type (e.g. 'arable' or 'dairy'), two variables characterizing the farm's type of soil ('fraction of dry soil' and 'fraction of neutral soil') and, coupled with the nitrate concentrations, the concentrations of a so-called **tracer**, a variable that simulates the effect of precipitation on the leaching of nitrate. While the definitions and roles of 'type of farm' and 'type of soil' are, in essence, easy to grasp for the layman, the meaning of the tracer is not obvious; therefore, and in view of the need for a correct interpretation of this variable for the basic model to be proposed later on, we shall give an intuitive description of it based on explanations provided by Leo Boumans (see sections 1.3 and 2.1.2 of Boumans and Fraters (2009) for more specialized references, which we have not consulted).

Consider an inert substance ─ a 'non decomposable salt'─ present in the soil in a certain concentration at the beginning of an 'observation period'. Due to precipitation and other factors, the salt tends to gradually sink into the soil and penetrate into the ground water. The **tracer concentration**, which just like nitrate concentration can be thought of as being measured in volumic mass, is an estimate of the amount of salt found in the ground water at the end of the observation period, the estimate being obtained with a simulation model that takes climate data (including precipitation data) and soil characteristics as input. Thus, ideally, the tracer concentration coupled to a given nitrate concentration measurement is proportional (the proportionality constant being determined by the relative abundance of nitrate and salt initially present in the soil) to the nitrate concentration *assuming that the amount of nitrate in the soil remains constant* (just like the amount of the inert substance does). In other words, the tracer concentration would be proportional to the nitrate concentration if only denitrification did not occur.

This interpretation of the tracer will play an important role in chapter 3, where we set up a simple but plausible model for describing and quantifying net decreases in nitrate levels within a given period. Chapter 4 describes the statistical methods used to estimate those net decreases and chapter 5 presents the estimates and their confidence intervals. The method of sample size determination is presented in chapter 6, together with numerical illustrations. Finally, chapter 7 provides some discussion, conclusions and recommendations. The next chapter is devoted to a discussion of the data set.

2 The data set

The data set provided to us consists of 3631 rows and seven columns of data, the columns representing seven variables:

farm code,
farm type (‘arable’, ‘dairy’, ‘factory’, ‘other’, ‘unknown’),
year (in which the measurements were taken),
nitrate concentration (mg/l),
tracer concentration,
fraction dry soil,
fraction neutral soil.

As we learned from Leo Boumans, each nitrate concentration measurement is obtained by determining the content of nitrate in a *mixture* of eight water samples from different, randomly chosen wells in a farm (one water sample per well); such a mixture of samples will be called a **composite sample**. Thus, a nitrate concentration measurement from a given farm can be thought of as an average of eight ‘replicate’ nitrate concentrations, and hence as an estimate of the mean nitrate concentration in that farm. Analyzing a mixture of water samples from different wells is of course cheaper than analyzing the water samples separately, but it does not allow us to estimate the within-farm variance unless there are at least two samples of composite samples.

If we use ‘farm code’ to identify farms, the data set contains measurements from 494 farms. However, there are 538 different farm code/farm type combinations, because in the course of time 44 of the farms have had a change of type. For this reason, we identify farms by their joint combination of code and type and thus consider that there are 538 farms in total.

Figure 1 (see the Figures section between the list of references and Appendix A) gives an overview of the sampling regime throughout the 1992–2008 period. Many farms have been sampled several years in a row. In 1996 no measurements were taken. The sampling regime was intensified towards the end of the period, especially in the last three years. ‘Dairy’ is by far the type of farm most represented in the data set. Data on ‘factory’ farms became available only from 1998 onwards. **‘Unknown’ farms** are farms whose type is missing in the original data file; there are only 11 of them and they will be discarded in the statistical analyses.[1]

The absolute and relative abundance of sampled farms by farm type can be better appreciated in the histograms of Figures 2 and 3. There appear to be some changes in the distribution of farm types throughout the years, which raises the question of whether the distribution of farm types represented in the sample reflects the actual distribution of farm types in the Netherlands. If this is not the case, then one can anticipate that an unbiased *national estimate* of a given quantity (e.g. mean nitrate concentration per farm in a given year) cannot be computed solely on the basis of this data set; instead, it must be computed as a weighted average of *stratified estimates* (the strata corresponding to the

[1] The measurement of nitrate in farms has been carried out in cooperation with the Landbouw Economisch Instituut (Agricultural Economic Institute), which provides the official classification of farms; not all farms had been classified when the data set was made available to us.

different farm types), with the weights accurately determined from census or registry data. One can anticipate in particular that the net decrease in nitrate levels follows different patterns in different farm types; for this reason, *all estimates of net decrease to be given in this report pertain to a given farm type stratum*.

The histograms of Figure 4 show the 'average sample size' throughout the 17-year period, i.e., the average number of measurements per farm per year, stratified by farm type. Except for the first two years, where four measurements were taken per farm, and for the third year, where the average number of measurements more than doubled, most years have an average number of two measurements per farm per year. These observations apply almost uniformly across farm types. The year 2000 is somewhat poorer, in that many farms have only a single measurement.

We have chosen to base our analysis on **aggregated data**, namely on the averages of the logarithm[2] of nitrate and tracer concentrations measured within each farm in the course of a year, rather than on the original data consisting of individual nitrate and tracer concentrations. This has the disadvantage that different observations (averages) may have different variances (as they may be based on different sample sizes), which needs to be accounted for in the computation of confidence intervals, but in the end it will make for a simpler analysis.

The box plots of Figures 5A and 5B show the distribution of the within-farm averages of log nitrate and log tracer concentrations throughout the 17-year period, stratified by farm. Since the number of original observations per farm remains approximately constant within a year, the aggregated observations (i.e., the within-farm averages of log nitrate and log tracer) from each year represent samples from a population with at most a small proportion of observations with greater or smaller variance than the rest. Comparing the box plots in the top panel of Figure 5A with those in the lower panel (for example) we see that there has been an overall decrease in nitrate levels but that this decrease has to a certain extent been followed by the tracer; consequently, it is not obvious whether there really has been a *net* decrease in nitrate ─ a decrease that resulted from policies on nitrate application and not from changes in precipitation patterns (which govern the levels of the tracer).

It is also seen from the box plots that the amount of spread in the distributions of log nitrate and log tracer sample averages can change over time; compare, for example, the data from arable farms in 2005 and 2007.

We have stated above that many farms have been sampled several years in a row. But an important characteristic of the data set is that most farms have *not* been measured for more than a couple of years. For example, none of the farms sampled in 1992 has been sampled in 2007. And while it is true that most farms sampled in 1993 and in 2008 have also been sampled in the preceding year, no net decreases in nitrate levels are likely to have occurred during 1992–1993 or 2007–2008. As explained in chapter 3, this makes it impossible to estimate the net changes in nitrate from one period to the next in a *direct* fashion.

Remark: Most of the analyses described in this report have been carried out with R software (http://www.r-project.org/); files with the R scripts can be obtained from the author.

[2] The reason for using the logarithm of the concentrations in place of the concentrations themselves will be given in chapter 3.

3 Basic model

Let $N_{i,j}(t)$ denote the *j*-th among a number of nitrate concentrations measured at time (year) *t* on a farm identified by label *i* of a given type and with given soil characteristics. Evidently, we can always stratify farms according to farm type and then according to 'proportion of dry soil' and 'proportion of neutral soil' jointly; we assume that the resulting categories of farm type and soil characteristics have been defined and that all quantities considered here pertain to a fixed combination of such categories — i.e., to a *stratum*. Let $T_{i,j}(t)$ stand for the tracer concentration coupled to $N_{i,j}(t)$. Then the following relation should give an approximate description of the change in nitrate concentration occurring from time *s* to time *t*:

$$\frac{N_{i,j}(t)}{N_{i,j}(s)} = \frac{T_{i,j}(t)}{T_{i,j}(s)} \frac{A_{i,j}(t)}{A_{i,j}(s)} \frac{B_{i,j}(t)}{B_{i,j}(s)} .$$

Indeed, according to the interpretation of the tracer given in the Introduction, if nitrate were an inert substance scattered in the soil then its concentration in the ground water at a given time would be equal to a constant times the concentration of the tracer in the ground water at that time and, consequently, its rate of change from time *s* to time *t* would be equal to the rate of change of the tracer during the same period. However, the joint actions of legislation and farm management (see section 1.3 of Boumans and Fraters (2009)) that *come into force* (they could have their *origin* at an earlier time) at times *s* and *t* perturb the relation between nitrate and tracer. The perturbations $A_{i,j}(t)$ and $A_{i,j}(s)$ that result from those actions should be *multiplicative*, since the nitrate concentration at time *t* should be a percentage (rather than an absolute) increase or decrease of the nitrate concentration at time *s*. At each time, the relation between nitrate and tracer concentration is further perturbed by the process of denitrification which, depending on the soil characteristics and other factors, represented by $B_{i,j}(t)$ and $B_{i,j}(s)$, will, from time *s* to time *t*, reduce the levels of nitrate content.

Fortunately, it is likely that the rate of denitrification within a farm will remain approximately constant, at least within a certain range of variation of nitrate levels and climate change, so that the above relation can be simplified by assuming that $B_{i,j}(s) = B_{i,j}(t)$:

$$\frac{N_{i,j}(t)}{N_{i,j}(s)} = \frac{T_{i,j}(t)}{T_{i,j}(s)} \frac{A_{i,j}(t)}{A_{i,j}(s)} .$$

This relation between **relative change in nitrate concentration** and **relative change in tracer concentration** implicitly assumes that the nitrate and tracer concentrations are always positive, in which case we can further write[3]

$$\{\log N_{i,j}(t) - \log N_{i,j}(s)\} - \{\log T_{i,j}(t) - \log T_{i,j}(s)\} = \log a_{i,j}(s,t) ,$$

[3] Throughout this report, log will always denote the natural logarithm.

say, where $a_{i,j}(s,t) = A_{i,j}(t) / A_{i,j}(s)$ reflects the *changes* in legislation and farm management that came into effect during the period of *s* to *t*.

In the sequel, we shall refer to $\log a_{i,j}(s,t)$, the difference of differences between the logarithms of the original nitrate and tracer measurements, as the **net effect of legislation from time *s* to time *t***, or more briefly as the **net effect**. Our analysis, and in particular the investigation of possible net decreases in nitrate, will be based mainly on the net effects and not on the original nitrate and tracer measurements. It is important to keep this in mind in order to correctly interpret a number of *parameters* to be introduced below: these parameters pertain to the *relative* change in nitrate levels that took place (at a given location and at a given farm) between two time points, the 'relative' referring to the synchronous change in tracer levels.

Now the measurements of nitrate and tracer concentrations taken on a sample of farms have to be regarded as (realizations of) random variables. And if the above relationship describes 'most' of the deterministic relationship between nitrate and tracer concentrations at a given farm without any systematic bias, so that only 'noise' remains, then one expects that the difference between differences on the left-hand side will have a probability distribution that is symmetric around the expected value of $\log a_{i,j}(s,t)$. Furthermore, if our data can be regarded as coming from a set of randomly sampled farms from a large population of farms, it follows that

$$\{\log N_{i,j}(t) - \log N_{i,j}(s)\} - \{\log T_{i,j}(t) - \log T_{i,j}(s)\} = \log \alpha_i(s,t) + \varepsilon_{i,j}(s,t), \qquad \textbf{(1)}$$

where $\beta_i(s,t) := \log \alpha_i(s,t)$ is a *parameter* that reflects changes in legislation and farm management in farm *i* and the random errors $\varepsilon_{i,j}(s,t)$ have mean zero but are otherwise arbitrary.

This model can be described as saying that the net effect of legislation from time *s* to time *t* in farm *i* is equal to a **mean, or median, net effect in farm *i***, namely $\beta_i(s,t)$, plus a random error.

As will be seen in chapter 4, this model turns out to be supported (though of course it is not 'proved') by the data. Although the distribution of the errors depends on *s*, *t*, and on the farm characteristics (mainly in its amount of spread), its symmetry around the mean seems to hold rather closely. It even appears that for a given farm type $\beta_i(s,t)$ and $\varepsilon_{i,j}(s,t)$ are (almost) independent of soil characteristics.

Our objective in the sequel will be to estimate and find confidence intervals for two **parameters of interest** pertaining to the total number of farms in the Netherlands, which will be denoted by *N*.[4] The first is the **mean net effect** from time *s* to time *t*:

$$\beta(s,t) := \frac{1}{N}\sum_{i=1}^{N}\beta_i(s,t) = \frac{1}{N}\sum_{i=1}^{N}\log \alpha_i(s,t). \qquad \textbf{(2A)}$$

The second is

[4] More precisely, *N* is the number of farms in a given farm stratum determined by farm type and possibly by farm type and soil characteristics jointly.

$$\delta(s,t) := \frac{1}{N}\sum_{i=1}^{N} 1-\alpha_i(s,t)\,, \tag{2B}$$

which will be called the **mean median net decrease in nitrate** from time *s* to time *t*.

To explain the meaning of these parameters, assume that model **(1)** is correct. Then $\beta_i(s,t) < 0$ means that legislation did have an effect in farm *i* during the period of *s* to *t*; consequently, $\beta(s,t) < 0$ means that the mean (over the whole population of farms) net effect per farm is negative. Naturally, the mean net effect may be positive, corresponding to a mean net *increase* in nitrate; and the net effects $\beta_i(s,t)$ may be positive for some farms even if the mean net effect $\beta(s,t)$ is negative.

Now $\beta_i(s,t) < 0$ also means that[5] $\log\alpha_i(s,t) = E[\log A_{i,j}(t)/A_{i,j}(s)] < 0$, which, thanks to the assumption that the distribution of $\log A_{i,j}(t)/A_{i,j}(s)$ is symmetric, is the same as $\alpha_i(s,t) = med(A_{i,j}(t)/A_{i,j}(s)) < 1$. And if $\alpha_i(s,t) < 1$ then $1-\alpha_i(s,t)$ represents the **median (proportional) decrease** in nitrate in farm *i* during the period of *s* to *t*. Consequently, $\delta(s,t)$ is the mean of the population's median decreases, whence the name *mean median net decrease in nitrate*. Again, $\delta(s,t)$ may be negative, corresponding to a mean median net *increase* in nitrate; and the median decrease $1-\alpha_i(s,t)$ may be negative for some farms even if $\delta(s,t)$ is positive.

Unfortunately, we will not be able to estimate $\delta(s,t)$. This is because in most cases our data set does not contain measurements on $\log N_{i,j}(t)$ and $\log N_{i,j}(s)$ (nor on $\log T_{i,j}(t)$ and $\log T_{i,j}(s)$) from the same farm *i*, so that $\beta(s,t)$ will have to be estimated indirectly (by equation **(3)** below) and not by means of individual estimates of the $\beta_i(s,t)$ s. Instead of $\delta(s,t)$, we will therefore need to estimate an *approximation* to $\delta(s,t)$. [*Note*: If $\{\log N_{i,j}(t) - \log N_{i,j}(s)\} - \{\log T_{i,j}(t) - \log T_{i,j}(s)\}$ *were* available for farm *i* then we could use them to compute estimates $\hat{\beta}_i(s,t)$ of $\beta_i(s,t)$, hence estimates $\hat{\alpha}_i(s,t) = e^{\hat{\beta}_i(s,t)}$ of $\alpha_i(s,t)$, and to estimate $\delta(s,t)$ by $\hat{\delta}(s,t) = N^{-1}\sum_{i=1}^{N} 1-\hat{\alpha}_i(s,t)$.]

From

$$1-\alpha_i(s,t) = 1-e^{\beta_i(s,t)} \approx -\beta_i(s,t)\,,$$
$$1-\alpha_i(s,t) = 1-e^{\beta_i(s,t)} \approx -\beta_i(s,t) - \frac{\beta_i(s,t)^2}{2}$$

for $|\,\beta_i(s,t)\,|$ 'small', we deduce

$$1-\alpha_i(s,t) = 1-e^{\beta_i(s,t)} \approx -\lambda\left\{\beta_i(s,t) + \frac{\beta_i(s,t)^2}{2}\right\} - (1-\lambda)\beta_i(s,t)$$

[5] As usual, *E*(*X*) and *med*(*X*) denote the mean (or expected value) and median of a random variable *X*, respectively.

or

$$1-\alpha_i(s,t)=1-e^{\beta_i(s,t)}\approx-\beta_i(s,t)-\lambda\frac{\beta_i(s,t)^2}{2}$$

for $0<\lambda<1$. Thus

$$\begin{aligned}\delta(s,t)&=\frac{1}{N}\sum_{i=1}^{N}1-\alpha_i(s,t)\\&\approx-\frac{1}{N}\sum_{i=1}^{N}\beta_i(s,t)-\frac{\lambda}{2}\frac{1}{N}\sum_{i=1}^{N}\beta_i(s,t)^2\\&=-\beta(s,t)-\frac{\lambda}{2}\beta(s,t)^2-\frac{\lambda}{2}\frac{1}{N}\sum_{i=1}^{N}\{\beta_i(s,t)-\beta(s,t)\}^2,\end{aligned}\qquad\textbf{(2C)}$$

for λ to be chosen. This provides an approximation to $\delta(s,t)$ in terms of $\beta(s,t)$ and of the population variance of the $\beta_i(s,t)$ s. As seen in Figure 6, which shows the function $x\rightarrow 1-e^x$ together with its first and second order approximations, the 'combined approximation' **(2C)** with $\lambda=2.75/4=0.688$ is very accurate in the likely range of the $\beta_i(s,t)$ s, which we expect to be typically not much bigger than 0.5 in absolute value (the reader who wants to anticipate a little can look at the estimates of the mean net effects $\beta(s,t)$ shown in Figures 12A-D and 13).

This approximation, however, is still beyond our reach because it is not possible to estimate $N^{-1}\sum_{i=1}^{N}\{\beta_i(s,t)-\beta(s,t)\}^2$ with the available data. Consequently, we will have to make do with the approximation

$$\delta(s,t)\approx\delta_0(s,t):=-\beta(s,t)-\frac{\lambda}{2}\beta(s,t)^2\qquad\textbf{(2D)}$$

Although $\delta_0(s,t)$ *overestimates* $\delta(s,t)$ somewhat, since $N^{-1}\sum_{i=1}^{N}\{\beta_i(s,t)-\beta(s,t)\}^2>0$, we think that the error incurred in this approximation will generally be small. For example, in the somewhat unfavourable situation in which $\beta(s,t)$ is around 0.5 and the typical value of the deviations $|\beta_i(s,t)-\beta(s,t)|$ is 0.5, we would expect the difference between $\delta(s,t)$ and $\delta_0(s,t)$ to be of about $0.688\times0.5\times0.5^2\approx0.1$.

For simplicity, we shall also refer to $\delta_0(s,t)$ in **(2D)** as the **mean median net decrease in nitrate**, as if $\delta_0(s,t)$ were exactly equal to $\delta(s,t)$; if necessary, we will explicitly draw the distinction between the *exact* and the *approximate* mean median net decrease.

Remarks: **(i)** The parameter $\beta(s,t)$ is really the more important of the two for the purpose of *detecting* changes in nitrate concentrations. $\delta(s,t)$ is useful for *quantifying* those changes, and hence it is useful for setting up the scenarios required for the sample size calculations of chapter 6. However, we emphasize that in place of $\delta(s,t)$, what we estimate is $\delta_0(s,t)$, which provides an *approximation* to the mean median net decrease in nitrate.

(ii) Ideally, in place of $1-\alpha_i(s,t)$ and of $\delta(s,t)$ one would rather estimate the *mean* net decrease in farm *i* and the *mean* mean net decrease (not the *median* net decrease in farm *i* and the mean *median* net decrease). However, this would require us to estimate the distribution of $A_{i,j}(t)/A_{i,j}(s)$, which is possible only if $N_{i,j}(t)/N_{i,j}(s)$ and $T_{i,j}(t)/T_{i,j}(s)$ are always available for farm *i*.

4 Statistical methods

According to model **(1)**,

$$\begin{aligned}\beta_i(s,t) &= E\{\log N_{i,j}(t) - \log N_{i,j}(s)\} - E\{\log T_{i,j}(t) - \log T_{i,j}(s)\} \\ &= E\{\log N_{i,j}(t) - \log T_{i,j}(t)\} - E\{\log N_{i,j}(s) - \log T_{i,j}(s)\},\end{aligned}$$

so the mean net effect in farm *i* amounts to the difference between two means, one pertaining to time *t* and another to time *s*. Consequently, a consistent point estimator of this parameter is simply

$$\hat{\beta}_i(s,t) = \Delta_i(t) - \Delta_i(s)$$

where $\Delta_i(\tau) = n_i(\tau)^{-1} \sum_{j=1}^{n_i(\tau)} \{\log N_{i,j}(\tau) - \log T_{i,j}(\tau)\}$ is the aggregated difference between log nitrate and log tracer and $n_i(\tau)$ the number of measurements in farm *i* at time τ.

Unfortunately, the $\Delta_i(s)$ and $\Delta_i(t)$ corresponding to the same *i* are frequently unavailable at different times *s* and *t*, so in most cases the within-farm mean net effects $\beta_i(s,t)$ cannot be estimated. The **mean net effect** from time *s* to time *t*, $\beta(s,t) := N^{-1} \sum_{i=1}^{N} \beta_i(s,t)$, however, can be estimated indirectly by

$$\hat{\beta}(s,t) = \frac{1}{n(t)} \sum_{i=1}^{n(t)} \Delta_i(t) - \frac{1}{n(s)} \sum_{i=1}^{n(s)} \Delta_i(s), \qquad \textbf{(3)}$$

where $n(\tau)$ is the number of farms sampled at time τ. This estimator makes full use of the data set, in contrast to an estimator based on the differences $\hat{\beta}_i(s,t) = \Delta_i(t) - \Delta_i(s)$, which would be restricted to using the relatively few farms with measurements both at time *s* and at time *t*. On the other hand, in order to compute its variance, and hence an estimate of its variance and a confidence interval for $\beta(s,t)$, one needs to account for the fact that in **(3)** there are farms which appear twice and that measurements from the same farm made at different time points are naturally paired (they could be dependent, for example).

The variance of $\hat{\beta}(s,t)$ and an estimate of it are derived in Appendix B. Here we simply denote the estimate of $Var\,\hat{\beta}(s,t)$ by $v(\hat{\beta}(s,t))$ and note that under weak assumptions on the errors of model **(1)** ─ e.g., that the variance of the $\varepsilon_{i,j}(s,t)$ s does not depend 'too much' on *i* and that the sample sizes are not 'too small' ─ the distribution of the variable

$$T(s,t) = \frac{\hat{\beta}(s,t) - \beta(s,t)}{\sqrt{v(\hat{\beta}(s,t))}}$$

is approximately standard normal, from which it follows in the usual way that an approximate 95% confidence interval for $\beta(s,t)$ is given by

$$\left[\ \hat{\beta}(s,t) - 1.96\sqrt{v(\hat{\beta}(s,t))},\ \hat{\beta}(s,t) + 1.96\sqrt{v(\hat{\beta}(s,t))}\ \right]. \quad \textbf{(4)}$$

When $\beta(s,t) = 0$, $T(s,t)$ becomes a t-statistic, namely a mixed form of the t-statistic for paired data and of the two-sample t-statistic. This t-statistic can be used to test the **null hypothesis** that $\beta(s,t) = 0$: the null hypothesis is rejected at the (approximate) 5% level if and only if $|\hat{\beta}(s,t)| > 1.96\sqrt{v(\hat{\beta}(s,t))}$; equivalently, the null hypothesis is rejected if and only if the interval **(4)** does not contain zero.

Finally, the mean median net decrease is estimated by $\hat{\delta}_0(s,t) = -\hat{\beta}(s,t) - \lambda\hat{\beta}(s,t)^2/2$, with $\lambda = 2.75/4$, and a confidence interval for it is obtained by replacing the upper and lower limits of the confidence interval for $\beta(s,t)$ into the function $x \to -x - \lambda x^2/2$. (Alternatively, a confidence interval for $\delta_0(s,t)$ could be obtained by the delta method, i.e., using the fact that the distribution of $\hat{\delta}_0(s,t)$ is approximately normal with mean $\delta_0(s,t)$ and variance $[1 + \lambda\beta(s,t)]^2 Var(\hat{\beta}(s,t))$. This produces symmetric confidence intervals; the method we use produces slightly asymmetric ones.)

This is all we need for estimating and finding confidence intervals for the parameters of interest. However, our estimates of $\beta(s,t)$ and $\delta(s,t)$ will be unbiased only if the main assumption behind model **(1)**, namely that the errors are symmetrically distributed around their mean, holds. It is therefore important to provide evidence based on the data that this assumption is plausible. We shall do this in two ways: first, by visually assessing the symmetry of the distributions of $\{\log N_{i,j}(t) - \log N_{i,j}(s)\} - \{\log T_{i,j}(t) - \log T_{i,j}(s)\}$ by means of box plots, for various choices of times *s* and *t*, or equivalently for various choices of *t* and **time lags** $t - s$, and separately for each farm type; secondly, by formally testing the symmetry of those distributions by means of the Cabilio-Masaro **symmetry test**, which is implemented in the R package 'lawstat' (Miao et al., 2008).[6]

A stronger assumption that is usually made in connection with models such as **(1)** is that the errors $\varepsilon_{i,j}(s,t)$ are normally distributed with mean zero (not just symmetrically distributed around zero), though their variances may still vary with *s*, *t* and *i*. This assumption is not at all necessary for obtaining unbiased estimates, and in our case it is not even required for the computation of confidence intervals for the parameters. However, because of the special characteristics of the normal distribution, normally distributed errors lend more credibility to the idea that everything that is deterministic or systematic in the data has been accounted for in the model and that the deviations between what the model predicts and what is observed are 'purely random'. For this reason, we shall also check the assumption of normality in model **(1)** by means of the Cramér-von Mises test, as implemented in the R package 'nortest' (Gross, 2009).

[6] In principle one could also use the Mira and Miao-Gel-Gastwirth symmetry tests, which are also implemented in 'lawstat'; however, the first of these gives results similar to those of the Cabilio-Masaro test and the second is highly biased (namely, it substantially underestimates the type I error) when the data come from a symmetric distribution with heavier tails than the normal distribution.

5 Model checking, estimates and confidence intervals

Recall that we have defined net effects only for positive nitrate and tracer concentrations. However, 130 nitrate concentrations equal to zero are found in 52 different farms (46 of such farms having at least one positive measurement) and we had to ignore them in our analysis (the corresponding tracer concentrations, when they exist, are always positive). It is important, therefore, to try to clarify the meaning of the individual and aggregated measurements obtained at a given farm. The individual nitrate and tracer measurements are censored: they only contribute to an aggregated average of logarithms conditionally on their being positive. Thus, the aggregated averages of log nitrate and log tracer per farm represent estimates of *conditional* medians (and expected values) and the exponentials of these estimates represent *overestimates* of the median nitrate and median tracer concentrations per farm. Consequently, the populations (defined in terms of the different farm types) to which the results of our analysis apply constitute sub-populations of farms in the Netherlands that are somewhat richer in nitrate than the sampled populations. However, given the relatively small number of censored measurements per farm, we think that the differences between the 'censored estimates' and 'uncensored estimates' that one might obtain by another approach are negligible.

5.1 Model checking

The first step in our analysis was to check the assumptions behind model **(1)**, as explained in the preceding chapter. Figures 7A-C show box plots of the samples of **net effects**, namely the differences of differences that appear on the left-hand side of **(1)**, for various choices of time *t* and **lag** $t-s$ and separately for each type of farm. In the top panel of Figure 7A, for example, we see box plots of data from arable farms and years separated by a lag of 1; the box plot on the left, with the label 1993 under it, has been constructed with the data from times $t=1993$ and $s=1992$, and the right-most box plot with the data from $t=2008$ and $s=2007$. In the bottom panel of Figure 7C, the left-most box plot represents net effects from 1992 to 1995 (lag 3) in dairy farms.

Overall, the box plots suggest that the net effects have a rather symmetric distribution. The left-hand panel of Figure 8 shows a histogram of the p-values of Cabilio-Masaro symmetry tests carried out on 51 different samples of net effects (each sample corresponding to a given lag, a given year and a given farm type). If all the samples came from truly symmetric distributions, one would expect the histogram to look roughly like the standard uniform density (which is equal to 1 in the interval $[0,1]$); if there were a systematic and strong deviation from the assumption of symmetry one would expect the histogram to be concentrated mainly near zero. In the present case, the histogram does not seem to suggest any systematic deviation from uniformity. Unfortunately, it is not possible to give an objective impression of how well the histogram approximates the uniform density in relation to the number of tests, since the p-values are somewhat dependent; we can only say that the application of the Kolmogorov-Smirnov test of uniformity (which requires independence) to this set of p-values yields a p-value of 0.46.

It is clear from the box plots in Figures 7A-C that there are observations, namely those that lie beyond the upper and lower bars, that could be regarded as **outliers**; see for instance the two right-most box plots in the bottom panel of Figure 7B. However, such observations could qualify as outliers only if the majority of net gains were (approximately) normally distributed. The many normal Q-Q (quantile-quantile) plots shown in Figures 9A-Q suggest that the net gains are *roughly* normally distributed *provided the 'outliers' are discarded*. Indeed, these Q-Q plots are robust against outlying observations, in so far as the median and the mean absolute deviation used to construct them are robust estimates of the mean and standard deviation (unlike the sample mean and standard deviation), and they show that except for the extreme observations (which *are* represented in the Q-Q plots) the data generally follow the 45° line. However, neither the number of outliers nor the goodness-of-fit as measured by the closeness to the line are the same for all data sets: while in some cases only a couple of observations qualify as outliers, in other cases a substantial number are too extreme to have come from a normal distribution.

The right panel of Figure 8 shows a histogram of the p-values of Cramér-von Mises normality tests carried out on 50 of the samples of net effects *without the outliers* ─ more precisely, without the observations that lie beyond the upper or lower bars of the box plots.[7] The histogram does not indicate any systematic deviation from uniformity, suggesting that without the 'outliers' the data look roughly normally distributed. A further overall check of normality is afforded by the Q-Q plot of the **standardized residuals**, i.e., the union of the standardized samples (the observations in the sample minus their mean divided by their standard deviation) without the 'outliers', shown in Figure 10; this clearly reveals systematic deviations from normality (even in the central portion of the data), though it could still be said that the discrepancies relative to a normal distribution are, all things considered, 'not that big'.

In any case, the fact remains that the distribution of net effects is generally more heavy-tailed than a normal distribution (incidentally, this 'heavy-tailedness' seems to change somewhat in time, judging by the first and last box plots in the bottom panel of Figure 7B, for example) and that even without putative outliers it often shows deviations from normality. Fortunately, the correctness of our estimates and confidence intervals does not rely on normality. The correctness of the *interpretation* of those estimates and confidence intervals *does* depend on the assumption of symmetry, but since symmetry seems to hold rather closely, we think that the estimates to be provided below are not only correct but also meaningful.

Having said that, however, it may be worthwhile to examine more closely the nitrate and tracer concentrations associated with the putative outlier net effects. The point is that if those concentrations were found to be genuine outliers (unrealistic nitrate concentrations resulting from measurement errors, for example) then, by definition, they should not enter in the calculation of our estimates and confidence intervals. The scatter plots of Figure 11 show the logarithms of nitrate and tracer concentrations (in grey) with the putative outliers (i.e. the logarithms involved in the calculation of the left-hand side of equation **(1)** for some s and t that yield 'too big' or 'too small' net effects) superimposed on them in different colours according to year of measurement, different farms being represented by different symbols. The majority of the 'outliers' come from dairy farms in the years 2005–2008. Many of them fall in the borders of the scatter plots (mainly in the lower range of nitrate), but the majority do not. In view of the 'regularity' of most of the putative outliers, we see no reason for not using them to compute our estimates.

[7] This is a standard procedure for removing outliers in *contaminated normal samples* (i.e., samples composed of a majority of normal observations with the same mean and variance and a minority of 'more extreme' observations). The upper and lower bars depicted in the box plots are at a distance of 1.5 times the inter-quartile range from the upper and lower sample quartiles.

Finally, let us mention that we have checked whether stratifying the data further in terms of 'fraction dry soil' and 'fraction neutral soil' makes the assumptions of symmetry and normality more plausible. Overall, symmetry and (approximate) normality remain plausible, but not more plausible than before, and the potential outliers remain. Since stratification reduces the number of observations that one can use to estimate the parameters, we conclude that nothing would be gained by stratifying the data beyond 'farm type'.

5.2 Estimates and confidence intervals

The graphs on Figures 12A-D show estimates and confidence intervals of the mean net effect and mean median net decrease over one-year periods separately for each farm type. Estimates are computed only if data from at least 15 farms are available in each of the two years. Figure 13 shows estimates and confidence intervals of the same parameters over the longer periods for which enough data (namely at least 15 farms in each of the start and end years) are available. The estimates and intervals of the mean median net decreases represented in the bottom panel of Figure 13 are given in Table 5.1 below.

Table 5.1 – Estimates of mean median net decreases in nitrate by farm type over the largest period over which a substantial number of measurements are available.

Farm type	Estimate of mean median net decrease	Approx. 95% confidence interval	Period
arable	0.09	(–0.28,0.39)	1992–2008
dairy	0.70	(0.65,0.72)	1992–2008
factory	0.08	(–0.31,0.38)	2005–2008
other	–0.15	(–0.77,0.31)	2006–2008

The clearest *long-term* — namely from 1992 to 2008 — net decrease has been observed in dairy farms, the biggest annual decreases occurring from 1997 to 1998 and from 2000 to 2001 (Figure 12B). The overall decrease in arable farms does not appear to be significant, but apparently there was a significant decrease from 2002 to 2003, which was compensated by *increases* from 1994 to 1995 and from 2007 to 2008. In factory farms the overall decrease is also not significant, but there is some evidence that an increase occurred from 2007 to 2008.

6 Sample size required to demonstrate a future net decrease

Let $\beta(s,t)$ denote the mean net effect of legislation from time *s* to time *t* on an arbitrary farm of a given type. Recall from chapter 3 that this parameter is equal to the mean (across all farms in a given stratum) of net effects: $\beta(s,t) := N^{-1}\sum_{i=1}^{N}\beta_i(s,t)$, where

$$\beta_i(s,t) = E[\Delta_i(t) - \Delta_i(s)],$$

in the notation of chapter 4. Recall also that $\delta_0(s,t) := -\beta(s,t) - 0.9\beta(s,t)^2/2$ represents the approximate mean median net decrease in nitrate levels from time *s* to time *t*. The problem we address in this chapter is that of determining the sample size (combination of number of farms *n* and number of measurements *p* per farm) that is required to demonstrate, with high probability, that $\delta_0(s,t)$ exceeds a certain number δ_0^* between 0 and 1, or that

$$\beta(s,t) < \frac{1}{\lambda}\left(\sqrt{1-2\lambda\delta_0^*} - 1\right), \qquad \textbf{(5)}$$

where $\lambda = 2.75/4$ and it is assumed that $\delta_0^* < (2\lambda)^{-1}$. (The term on the right here is the biggest root of the quadratic equation $\delta_0(s,t) = -\beta(s,t) - \lambda\beta(s,t)^2/2$; see chapter 3 and Figure 7.)

One way of providing evidence that **(5)** holds is to compute a 95% confidence interval for $\beta(s,t)$ and to show that the interval is located to the left of $\lambda^{-1}\left(\sqrt{1-2\lambda\delta_0^*} - 1\right)$. Therefore, in order to compute the sample size that is required to actually find evidence in that direction one needs to make an assumption about what the real value of $\delta_0(s,t)$ is and to compute the probability that the interval does lie to the left of $\lambda^{-1}\left(\sqrt{1-2\lambda\delta_0^*} - 1\right)$.

Suppose that we have an estimate of $\beta(s,t)$ obtained from *n* farms of a given type; denote these estimates by $\hat{\beta}_1(s,t),\ldots,\hat{\beta}_n(s,t)$. Then an estimate of $\beta(s,t)$ can be computed as

$$\hat{\beta}(s,t) = \frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_i(s,t).$$

Evidently, the $\hat{\beta}_i(s,t)$ are within-farm sample averages of net effects, so the distribution of

$$\frac{\hat{\beta}(s,t)-\beta(s,t)}{\sqrt{v(\hat{\beta}(s,t))}},$$

where $v(\hat{\beta}(s,t))$ stands for an estimate of $Var(\hat{\beta}(s,t))$, is approximately standard normal and yields the required (approximate) 95% one-sided (open on the left) confidence interval:

$$\left]-\infty,\ \hat{\beta}(s,t)+1.645\sqrt{v(\hat{\beta}(s,t))}\ \right]$$

Suppose that in reality $\beta(s,t) < \lambda^{-1}\left(\sqrt{1-2\lambda\delta_0^*}-1\right)$. Then the probability that the interval for $\beta(s,t)$ will be located to the left of $\lambda^{-1}\left(\sqrt{1-2\lambda\delta_0^*}-1\right)$ is

$$P\left(\hat{\beta}(s,t)+1.645\sqrt{v(\hat{\beta}(s,t))} < \frac{\sqrt{1-2\lambda\delta_0^*}-1}{\lambda}\right) \approx \Phi\left(-1.645+\frac{\frac{\sqrt{1-2\lambda\delta_0^*}-1}{\lambda}-\beta(s,t)}{\sqrt{v(\hat{\beta}(s,t))}}\right)$$

$$\approx \Phi\left(-1.645+\frac{\frac{\sqrt{1-2\lambda\delta_0^*}-1}{\lambda}-\beta(s,t)}{\sqrt{Var(\hat{\beta}(s,t))}}\right)$$

where as usual Φ stands for the standard normal distribution function. It is this probability, or **power**, that we need to estimate for a given choice of **sampling characteristics**, namely a choice of the number ***n*** of farms sampled and of the number ***p*** of measurements taken per farm.

In order to do this we need to define $\hat{\beta}(s,t)$ explicitly on the basis of the sampling characteristics and to compute $Var(\hat{\beta}(s,t))$, which will be a function of *n* and *p*. We shall only consider the two simplest types of sampling characteristics: in **Case I**, the data from times *s* and *t* come from the same *n* farms; in **Case II**, the data from time *s* and the data from time *t* come from different farms. In both cases, the number *n* of farms to be sampled per year and the number *p* of measurements taken per farm entail a total of $2np$ measurements. Details on how $Var(\hat{\beta}(s,t))$ is computed in each case are given in Appendix C; here we only need to know that the **power attained with *n* farms and *p* measurements per farm** is given approximately by

$$\pi(n,p) := \Phi\left(-1.645+\frac{\frac{\sqrt{1-2\lambda\delta_0^*}-1}{\lambda}-\beta(s,t)}{\sqrt{\frac{1}{n}\left(\Sigma_1^2+\frac{1}{p}\Sigma_2^2\right)}}\right), \tag{6}$$

where Σ_1^2 and Σ_2^2 are the so-called **between-farms** and **within-farms** variances, defined in Appendix C, which can be estimated from the data.

6.1 Numerical examples

The estimation of Σ_1^2 and Σ_2^2, which determine the power function in **(6)**, is described in Appendix C separately for **Case I** and **Case II**; here we provide numerical illustrations. We first consider **Case I**, which should be of interest in detecting short-term changes in nitrate, since within a period of a couple of years it is more likely that the farms sampled are the same at times *s* and *t*.

Table 6.1 shows estimates of these variances obtained with data from a selection of dairy farms that were sampled in two consecutive years, namely in 1992 and 1993, in 2005 and 2006, in 2006 and 2007, and in 2007 and 2008. The first and second year of sampling are indicated in the table by *s* and *t* (following the notation used in Appendix C), the number of farms and the number of measurements characterizing the sampling regime of each period by *n* and *p*, and the estimate of the mean net effect from year *s* to year *t* due to legislation by $\hat{\beta}(s,t)$. The other columns show the estimates of the within- and between-farms variances and of the 'total variance' $\Sigma_1^2 + p^{-1}\Sigma_2^2$. As explained in Appendix C, the within-farms variance $\hat{\Sigma}_2^2$ cannot be obtained directly from the available data; instead, it is computed as the sum of two within-farms variances, $\hat{\Sigma}_2^2(s)$ and $\hat{\Sigma}_2^2(t)$, which are computed separately with data from year *s* and data from year *t*.

Table 6.1 – Estimates of net effects ($\hat{\beta}(s,t)$), between-farms variances ($\hat{\Sigma}_1^2$) and within-farms variances ($\hat{\Sigma}_2^2$, $\hat{\Sigma}_2^2(s)$, $\hat{\Sigma}_2^2(t)$) corresponding to two consecutive years, *s* and *t*, based on data from dairy farms in the situation where each farm is sampled in both years (Case I). *n* indicates the number of farms with measurements available in both years and *p* the number of composite samples per farm.

s	*t*	*n*	*p*	$\hat{\beta}(s,t)$	$\hat{\Sigma}_1^2$	$\hat{\Sigma}_2^2$	$\hat{\Sigma}_2^2(s)$	$\hat{\Sigma}_2^2(t)$	$\hat{\Sigma}_1^2 + p^{-1}\hat{\Sigma}_2^2$
1992	1993	63	4	0.02	0.02	0.22	0.11	0.11	0.08
2005	2006	44	2	-0.10	0.06	0.59	0.33	0.26	0.35
2006	2007	113	2	0.08	0.12	0.63	0.38	0.25	0.43
2007	2008	137	2	0.01	0.00	0.77	0.25	0.52	0.33

The first thing to note is that there are some apparent discrepancies between the estimates of $\hat{\beta}(s,t)$ obtained using the method of chapter 4, which are represented in the top panel of Figure 12D, and the estimates shown in Table 6.1, which are based only on farms for which measurements were available both at time *s* and time *t*. While for $s = 1992$ and $t = 1993$ we obtain $\hat{\beta}(s,t) = 0.11$ with a confidence interval of $[0.01, 0.20]$ using the whole data, Table 6.1 shows us $\hat{\beta}(s,t) = 0.02$; and while for $s = 2005$ and $t = 2006$ we obtain $\hat{\beta}(s,t) = 0.01$ with a confidence interval of

$[-0.27, 0.24]$ the table shows us $\hat{\beta}(s,t) = -0.10$. These discrepancies, however, can be attributed to the uncertainty implied by the confidence intervals.
The estimates $\hat{\Sigma}_1^2$ and $\hat{\Sigma}_2^2$ given in Table 6 can be substituted into equation **(6)** above to obtain estimates of power. It seems, however, that these estimates are also quite uncertain. The estimates $\hat{\Sigma}_2^2(s)$ and $\hat{\Sigma}_2^2(t)$ (and hence $\hat{\Sigma}_2^2$) in the periods of 2005–2006 and 2006–2007 are consistent with each other; however, in the period of 1992–1993 they are both much smaller (by roughly 60%). As to $\hat{\Sigma}_1^2$, it is more than five times bigger in 2006–2007 than in 1992–1993, and two times bigger in 2006–2007 than in 2005–2006. Also, in 2007–2008 the estimate we obtain for Σ_1^2 is actually negative (–0.16), so we set $\hat{\Sigma}_1^2$ to zero. Thus the estimates seem to be quite uncertain, which should be kept in mind when assessing estimates of power gotten from equation **(6)** with Σ_1^2 and Σ_2^2 replaced by $\hat{\Sigma}_1^2$ and $\hat{\Sigma}_2^2$. On the other hand, there is a striking agreement between the $\hat{\Sigma}_2^2(s)$ and $\hat{\Sigma}_2^2(t)$ estimates in the 1992–1993 period, indicating the benefits of having four rather than just two measurements per farm.

Table 6.2 – Estimates of the power that a statistical test has to detect a net decrease/increase in nitrate concentration from any given year to another year in the situation where the actual decrease is of 20%. n denotes the number of farms sampled per year and p the number of composite samples per farm, so that in the two years a total of 2np measurements need to be carried out. Two cases are considered: in Case I all farms are sampled in both years; in Case II no farm is sampled in both years. The statistical test is to be carried out at the 5% level. The estimates of power are based on a choice of variance parameters suggested by the variance estimates of Tables 2 and 4, which are based on data from dairy farms. The nitrate data from these farms have medians in the range of 40 to 160 (mg/l). The estimates of power are uncertain, mainly due to the uncertainty in the estimates of the variances. For example, a prescription of a sample size of 100 may be more appropriately interpreted as a sample size in the range of 75 to 125.

n	*p*	power in Case I	power in Case II
20	2	0.38	0.18
20	4	0.53	0.20
20	6	0.61	0.20
50	2	0.68	0.31
50	4	0.85	0.34
50	6	0.92	0.35
100	2	0.91	0.50
100	4	0.99	0.54
100	6	1.00	0.56
200	2	1.00	0.75
200	4	1.00	0.80
200	6	1.00	0.18

The third column of Table 6.2 provides some estimates of power for illustration purposes, assuming that the correct values for Σ_1^2 and Σ_2^2 are $\Sigma_1^2 = 0.12$ and $\Sigma_2^2 = 0.8$, which can be thought of as the worse-case scenario suggested by Table 6.1, that the true mean median net decrease from one year to the next is $\delta_0 = -\beta - 0.9\beta^2/2 = 0.20$ (i.e. a 20% decrease) and that one would like to demonstrate that $\delta_0 > \delta_0^* = 0$ (i.e., that *there is* a decrease).

The estimates of Σ_1^2 and Σ_2^2 can be obtained as $\hat{\Sigma}_1^2 = \hat{\Sigma}_1^2(s) + \hat{\Sigma}_1^2(t)$, $\hat{\Sigma}_2^2 = \hat{\Sigma}_2^2(s) + \hat{\Sigma}_2^2(t)$ from the estimates of the variances presented in Table 6.3. The differences between the $\hat{\beta}(s,t)$ estimates in this table and those obtained in **Case I** are due to differences in the samples (**Case I** only considers farms with measurements in both years), and so are the differences between estimates of the within-farms variances, $\hat{\Sigma}_2^2(s)$ and $\hat{\Sigma}_2^2(t)$. The differences between the $\hat{\Sigma}_1^2 + p^{-1}\hat{\Sigma}_2^2$ estimates obtained in **Case I** and **Case II** are explained partly by sampling variation (i.e., by the use of different samples) but mainly by the fact that the pairing between measurements from the same farm in different years has been ignored, which, as expected, contributes to greater variation in the $\hat{\beta}(s,t)$ estimates. The agreement between $\hat{\Sigma}_1^2(s)$ and $\hat{\Sigma}_1^2(t)$ and between $\hat{\Sigma}_2^2(s)$ and $\hat{\Sigma}_2^2(t)$ in the 1992–1993 period indicates once more the benefits of having four measurements per farm.

We now consider **Case II** with data from the same years. **Case II** should be of interest in detecting long-term changes in nitrate, for if *s* and *t* are separated by a decade or so, many farms that existed at time *s* will have been replaced by other farms at time *t*.

The last column of Table 6.2 refers to **Case II** and assumes that $\Sigma_1^2 = 1.30$ and $\Sigma_2^2 = 0.85$, which can be thought of as the worse case scenario suggested by Table 6.3; the other settings are the same as those used in **Case I**. According to these estimates of power, in **Case II** one needs to quadruple the number of farms needed in **Case I** in order to achieve comparable power.

An Excel sheet that computes power estimates for a given choice of Σ_1^2, Σ_2^2, δ, δ_0, n and p can be obtained from the author.

Table 6.3 - Estimates of net effects ($\hat{\beta}(s,t)$), between-farms variances ($\hat{\Sigma}_1^2(s)$, $\hat{\Sigma}_1^2(t)$) and within-farms variances ($\hat{\Sigma}_2^2(s)$, $\hat{\Sigma}_2^2(t)$) corresponding to two consecutive years, *s* and *t*, based on data from dairy farms, in the situation where different farms are sampled in different years (Case II). $n(s)$ and $n(t)$ indicate the numbers of farms with measurements in years *s* and *t*, and *p* the number of composite samples per farm.

s	*t*	$n(s)$	$n(t)$	*p*	$\hat{\beta}(s,t)$	$\hat{\Sigma}_1^2(s)$	$\hat{\Sigma}_1^2(t)$	$\hat{\Sigma}_2^2(s)$	$\hat{\Sigma}_2^2(t)$	$\hat{\Sigma}_1^2 + p^{-1}\hat{\Sigma}_2^2$
1992	1993	67	65	4	0.03	0.21	0.24	0.11	0.11	0.50
2005	2006	63	127	2	-0.05	0.69	0.61	0.30	0.43	1.66
2006	2007	127	160	2	0.01	0.61	0.69	0.43	0.29	1.66
2007	2008	160	143	2	0.09	0.69	0.39	0.29	0.55	1.51

7 Discussion, conclusions and recommendations

The results presented in chapter 5 indicate that there has been a *net* decrease (i.e. a decrease attributable to policies governing nitrate application and farm management) in the levels of nitrate in the Netherlands during the period of 1992 to 2008, in agreement with the general conclusion of Boumans and Fraters (2009). As might be expected, the strength of evidence in favour of a decrease and the estimates quantifying possible decreases vary with time and according to the type of farm; the clearest and biggest net decreases have occurred in dairy farms.

Naturally, there is a time lag between the moment when new legislation is introduced and the moment of its biggest impact, and this time lag will probably depend both on the type of farm and on the moment at which legislation has been enforced. The graphs showing estimates of the net changes in nitrate from one year to the next (Figures 12A-D) provide an indication of the years in which the biggest net decreases have occurred, and may thus prove useful to experts interested in identifying the relevant time lags in the face of the history of nitrate legislation (summarized for instance in section 1.2 of Boumans and Fraters (2009)).

Despite some misgivings arising from variations in the quality of the data and in the sampling regime, we think that our estimates and conclusions are overall correct because the methods we have employed are based on rather weak assumptions and make use of (and in a sense validate) the theoretical connection between nitrate measurements and tracer measurements. Nevertheless, it seems worth considering a possible revision of the current sampling plan in the light of the shortcomings pointed out in chapters 2 and 4 before embarking on future data collection campaigns, as better data will almost certainly result in sharper and more reliable conclusions.

The parameters we have used to quantify net changes are the medians and means of medians of chapter 3, which are not as easy to interpret as means and means of means. Estimating the latter, however, would require considerably more and better data. More precisely, the model for net effects described in chapter 3 involves a parameter, the *median* net decrease in nitrate in a farm, and a symmetric probability distribution; the parameter can be estimated, but from its estimate we cannot derive an estimate of the *mean* net decrease without further information on the probability distribution, and information on this distribution requires more observations within a farm and observations from the same farm in successive years. Thus, a revision of the current sampling plan becomes even more important in case more interpretable estimates are required in the future.

The method of sample size determination presented in chapter 6 is quite standard and its reliability depends on the reliability of the estimates of the between-farms and within-farms variances that are needed in order to compute power. Unfortunately, it is difficult to properly assess the quality of these variance estimates. It appears from our results that the between-farms variance has increased in recent years, which could be the result of a genuine change in the population of farms in the Netherlands or a result of changes in the way farms are selected, and that the within-farm variances are somewhat uncertain. Consequently, we do not expect our estimates of power and sample sizes to be very accurate. However, we expect them to be at least of the correct order of magnitude: for example, if we estimate that 100 farms and 2 observations per farm yield about 95% power then we expect that the actual number of farms required to attain such power could be 75 or 125 instead, but not 25 or 250, say. Again, in order to improve the estimation of the within-farms variances, and hence of power and sample size, more observations from the same farm are required; and in order to properly study possible time changes in the variances, the same farms would have to be sampled in consecutive years.

rivm

Acknowledgements

The author is grateful to Leo Boumans (MEV/CMM) for information on the processes underlying the nitrate and tracer measurements and for many useful conversations on the problems treated here.

References

Boumans, L.J.M. and Fraters, B. (2009). A legislation induced decrease in nitrate leaching in the sandy areas of the Netherlands during the 1992–2006 period.

Gross, J. (2009). Package ‘nortest’. Available at http://cran.r-project.org/web/packages/nortest.

Miao, W., Gel, Y. R., and Gastwirth, J. L. (2008). lawstat: An R Package for Law, Public Policy and Biostatistics. Journal of Statistical Software, Vol. 28, Issue 3, available at. http://www.jstatsoft.org/.

rivm

Figures

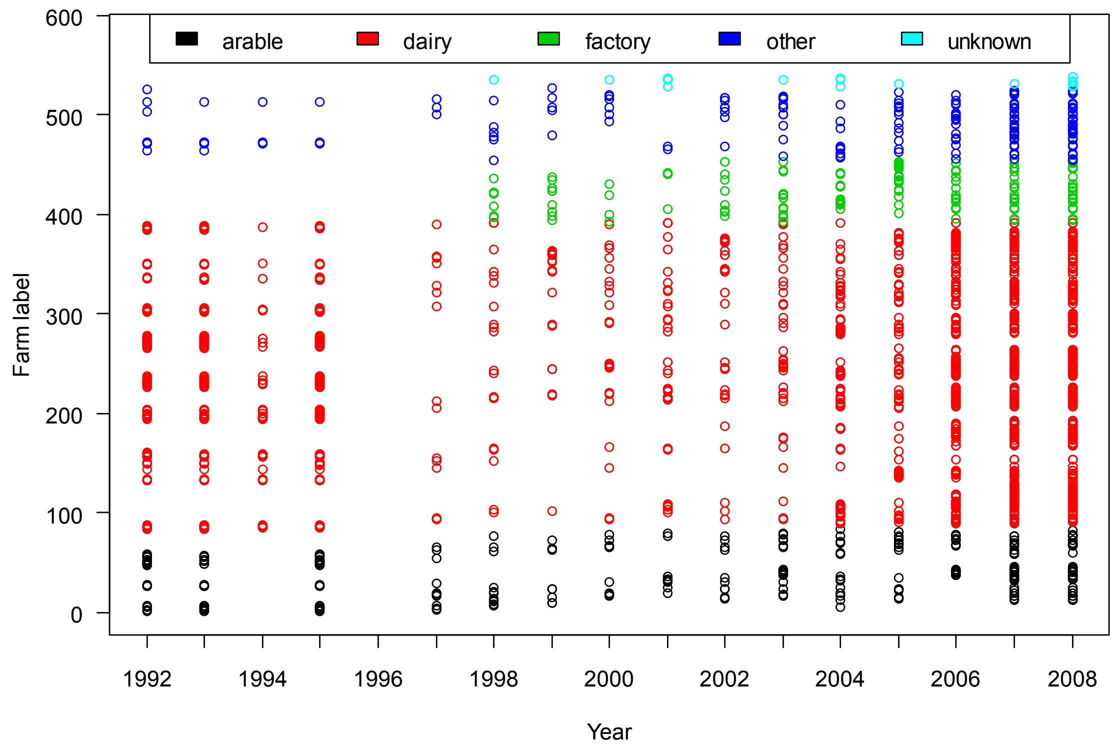


Figure 1 Sampling regime throughout the 1992–2008 period.

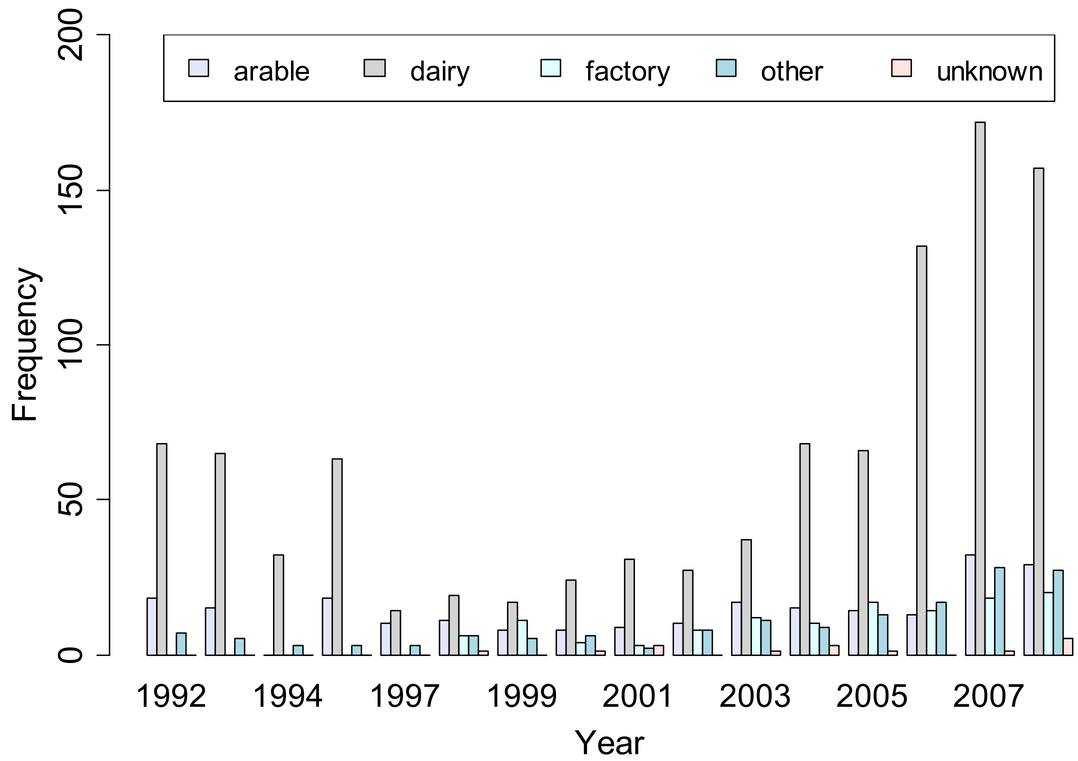


Figure 2 Number of sampled farms by year and farm type.

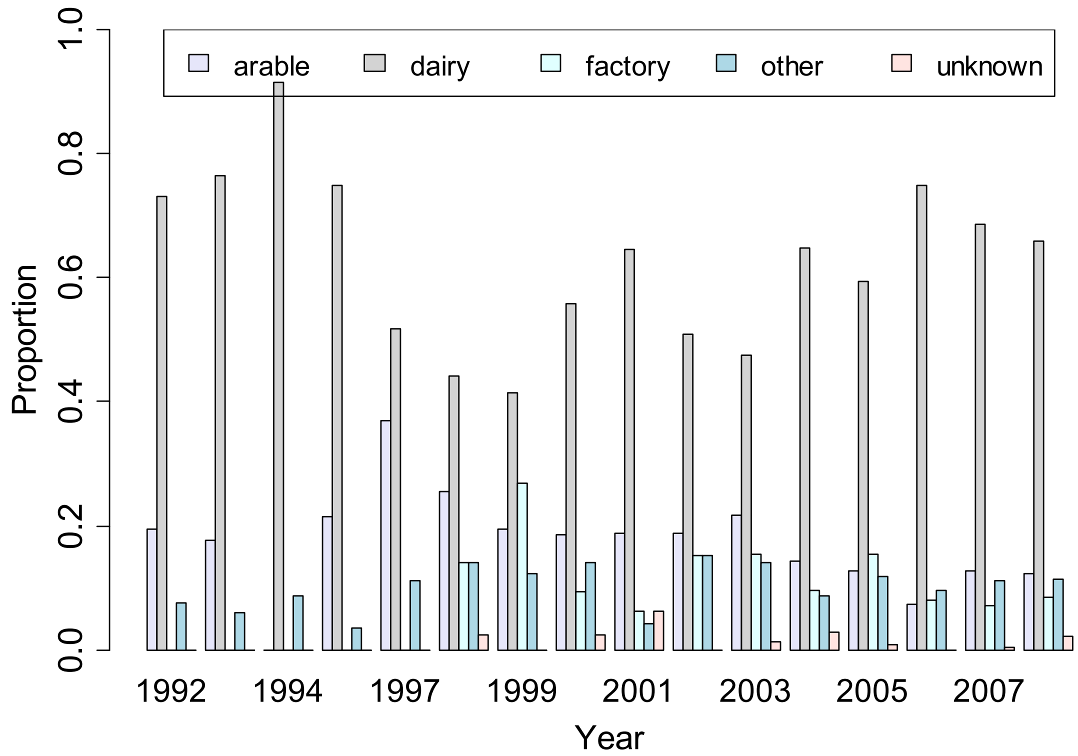


Figure 3 Proportions of sampled farms by farm type during the 1992–2008 period.

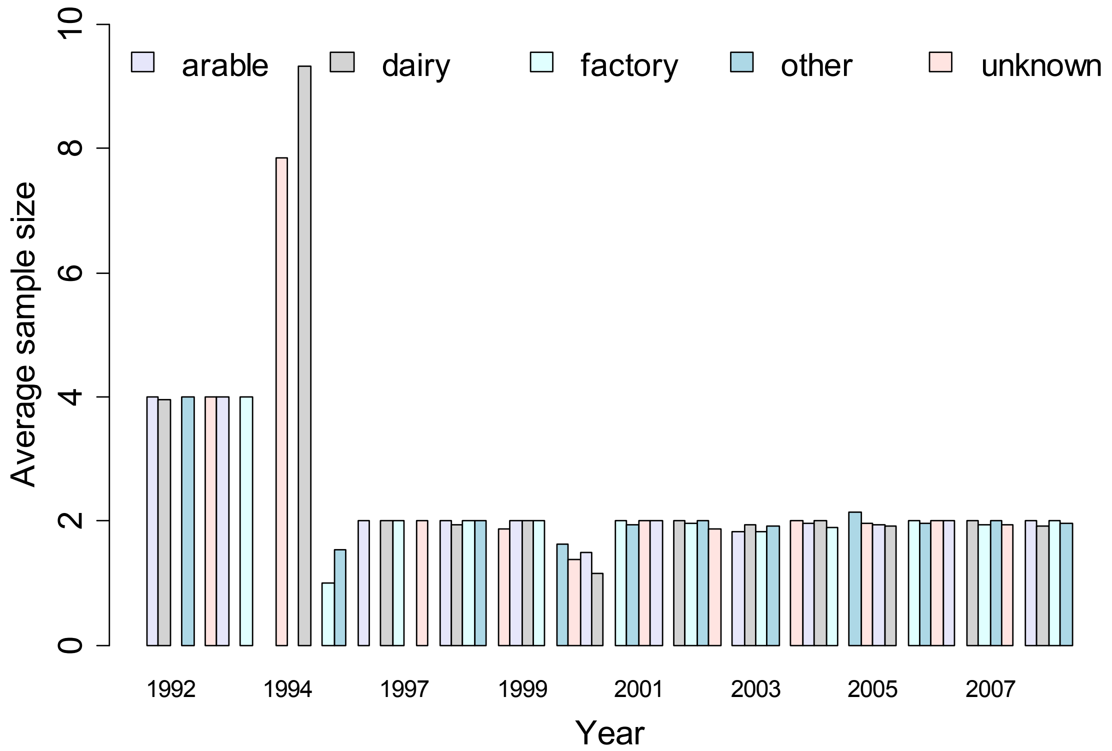
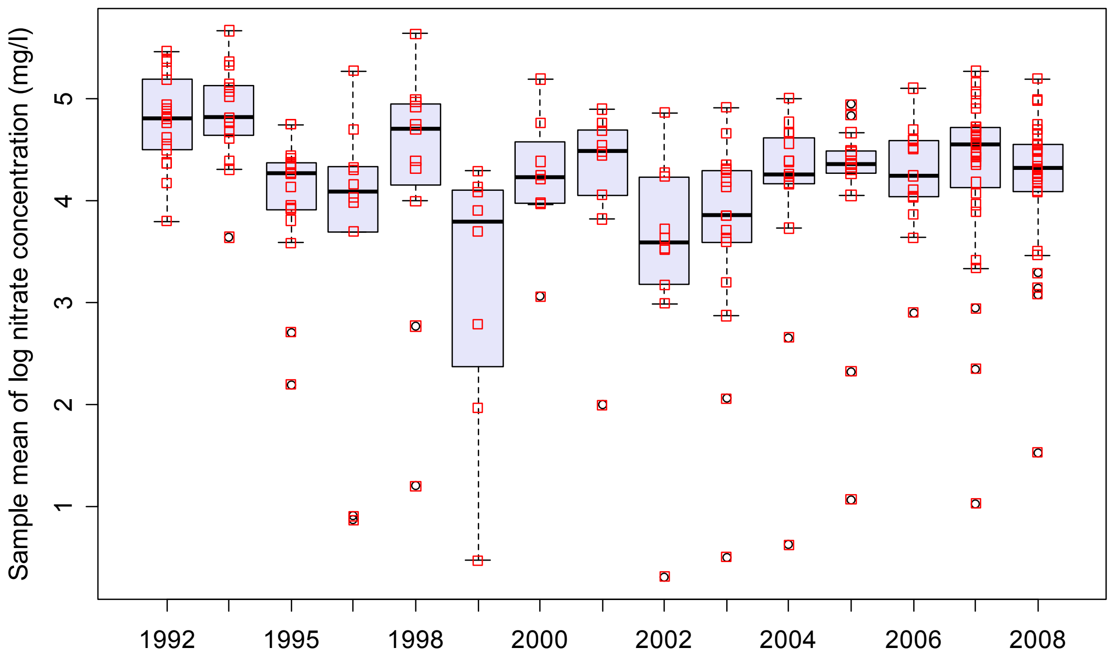


Figure 4 Sample size (number of composite measurements) by farm type.



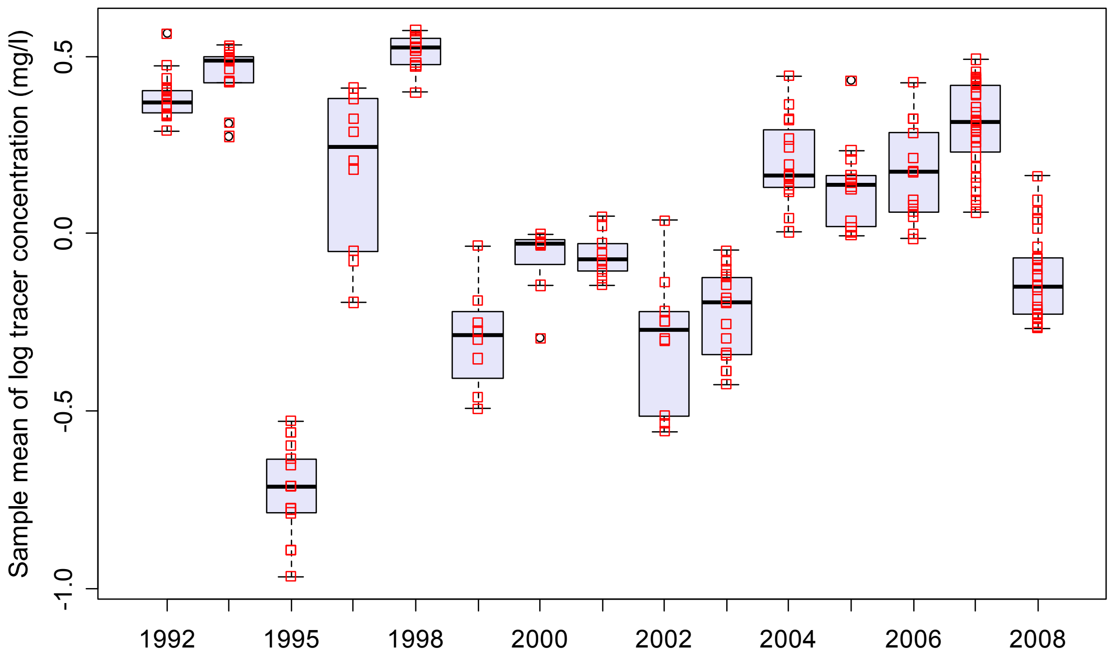


Figure 5A Box plots of sample means of log nitrate and log tracer concentrations by year in arable farms.

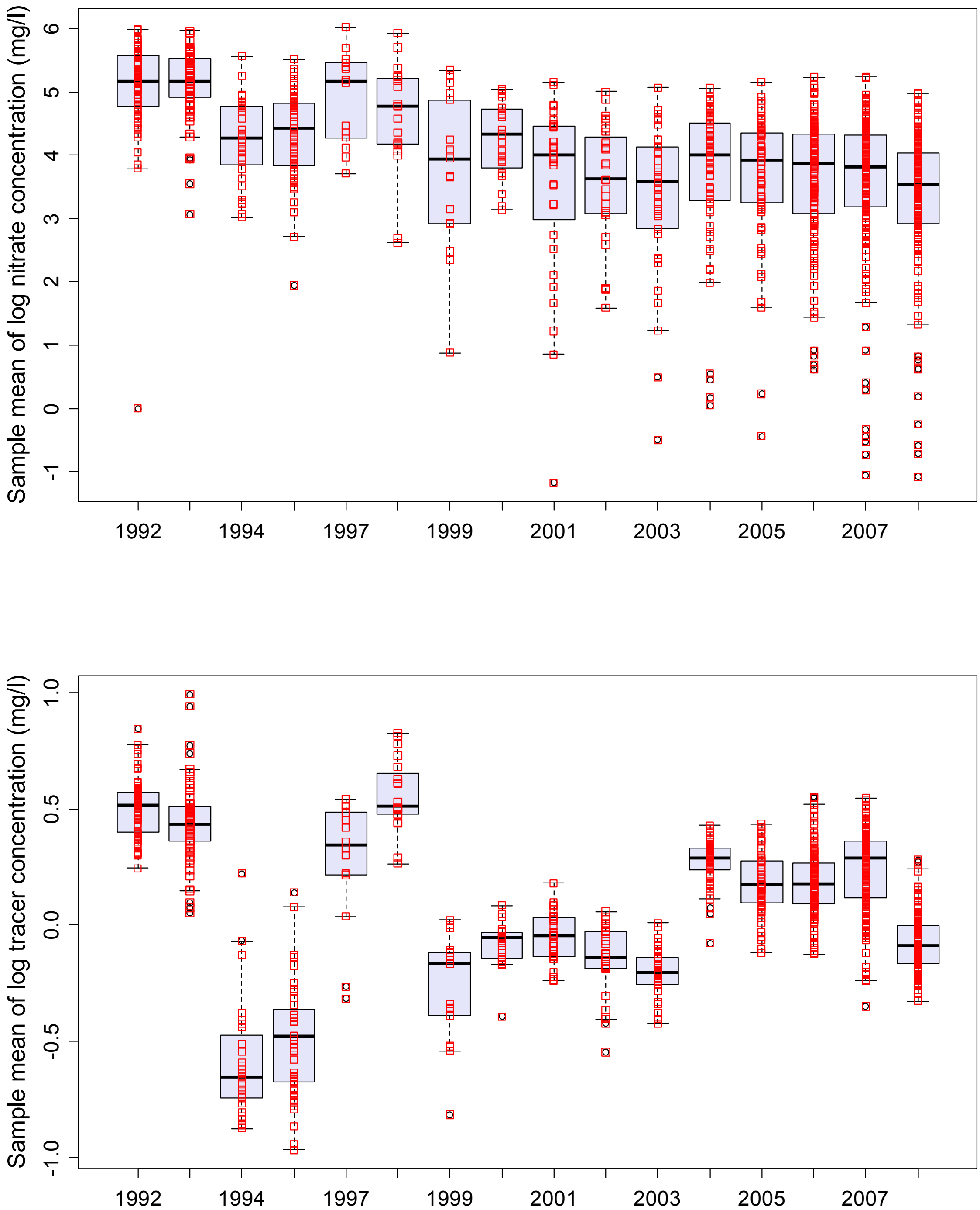


Figure 5B Box plots of sample means of log nitrate and log tracer concentrations by year in dairy farms.

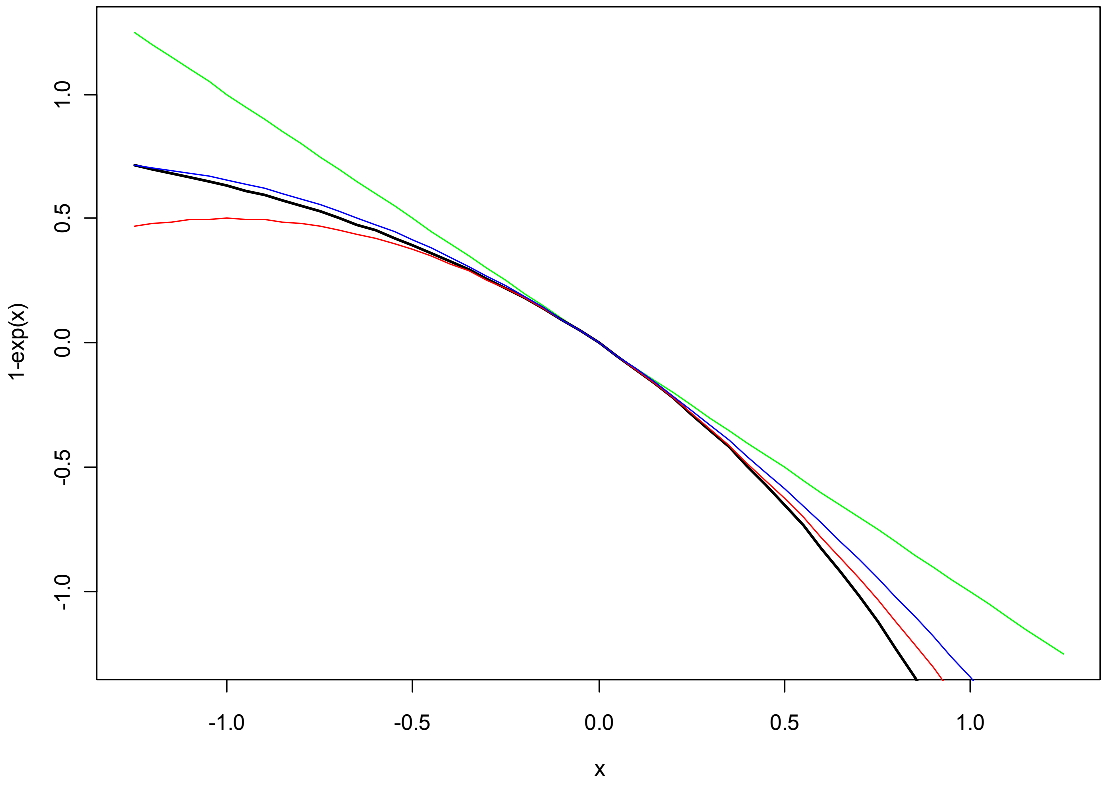
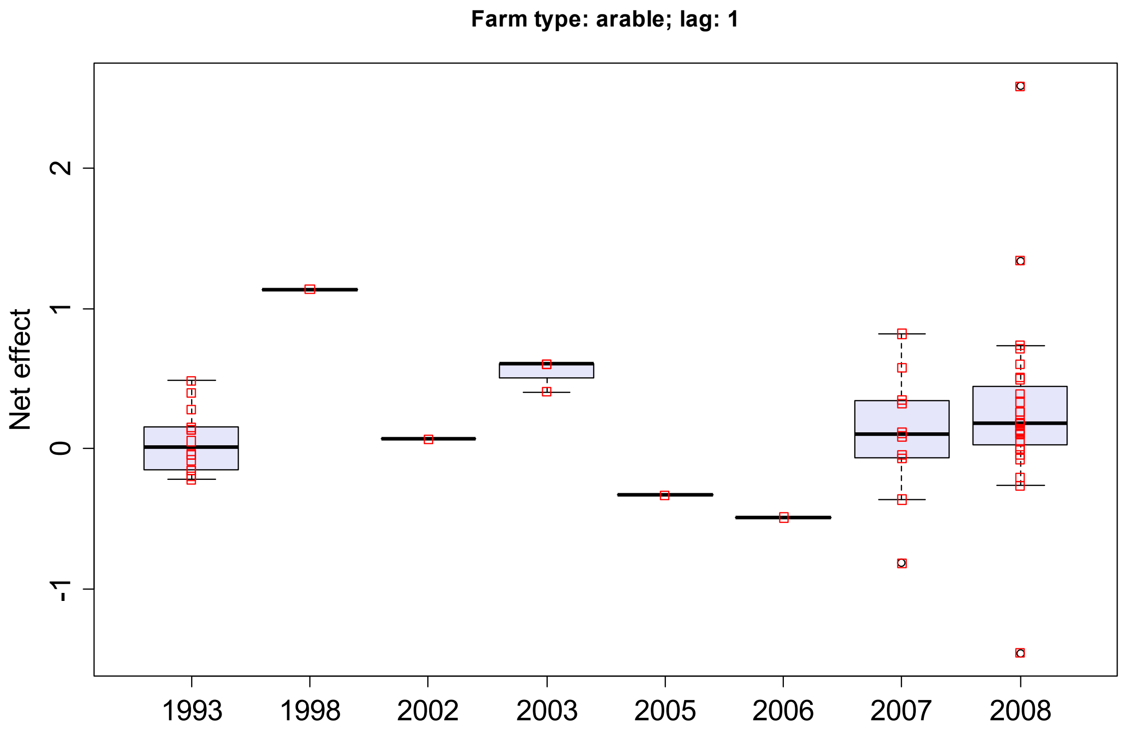


Figure 6 Illustration of the first (green) and second (red) order approximations and of the approximation adopted in chapter 3 (blue) to the function x →1 − exp(x) over the likely range of the mean net effects, namely the interval]-1.25,1.25[.



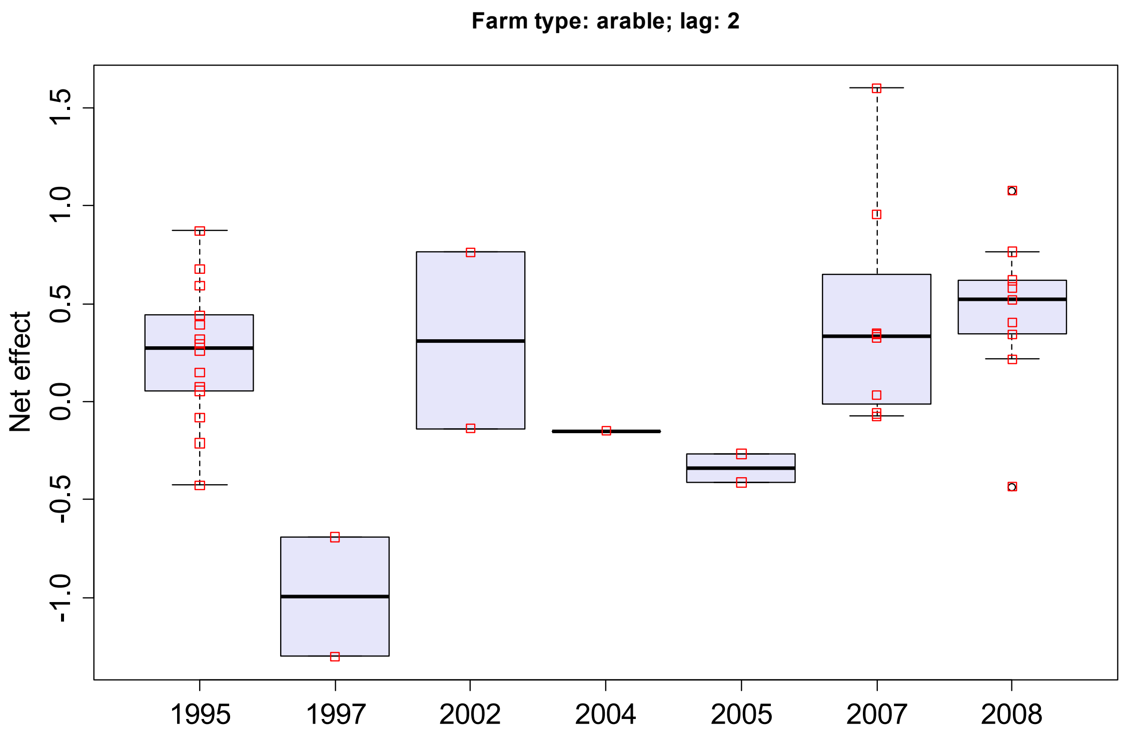
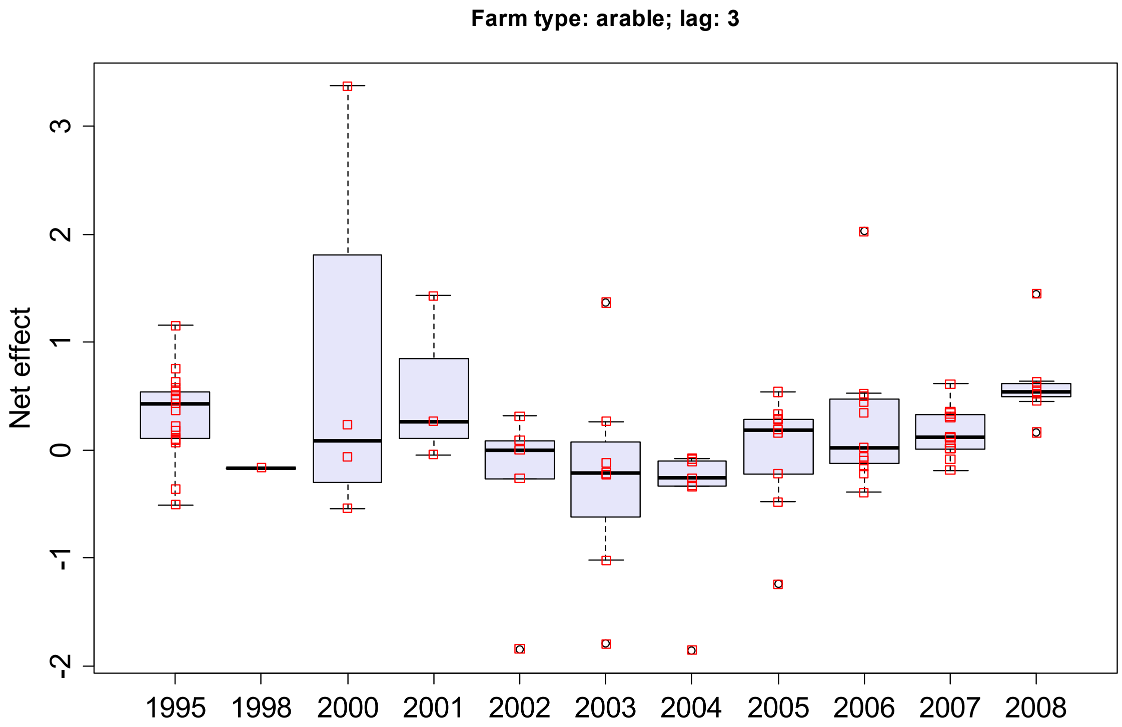


Figure 7A Box plots of net effects corresponding to several lags and years.



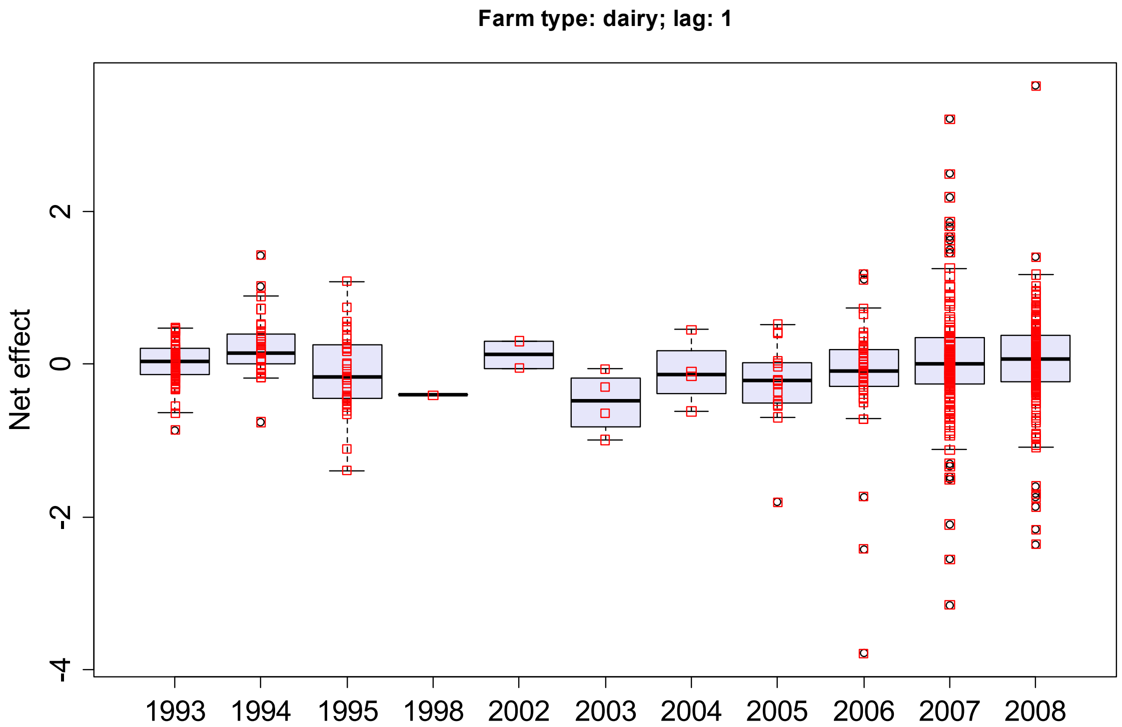
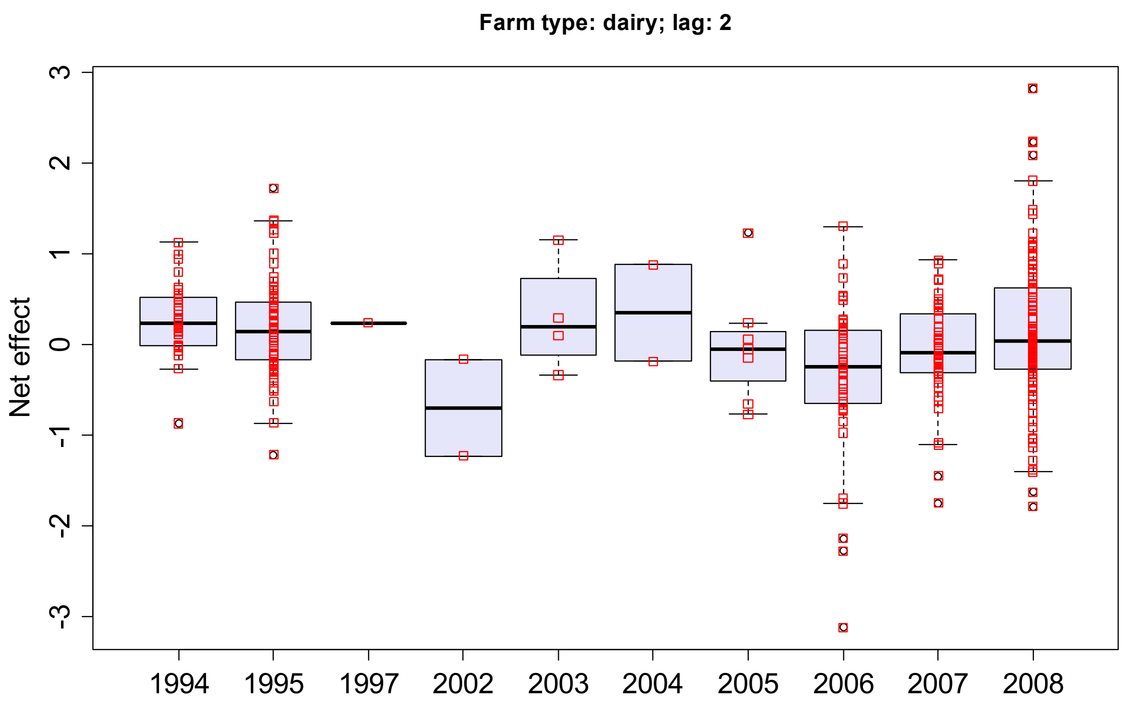


Figure 7B Box plots of net effects corresponding to several lags and years.



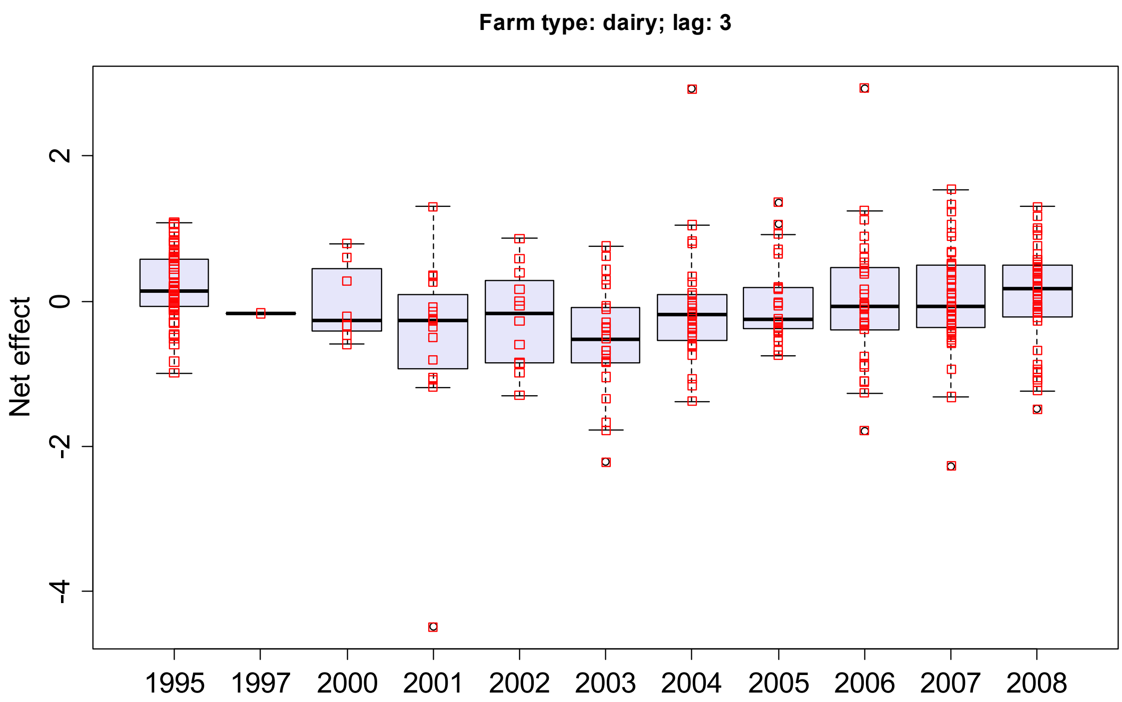


Figure 7C Box plots of net effects corresponding to several lags and years.

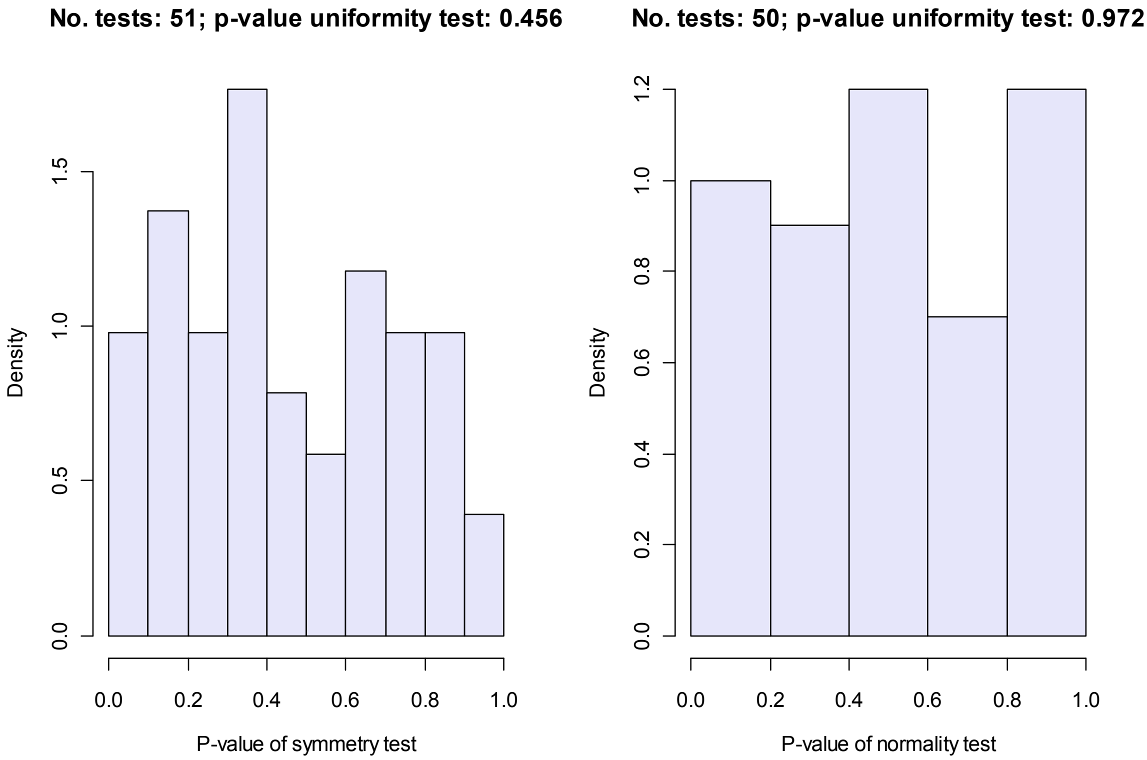
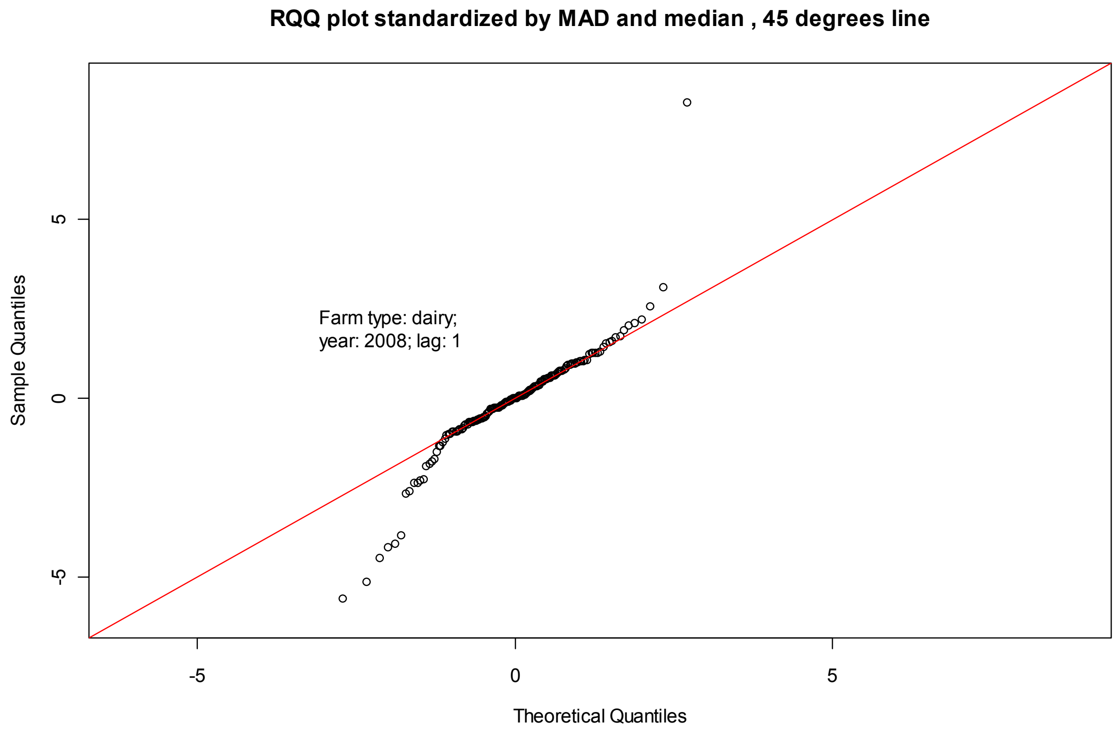


Figure 8 Histograms of p-values from symmetry (left) and normality (right) tests carried out on 51 and 50 samples, respectively, of net effects corresponding to various years and time lags.



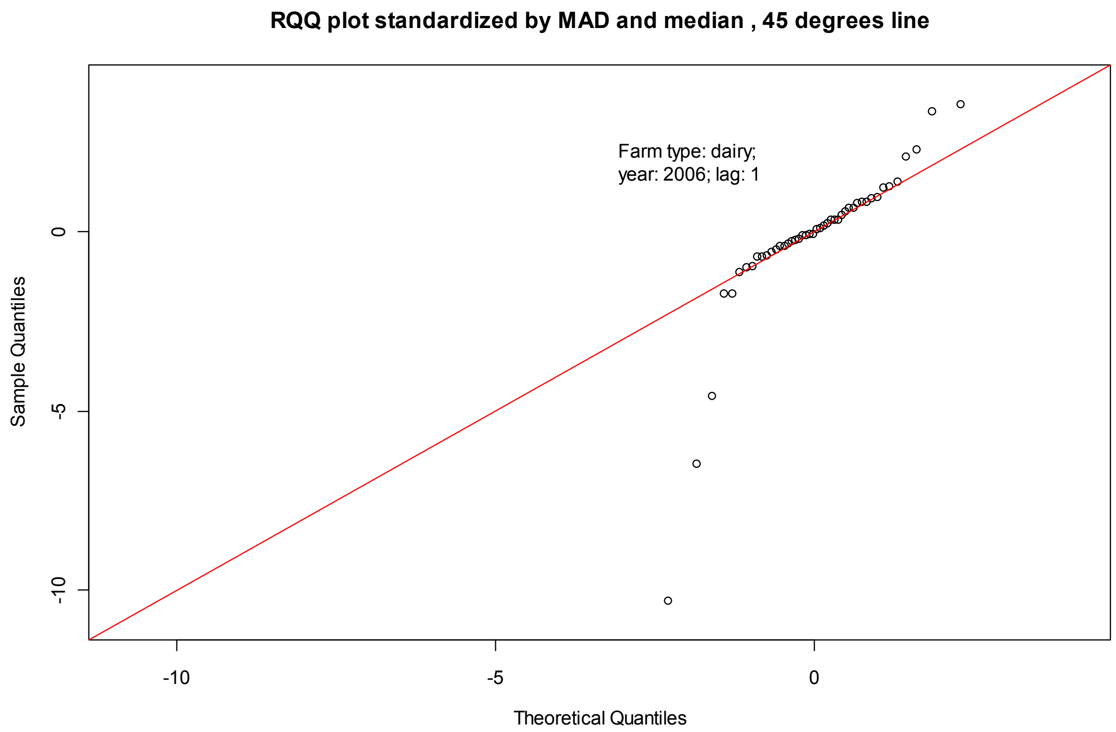
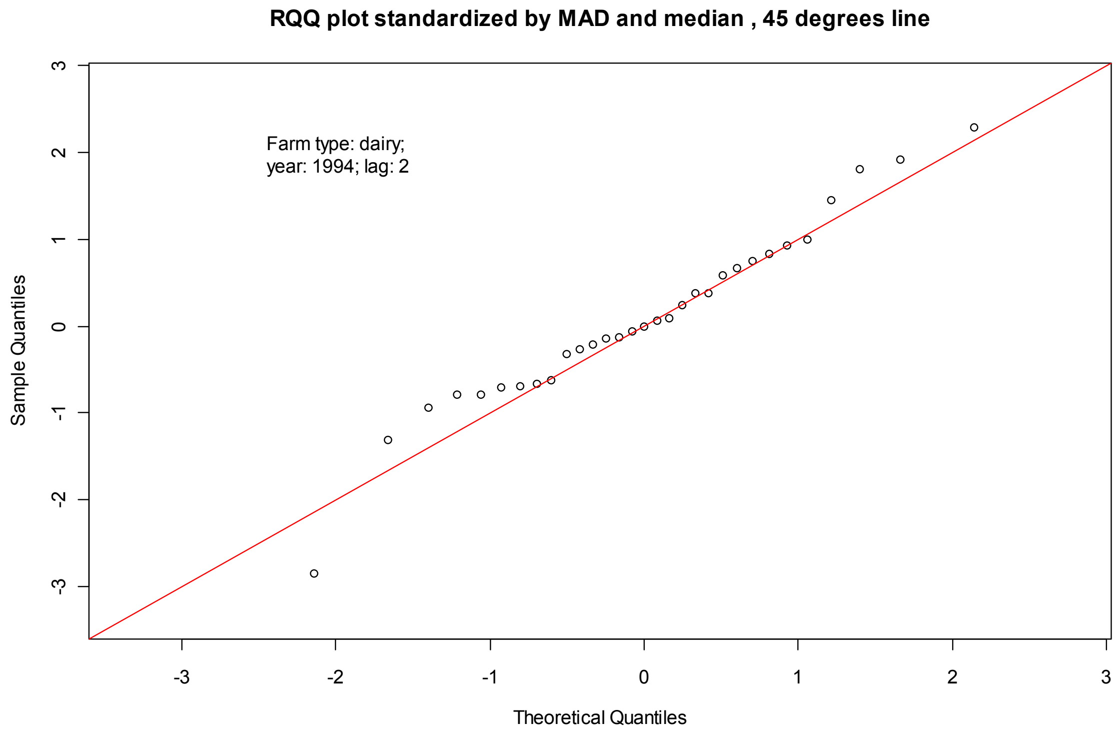


Figure 9A QQ plots for checking the normality of net effects corresponding to several lags and years.



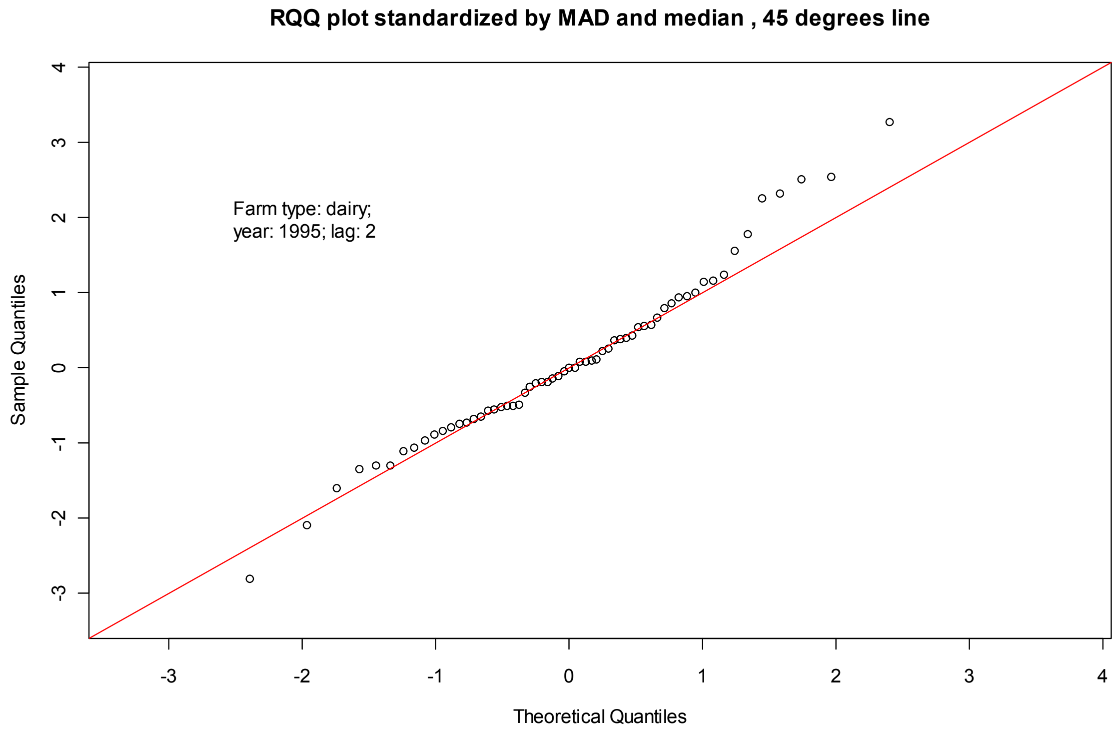


Figure 9B QQ plots for checking the normality of net effects corresponding to several lags and years.

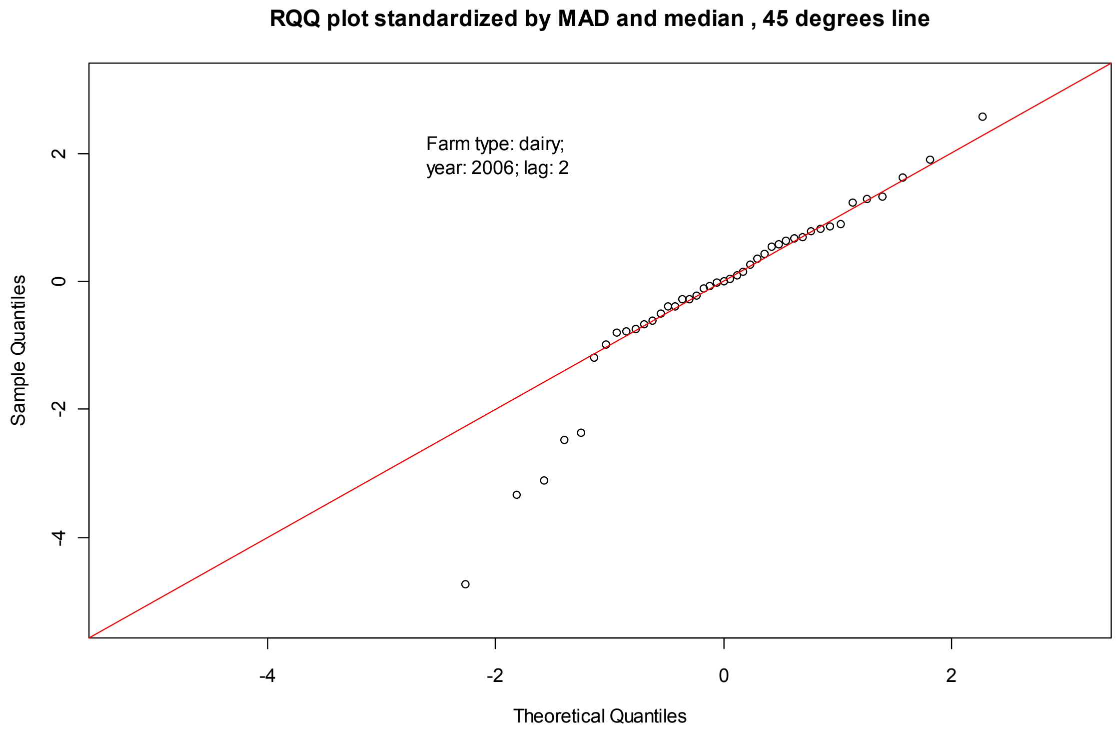
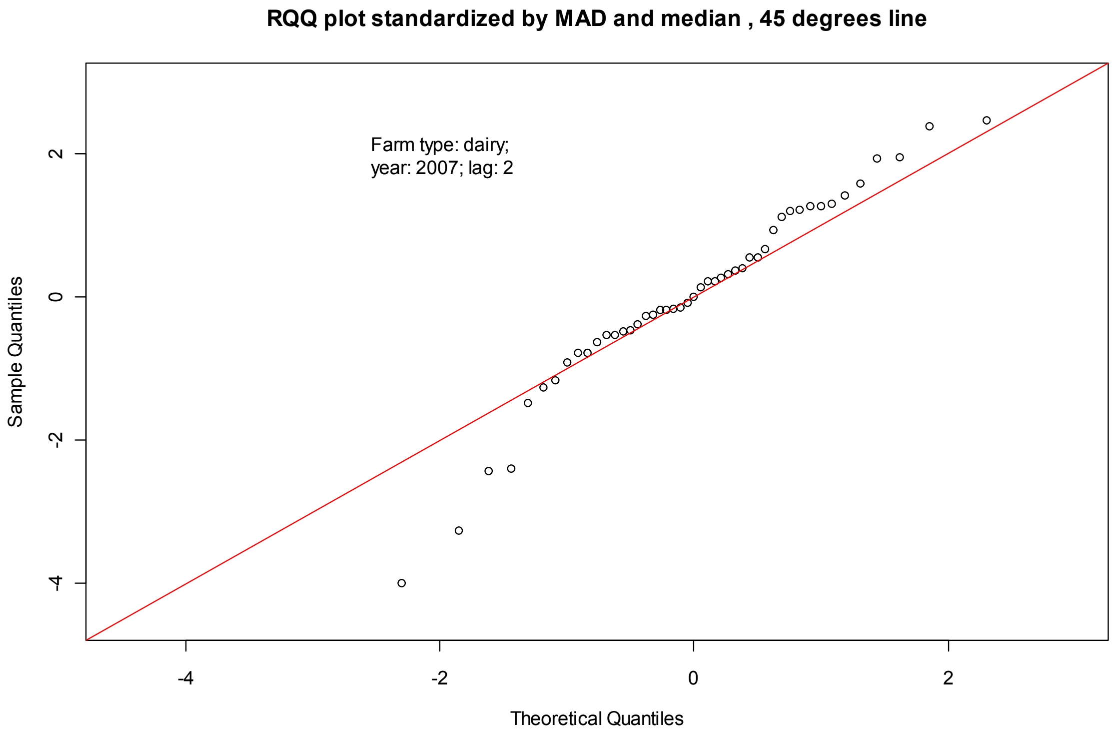


Figure 9C QQ plots for checking the normality of net effects corresponding to several lags and years.



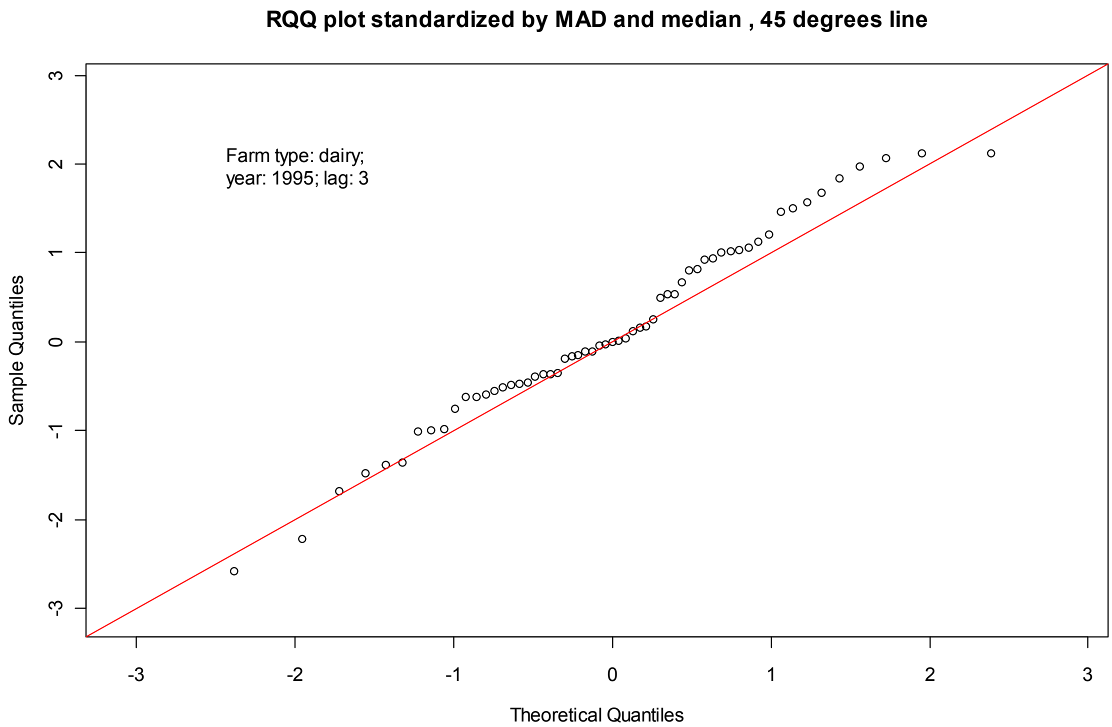
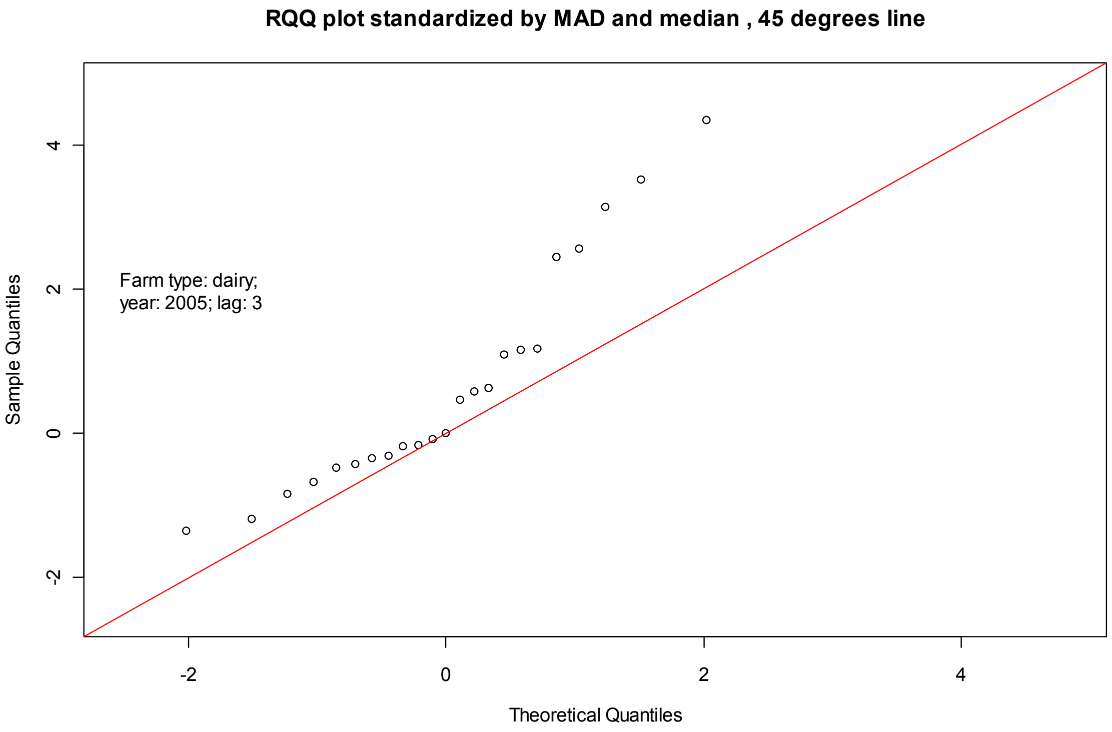


Figure 9D QQ plots for checking the normality of net effects corresponding to several lags and years.



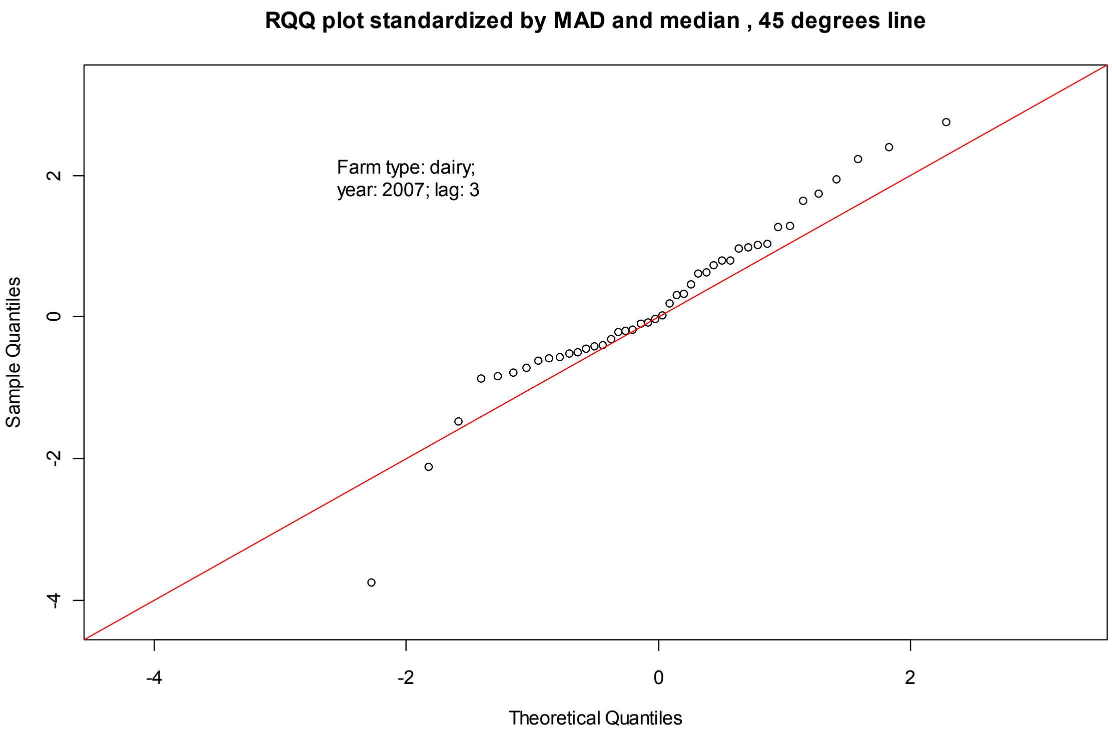
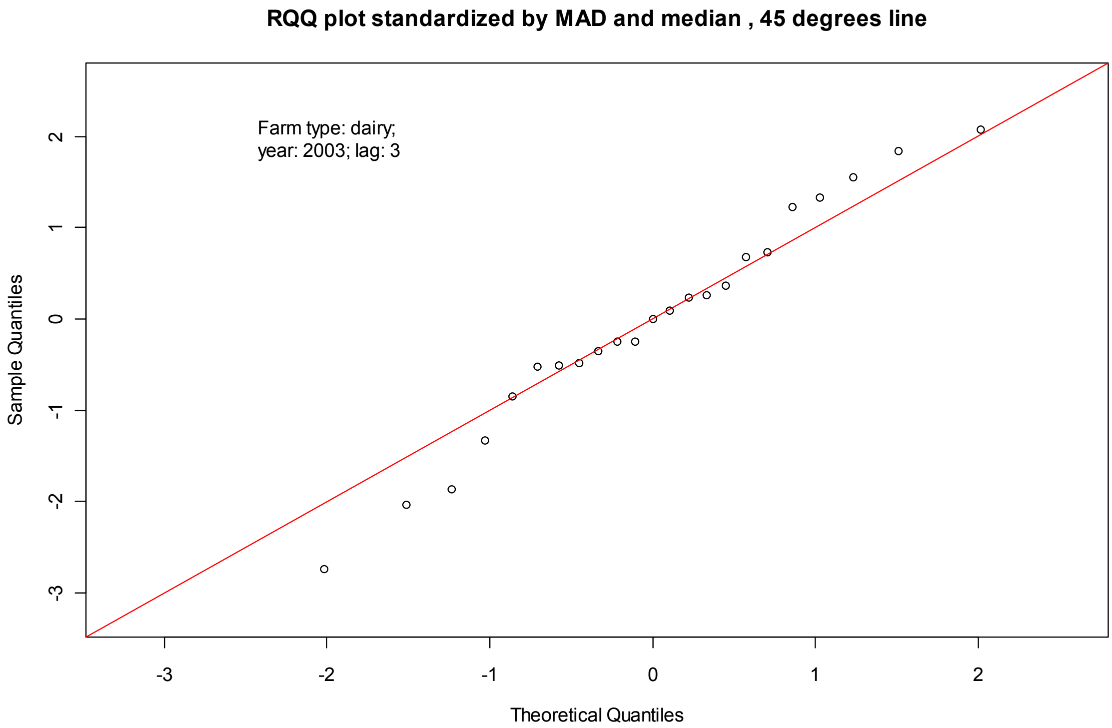


Figure 9E QQ plots for checking the normality of net effects corresponding to several lags and years.



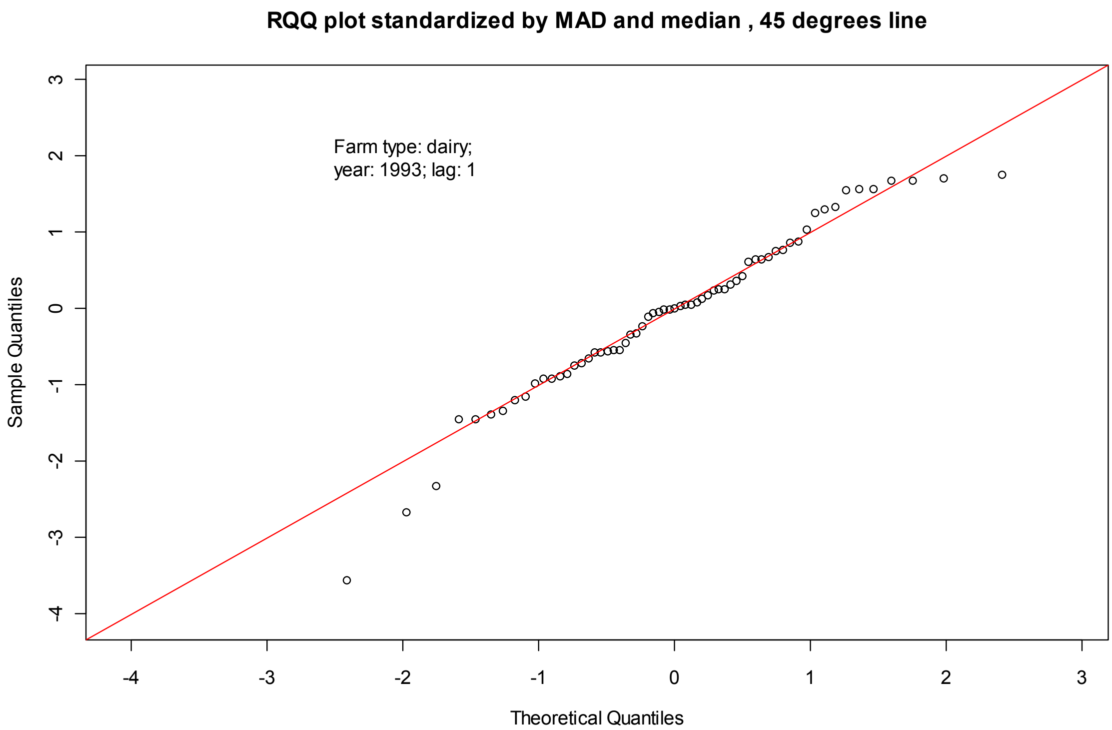
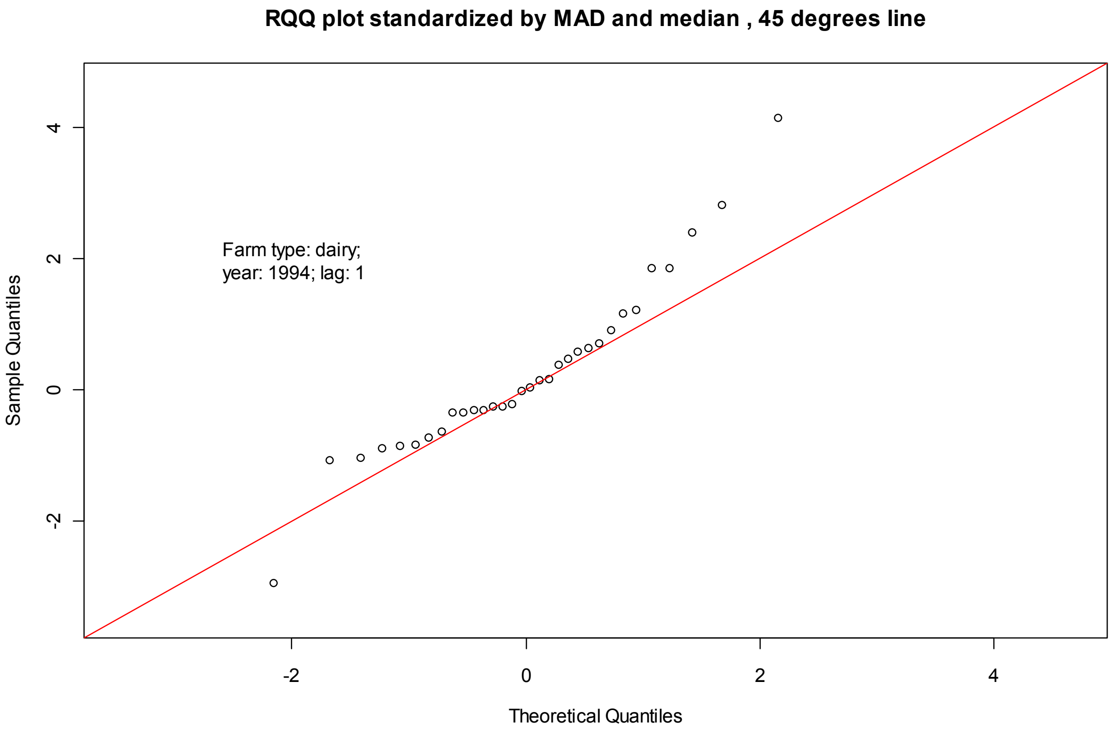


Figure 9F QQ plots for checking the normality of net effects corresponding to several lags and years.



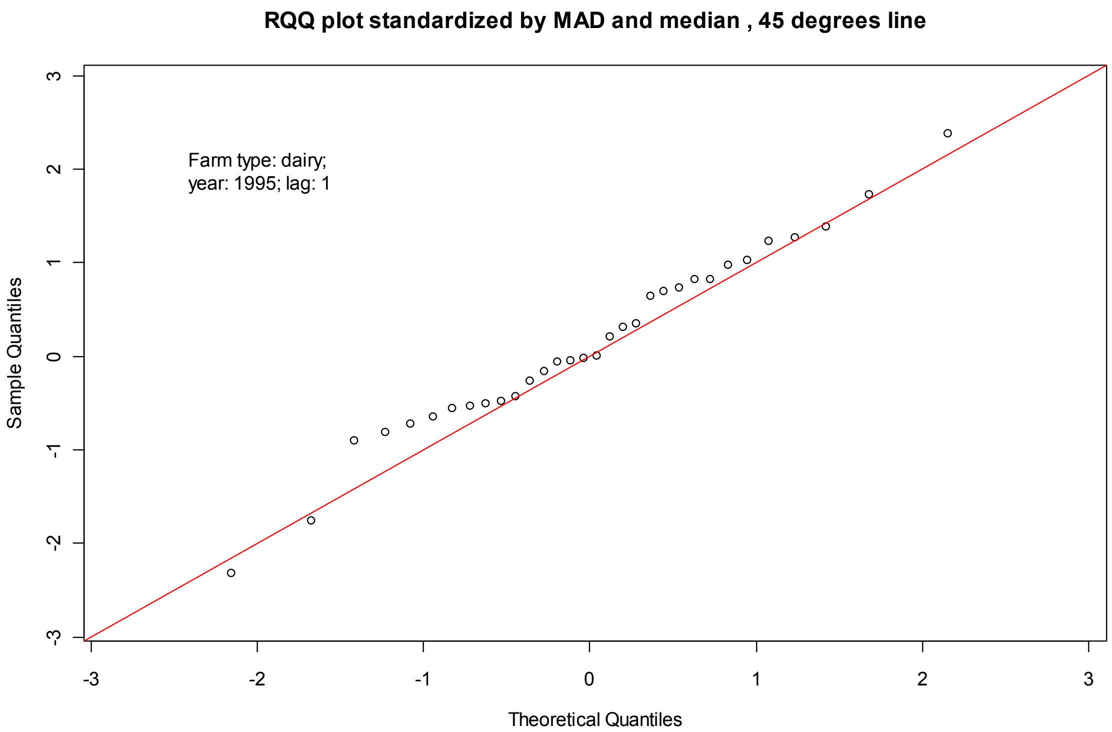
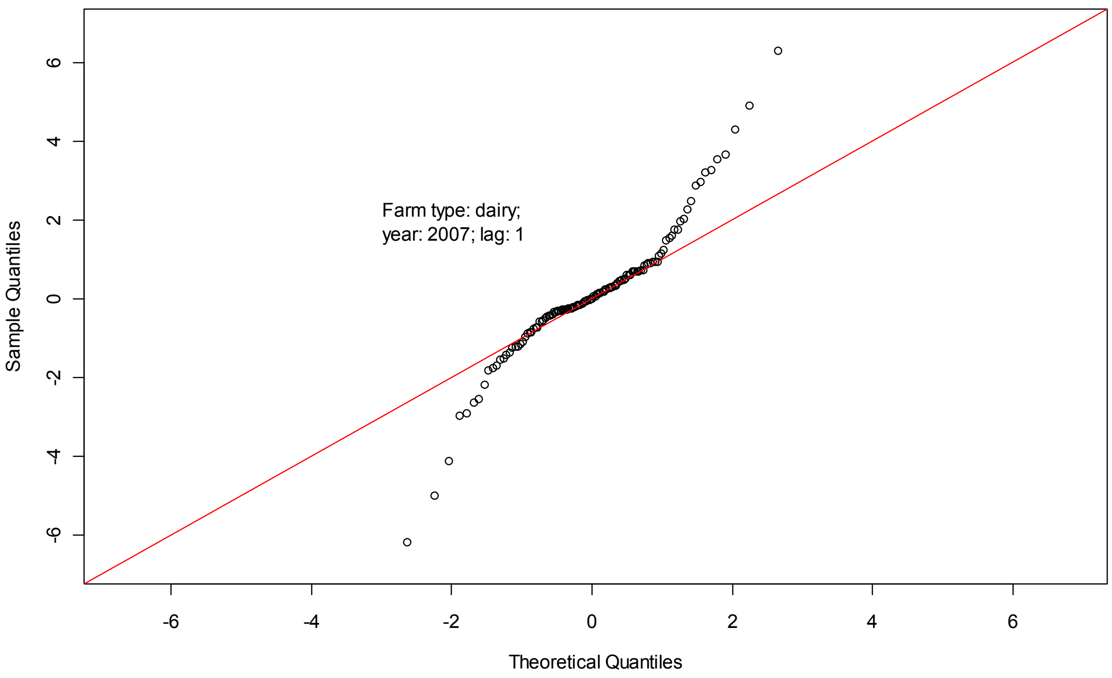


Figure 9G QQ plots for checking the normality of net effects corresponding to several lags and years.

RQQ plot standardized by MAD and median , 45 degrees line



RQQ plot standardized by MAD and median , 45 degrees line

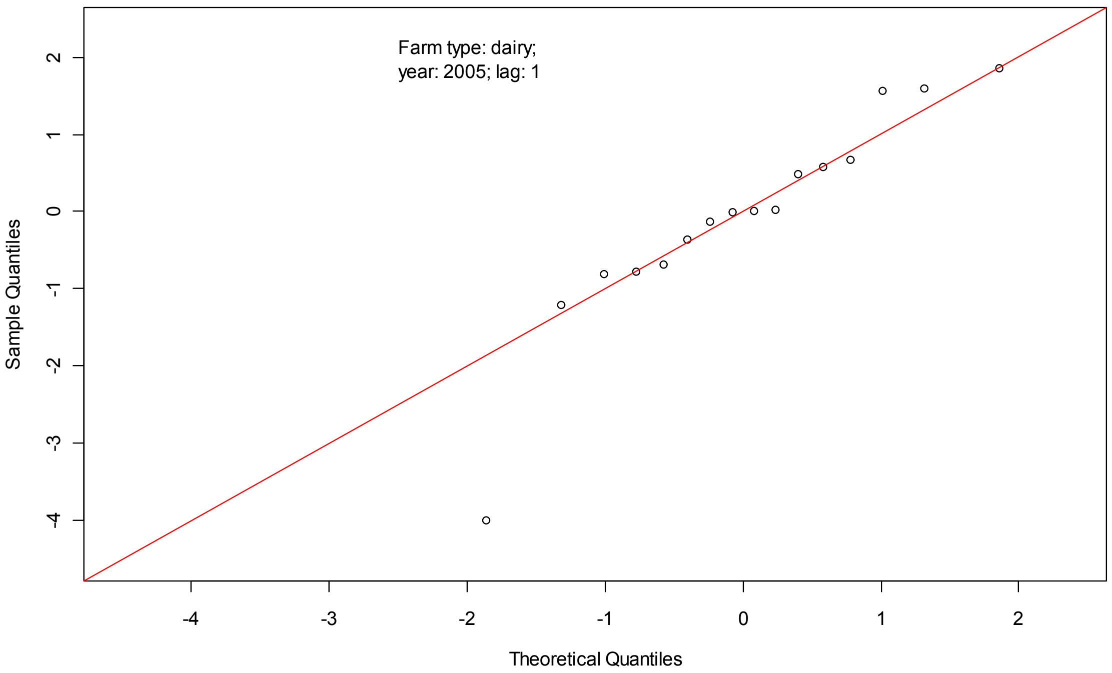
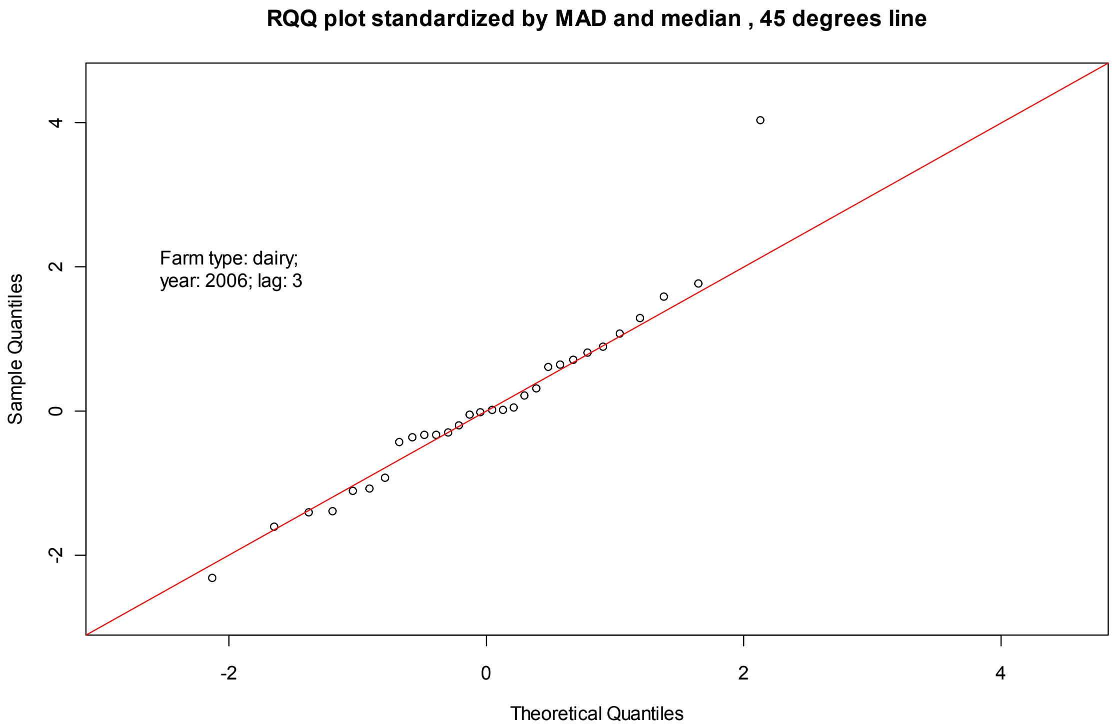


Figure 9H QQ plots for checking the normality of net effects corresponding to several lags and years.



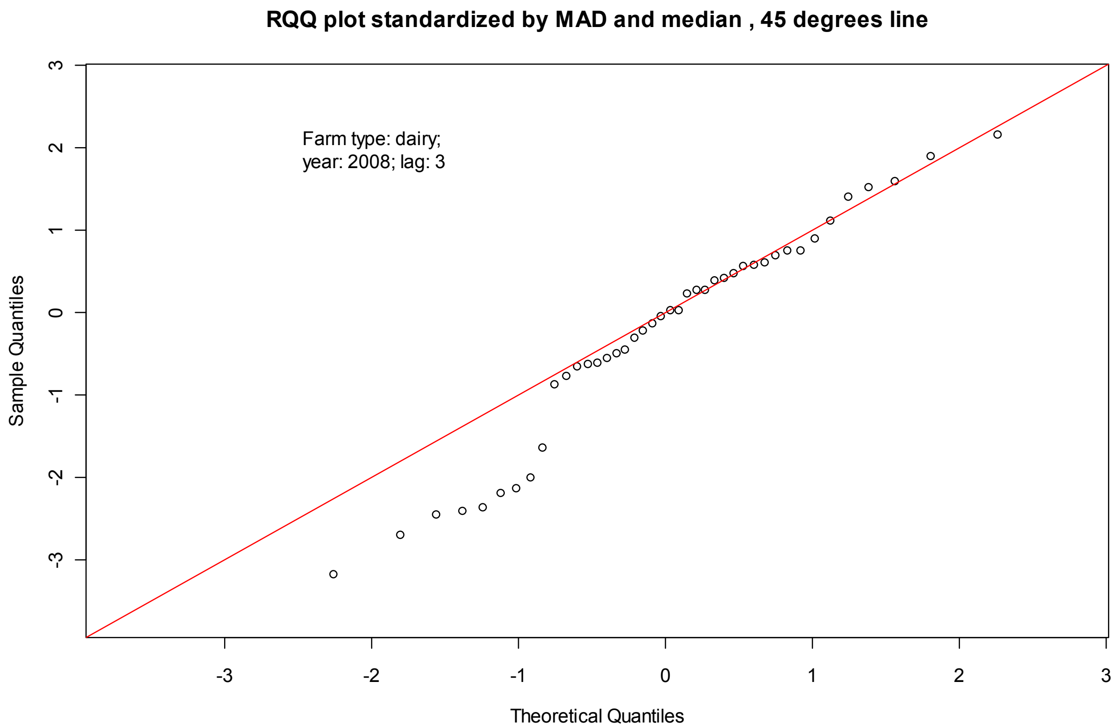
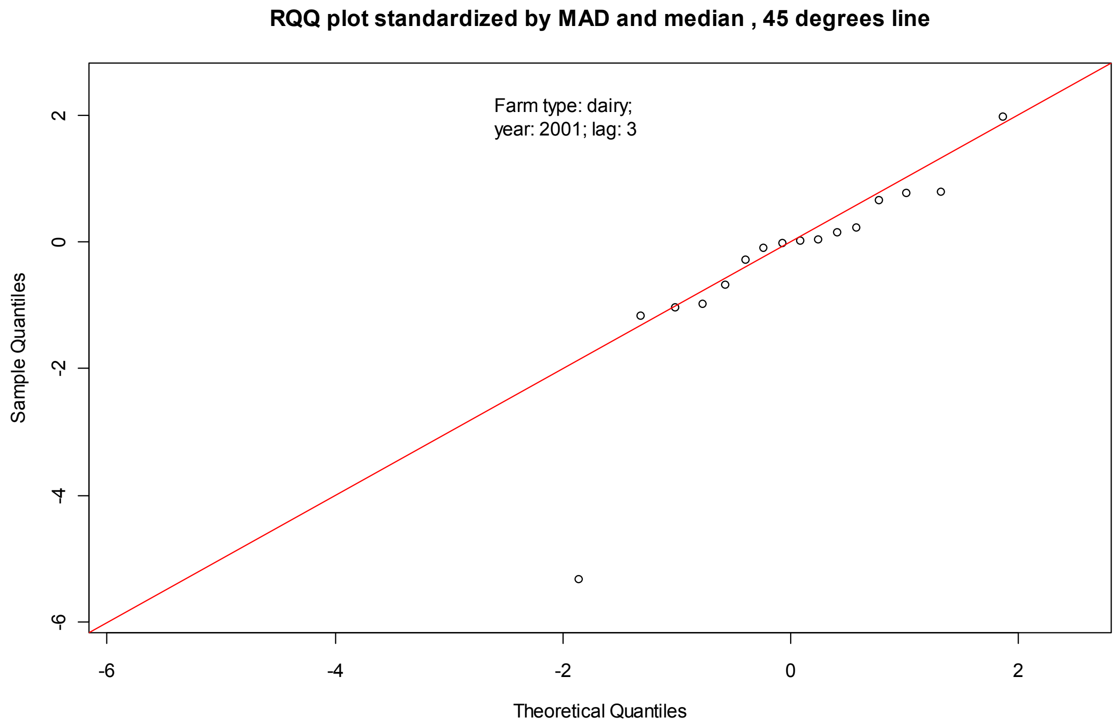


Figure 9l QQ plots for checking the normality of net effects corresponding to several lags and years.



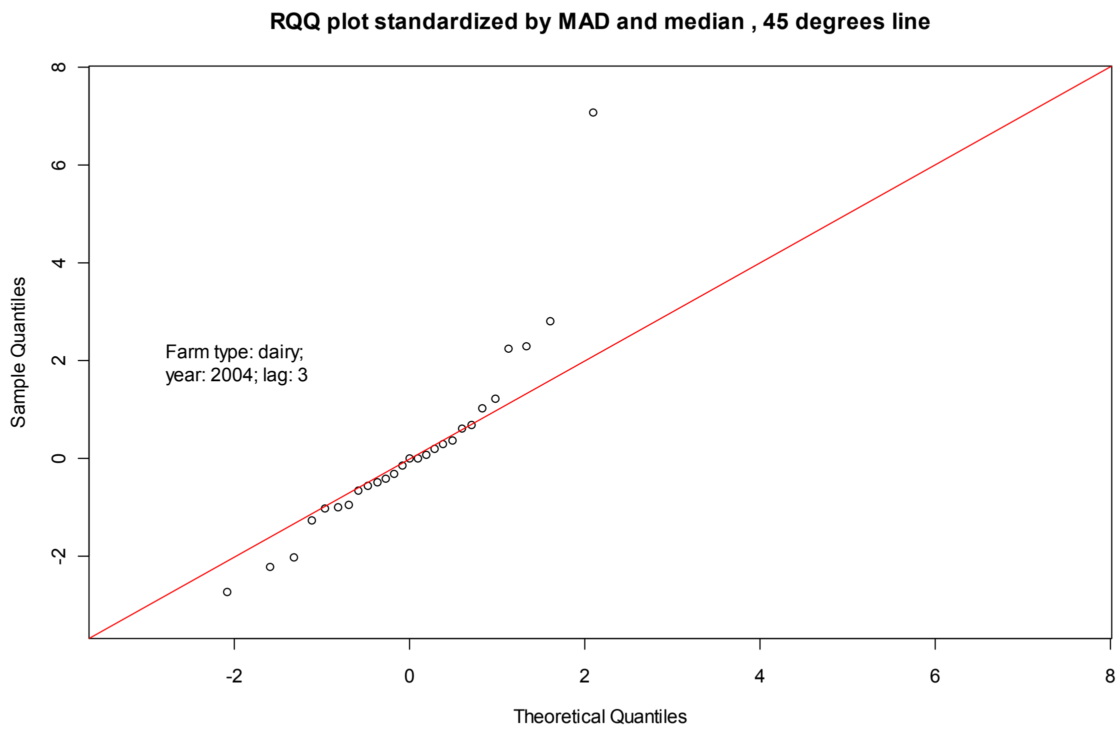
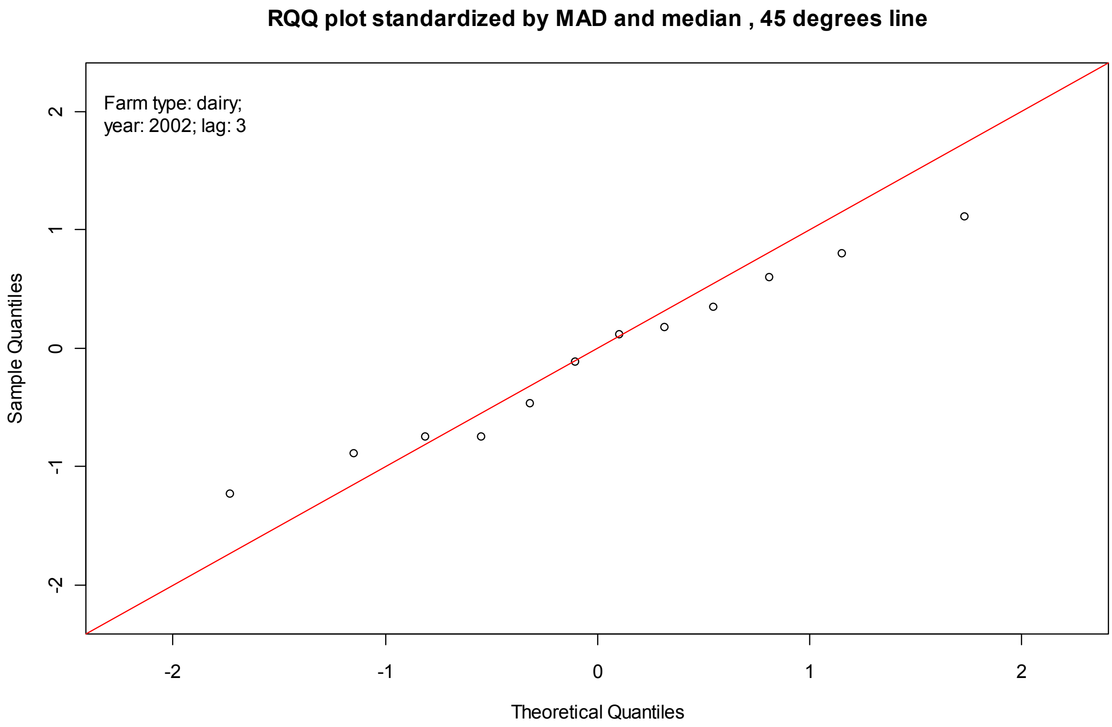


Figure 9J QQ plots for checking the normality of net effects corresponding to several lags and years.



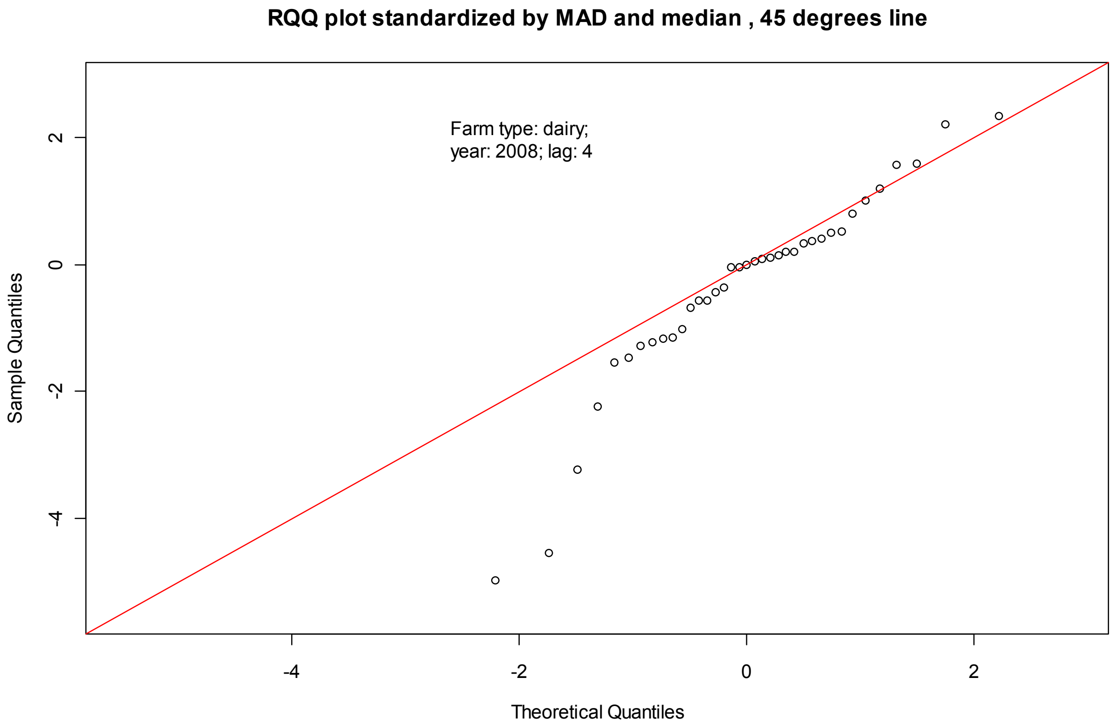
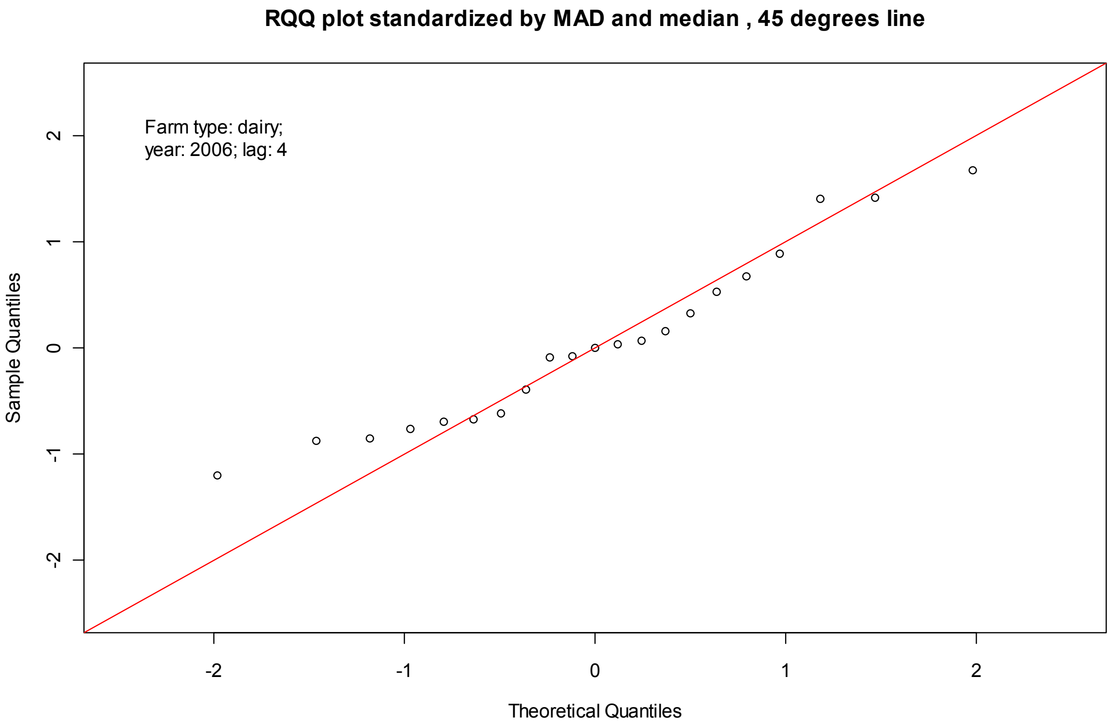


Figure 9K QQ plots for checking the normality of net effects corresponding to several lags and years.



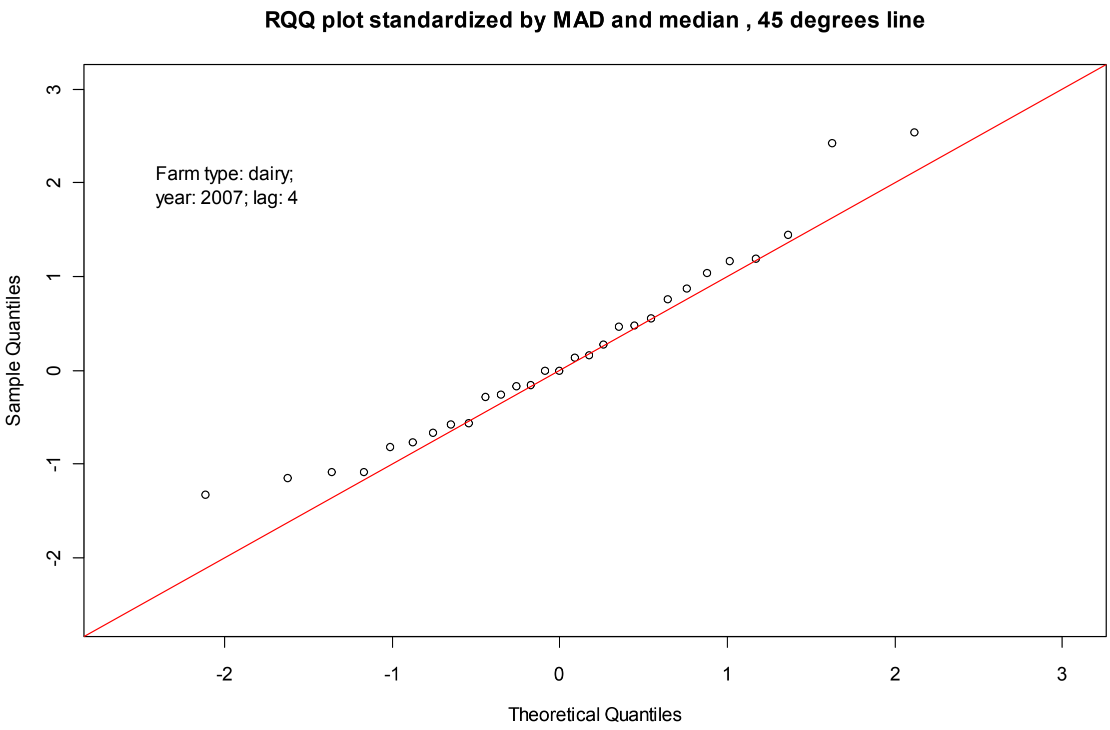
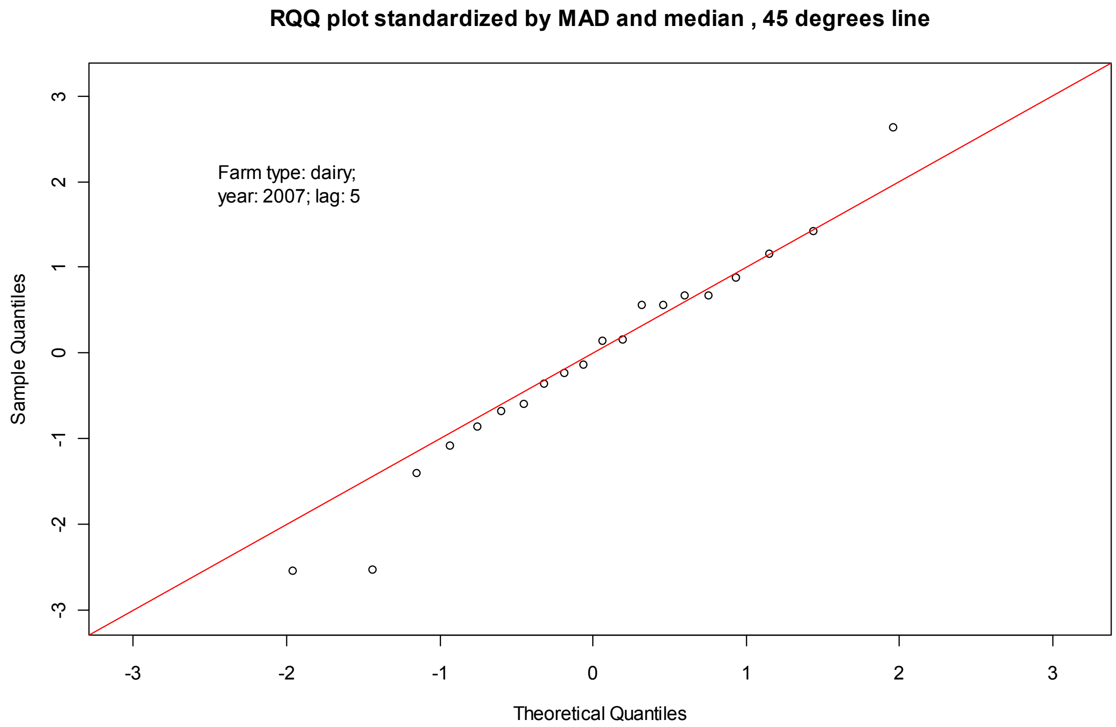


Figure 9L QQ plots for checking the normality of net effects corresponding to several lags and years.



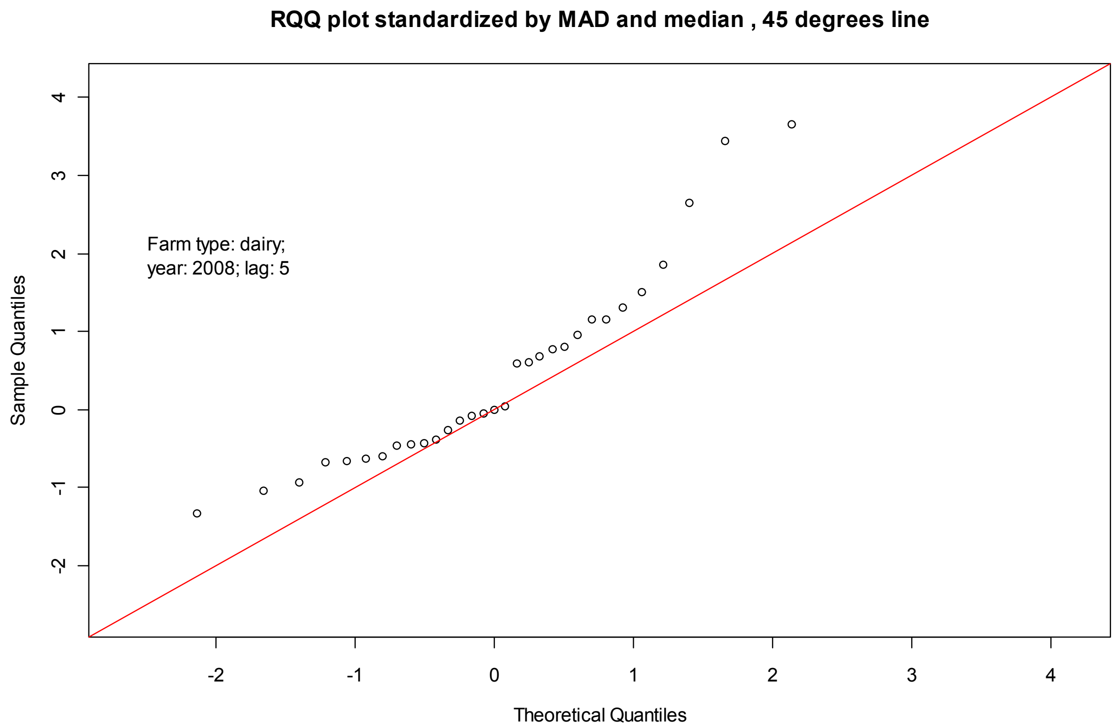
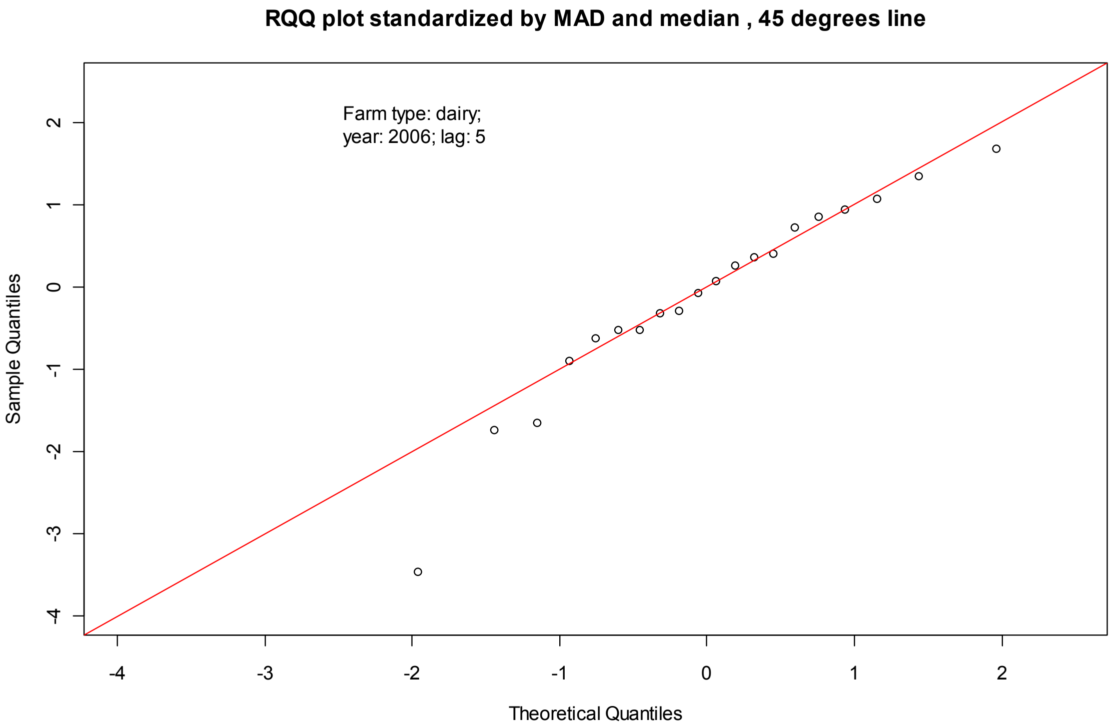


Figure 9M QQ plots for checking the normality of net effects corresponding to several lags and years.



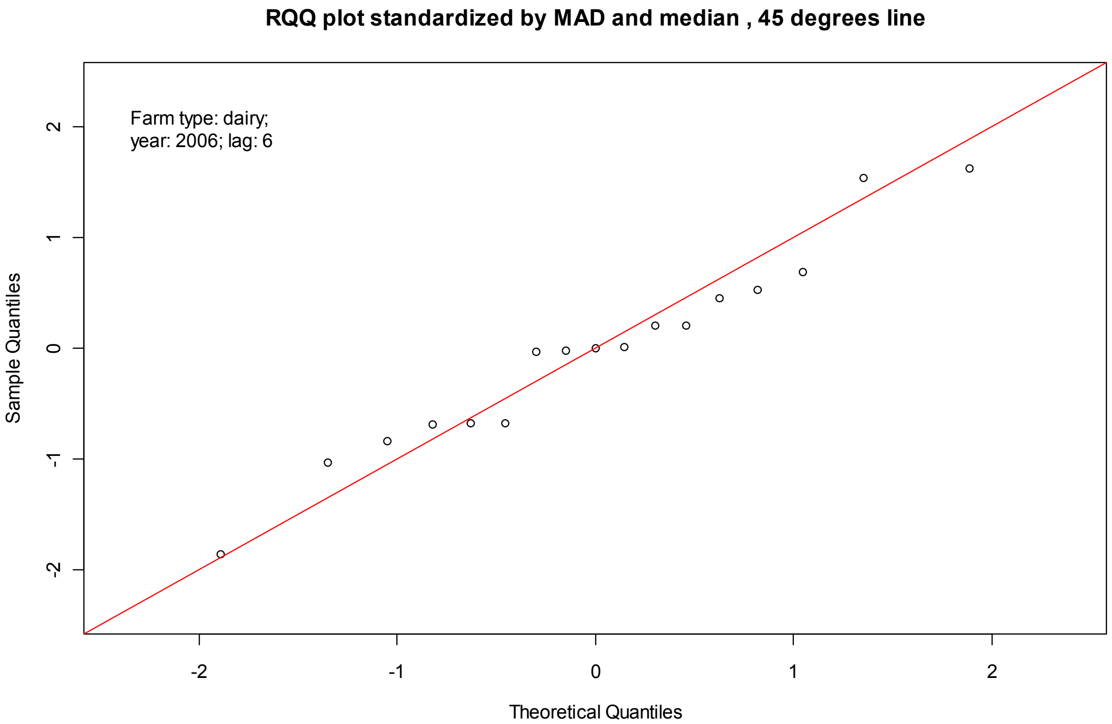
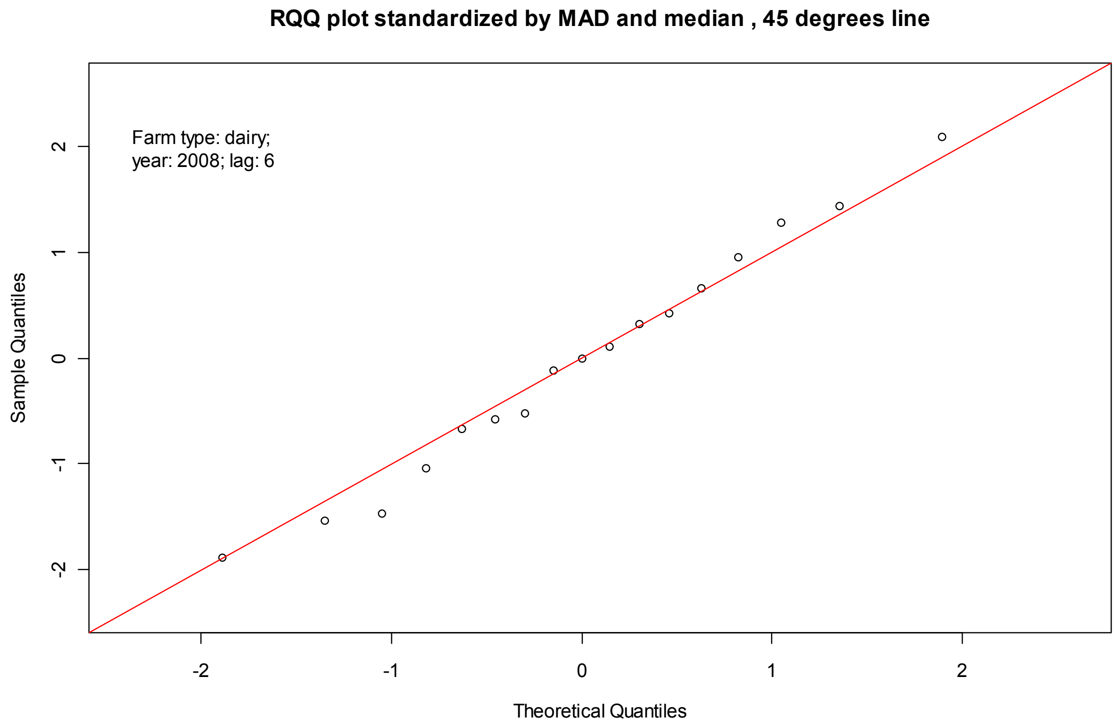


Figure 9N QQ plots for checking the normality of net effects corresponding to several lags and years.



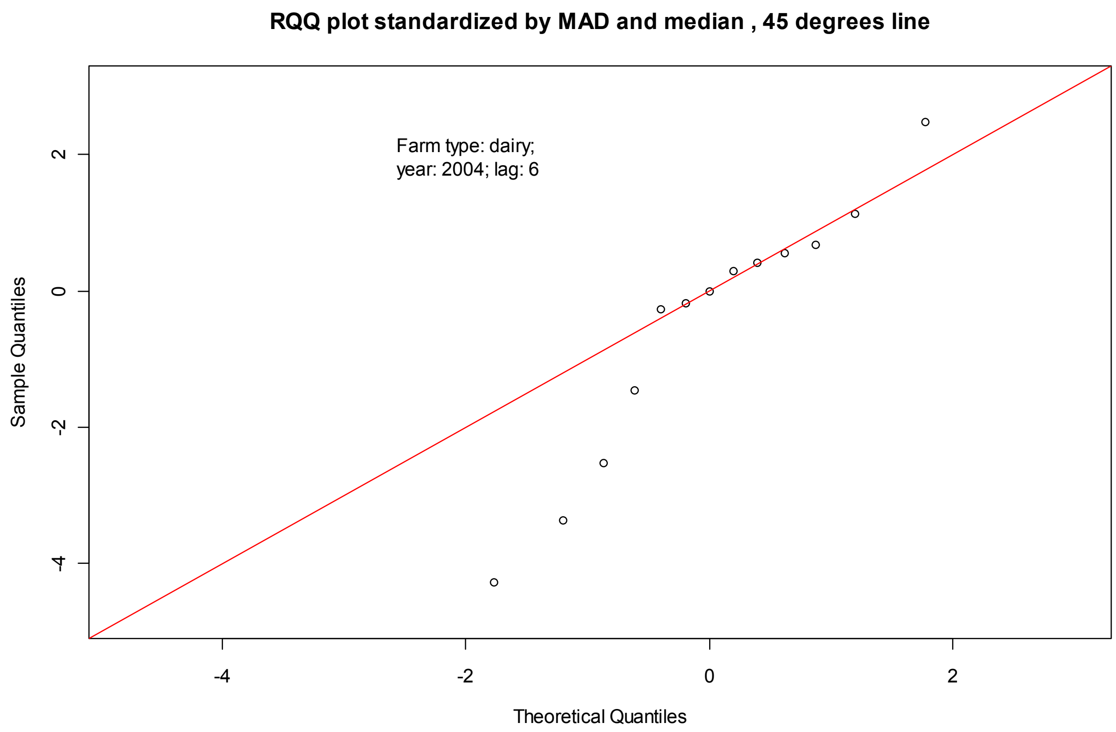
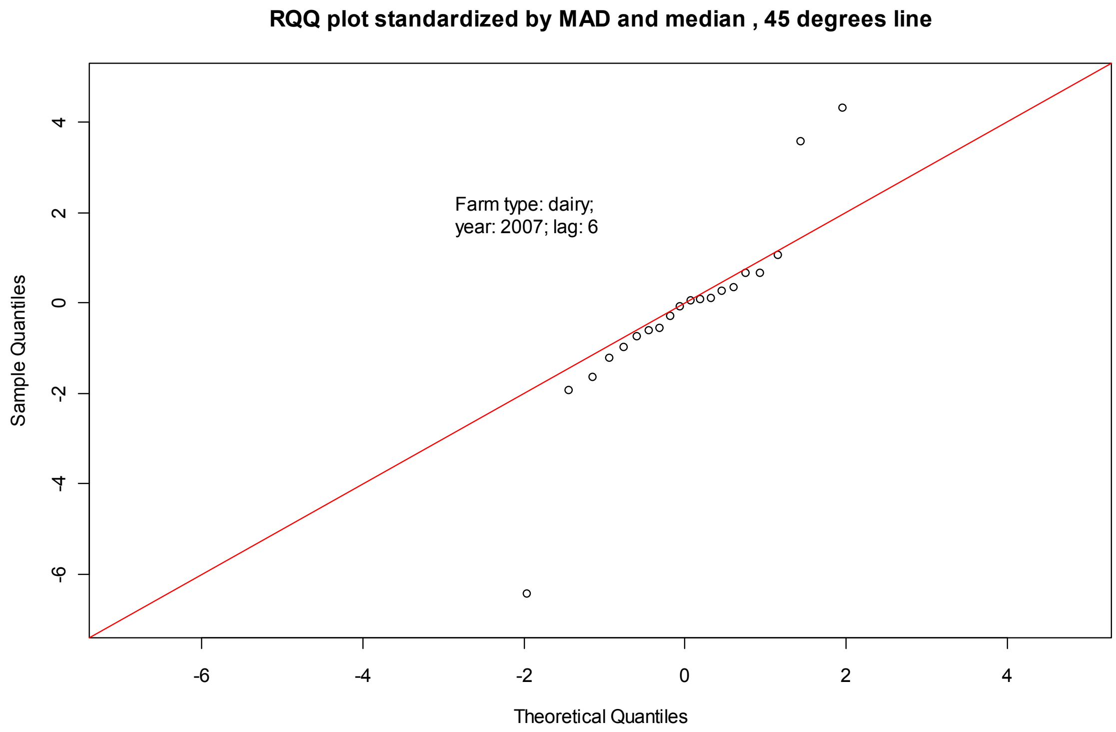


Figure 9O QQ plots for checking the normality of net effects corresponding to several lags and years.



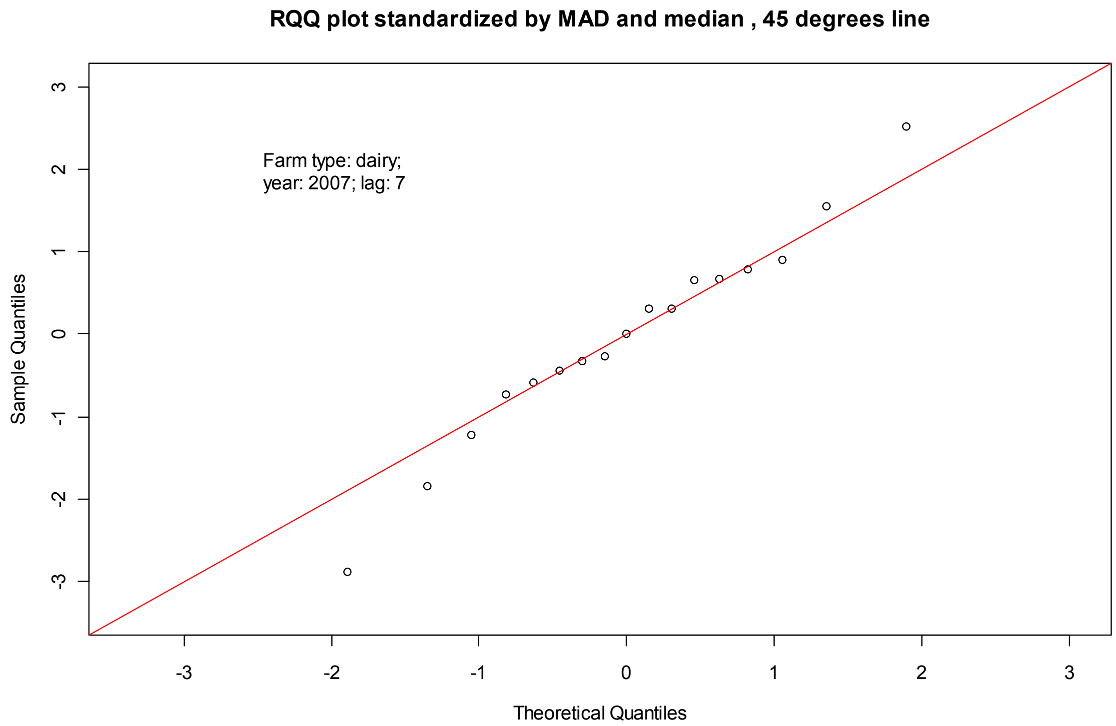
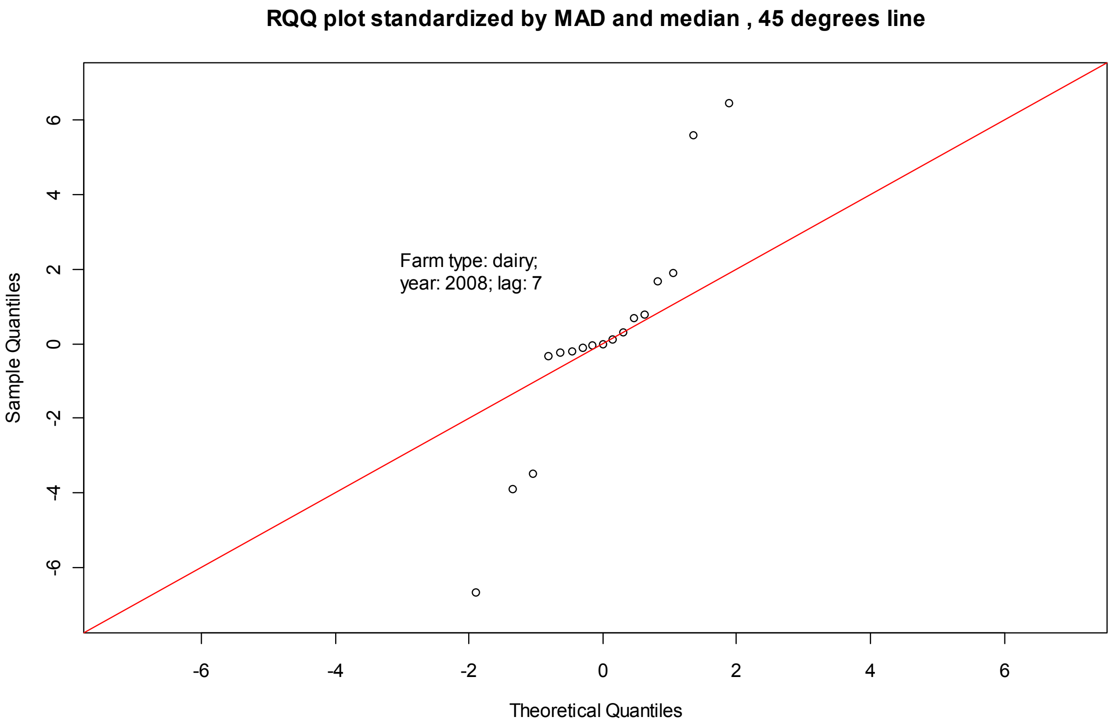


Figure 9P QQ plots for checking the normality of net effects corresponding to several lags and years.



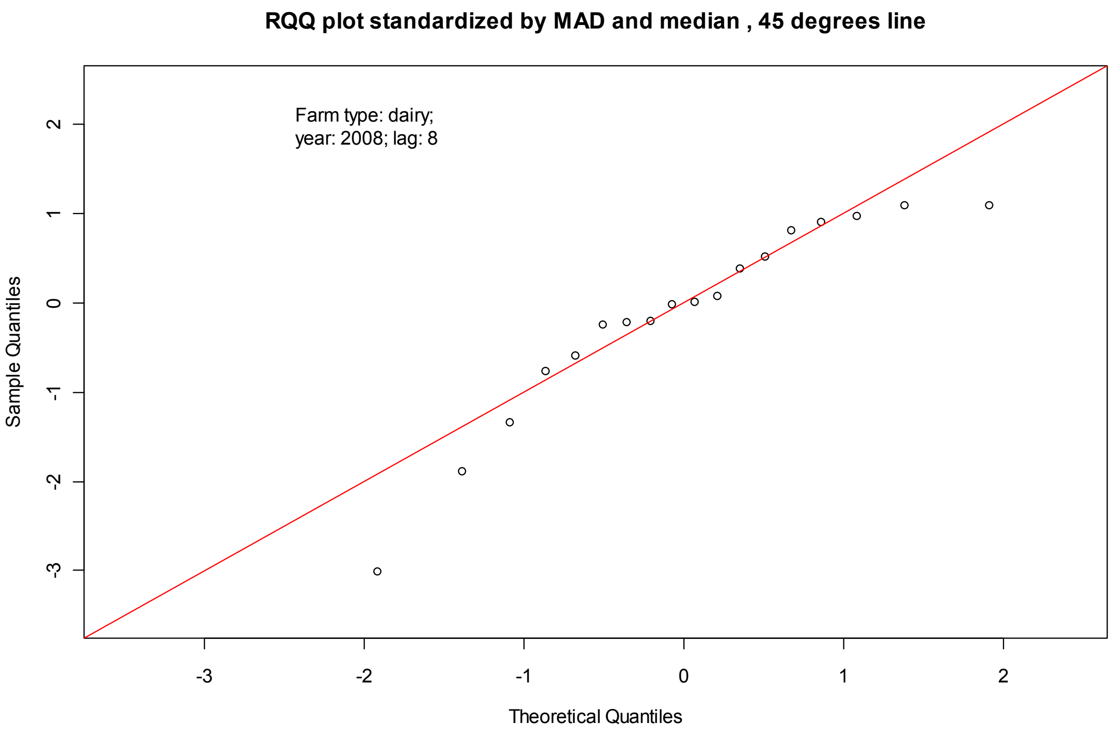


Figure 9Q QQ plots for checking the normality of net effects corresponding to several lags and years.

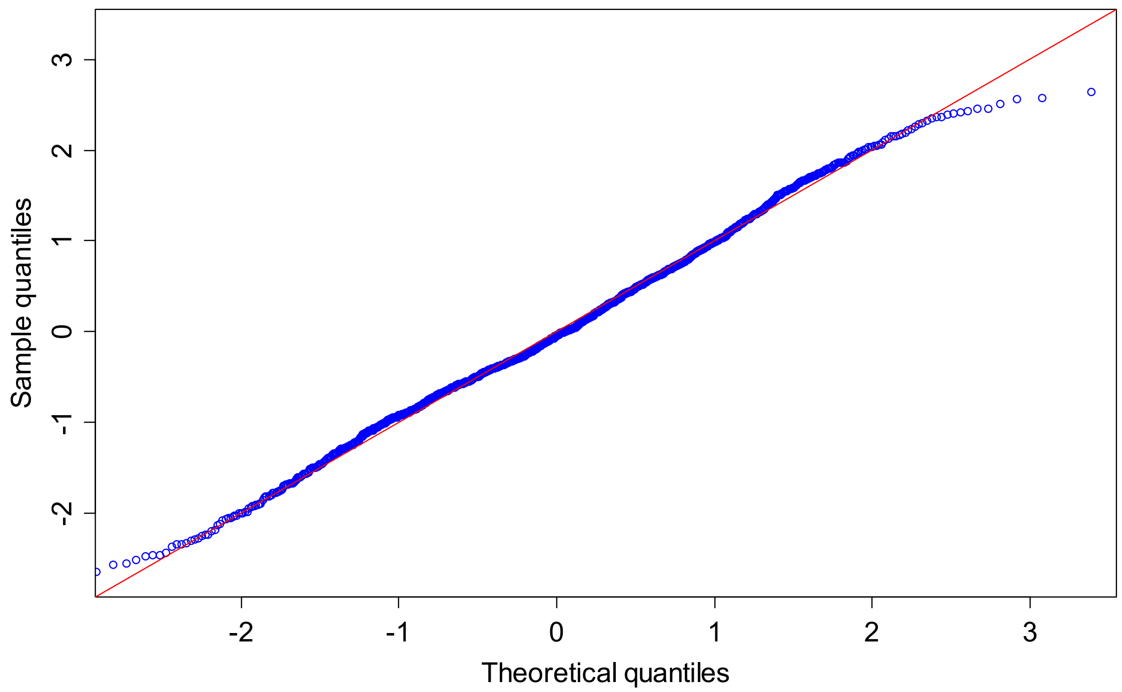


Figure 10 QQ plot for checking the normality of the 'standardized residuals' of samples of net effects corresponding to various years and time lags from different farm types.

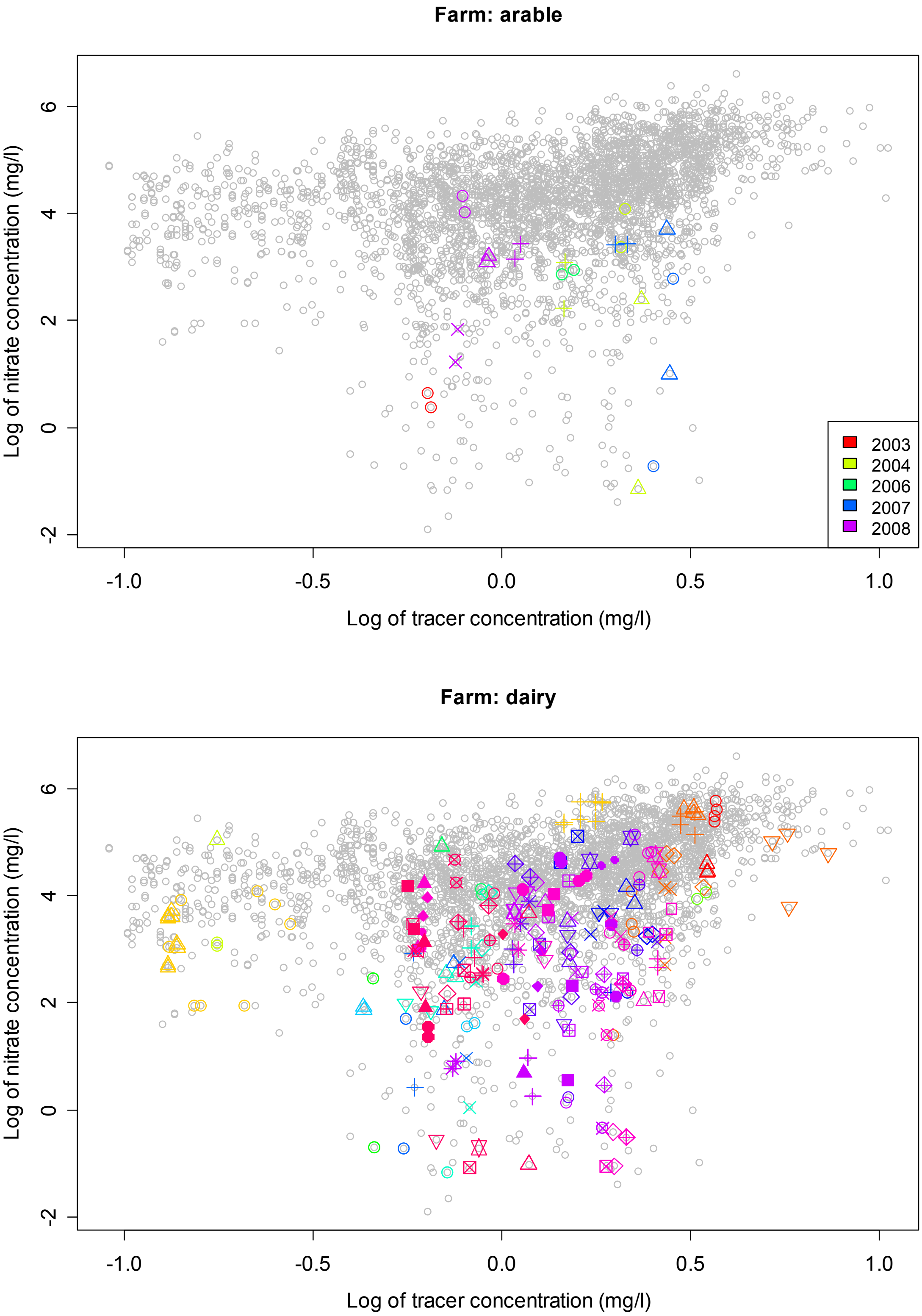


Figure 11 Scatter plots of the logarithms of nitrate and tracer concentrations (in grey) corresponding to net effects over one year periods that might be qualified as outliers (in various colours according to year).

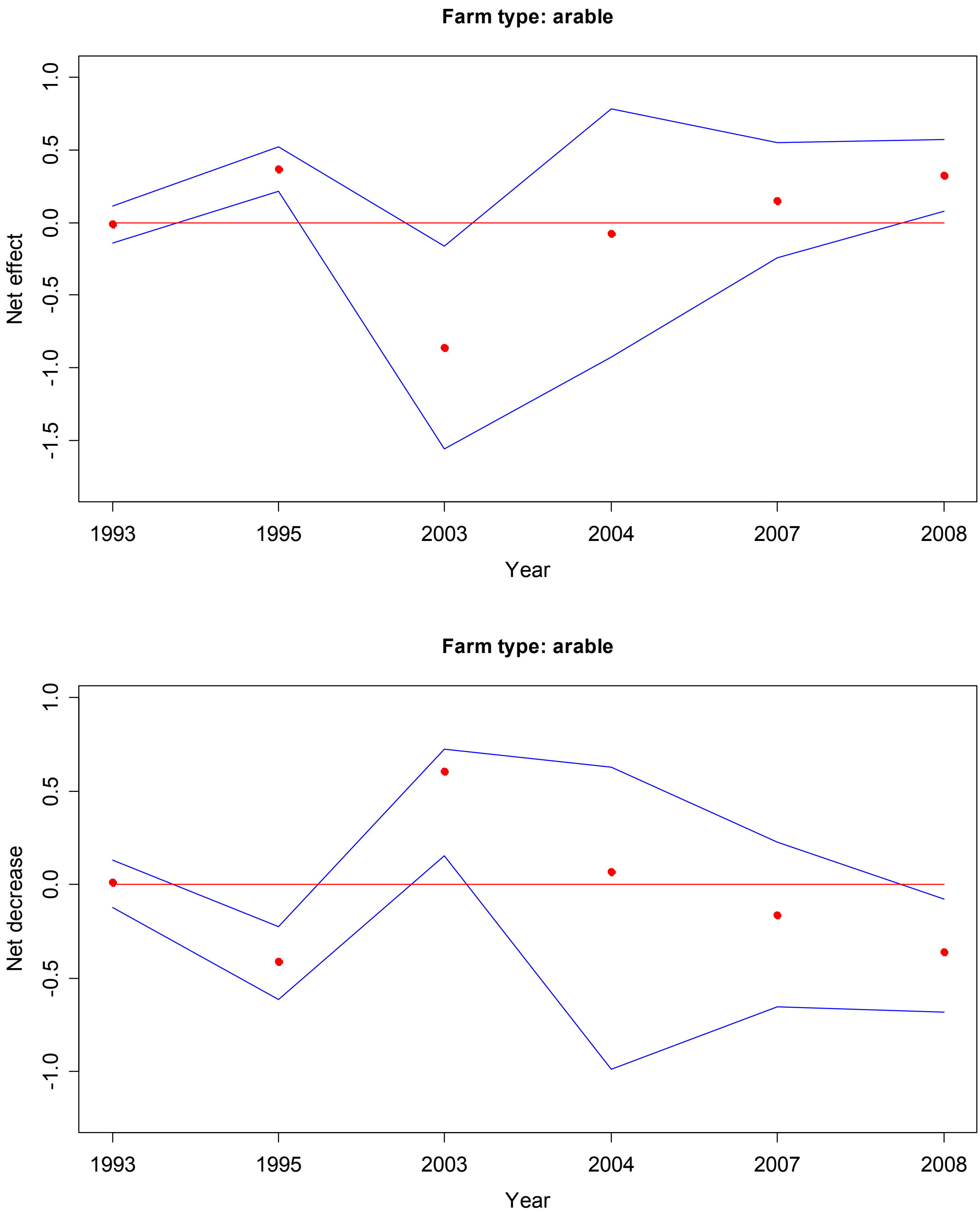


Figure 12A Point estimates (red solid circles) and approximate 95% confidence intervals (blue lines) of mean net effects and mean median net decreases over one-year periods.

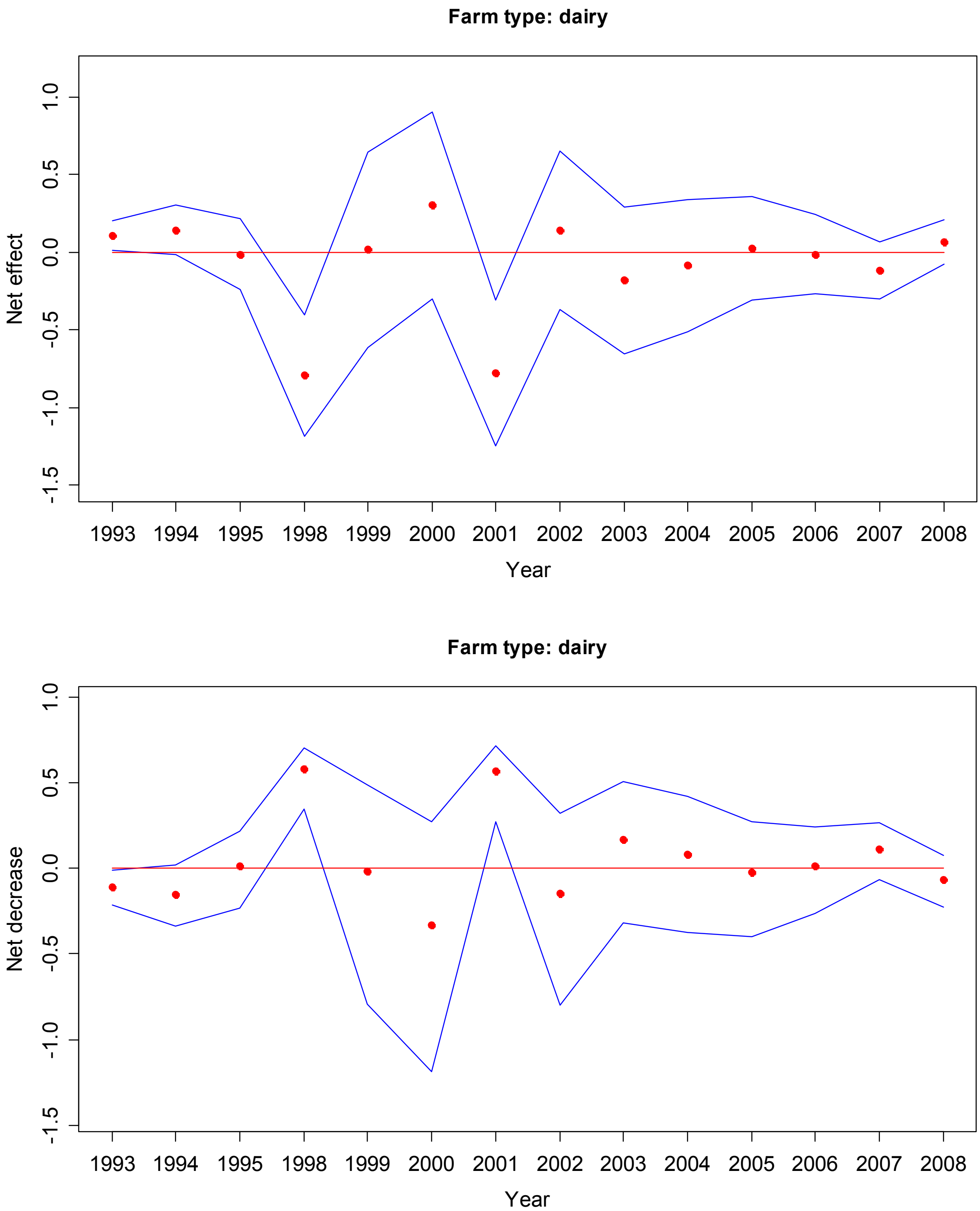
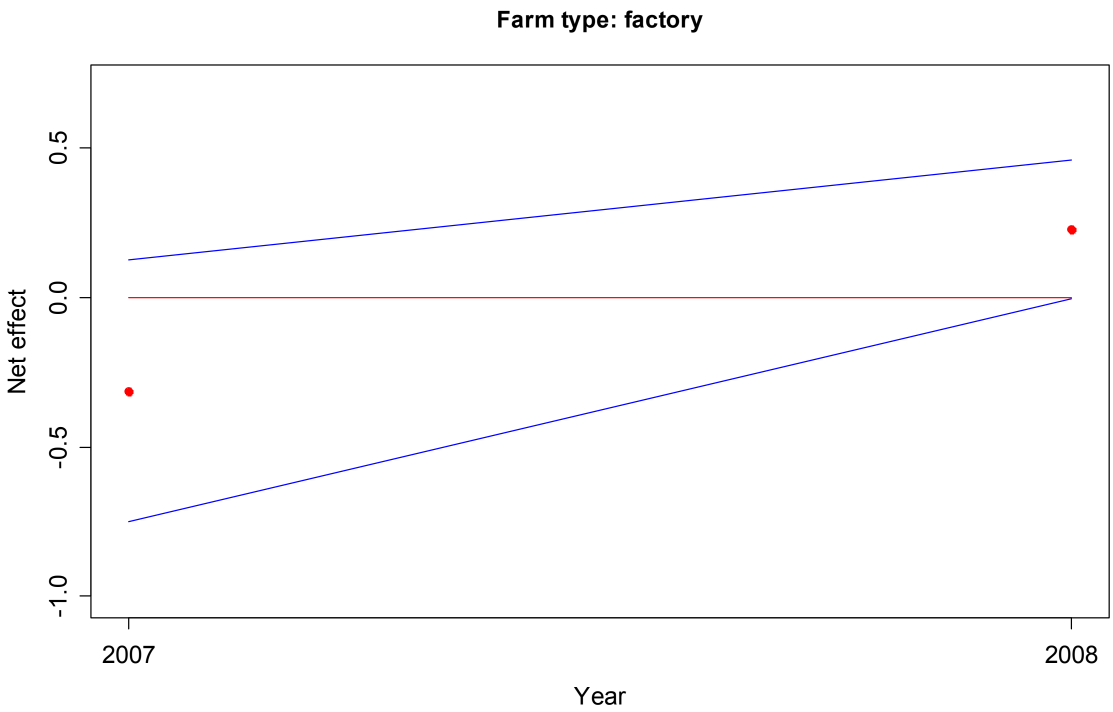


Figure 12B Point estimates (red solid circles) and approximate 95% confidence intervals (blue lines) of mean net effects and mean median net decreases over one-year periods.



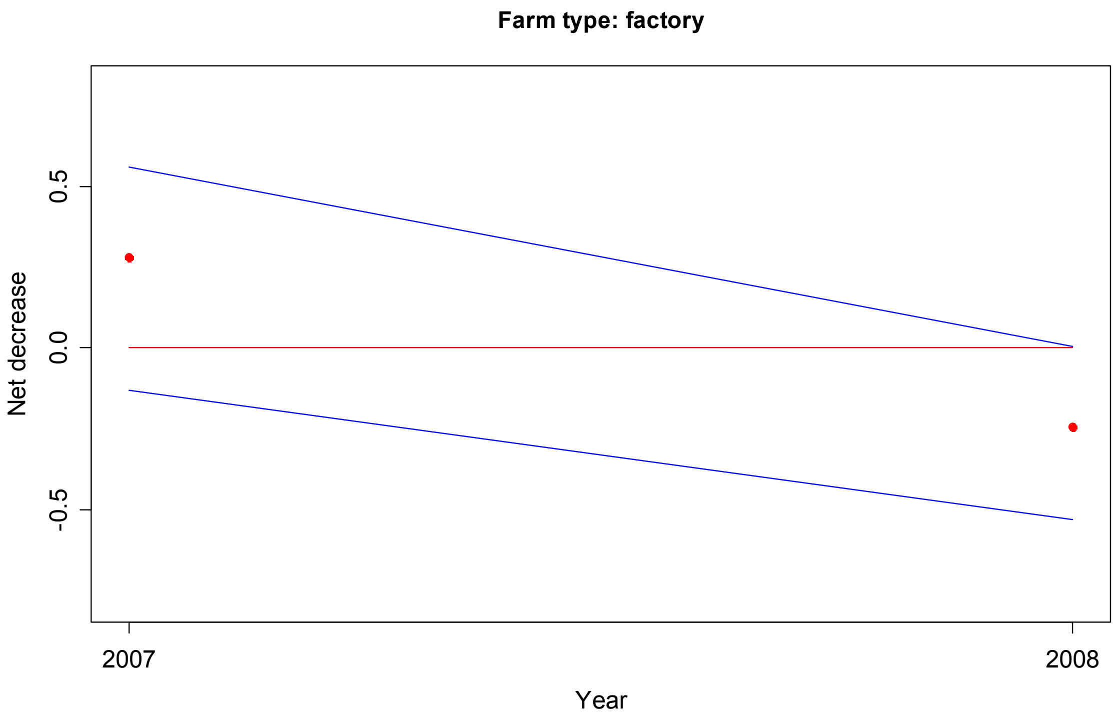
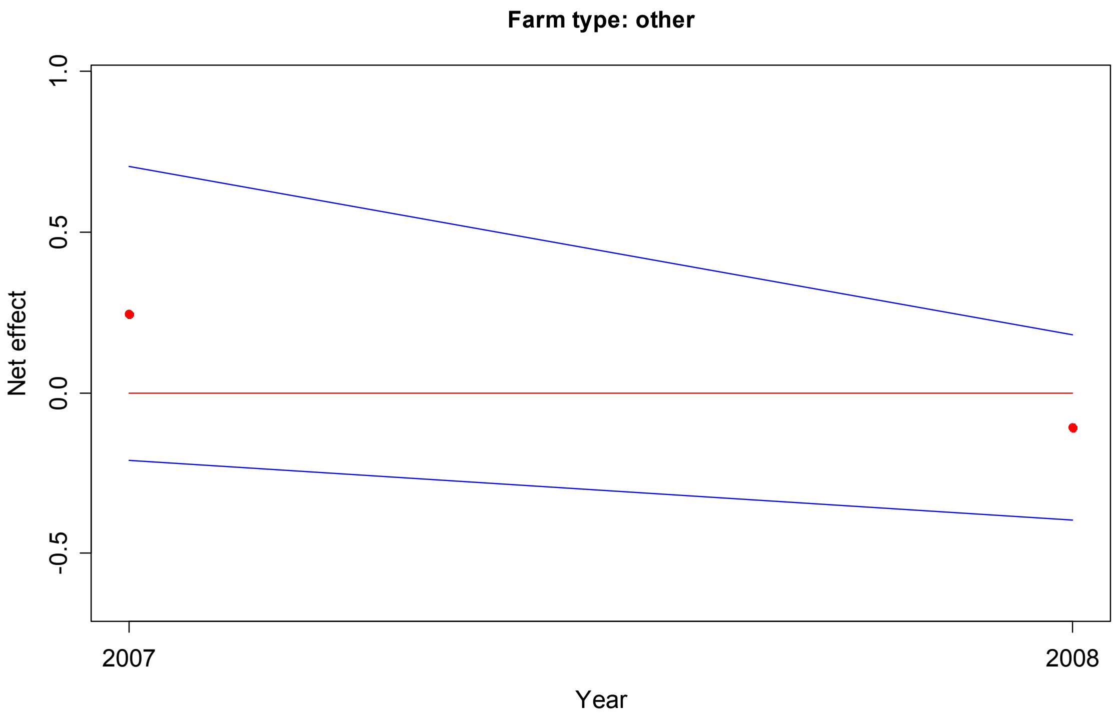


Figure 12C Point estimates (red solid circles) and approximate 95% confidence intervals (blue lines) of mean net effects and mean median net decreases over one-year periods.



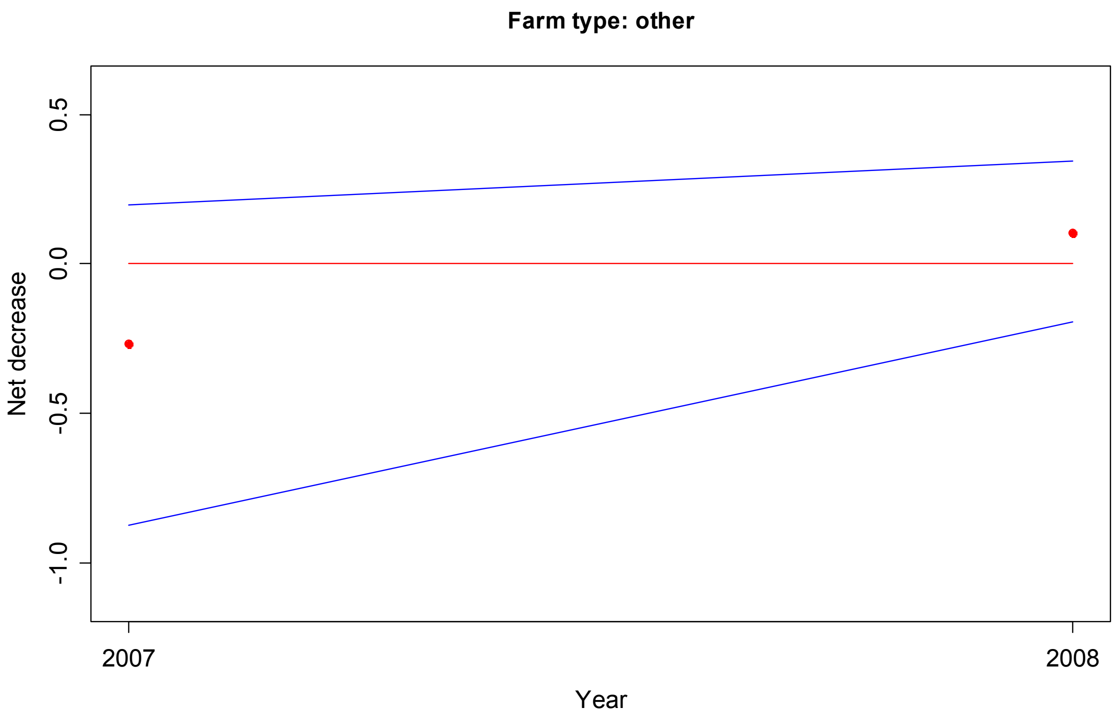


Figure 12D Point estimates (red solid circles) and approximate 95% confidence intervals (blue lines) of mean net effects and mean median net decreases over one-year periods.

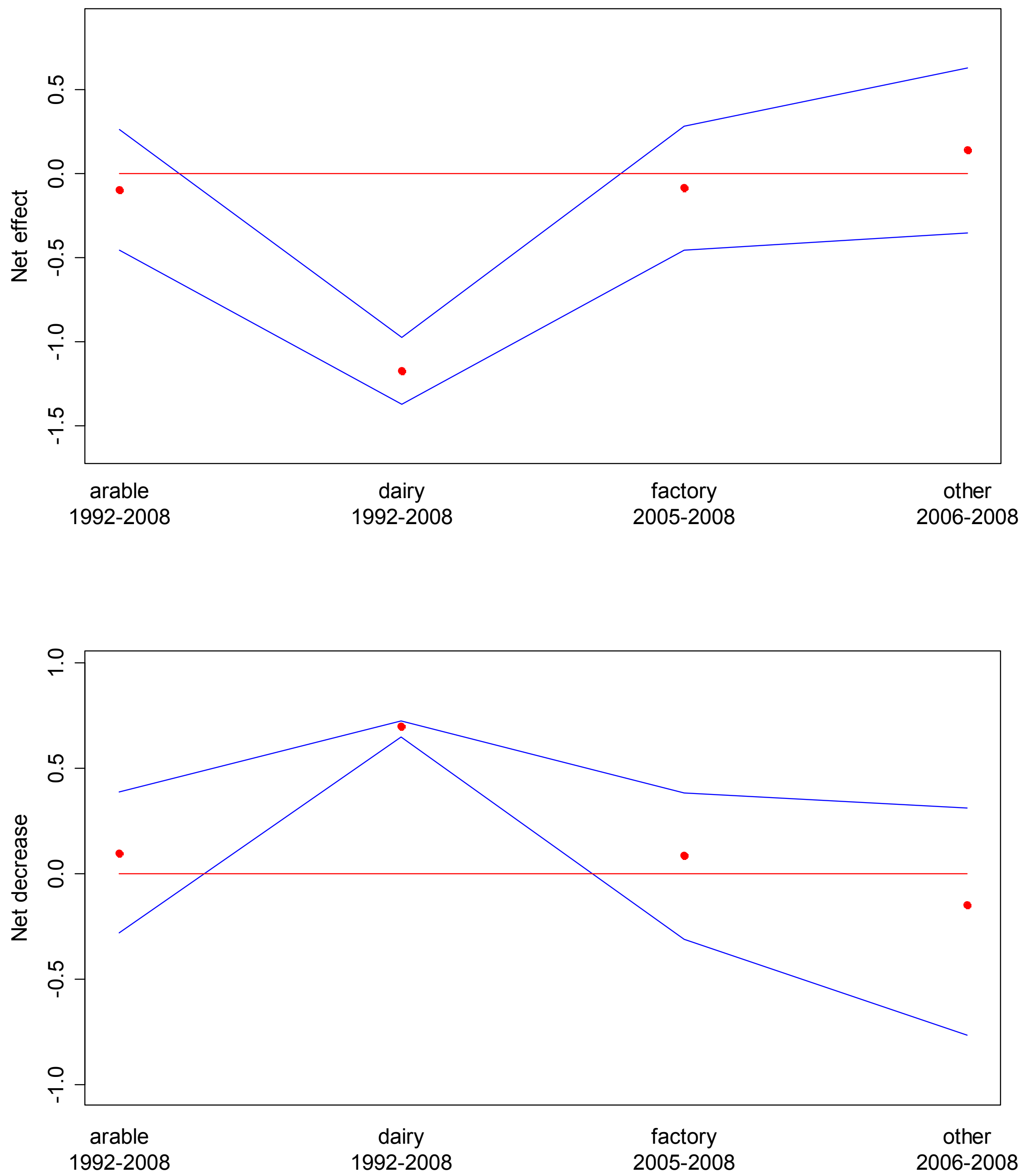


Figure 13 Point estimates (red solid circles) and approximate 95% confidence intervals (blue lines) of 'long-term' mean net effects and mean median net decreases.

Appendix A: Arguments for a reanalysis of the data of Boumans and Fraters

I have been asked by Leo Boumans to carry out a calculation to determine the sample size that would be required in order to detect and estimate a 'net decrease' (or 'compensated decrease' as Boumans and Fraters (2009) call it) in the average nitrate concentration in ground water per farm in the Netherlands. The *net* decrease refers to the putative positive difference, *accounting for variations in precipitation and in other variables known to influence the amount of nitrate in ground water*, between the average nitrate concentration within a time period in the past and the average nitrate concentration within a time period in the future. Ideally, the net decrease reflects the decrease in nitrate that is due solely to the introduction of new policies governing the application of nitrate in the Netherlands.

A sample size calculation is based on the probability of *producing evidence* that a certain proposition is true. For example, it may be true or not that "no more than 15% of RIVM's current employees are 'regular' smokers". In order to ascertain the truth of this proposition one would really have to interview all current employees. However, instead of carrying out a census, one might very well draw a random sample from RIVM's current employees and use that sample to *estimate* and find a *confidence interval* for the proportion of smokers. If the confidence interval turned out to be located below 15%, we would have found evidence for the truth of the proposition; and if it turned out to be above 15%, we would have found evidence that the proposition is false. However, in neither case would we be absolutely certain of having discovered the true status of the proposition: because of sampling variability, there is always a chance (though usually a small chance) that the evidence provided by the confidence interval is wrong. And, of course, there is still a third possibility: if the confidence interval turned out to include 15%, the proposition would be 'undecidable' on the basis of the particular sample drawn.

In order to avoid this third possibility one often carries out a *power* or *sample size* calculation. In essence, this calculation determines the sample size that is required to be able, with high probability or 'power', to produce evidence about the truth of a proposition in a future draw of a sample.

Unfortunately, a sample size calculation is not as miraculous as it may sound. Even in the very simple example now given, where we are interested in making statements about the value of a proportion and the underlying probability model is clearly specified, a *sharp* sample size calculation (one that would avoid sampling 'in excess') requires us to make *some assumptions* about the value of the proportion (which is of course unknown: it is what we want to estimate). In general, a sample size calculation requires us to consider *plausible scenarios based on an unknown parameter* (e.g. the parameter 'proportion of regular smokers' in our example) and in return it will give us an estimate of the probability that a sample of a given size will, under each such scenario, furnish evidence for the proposition.

It should be clear from the above that a sample size calculation requires *four main ingredients*. First, one needs to formulate a proposition or *hypothesis*; secondly, this hypothesis must refer to a given population parameter or set of parameters (i.e., population quantities such as 'proportion of regular smokers'); thirdly, there must be a well-defined procedure (specified prior to the examination of the sample) by which one takes the information contained in the sample and produces a statement or decision about the truth of the proposition; and finally, one must consider plausible scenarios describing what is ‘actually happening in reality’ and for each such scenario one must be able to compute the probability that the well-defined procedure will indeed provide the evidence required.

In their report, Boumans and Fraters (2009) provide evidence for a net decrease in the average nitrate concentration in the period of 1992 to 2007 on the basis of a statistical model that describes the nitrate concentration as a function of the tracer concentration and other variables known to influence the amount of nitrate in ground water. In my opinion, this statistical model has not been explicitly formulated by Boumans and Fraters (2009) in terms of the parameters and probability distributions involved; as a result, the conclusions drawn by the authors, at least as they are currently formulated, do not seem to be based on parameter estimates or on tests about parameters. Consequently, I am not able in particular to find a well-defined procedure, or a set of well-defined procedures, in the analysis of Boumans and Fraters (2009) which can be used to test the hypothesis of interest (whether there has been an average net decrease of nitrate concentration or not) in a future draw of the sample.

My opinion is thus that the model and analysis used by Boumans and Fraters (2009) are not sufficiently well defined to provide us with the basic ingredients for a sample size calculation.

On the other hand, I think that the model used by Boumans and Fraters (2009) has lots of parameters and makes what I think are questionable assumptions about the relationship between nitrate concentration and the other variables and about the probability distributions of the random effects. Consequently, even if the model is explicitly formulated, with lists of parameters and probability distributions, I have doubts about (a) whether the model is able to properly describe the data and thus produce realistic sample size determinations, and (b) whether the model allows the ‘plausible scenarios’ needed in sample size calculations to be specified in a meaningful way.

Appendix B: Derivation of the variance of the estimator of *β*(*s*,*t*) and of an estimator thereof

Let I_s and I_t denote the set of indices of the farms sampled at times *s* and *t*, so that $I_s \cap I_t$ is the set of indices of the farms sampled at both times, $I_s^{\,C} \cap I_t$ the set of indices of farms sampled at time *t* but not at time *s*, etc. Then the estimator $\hat{\beta}(s,t)$ of chapter 4 can be written as

$$\hat{\beta}(s,t) = \sum\nolimits_{i \in I_s \cap I_t} \frac{\Delta_i(t)}{\# I_t} - \frac{\Delta_i(s)}{\# I_s} +$$
$$\frac{\# I_s^{\,C} \cap I_t}{\# I_t} \left\{ \frac{1}{\# I_s^{\,C} \cap I_t} \sum\nolimits_{i \in I_s^{\,C} \cap I_t} \Delta_i(t) \right\} -$$
$$\frac{\# I_s \cap I_t^{\,C}}{\# I_s} \left\{ \frac{1}{\# I_s \cap I_t^{\,C}} \sum\nolimits_{i \in I_s \cap I_t^{\,C}} \Delta_i(s) \right\}$$

where as usual $\# I$ denotes the cardinal (the number of elements) of the set *I*. Because of the independence between observations from different farms, the variance of $\hat{\beta}(s,t)$ is therefore

$$Var\,\hat{\beta}(s,t) = Var\left\{ \sum\nolimits_{i \in I_s \cap I_t} \frac{\Delta_i(t)}{\# I_t} - \frac{\Delta_i(s)}{\# I_s} \right\} +$$
$$\left(\frac{\# I_s^{\,C} \cap I_t}{\# I_t} \right)^2 \frac{1}{(\# I_s^{\,C} \cap I_t)^2} Var\left\{ \sum\nolimits_{i \in I_s^{\,C} \cap I_t} \Delta_i(t) \right\} +$$
$$\left(\frac{\# I_s \cap I_t^{\,C}}{\# I_s} \right)^2 \frac{1}{(\# I_s \cap I_t^{\,C})^2} Var\left\{ \sum\nolimits_{i \in I_s \cap I_t^{\,C}} \Delta_i(s) \right\}.$$

This expression involves the variances of three sums. *If the summands involved in each of the three sums have the same variances* then the variance of the sum equals the number of summands times the variance of the summands and the latter can be estimated by the sample variance of the summands. Thus the variance of $\hat{\beta}(s,t)$ can be estimated by

$$v(\hat{\beta}(s,t)) = \#(I_s \cap I_t)\, v\left\{ \frac{\Delta_i(t)}{\# I_t} - \frac{\Delta_i(s)}{\# I_s} : i \in I_s \cap I_t \right\} +$$

$$\left(\frac{\# I_s^{\,C} \cap I_t}{\# I_t} \right)^2 \frac{v\{\Delta_i(t) : i \in I_s^{\,C} \cap I_t\}}{\# I_s^{\,C} \cap I_t} +$$

$$\left(\frac{\# I_s \cap I_t^{\,C}}{\# I_s} \right)^2 \frac{v\{\Delta_i(t) : i \in I_s \cap I_t^{\,C}\}}{\# I_s \cap I_t^{\,C}},$$

where $v\{X_i : i \in I\}$ denotes the (unbiased) sample variance of a sample $\{X_i : i \in I\}$.

In our case, a complication can arise when the summands are based on different 'sampling regimes' and hence are likely to have different variances: as the reader will recall from chapter 2, the nitrate and tracer observations we work with here are within-farm averages, so it can happen that $\Delta_i(t)/\# I_t - \Delta_i(s)/\# I_s$, say, is based on an average of four observations at time *s* and on eight observations at time *t* in certain farms, while it is based on an average of four observations at time *s* and on nine observations at time *t* in other farms (see Figure 4, where the years of 1993 and 1994 are seen to have different average within-farm sample sizes).

To deal with this situation, consider an arbitrary sum of variables Y_j with indices on a set *J* that can be decomposed into *p* disjoint subsets $J_1, \dots J_p$ representing farms with the same sampling regime, so that

$$\sum\nolimits_{j \in J} Y_j = \sum\nolimits_{j \in J_1} Y_j + \cdots + \sum\nolimits_{j \in J_p} Y_j$$

is a sum of independent sums each of which involves identically distributed summands. If a sum is based on at least two observations its variance can be estimated in an unbiased way by the number of summands times the sample variance of the summands, and the variance of $\sum\nolimits_{j \in J} Y_j$ can then be estimated by the sum of the individual estimates. If a sum has only one observation then its variance cannot be estimated, and in this case, we have taken as its variance the maximum of the other sample variances; this is a somewhat *ad hoc* procedure but it is applied in only a few cases and has no visible effect on the confidence intervals.

Appendix C: Calculation of the formula of power needed in chapter 6

Case I: all farms are sampled at both times

Let N be the total number of 'candidate farms' of a given type in the Netherlands. In order to simplify the notation, denote the mean net effect at farm i by μ_i. The mean net effect in farms of that type is then

$$\mu = \frac{1}{N}\sum_{i=1}^{N}\mu_i$$

and the (population) variance of the mean net effect is

$$\Sigma_1^2 = \frac{1}{N}\sum_{i=1}^{N}(\mu_i - \mu)^2 ,$$

which will be referred to as the **between-farms variance**.

Next, suppose that within each farm we measure the net effect at p randomly chosen wells and that each of the p mean net effects measured at farm i has variance σ_i^2, and put

$$\Sigma_2^2 = \frac{1}{N}\sum_{i=1}^{N}\sigma_i^2 ,$$

which will be referred to as the (average) **within-farms variance**.

In order to estimate the mean net effect μ one draws randomly n farms from the population of N candidate farms in the Netherlands, computes the net effect from p wells in each of these farms, computes the average of each such set of p mean net effects, and then computes an estimate of μ as the average of all the n averages. This estimate, which is denoted by $\hat{\mu}$, is unbiased and its variance satisfies

$$Var(\hat{\mu}) \approx \frac{1}{n}\left(\Sigma_1^2 + \frac{1}{p}\Sigma_2^2\right). \qquad \textbf{(c.1)}$$

To prove **(c.1)**, denote by $X_{i,j}$ the j-th net effect at farm i ($i = 1,\ldots,N$, $j = 1,\ldots,p$). By assumption, $\mu_i = EX_{i,j}$, $\mu = N^{-1}\sum_{i=1}^{N}\mu_i$, $\sigma_i^2 = VarX_{i,j}$, $\Sigma_1^2 = N^{-1}\sum_{i=1}^{N}(\mu_i - \mu)^2$ and

$\Sigma_2^2 = N^{-1}\sum_{i=1}^{N}\sigma_i^2$. Let $I_1,\ldots,I_n$ denote the indices of the *n* farms represented in the sample drawn randomly without replacement from the *N* farms; by definition of random sample (with or without replacement) $P(I_i = k) = 1/N$ for all *i* and all *k*. The estimate of μ is

$$\hat{\mu} = n^{-1}\sum_{i=1}^{n}\hat{\mu}_{I_i},$$

where

$$\hat{\mu}_i = p^{-1}\sum_{j=1}^{p}X_{i,j}.$$

Clearly, $E(\hat{\mu}_{I_i}) = \mu$ for all *i*, so $E(\hat{\mu}) = \mu$ (i.e., $\hat{\mu}$ is unbiased). Moreover,

$$\begin{aligned}
Var(\hat{\mu}_{I_i}) &= E(\hat{\mu}_{I_i}^2) - E^2(\hat{\mu}_{I_i}) = \frac{1}{N}\sum_{i=1}^{N}E(\hat{\mu}_i^2) - \mu^2 \\
&= \frac{1}{N}\sum_{i=1}^{N}\{Var(\hat{\mu}_i) + E^2(\hat{\mu}_i)\} - \mu^2 = \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{\sigma_i^2}{p} + \mu_i^2\right\} - \mu^2 \\
&= \Sigma_1^2 + \frac{1}{p}\Sigma_2^2,
\end{aligned}$$

and since the correlation between $\hat{\mu}_{I_i}$ and $\hat{\mu}_{I_j}$ ($i \neq j$) is very small (because *N* is large)

$$Var(\hat{\mu}) \approx \frac{1}{n^2}\sum_{i=1}^{n}Var(\hat{\mu}_{I_i}) = \frac{1}{n}\left(\Sigma_1^2 + \frac{1}{p}\Sigma_2^2\right).$$

Remark: A 'finite sample size correction' is obtained by noting that

$$Cov(\hat{\mu}_{I_i}, \hat{\mu}_{I_j}) = -\frac{1}{N-1}\Sigma_1^2$$

for $i \neq j$, which yields

$$\begin{aligned}
Var(\hat{\mu}) &= \frac{1}{n^2}\sum_{i=1}^{n}Var(\hat{\mu}_{I_i}) + \frac{1}{n^2}\sum_{i\neq j}Cov(\hat{\mu}_{I_i}, \hat{\mu}_{I_j}) \\
&= \frac{1}{n}\left(\Sigma_1^2 + \frac{1}{p}\Sigma_2^2\right) - \frac{(n-1)}{n}\frac{1}{(N-1)}\Sigma_1^2.
\end{aligned}$$

It remains to show how estimates of the between- and within-farms variances Σ_1^2 and Σ_2^2 can be obtained from the data.

We first note (see the proof at the end of this appendix) that $(n-1)^{-1}\sum_{i=1}^{n}(\hat{\mu}_{I_i}-\hat{\mu})^2$ is an approximately unbiased estimator of the variance of $\hat{\mu}$ when this estimator is based on one farm (i.e. when $n=1$):

$$E\frac{1}{n-1}\sum_{i=1}^{n}(\hat{\mu}_{I_i}-\hat{\mu})^2 \approx \left\{\Sigma_1^2+\frac{1}{p}\Sigma_2^2\right\}. \qquad \textbf{(c.2)}$$

Thus, assuming that an unbiased estimator $\hat{\Sigma}_2^2$ of Σ_2^2 is available, one can estimate Σ_1^2 by

$$\hat{\Sigma}_1^2 := \frac{1}{n-1}\sum_{i=1}^{n}(\hat{\mu}_{I_i}-\hat{\mu})^2-\frac{1}{p}\hat{\Sigma}_2^2 .$$

The variance at a randomly chosen farm I_i, namely $\sigma_{I_i}^2$, can be estimated as usual by

$$\hat{\sigma}_{I_i}^2 = \frac{1}{p-1}\sum_{j=1}^{p}(X_{I_i,j}-\hat{\mu}_{I_i})^2 .$$

Since $E\hat{\sigma}_{I_i}^2 = \frac{1}{N}\sum_{k=1}^{N}E[\hat{\sigma}_k^2 \mid I_i=k] = \frac{1}{N}\sum_{k=1}^{N}\sigma_k^2 = \Sigma_2^2$,

$$\hat{\Sigma}_2^2 = \frac{1}{n}\sum_{i=1}^{n}\hat{\sigma}_{I_i}^2$$

is an unbiased estimator of Σ_2^2 based on *n* randomly sampled farms. The problem with this estimator, however, is that the *j*-th net effect at farm *i* is of the form

$$\begin{aligned}X_{i,j} &= \{\log N_{ij}(t)-\log N_{ij}(s)\}-\{\log T_{ij}(t)-\log T_{ij}(s)\}\\ &= \{\log N_{ij}(t)-\log T_{ij}(t)\}-\{\log N_{ij}(s)-\log T_{ij}(s)\}\\ &=: D_{ij}(t)-D_{ij}(s),\end{aligned}$$

and unfortunately *we are not able to compute this from the data*. Indeed, we do not know, for a fixed farm *i*, whether the measurements at time *t* are in some way coupled to those at time *s* (whether they come from the same wells, for instance), and, if they are coupled, which measurement (which index *j*) at time *t* corresponds to which measurement at time *s*.[8] For this reason, we will have to assume that the two differences $D_{ij}(s)$ and $D_{ij}(t)$ are uncorrelated, so that, in obvious notation,

[8] Since $\hat{\mu}_i = p^{-1}\sum_{j=1}^{p}D_{i,j}(t)-p^{-1}\sum_{j=1}^{p}D_{i,j}(s)$, and since these two averages *can* be computed from the data, this problem does not occur in the calculation of $(n-1)^{-1}\sum_{i=1}^{n}(\hat{\mu}_{I_i}-\hat{\mu})^2$.

$$\sigma_i^2 = Var X_{ij} = Var D_{ij}(t) + Var D_{ij}(s) = \sigma_i^2(s) + \sigma_i^2(t)$$

and

$$\Sigma_2^2 = \frac{1}{N}\sum_{i=1}^{N} \sigma_i^2 = \Sigma_2^2(s) + \Sigma_2^2(t).$$

In order to estimate each of the two variances involved here we then take (for $\tau = s,\ t$)

$$\hat{\Sigma}_2^2(\tau) = \frac{1}{n}\sum_{i=1}^{n} \hat{\sigma}_{I_i}^2(\tau),$$

with

$$\hat{\sigma}_{I_i}^2(\tau) = \frac{1}{p-1}\sum_{j=1}^{p} (D_{I_i,j}(\tau) - \overline{D}_{I_i}(\tau))^2, \qquad \overline{D}_{I_i}(\tau) = \frac{1}{p}\sum_{j=1}^{p} D_{I_i,j}(\tau).$$

Remark: If $D_{ij}(s)$ and $D_{ij}(t)$ are correlated, they are likely to be *positively* correlated; and if that is the case then our estimates of power will be somewhat conservative (i.e., the true powers will be somewhat larger than our estimates).

Proof of (c.2)

We have

$$E[\hat{\mu}_{I_i}{}^2] = \frac{1}{N}\sum_{j=1}^{N} E[\hat{\mu}_{I_i}{}^2 \mid I_i = j] = \frac{1}{N}\sum_{j=1}^{N} E[\hat{\mu}_j{}^2] = \frac{1}{N}\sum_{j=1}^{N} Var[\hat{\mu}_j] + E^2[\hat{\mu}_j]$$

$$= \frac{1}{N}\sum_{j=1}^{N} \frac{\sigma_j^2}{p} + \frac{1}{N}\sum_{j=1}^{N} \mu_j^2 = \frac{1}{p}\Sigma_2^2 + \frac{1}{N}\sum_{j=1}^{N} (\mu_j - \mu)^2 + \mu^2$$

$$= \frac{1}{p}\Sigma_2^2 + \Sigma_1^2 + \mu^2,$$

$$E\hat{\mu}^2 = Var\hat{\mu} + E^2(\hat{\mu}) \approx \frac{1}{n^2}\sum_{i=1}^{n} Var\hat{\mu}_{I_i} + \mu^2$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} E(\hat{\mu}_{I_i}{}^2) - \frac{1}{n^2}\sum_{i=1}^{n} E^2(\hat{\mu}_{I_i}) + \mu^2$$

$$= \frac{1}{n}\left\{\frac{1}{p}\Sigma_2^2 + \Sigma_1^2 + \mu^2\right\} - \frac{1}{n}\mu^2 + \mu^2,$$

so

$$E\frac{1}{n}\sum_{i=1}^{n}(\hat{\mu}_{I_i}-\hat{\mu})^2=\frac{1}{n}\sum_{i=1}^{n}E[\hat{\mu}_{I_i}{}^2]-E[\hat{\mu}^2]\approx\left\{\frac{1}{p}\Sigma_2^2+\Sigma_1^2+\mu^2\right\}-$$
$$\frac{1}{n}\left\{\frac{1}{p}\Sigma_2^2+\Sigma_1^2+\mu^2\right\}-\frac{(n-1)}{n}\mu^2$$
$$=\frac{(n-1)}{n}\left\{\frac{1}{p}\Sigma_2^2+\Sigma_1^2\right\}.$$

Thus

$$E\frac{1}{n-1}\sum_{i=1}^{n}(\hat{\mu}_{I_i}-\hat{\mu})^2\approx\left\{\Sigma_1^2+\frac{1}{p}\Sigma_2^2\right\}.$$

Case II: different farms are sampled at different times

In this case, the estimator of $\beta(s,t):=N^{-1}\sum_{i=1}^{N}\beta_i(s,t)$ is

$$\hat{\beta}(s,t)=\frac{1}{n}\sum_{i=1}^{n}\Delta_i(t)-\frac{1}{n}\sum_{i=1}^{n}\Delta_i(s)=:\hat{\beta}^{(t)}-\hat{\beta}^{(s)},$$

where $\Delta_i(\tau)=p^{-1}\sum_{j=1}^{p}\{\log N_{i,j}(\tau)-\log T_{i,j}(\tau)\}$ is the aggregated difference between log nitrate and log tracer at time τ, *p* is the number of measurements per farm, *n* is the number of farms to be sampled, and *the* $\Delta_i(s)$ *s are independent of the* $\Delta_i(t)$ *s*. The variance of $\hat{\beta}(s,t)$ is then $Var\{\hat{\beta}^{(t)}-\hat{\beta}^{(s)}\}=Var\{\hat{\beta}^{(t)}\}+Var\{\hat{\beta}^{(s)}\}$, and in order to estimate this we only need to apply the results of **Case I** separately to the estimators $\hat{\beta}^{(s)}$ and $\hat{\beta}^{(t)}$.

More precisely, as before let *N* denote the total number of 'candidate farms' of a given type in the Netherlands and let $\mu_i(\tau)$ be the mean of $D_{ij}(\tau)=\log N_{i,j}(\tau)-\log T_{i,j}(\tau)$, a variable computed with data from farm *i* at time τ. Similarly to **Case I**, we set

$$\mu(\tau)=\frac{1}{N}\sum_{i=1}^{N}\mu_i(\tau)$$

$$\Sigma_1^2(\tau)=\frac{1}{N}\sum_{i=1}^{N}(\mu_i(\tau)-\mu(\tau))^2 \quad\text{and}\quad \Sigma_2^2(\tau)=\frac{1}{N}\sum_{i=1}^{N}\sigma_i^2(\tau),$$

where $\sigma_i^2(\tau)$ is the variance of $\log N_{i,j}(\tau)-\log T_{i,j}(\tau)$. In order to estimate $\mu(\tau)$ we take

$$\hat{\mu}(\tau) = \frac{1}{n}\sum_{i=1}^{n} \Delta_{I_i}(\tau),$$

where $I_1, \ldots, I_n$ are as defined in **Case I** (note that $\hat{\mu}(\tau)$ equals $\hat{\beta}^{(s)}$ or $\hat{\beta}^{(t)}$ when $\tau = s$ or $\tau = t$). This estimator is unbiased and its variance, just as before, satisfies

$$Var(\hat{\mu}(\tau)) \approx \frac{1}{n}\left(\Sigma_1^2(\tau) + \frac{1}{p}\Sigma_2^2(\tau) \right).$$

To estimate $\Sigma_1^2(\tau)$ and $\Sigma_2^2(\tau)$, we do exactly as in **Case I**. First, we make use of the relation

$$E\frac{1}{n-1}\sum_{i=1}^{n} (\Delta_{I_i}(\tau) - \hat{\mu}(\tau))^2 \approx \left\{ \Sigma_1^2(\tau) + \frac{1}{p}\Sigma_2^2(\tau) \right\}.$$

Assuming that an unbiased estimator $\hat{\Sigma}_2^2(\tau)$ and $\Sigma_2^2(\tau)$ is available, one can thus estimate $\Sigma_1^2(\tau)$ by

$$\hat{\Sigma}_1^2(\tau) := \frac{1}{n-1}\sum_{i=1}^{n} (\Delta_{I_i}(\tau) - \hat{\mu}(\tau))^2 - \frac{1}{p}\hat{\Sigma}_2^2(\tau).$$

Secondly, $\sigma_{I_i}^2(\tau)$ is estimated by

$$\hat{\sigma}_{I_i}^2(\tau) = \frac{1}{p-1}\sum_{j=1}^{p} (D_{I_i,j}(\tau) - \Delta_{I_i}(\tau))^2 \ .$$

which yields

$$\hat{\Sigma}_2^2(\tau) = \frac{1}{n}\sum_{i=1}^{n} \hat{\sigma}_{I_i}^2(\tau),$$

the required unbiased estimator of $\Sigma_2^2(\tau)$.

To sum up, we have

$$\begin{aligned} Var\{\hat{\beta}^{(t)} - \hat{\beta}^{(s)}\} &= Var\{\hat{\beta}^{(t)}\} + Var\{\hat{\beta}^{(s)}\} \\ &\approx \frac{1}{n}\left(\Sigma_1^2(s) + \frac{1}{p}\Sigma_2^2(s) \right) + \frac{1}{n}\left(\Sigma_1^2(t) + \frac{1}{p}\Sigma_2^2(t) \right) \\ &= \frac{1}{n}\left(\left\{ \Sigma_1^2(s) + \Sigma_1^2(t) \right\} + \frac{1}{p}\left\{ \Sigma_2^2(s) + \Sigma_2^2(t) \right\} \right) \\ &=: \frac{1}{n}\left(\Sigma_1^2 + \frac{1}{p}\Sigma_2^2 \right), \end{aligned}$$

and this can be estimated by

$$v\{\hat{\beta}^{(t)} - \hat{\beta}^{(s)}\} = \frac{1}{n}\left(\left\{\hat{\Sigma}_1^2(s) + \hat{\Sigma}_1^2(t)\right\} + \frac{1}{p}\left\{\hat{\Sigma}_2^2(s) + \hat{\Sigma}_2^2(t)\right\}\right) =: \frac{1}{n}\left(\hat{\Sigma}_1^2 + \frac{1}{p}\hat{\Sigma}_2^2\right).$$

It is the latter estimate that needs to be introduced in equation **(6)** in place of $\frac{1}{n}\left(\Sigma_1^2 + \frac{1}{p}\Sigma_2^2\right)$.

rivm

Appendix D: Preliminary version of the report of Boumans and Fraters

A legislation induced decrease in nitrate leaching in the sandy areas of the Netherlands during the 1992–2006 period

L.J.M. Boumans and B. Fraters

groundwater quality trend, mixed model, REML, Nitrates Directive, Water Framework Directive. agriculture, nitrogen appliance, nitrogen surplus, monitoring

Abstract

The Netherlands has, since 1987, turned the increase of nitrogen appliance in Dutch agriculture into a decrease because of legislation. Monitoring efforts were started at farms in 1992 to visualise the effect of this decrease in nitrogen appliances. Changes in; weather conditions (precipitation surplus), in the assembly of monitored farms and national area per farm type during the monitoring period, complicate the visualisation of this effect. A statistical method, REML or residual maximum likelihood, compensating for these complications, is used for the visualisation. It is noted that since 1992 nitrate leaching has decreased by about 50%.

D1. Introduction

D.1.1 EU legislation and monitoring

The European "Directive concerning the protection of waters against pollution caused by nitrates from agricultural sources" (the Nitrates Directive; EU, 1991) aims at reducing existing nitrate levels in groundwater and surface water and at preventing further pollution. Member states are obliged to establish Action Programmes every four years and to monitor and evaluate their effectiveness. For this reason, the agricultural appliance of nitrogen (N) and nitrate leaching are monitored in the Netherlands and in other European countries (Fraters et al., 2005). The Netherlands is unique in monitoring nitrate leaching by sampling on-farm water, the upper metre of groundwater, throughout the country (Fraters et al., 1998). This on-farm water will show the clearest and quickest response to the Action Programme measures. The effect of measures from the first Action Programme (1996–2000) is expected in the on-farm water quality of the Sand region between 1998 and 2002. Groundwater quality of phreatic aquifers at a depth of more than 5 m below groundwater level will only show the effects of the measures after one or more decades (Broers, 2004). Moreover, these effects will be hard to detect due to the mixing of groundwater of different ages and origins as well as soil physical-chemical processes (Wendland et al., 2004; Cherry et al., 2008; Fraters et al., 2006). The Sand region (Figure D1) is the largest and the most vulnerable region in the Netherlands for nitrate leaching to groundwater and fresh surface waters. The monitoring programme is applied to about 40% of this area, which is in use for professional agricultural activities. Farms are selected per farm type for monitoring on-farm groundwater. Four farm types are distinguished, being arable, factory, dairy and others.
In this article the nitrate leaching concentration is synonymous with the nitrate concentration in the uppermost metre of groundwater.

D.1.2 Dutch legislation

In 1987, before the adoption of the Nitrates Directive in 1991, the increasing trend of nitrogen surpluses in agriculture in the Netherlands changed into a decreasing trend. This change in 1987 is attributed to the introduction of milk quota system in the European Union in 1984 and a systematic reduction of the quotas in subsequent years. (Zwart et al., 2008). In other EU Member States a trend reversal also occurred in the same period (Austria, Schwaiger, K., 2005; Germany and Europe, Wolter and Mohaupt, 2005; Denmark, Grant and Blicher-Mathiesen, 2004). In 1998, a Minerals Accounting System (Ondersteijn et al., 2002) was introduced in the Netherlands to regulate the appliance of both fertiliser and manure nitrogen through loss standards for farms with more than 2.5 livestock units per ha. In the 1998–2002 period, loss standards were tightened and stricter standards were introduced for well-drained soils prone to nitrate leaching (about 7% of the agricultural area). Loss standards for nitrogen were 14–43% lower in 2002 compared to 1998.

In addition to the milk quota and tightening of N loss standards, the application of N has been prohibited between the beginning of September and 1 February since 1996. To decrease overland flow of minerals to surface water, the application of manure has been prohibited to snow-covered ground since 1994. In 1998, this was extended to frozen or partly frozen ground. The application of manure and artificial fertiliser to water-saturated, flooded, frozen or snow-covered ground has been prohibited since 1999.

To minimise ammonia volatisation and overland flow, it is obligatory to cover the applied manure with soil (Zwart et al., 2008). The atmospheric ammonia emissions from agricultural sources into the atmosphere decreased by about 45% between 1990 and 2007. This decrease in atmospheric N emissions will increase nitrate leaching in agriculture.

Milk quota, N loss standards and regulations for period and method of N appliance have resulted in a decrease in the amount of N applied (Van den Ham et al., 2007). The decrease in annual N appliance in the Sand region of the Netherlands during the period 1990–2005 is shown in Figure D2.
The area of land used for agriculture decreased by 4.3% and the number of farms by 30% in the 1992–2006 period. The number of cattle and pigs both decreased by 21%, the number of poultry decreased by 3.6%. Manure nitrogen and phosphorus production by livestock decreased by 30% and 27%, respectively, due to a combination of the decrease in number of livestock and excretion per head as a consequence of lower nitrogen and phosphorous contents in fodder and an improved fodder conversion. As a result, the N surplus and phosphorus surplus in Dutch agriculture decreased, respectively, by 36% and 47% (Zwart et al., 2008).

D.1.3 Legislation and other effects on nitrate leaching

We assume that a decrease in the amount of N appliance is the best indicator for legislation effects on agriculture. It is also assumed that a decrease in the amount of N appliance that is accompanied by a decrease in the nitrate leaching concentration in on-farm water is the best indicator for legislation induced agricultural effects on the environment. These assumptions can be criticised.

It is well known that there is less leaching (more denitrification) in the case of permanent crops like grass than in temporary crops like maize. There is also more nitrate leaching in the case of more grazing (Ryden et al., 1984), especially grazing in late summer and autumn (Sauer and Harrach, 1996). It is therefore possible to decrease nitrate leaching without decreasing the amount of N appliance. If the loss of N is statutorily decreased, in practice, farmers will have to decrease the amount of N appliance and manage their remaining N more efficiently. An efficient use of N also influences the period and method of N appliance and grazing. It is difficult to disentangle legislation effects from farm management effects. For simplicity and because there was no reason to do otherwise, we assume that all changes in farm management in the period of decreasing amounts of N appliance are induced by policy. This simplicity is also mentioned by Oenema et al., 1998: “The main cause of the increased nitrogen losses from agriculture and increased nitrate concentrations in groundwater is the fact that the inputs of fertilisers and animal manures to agricultural land have increased much more than the output of nitrogen in harvested products”. The Nitrates Directive also sets a maximum value for N application and not for N losses.

Besides the amount of N appliance on a farm, it is well known that nitrate leaching is also affected by the precipitation surplus (Boumans et al., 2005), soil type and drainage or groundwater level (D’Heane et al., 2003). These effects are related to dilution and denitrification of leaching nitrate. There are also possible environmental effects on nitrate leaching. It can be imagined that global warming, increasing atmospheric carbon dioxide content and more rain increase crop growth and affects nitrogen volatisation, denitrification and changes in the soil N pool (Jensen and Veihe, 2009). These changes in turn also affect nitrate leaching (Schröder et al., 2003). Measured nitrate leaching concentrations are also affected by the monitoring itself (methods, assembly of farms, etc.).

Summarising, the following effects on measured nitrate leaching concentrations are discerned:

1. Legislation effect on agriculture; a decreasing amount of N appliance;
2. Farm management effects to use the remaining N more efficiently;
3. Well-known environmental effects;
4. Possible environmental effects;
5. Monitoring effects.

D.1.4 Visualisation of effects of Action Programmes for the EC

Fraters et al. (2005) discern two approaches used in Europe for visualising the effect of the Action Programme, being 'up-scaling' and 'interpolation'. The up-scaling approach is used in Denmark, Sweden and the United Kingdom. Also, for a non-member EU state, Switzerland's effects of legislation are visualised with the up-scaling approach (Decrem, 2007). Empirical and deterministic models together with data on national changes in agricultural practice are used to upscale the results of experimental sites. Børgesen et al., 2001 use a statistical program for up-scaling the results of a deterministic simulation model to estimate the effects of legislation on nitrate leaching. Grant et al. (2006) visualise the effect on nitrate leaching in Denmark with empirical and deterministic mass-balance models. Models are tuned with the aid of field experiments and by monitoring seven small agricultural catchments, which are located throughout Denmark in order to represent variation in soil type, rainfall and agricultural practice. Hoffmann et al. (2000) visualise nitrogen leaching in Swedish agriculture for the 1865–1985 period, using a deterministic model and taking changes in soil organic nitrogen into account. Wolf et al. (2004) describe a deterministic model, STONE, which also calculates nitrate leaching at the regional and national scale in the Netherlands. This STONE model is used to evaluate effects of legislation but is not used to report to the EC. Lord and Anthony (2000) describe a national agri-environmental database and nitrate modelling system to support the United Kingdom government's nitrate policy development.

The interpolation method is used by Austria, Belgium, Germany, Ireland and the Netherlands. Monitoring data on nitrate leaching from a random sample of locations, e.g., farms and monitoring wells are used, instead of results from experimental sites. A statistical (interpolation) model and national scale monitored changes in agricultural practice are used to visualise the effect of the Action Programme (Fraters et al., 2005)

The up-scaling models are made for crop-soil type combinations. Our monitoring is not performed on individual crop-soil type combinations but on a farm scale. Groundwater samples of different crop-soil type combinations at randomly selected locations are mixed. Therefore, crop-soil type models cannot be used directly to model our monitoring results. Besides this problem for using these up-scaling crop-soil type models for monitoring data on a farm scale, there is also the problem of extrapolating model parameters, such as a mineralisation coefficient, from laboratory or experimental field scale to a national scale. These points are remarked by Wolf et al. (2005): "testing of a large-scale model, like STONE, on measured data from field experiments can hardly be expected to be satisfactory and second, calibration of a large-scale model on well-managed experiments may be wrong for practical applications." The problem of up-scaling for deterministic agro environmental models is also noted by Diekkrüger et al. (1995) and Boesten (2000). They both conclude that the modellers subjectivity varies so much that it overrules conceptual differences between (up-scaling) models in many cases. Input requirements for interpolation models are smaller than for up-scaling models and can be obtained relatively easily from increasingly computerised farm records (Cherry et al., 2008). Nationwide representative sampling of nitrate leaching in on-farm water (Fraters et al., 2005) instead of field or laboratory experiments, allows for an interpolation approach with statistical methods to visualise a decrease in nitrate leaching. Statistical or interpolation models for observational data are made with 'expert judgement' and therefore will be more subjective than up-scaling models in the way the relation between N appliance and leaching is formulated but they do not need to be up-scaled, because the same kind of data that is used for calibration is also used for the prediction of unsampled locations. For example, a statistical model is calibrated or parameters are estimated by relating monitored farm nitrate leaching to information on the distribution of soil types on a farm according to a certain soil map. Next, soil type information for non-monitored areas is extracted from the same soil map and used, together with the estimated parameters, to predict nitrate leaching on these non-monitored areas.

D.1.5 Goal

Our goal is to visualise the effect of legislation (Action Programme), on nitrate leaching in the Sand region of the Netherlands with on-farm groundwater quality monitoring data.

This paper presents and discusses a statistical (interpolation) model which is used in the Netherlands for this visualisation. The model is derived from and applied to data from the Minerals Policy Monitoring Network, a stratified random sample of farms, for the 1992–2006 period. To visualise the effects of legislation on measured nitrate leaching, the interpolation model must show a plausible relation between a decreasing amount of N appliance and decreasing nitrate leaching concentrations, which are compensated for the environmental and monitoring effects.

rivm

D2. Materials and Methods

D.2.1 Data used in calculations

D.2.1.1 Farm, farm type, area per farm type, composite nitrate concentrations

Data used are collected in the framework of the Minerals Policy Monitoring Programme in the Sand region of the Netherlands. The method of groundwater sampling and the chemical analysis are described by Fraters et al. (2005). The annual number of farms visited, the annual number of composite groundwater samples and the measured nitrate concentrations that are used to calculate a decrease in nitrate leaching are shown respectively in Table D1Table D, Table D and Table D3**.** Table D4 shows the national areas occupied by the farm types in the Sand region.
In addition to these data, climatic and soil drainage data were collected and transformed for usage in a statistical model, see next paragraph.

D.2.1.2 Precipitation Excess

Precipitation excess was quantified in the following two steps. Firstly, the downward movement of a tracer was simulated using national climatic monitoring data for precipitation and evapotranspiration of 16 districts and a soil simulation model ONZAT (OECD, 1989). The tracer was applied each day to the soil surface of a standard soil profile with grass vegetation for eight different drainage types. This resulted in a groundwater head and a tracer concentration in the uppermost metre of groundwater for different sampling dates. Figure D3 gives an example of the calculation results for a drainage type in a district.
From Figure D3 concentrations in the upper meter of groundwater can be seen to vary from year to year by a factor of two, and sometimes even three, due to variations in precipitation. The tracer concentration is inversely related to the precipitation excess.

Secondly, for each temporary well, the measured groundwater head, sampling district and sampling date were used to obtain an accompanying calculated tracer concentration from the simulation results (Boumans et al., 2001). The calculated tracer concentrations for each temporary well were averaged for each composite on-farm groundwater sample, resulting in one mean simulated tracer concentration for each composite sample per farm for the 1992–2005 period.

D.2.1.3 Soil drainage class

De Vries and Denneboom (1992) described drainage classes (Gt) on a 1:50 000 soil map. The farms were compared to the appropriate drainage class by means of overlaying the Gt map and the digitised farm surface on the topographical map. The different Gt classes were grouped as intermediate drained (Gt V, Gt V* and Gt VI), well-drained (Gt VII, Gt VII* and Gt VIII), and poorly drained (remaining classes: Gt I, Gt II, Gt II*, Gt III and Gt III*).

D.2.2 Modelling

D.2.2.1 Fitting and estimating nitrate concentrations

The following statistical model was formulated by expert judgement and fitted with the monitoring and environmental data.

Response: measured nitrate concentration in a composite sample (NO_3 mg/l)
Fixed effects: ***constant + year*** + ***farm type*** + (***fraction well-drained soils)*(precipitation excess)***$^{-1}$
Random effects: ***individual farm***, ***precipitation excess*** per individual farm, and (***fraction well-drained soils)* (precipitation excess)***$^{-1}$ per individual farm.

The variables are described in the former section. Please note that the value of ***precipitation excess*** equals the reciprocal value of the simulated tracer concentration.

The fitting was performed with the REML (REsidual Maximum Likelihood) procedure of Genstat Tenth Edition

The effect of each farm type and each year on the nitrate concentration (Response) was calculated for the condition that the other variables (fraction well-drained soils and precipitation excess) have the same mean value for each farm type and year and taking into account the unequal sampling of the individual farms.

Next, a compensated nitrate concentration for each year was estimated (Welham et al., 2004) by summing the calculated effects of year with the effects of farm type, which are weighed with the annual national areas per farm type in the Sand region (see Table D4).

D.2.2.2 Estimation of a compensated fraction of farms which exceed the EU groundwater value of 50 mg/l.

The fraction of farms, per year and farm-type, exceeding the EU standard was related to annual mean of farm mean nitrate concentrations per farm type, using a generalised linear regression model with a binomial distribution and a logit link function ($P < 0.001$). Extreme influences were investigated by the modified Cook's statistic (Payne et al., 2008a). The annual mean concentrations were transformed to equalise influences. Besides the annual mean per farm type, the year of measurement and farm type did not show a significant influence upon the fraction exceedance ($P > 0.05$). Therefore, the derived model was used to estimate annual exceedances. The compensated nitrate concentrations, which were calculated according to section D.2.2.1, were used with this model to estimate a compensated fraction of farms exceeding the EU groundwater value of 50 mg/l.

rivm

D3. Results

D.3.1. Compensated nitrate leaching concentrations

Compensated nitrate concentrations in the upper metre of groundwater and N appliance on agricultural land both decrease in the 1992–2006 period, see Figure D4. There are no data for nitrate concentrations in 1996, therefore no line is drawn between 1995 and 1997. The time lag between the N appliance and the nitrate leaching is something like one to two years. In 1998 there is a discrepancy between the time pattern of the appliance and the leaching. Compensated nitrate concentrations decreased by 50% between 1992 and 2006 and the fraction of the agricultural area in the Sand region exceeding the EU nitrate standard decreased by 25% from 93% to 66% (Figure D5). The decrease mainly occurs in the period from 1997 until 2004. The years 1992, 1993 and 1994 show the three highest compensated nitrate concentrations, while the years 2004, 2005 and 2006 show the three lowest values.

D.3.2. The nitrate leaching model

The estimated standard errors for the variance of the random effects in the model indicate the discerned random effects to be significant (Table D). Residuals show to be non normal or non Gaussian and variance is dependent upon fitted values (heteroscedastic), see Figure D6Figure D6. Factory farming leads to the highest groundwater nitrate concentrations and arable farming to the lowest (Table D6).

D4. Discussion

D.4.1. Plausibility of the found decrease in nitrate leaching concentrations caused by legislation

D4.1.1 N appliance and nitrate leaching

With the statistical model, annual mean nitrate leaching concentrations are calculated that are compensated for the well-known environmental and monitoring effects. These are precipitation, soil drainage (dilution and denitrification), a changing assembly of monitored farms and changing areas per farm type in the Sand region. The time trend of these compensated annual concentrations should visualise the effect of legislation.
According to the decrease in amount of N appliance in the 1993–2002 period, a decrease of nitrate leaching concentrations is expected shortly after 1993. The decrease in nitrate leaching concentrations starts somewhere between 1994 and 1996 and lasts until a few years after the last decrease in the N appliance in 2002, see Figure D4Figure D4. The N appliance gradually decreases over time. We therefore expect an equally gradual decrease in the compensated nitrate concentrations because we have tried to filter out other effects. The compensated nitrate concentrations, as shown in Figure D4Figure D4, indeed show a similar trend as the N appliance and are therefore plausible. Information about the time sequence of the years is not a property of the model, therefore this smooth pattern is an indication that the model works well or that the model contains all important effects. The discrepancy for the year 1998 could correspond with the swine fever in 1997. In 1997 there was an export embargo and all newly-born pigs and over 9.6 million other pigs were destroyed. Maybe there was less manure or there was a delay in the appliance of swine manure, as this manure can contain the virus. However, such a dip in N appliance in 1997 cannot be found in Figure D2Figure D2.

The compensated concentrations are calculated from measured concentrations. The annual means of measured concentrations are shown in FigureD7Figure D7, together with the non-weighed and the compensated nitrate concentrations. Because factory farming gives the highest nitrate concentrations and the area factory farming in the Sand region is lowest during the whole period, the compensated concentrations are lower than the non-weighed concentrations (see Table D4Table D and FigureD7Figure D7).

The largest difference between measured and compensated nitrate concentrations can be found for 1994 and 1995. This can be attributed to 1993 being an extremely wet year, causing more dilution and/or more denitrification due to a rise in groundwater level and consequently longer periods of anaerobic conditions in soil horizons with higher organic carbon contents.

D.4.1.2 Possible environmental effects that are not in the model but may have caused the decrease in nitrate leaching

The model takes into account the well-known environmental and monitoring effects on nitrate leaching. The decreasing nitrate concentrations could also be (partly) caused by an increase in crop uptake of N. The harvest differs from year to year due to variable weather conditions but the nitrogen output in 2000–2002 period is about 17% lower compared to the 1992–1994 period (Zwart et al., 2008). It is

plausible that the nitrogen crop uptake decreased during the 1992–2006 period because of less N appliance and therefore, crop uptake cannot be the cause of a decrease in nitrate leaching.
There is a possibility that the soil N pool changes. The amount of N in the soil pool is so large compared to the amount of N that leaches in a year that relevant changes in the soil N pool cannot be measured with enough accuracy (Kroeze et al., 2003). There are no indications for N accumulation of the soil pool in sandy soils (Kroeze et al., 2003). Schröder et al. (2007) estimated a net nitrogen mineralisation of 20 kg per ha for reclaimed peat soils in the Sand region. For grassland on peat soils, which occur to a small extent in the Sand region, a net mineralisation of 160 kg per ha is estimated. If the soil pool N progressively increases, it could cause a decrease in nitrate leaching. A decrease in N appliance will probably lead to a lower equilibrium level of the soil N pool and therefore to a temporary net mineralisation and increased nitrate leaching, if not counterbalanced by crop uptake. Therefore the soil N pool was not monitored and cannot be the reason for the decrease in nitrate leaching. Deterministic and empirical models used in the up-scaling approach, model the effect of N surplus or N appliance on nitrate leaching in the following years to visualise the effect of legislation (Grant et al., 2006; Hoffman et al., 2000 and Wolf et al., 2005). The deterministic and empirical models, using a mass balance approach for Denmark, are in agreement with the monitored decrease in nitrate leaching in terms of percentage, but independent estimates showed that 5–18% of the N surplus was not accounted for in the mass balance, indicating that some of the loss processes were underestimated (Grant et al., 2006). Heidmann and Søegaard (2002) investigated whether changes in the soil N pool could be the reason, but this was not confirmed.
Besides a decreasing amount of N appliance, measures have been taken to decrease the volatisation of ammonia. This decrease in volatisation will lead to more nitrate leaching on farms and less leaching in neighbouring nature areas. A decrease in nitrate leaching in nature areas is indeed found, see PBL. Therefore, an increase in N volatisation cannot be the cause for the decrease in nitrate leaching and was not modelled,

D.4.2. Statistical model

D.4.2.1 Modelling ‘year’ instead of N appliance

The most important variable related to legislation and nitrate leaching, being the amount of N appliance, is not in our interpolation model. The effect of farm management is also not included in the model. This is in contrast to up-scaling models. To visualise a decrease in nitrate leaching, an interpolation model, with N appliance as an independent variable could estimate the nitrate leaching for each year. The reason for not doing so is that it is not realistic to assume that there is a simple linear relation between the amount of N appliance and nitrate leaching in the following years, even if we compensate for environmental and farm management effects. Therefore the variable ‘year’ (section D.2.1.1) is incorporated into the model instead of N appliance and farm management. The variable year counts 14 values. The most common functions used to model groundwater quality changes are a simple linear function or step changes occurring at particular points in time (Loftis, 1996). The variable year models 13 step changes. Thirteen step changes is much and an alternative could be to cluster the years into groups of say 5 years but there are events in individual years, for example, the year of the swine fever (1997) or the very wet year, 1993, for which we want to see effects. Clustering data into groups of more years will lead to a more smooth pattern in time of compensated concentrations but we want to see if there are abnormal years or if there is reason to compensate for more effects than we have already done. A geohydrological consideration is that sampling the upper metre of groundwater is sampling the precipitation excess of one year in case of a hydrological mean wet year. A practical consideration is that annual updates of compensated nitrate leaching concentrations are published for the government. Because of all these reasons, we decided to model each year individually.

An alternative, for considering 'year' as a fixed effect, is to model 'year' as a random effect. That is, we could consider each year as a random selected individual from a population characterised by a mean and a standard deviation only. The step changes of individual years are then estimated as 'shrunk' values, which are closer to the overall mean. This is called an empirical Bayes estimate, see Davidian and Giltinan (1995, p76).This results in less difference between the estimated annual mean nitrate concentrations. But we want to see maximal differences between years and therefore 'year' was not modelled as random effect but as a fixed effect

D.4.2.2 Choice of modelling technique

For the choice of statistical modelling technique the following points were considered:

1. Nitrate concentrations of composite samples from one farm are probably more the same or related than of samples from different farms. This is indicated by Table D5.Table D
2. There is a linear relation between the precipitation surplus and the drainage class on the nitrate concentration, which is probably different on different farms.
3. The assembly of monitored farms differs between years, see Table D2.

Following Payne et al., 2008b, p460, the REML method with 'random coefficient regression' was chosen to estimate nitrate concentrations for each year, for the hypothetical case that each farm was monitored each year with the same number of samples and by constant mean precipitation excess and fraction wet soils and all farm types being equally present.

D.4.2.3 One effect for farm type and one effect for year

The model formulates no specific year effect per farm type but one effect for each farm type is estimated and one effect for each year. The nitrate leaching concentration for a farm type and a year is calculated by summing the estimated mean of a farm type and the decrease or increase in a year. It is assumed that all farm types have the same decrease or increase in a certain year. This assumption can be questioned because there are indications for different decreases for farm type and year combinations (Table D3). Due to few or no observations for certain farm types in some years, it is not possible to calculate compensated nitrate concentrations for all farm/year combinations for the whole 1992–2006 period. This drawback is mitigated in FigureD4Figure D4 because yearly mean values over the farm types have been calculated.

D.4.2.4 Statistical confidence

Regression methods minimise squared deviations and, under the assumptions of independence, constant variance and normality. Confidence intervals can be calculated for estimations. Found deviations have no normal distribution and no constant variance, see FigureD6Figure D6. This is because nitrate concentrations cannot have negative a value and because measured nitrate concentrations differ more between farms as the overall mean concentration is higher. If the relation between the mean nitrate concentration of a farm type and the variation of the mean nitrate between the farms is known, more efficient methods can be used. But this relation is unknown. If the nitrate concentrations are transformed to their square root, the model residuals show more agreement with the assumptions of constant variance and normality, see FigureD8. The findings of this model using transformed nitrate concentrations with regard to compensated nitrate concentrations and significance of effects are in line with the model using untransformed concentrations. Back transformation of estimated transformed nitrate concentrations results in median values that are lower than the estimated untransformed mean concentrations. If there is theoretical knowledge about the relation between the median and the mean, then these medians can be transformed to mean nitrate concentrations and it is possible to calculate confidence intervals. But this theoretical knowledge does not exist and the REML method is used with untransformed sample nitrate concentrations to estimate compensated nitrate concentrations for each year.

D5. Conclusion

A legislation induced decrease in nitrate leaching from agricultural land occurred in the Sand region of the Netherlands in the 1992–2006 period. There is a decreasing trend in the amount of N appliance of about 40% in this period. This decreasing trend in N appliance is followed, with a time lag of about one to two years, by a decreasing trend of about 50% in nitrate leaching concentrations ,which are compensated for changing precipitation excess, soil drainage, areal fractions per farm type and the assembly of monitored farms. The compensated percentage of farms exceeding the EU standard for nitrate in waters of 50 mg/l decreased from 93% in 1992 to 66% in 2006. Non compensated or measured on-farm nitrate leaching concentrations decreased about 70% in this period.

rivm

D6. References

Boesten, J.J.T.I., 2000. Modeller subjectivity in estimating pesticide parameters for leaching models using the same laboratory dataset. Agric Water Mgmt 44,389-409.

Borgesen CD, Djurhuus J and Kyllingsbak A (2001) Estimating the effect of legislation on nitrogen leaching by upscaling field simulations. Ecological Modelling 136:31-48.

Boumans, L.J.M., Fraters, B. and Van Drecht, G. (2001). Nitrate in the upper groundwater of De Marke and other farms. Neth. J. Agric. Sci. 49, 163–177.

Boumans, L.J.M., Fraters, B. and van Drecht, G. (2005). Nitrate leaching in agriculture to upper groundwater in the sandy regions of the Netherlands during the 1992-1995 period. Environ. Monit. Assess. 102, 225-241.

Broers, H.P. The spatial distribution of groundwater age for different geohydrological situations in the Netherlands: implications for groundwater quality monitoring at the regional scale. Journal of Hydrology, Volume 299, Issues 1-2, 1 November 2004, Pages 84-106.

Cherry KA, Shepherd M, Withers PJA and Mooney SJ (2008) Assessing the effectiveness of actions to mitigate nutrient loss from agriculture: A review of methods. Science of the Total Environment 406:1-23.

Davidian, M. and Giltinan, D.M. (1995). Nonlinear models for repeated measurement data. Monographs on statistics and applied probability 62, US: Chapman and Hall/CRC.

Decrem M., Spiess E., Richner W. and Herzog F. (2007) Impact of Swiss agricultural policies on nitrate leaching from arable land. Agronomy for Sustainable Development 27:243-253.

Diekkrügger, B., Söndgerath, D., Kersebaum, K.C. and Mcvoy, C. W. (1995). Validity of agroecosystem models a comparison of results of different models applied to the same data set. Ecological Modelling 81(1-3), 3-29.

D'Haene, K., Moreels, E., de Neve, S., Daguilar, B.C., Boekx, P., Hofman, G. and van Cleemput, O. 2003. Soil properties influencing the denitrification potential of Flemish Agricultural soils. Biol Fertil Soils 38:358-366.

EU (1991) Directive of the Council of December 12, 1991 concerning the protection of water against pollution caused by nitrates from agricultural sources. European Union, Brussels, 91/676/EEC.

Fraters, D., Boumans, L.J.M., van Drecht, G., de Haan, T. and de Hoop, W. (1998). Nitrogen monitoring in groundwater in the regions of the Netherlands. Environ Pollut 102,s1, (479-485).

Fraters, D., Boumans L.J.M., van Leeuwen, T.C. and de Hoop W.D. (2005). Results of 10 years of monitoring nitrogen in the sandy region in The Netherlands. Water Science & Technology, 5(3-4),239-247.

Fraters, B., Boumans, L.J.M., Elzakker, B.G. van, Gast L.F.L., Griffioen, J., Klaver, G.T., Nelemans, J.A., Velthof, G.L. and Veld, H. (2006). A new compliance checking level for nitrate in groundwater? Feasibility study on monitoring the upper five metres of groundwater. RIVM-rapport 680100005.

Grant R. and Blicher-Mathiesen, G. (2004). Danish policy to reduce diffuse nitrogen emissions from agriculture to the aquatic environment. Water Science & Technology 49(3),91-100.

Grant, R., Nielsen, K. and Waagepetersen, J. (2006). Reducing nitrogen loading of inland and marine waters. Evaluation of Danish policy measures to reduce nitrogen losses from farmland. Ambio, 35(3)117-123.

Heidmann, T. and Søegaard K. 2002, Change in the N-content of soils (in Danish). Background paper for the mid-term evaluation of Action Plan II. Danish Institute of Agricultural Sciences, Tjele, Denmark.

Hoffmann, M., Johnsson, H., Gustafson, A. and Grimvall, A. (2000). Leaching of nitrogen in Swedish agriculture – a historical perspective. Agric Ecosyst Environ 90(3),277-290.

Jensen, N.H. and Veihe, A. (2009) Modelling the effect of land use and climate change on the water balance and nitrate leaching in eastern Denmark. Journal of Land Use Science 4:53-72.

Kroeze, C., Aerts, R., van Breemen, N., van Dam, D., van der Hoek, K., Hofschreuder, P., Hoosbeek, M., de Klein, J., Kros, H., van Oene, H., Oenema, O., Tietema, A., van der Veen R., and de Vries, W. (2003). Uncertainties in the fate of nitrogen I: An overview of sources of uncertainty illustrated with a Dutch case study. Nutr Cycling Agroecosyst 66,71-102.

Loftis, J. C. (1996). A global look at trends in ground-water quality. Hydrological Processes, 10:335-355.

Lord, E.I. and Anthony, S.G. 2000. MAGPIE: A modeling framework for evaluating nitrate losses at national and catchment scales. Soil Use and Management 16,167-174.

OECD: 1989. Compendium of environmental exposure assessment methods for chemicals. OECD Environ. Monogr. 27, pp. 181–188.

Oenema, O., Boers, P.C.M., Van Eerdt, M.M., Fraters, B., Van Der Meer, H.G., Roest, C.W.J., Schroder and J.J., Willems, W.J. (1998) Leaching of nitrate from agriculture to groundwater: The effect of policies and measures in the Netherlands. Environmental Pollution 102:471-478.

Ondersteijn, C.J.M., Beldman, A.C.G., Daatselaar, C.H.G., Giesen, G.W.J. and Huirne, R.B.M. (2002). The Dutch Mineral Accounting System and the European Nitrate Directive: implications for N and P management and farm performance. Agric Ecosyst Environ 92(2-3),283-296.

Payne, R.W., Harding, S.A., Murray, D.A., Soutar, D.M., Baird, D.B., Glaser, A.I., Channing, I.C., Welham, S.J., Gilmour, A.R., Thompson, R. and Webster R. (2008a). GenStat® Release 11 Reference Manual Part 3: Procedure Library PL19; VSN International, 5 The Waterhouse, Waterhouse Street, Hemel Hempstead, Hertfordshire HP1 1ES, UK.

Payne, R.W., Harding, S.A., Murray, D.A., Soutar, D.M., Baird, D.B., Glaser, A.I., Channing, I.C., Welham, S.J., Gilmour, A.R., Thompson, R. and Webster R. (2008b). GenStat® Release 11 Reference

Manual. The Guide to GenStat® Release 11, Part 2: Statistics; VSN International, 5 The Waterhouse, Waterhouse Street,Hemel Hempstead, Hertfordshire HP1 1ES, UK.

PBL, http://www.milieuennatuurcompendium.nl/indicatoren/nl0276-Grondwaterkwaliteit-onder-bossen.html?i=25-107 (in Dutch).

Ryden, J,C,, Ball, P.R. and Garwood, E.A. (1984) Nitrate leaching from grassland. Nature 311:50-53.

Sauer, S. and Harrach, T. (1996) Leaching of nitrogen from pastures at the end of the grazing season. Journal of Plant Nutrition and Soil Science 159:31-35.

Schröder, J.J., Aarts, H.F.M., ten Berge, H.F.M., van Keulen, H. and Neeteson, J.J. 2003. An evaluation of whole farm nitrogen balances and related indices for efficient nitrogen use. Euro J Agron 20:33-44.

Schröder, J.J., G.L. Velthof, J.R. Van der Schoot and W. Van Dijk. 2007. Effect van nalevering op het stikstofoverschot van akker- en tuinbouwbedrijven en van melkveebedrijven. Nota 492. Plant Research International, Wageningen.

Schwaiger K. (2005) Monitoring effectiveness of the EU Nitrates Directive Action Programmes: Approach by Austria. In: B. Fraters, K. Kovar, W.J. Willems, J. Stockmarr and R. Grant. Monitoring effectiveness of the EU Nitrates Directive Action Programmes Results of the international MonNO3 workshop in the Netherlands, 11-12 June 2003; RIVM report 500003007/2005.

Van den Ham A., Daatselaar G.H.G., Doornerwaard, G.J. and de Hoop, D. W. 2007. Bodemoverschotten op landbouwbedrijven. LEI-rapport 3.07.05 (In Dutch).

Welham, S., Cullis, B., Gogel, B., Gilmour, A. and Thompson, R. (2004). Prediction in linear mixed models. Aust. N.Z.J. Stat. 46(3),325-347.

Wolf, J., Hack-ten Broeke, M.D. and Rötter, R. (2005). Simulation of nitrogen leaching in the Sandy soils in the Netherlands with the ANIMO model and the integrated modeling system STOME. Agric Ecosyst Environ 105, 523-540.

Wolter, R. and Mohaupt V. (2005) Monitoring effectiveness of the EU Nitrates Directive Action Programmes: Approach by Germany. In: B. Fraters, K. Kovar, W.J. Willems, J. Stockmarr, R. Grant; Monitoring effectiveness of the EU Nitrates Directive Action Programmes Results of the international MonNO3 workshop in the Netherlands, 11-12 June 2003; RIVM report 500003007/2005.

Zwart, M.H. , Hooijboer, A.E.J., Fraters, B., Kotte, M., Duin, R.N.M., Daatselaar, C.H.G., Olsthoorn, C.S.M. and Bosma, J.N. (2008). Agricultural practice and water quality in the Netherlands in the 1992-2006 period. RIVM Report 680716003.

D7. Tables and figures

Table D1 Number of sampled farms per farm type and per year in the Sand region in the 1992–2006 period.

Year	Arable	Factory	Dairy	Other types
1992	18	0	68	7
1993	15	0	65	5
1994	0	0	32	3
1995	18	0	62	4
1996	0	0	0	0
1997	10	0	14	3
1998	11	6	19	6
1999	8	11	17	5
2000	8	3	24	6
2001	10	3	29	3
2002	10	6	28	7
2003	17	11	38	12
2004	15	11	69	8
2005	13	16	68	14
2006	14	9	121	12

Table D2 Number of composite samples of on-farm groundwater per farm type per year in the Sand region in the 1992–2006 period.

Year	Arable	Factory	Dairy	Other types
1992	72	0	272	28
1993	60	0	260	20
1994	0	0	252	28
1995	18	0	96	7
1996	0	0	0	0
1997	20	0	28	6
1998	22	12	38	12
1999	16	22	34	10
2000	14	5	33	7
2001	20	6	58	6
2002	20	12	55	13
2003	31	22	76	23
2004	30	22	143	15
2005	28	31	135	28
2006	28	18	241	24

Table D3 Mean of measured nitrate concentrations (mg/l) of on-farm groundwater per farm type per year in the Sand region in the 1992–2006 period.

	Arable	Factory	Dairy	Other types
1992	134		197	223
1993	143		194	179
1994			92	94
1995	65		95	100
1996				
1997	70		173	193
1998	113	208	133	157
1999	39	128	77	143
2000	74	149	75	126
2001	81	166	64	36
2002	48	113	52	85
2003	58	58	46	65
2004	77	153	63	83
2005	78	173	55	87
2006	71	126	55	95

Table D4 Agricultural area (in 1000 ha) per farm type per year in the Sand region of the Netherlands (Zwart et al., 2008) and nitrogen appliance (in kg per ha) with manure and artificial fertiliser (Van den Ham et al., 2007).

	Arable	Factory	Dairy	Other types	Manure N	Artificial N
1990					327	203
1991					326	197
1992	131	21	472	151	346	192
1993	125	22	472	156	338	189
1994	121	22	468	157	320	171
1995	126	23	467	156	293	196
1996	125	26	459	162	322	175
1997	126	26	460	164	283	183
1998	126	28	452	163	242	183
1999	127	31	441	173	242	179
2000	119	32	441	164	219	153
2001	120	33	446	171	219	122
2002	127	28	444	161	196	109
2003	130	29	439	162	201	117
2004	127	32	433	165	192	129
2005	123	32	421	166	188	110
2006	123*	32*	421*	166*		

* no data available, but these values were used in the calculations

Table D5 Estimated variance components and accompanying standard errors for the random effects of the Nitrate Leaching Model.

Random effects	**component**	**standard error**
farm	1566	179
precipitation excess^{-1} per farm	1664	401
precipitation excess^{-1} times fraction well-drained soils per farm	2763	773
residual	1729	56

Table D6 Estimated mean nitrate concentrations (mg/l) in the Sand region of the Netherlands per farm type; mean for the 1992–2006 period.

Farm type	Arable farms	Factory farms	Dairy farms	Other farm types
Estimated NO_3 (mg/l)	92	161	104	127

Standard errors of differences: average 8.0; maximum 9.4; minimum 6.3.

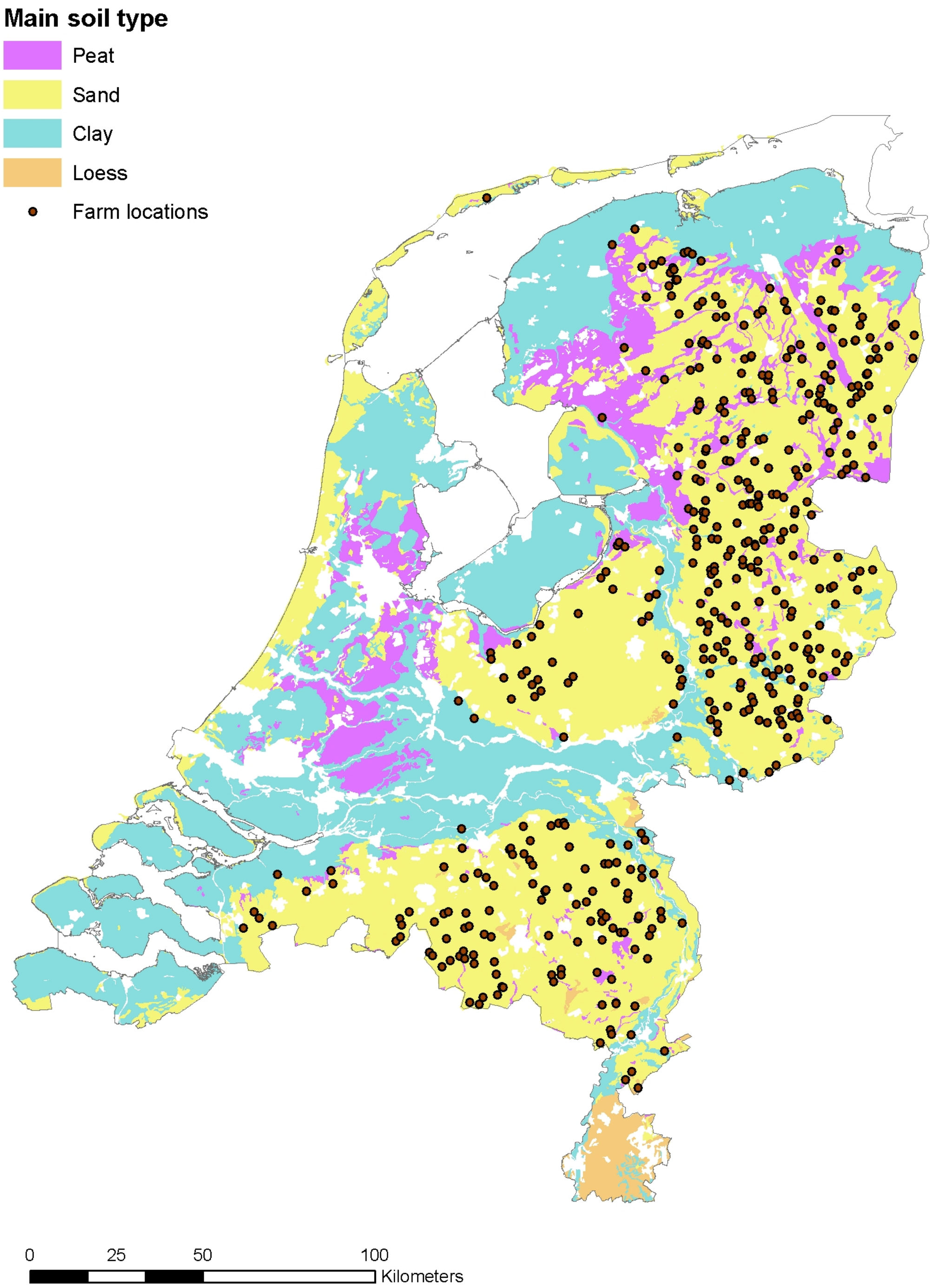


Figure D1 Soil type regions of the Netherlands.

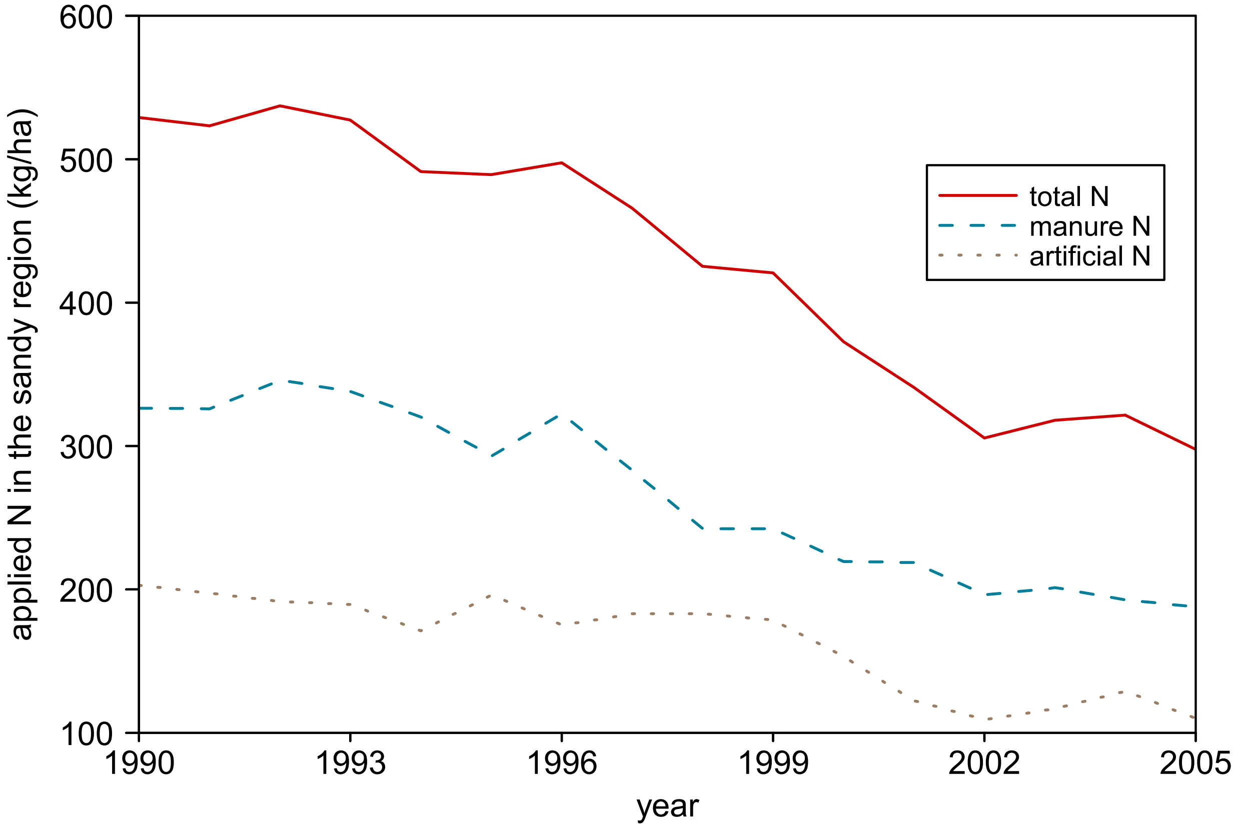


Figure D2: Trend of nitrogen appliance on agricultural land in the Sand region in the Netherlands in the 1990–2005 period (Van den Ham et al., 2007).

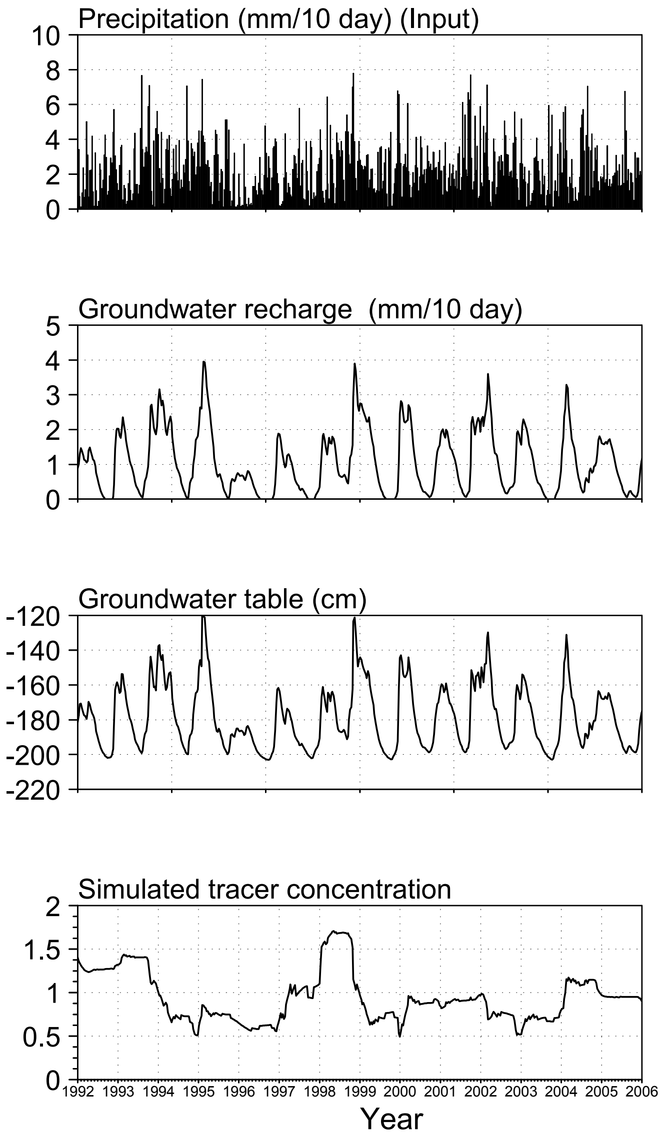


Figure D3: Calculation of weather effects on tracer concentration in the uppermost metre of groundwater.

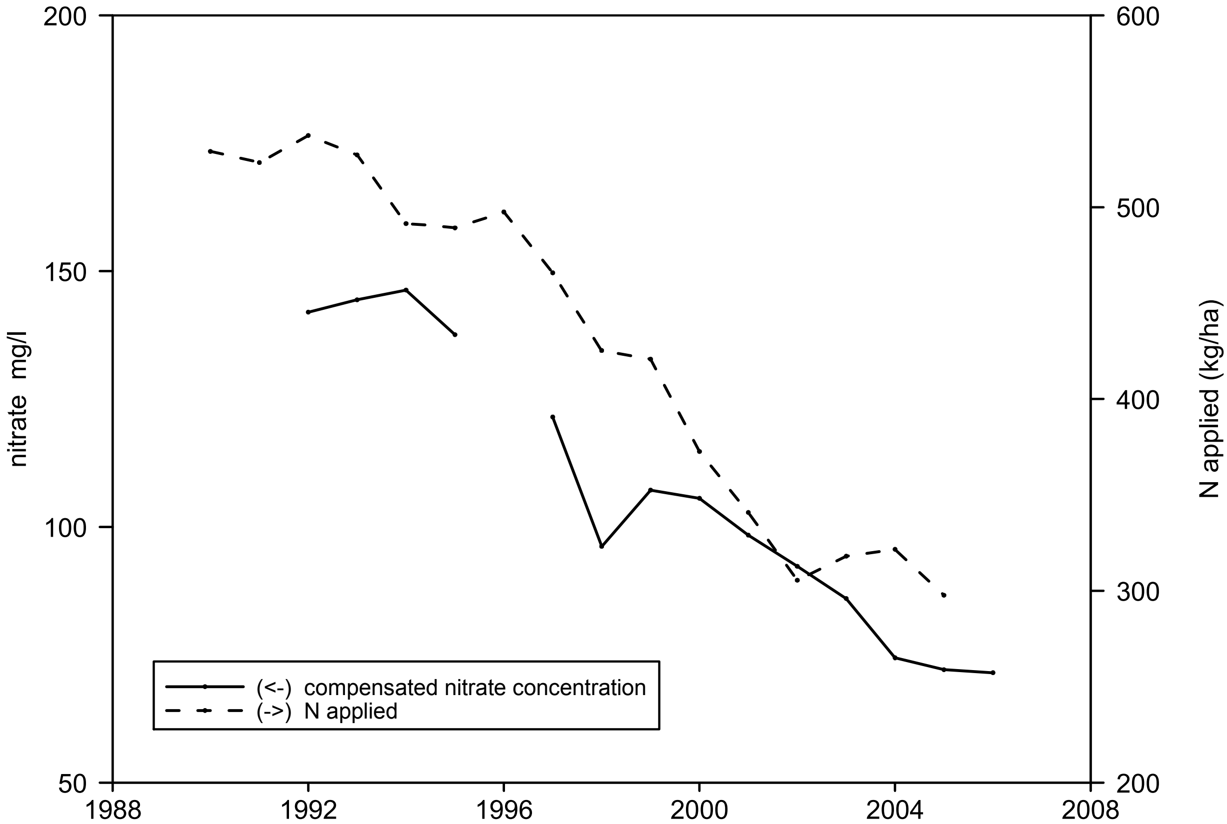


Figure D4 Compensated annual mean nitrate concentrations in the upper metre of groundwater (mg/l) and nitrogen appliance on agricultural land (kg/ha) in the Sand region of the Netherlands during the 1990–2006 period.

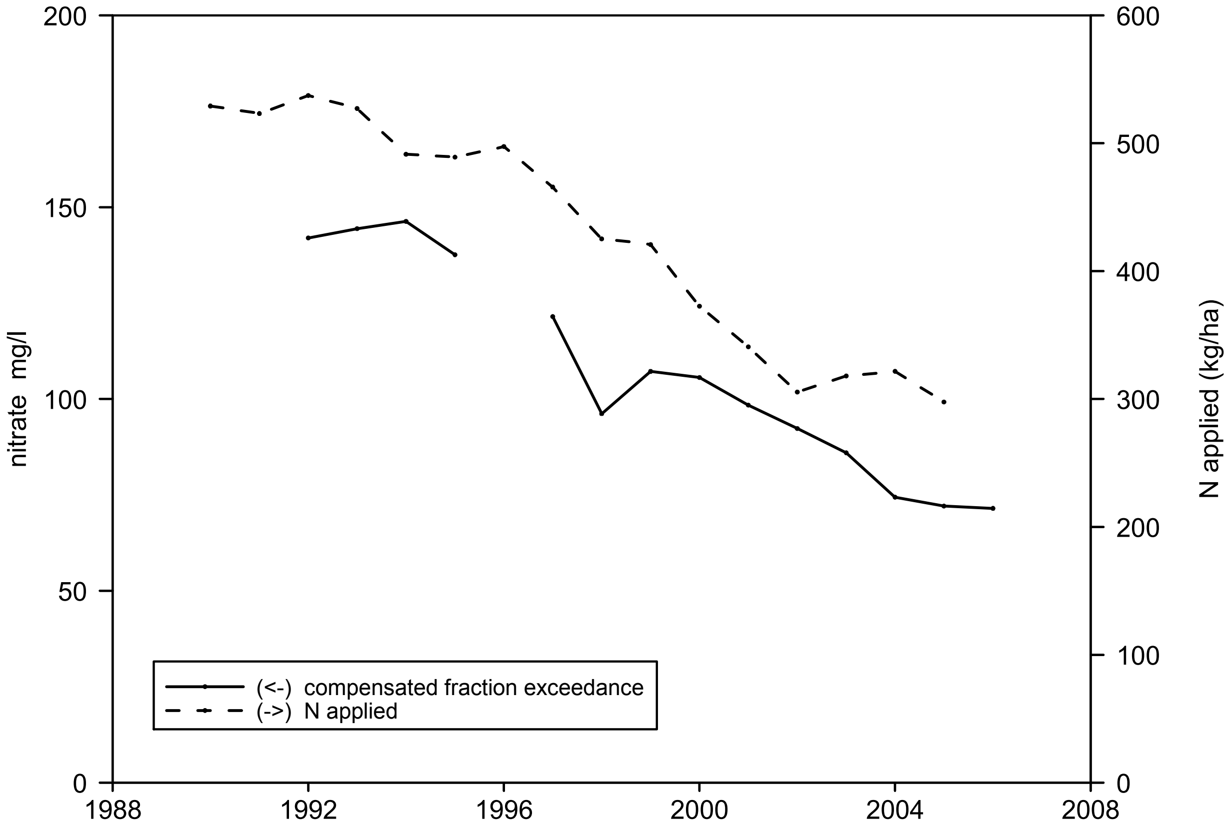


Figure D5 Balanced and weighed annual mean fraction of farms exceeding EU nitrate standard in the upper metre of groundwater and N appliance in the Sand region of the Netherlands in the 1990–2006 period.

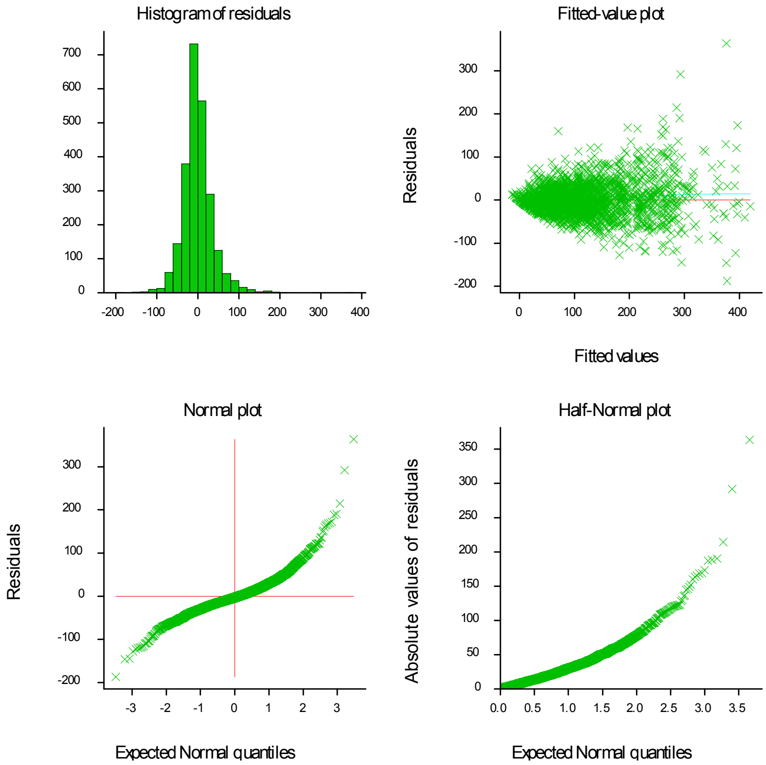


Figure D6 Residuals diagnostic for modelled untransformed nitrate concentrations.

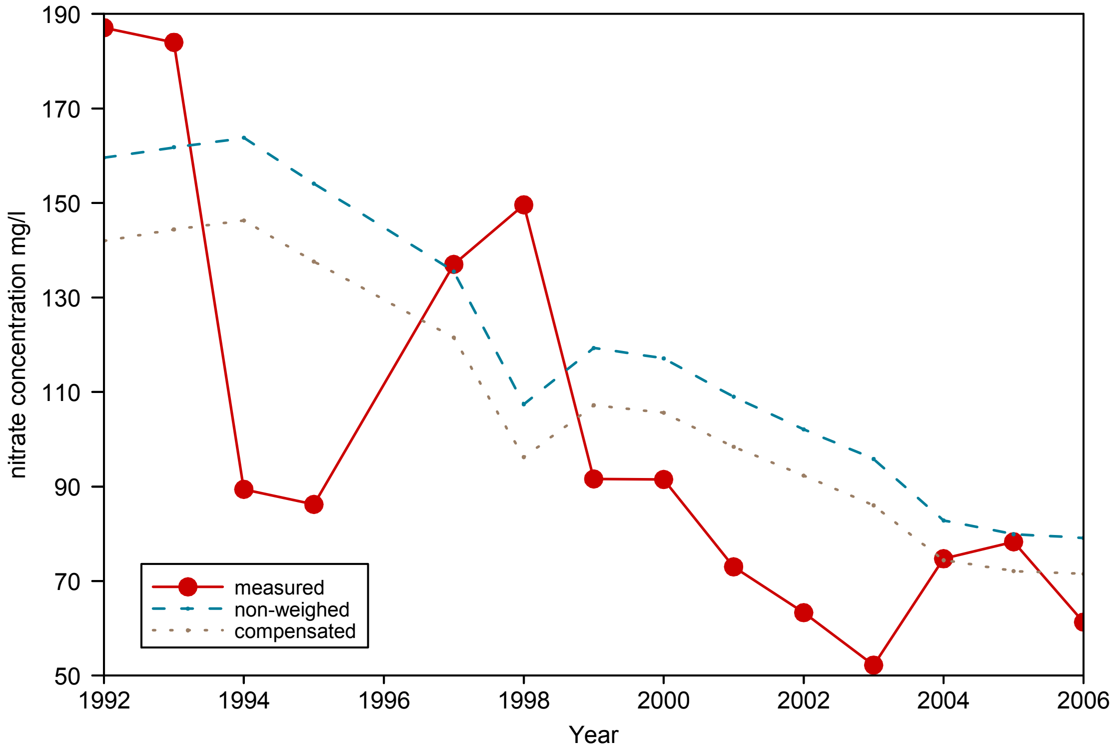


Figure D7 Annual mean of measured nitrate concentrations, non-weighed nitrate concentrations and compensated nitrate concentrations in the upper metre of groundwater in the Sand region in the Netherlands in the 1992–2006 period.

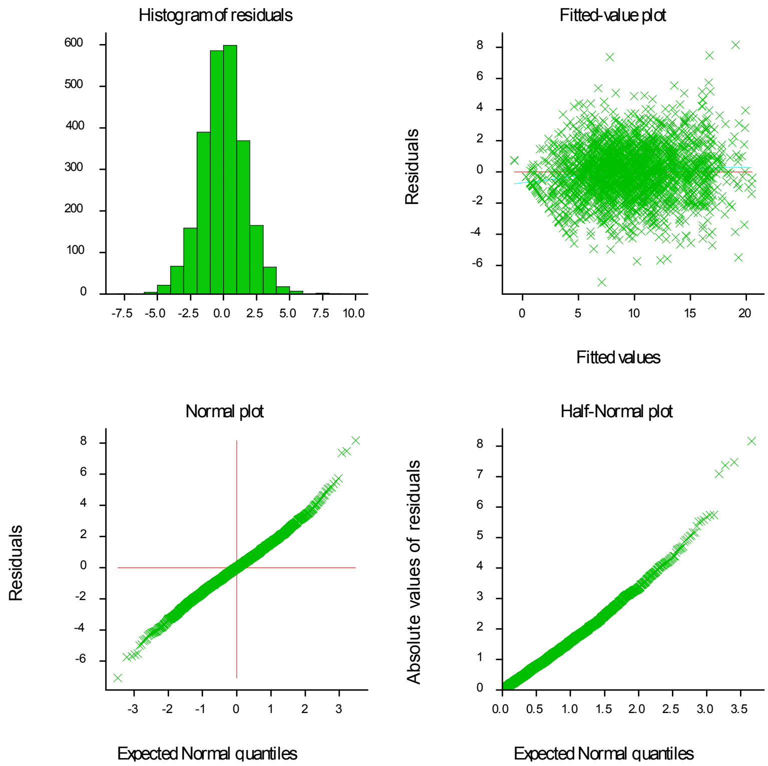


Figure D8 Residuals diagnostic for modelled square root transformed measured nitrate concentrations.

RIVM

National Institute
for Public Health
and the Environment

P.O. Box 1
3720 BA Bilthoven
The Netherlands
www.rivm.nl