Characterising the within-field scale spatial variation of nitrogen in a grassland soil to inform the efficient design of in-situ nitrogen sensor networks for precision agriculture

R. Shaw[a*], R.M. Lark[b], A.P. Williams[a], D.R. Chadwick and D.L. Jones[a]

[a]*School of Environment, Natural Resources & Geography, Bangor University, Gwynedd, LL57 2UW, UK*

[b]*British Geological Survey, Keyworth, Nottingham, NG12 5GG*

[*]*Corresponding author*

E-mail address: rory.shaw@bangor.ac.uk: Tel: +44 1248 382579

ABSTRACT

The use of in-situ sensors capable of real-time monitoring of soil nitrogen (N) may facilitate improvements in agricultural N-use efficiency (NUE) through better fertiliser management. The optimal design of such sensor networks, consisting of clusters of sensors each attached to a data logger, depends upon of the spatial variation of soil N and the relative cost of the data loggers and sensors. The primary objective of this study was to demonstrate how in-situ networks of N sensors could be optimally designed to enable the cost-efficient monitoring of soil N within a grassland field (1.9 ha). In the summer of 2014, two nested sampling campaigns (June & July) were undertaken to assess spatial variation in soil amino acids, ammonium (NH_4^+) and nitrate (NO_3^-) at a range of scales that represented the within (less than 2 m) and between (greater than 2 m) data logger/sensor cluster variability. Variance at short range (less than 2 m) was found to be dominant for all N forms. Variation at larger scales (greater than 2 m) was not as large but was still considered an important spatial component for all N forms, especially NO_3^-. The variance components derived from the nested sampling were used to inform the efficient design of theoretical in-situ networks of NH_4^+ and NO_3^- sensors based on the costs of a commercially available data logger and ion-selective electrodes (ISEs). Based on the spatial variance observed in the June nested sampling, and given a budget of £5000, the NO_3^- field mean could be estimated with a 95% confidence interval width of 1.70 µg N g^{-1} using 2 randomly positioned data loggers each with 5 sensors. Further investigation into "aggregate-scale" (less than 1 cm) spatial variance revealed further large variation at the sub 1-cm scale for all N forms. Sensors, for which the measurement represents an integration over a sensor-soil contact area of diameter less than 1 cm, would therefore, be subject to further spatial variability and local replication at scales less than 1 cm would be needed to maintain the precision of the resulting field mean estimation. Adoption of in-situ sensor networks will depend upon the development of suitable low-cost

sensors, demonstration of the cost-benefit and the construction of a decision support system that utilises the generated data to improve the NUE of fertiliser N management.

Keywords: Fertilizer management; soil heterogeneity; dissolved organic nitrogen; precision agriculture; nitrogen-use efficiency; nutrient cycling

1. Introduction

Improving nitrogen (N) use efficiency (NUE) remains one of the key challenges for global agriculture (Cassman et al., 2002; Robertson and Vitousek, 2009) and is essential for the success of sustainable intensification (Tilman et al., 2011). The deleterious environmental effects and economic costs of diffuse N pollution from farmland in Europe, where N has been applied in excess of crop requirement, are well documented (Sutton et al., 2011).

One often-cited approach to reduce N losses and improve NUE, is to ensure synchronicity of N supply with crop demand (Shanahan et al., 2008; Robertson and Vitousek, 2009), although, achieving this in practice is challenging due to the complex nature of the soil-plant system. Precision agriculture (PA) attempts to address this issue by reducing uncertainties surrounding the measurement of key variables to determine optimum N fertiliser management (Pierce and Nowak, 1999; Dobermann et al., 2004). Temporal variations in growing conditions, both within and between seasons may lead to considerable differences in optimum N fertiliser requirement and hence, inefficiencies in N fertiliser-use if temporal variations are not considered (Lark and Wheeler, 2003; McBratney et al., 2005; Shahandeh et al., 2005; Shanahan et al., 2008; Deen et al., 2014). However, conventional soil sampling techniques, coupled with laboratory analysis is expensive, labour-intensive, and time-consuming and cannot provide real-time data of sufficient resolution to accurately inform PA management (Sylvester-Bradley et al., 1999; Kim et al., 2009).

A number of different approaches have been used to address this issue. Crop canopy sensing techniques, for determination of plant N status, are now in commercial use and can be used to inform variable rate fertiliser application (e.g. wheat, maize; Raun and Johnson, 1999; Diacono et al., 2013). Whilst the advantages of this approach in some situations have been evidenced (Diacono et al., 2013), plant N status and yield is the product of many variables and may not always correlate with soil mineral N status. On-the-go soil sampling

for NO_3^-, using electrochemical sensor platforms attached to agricultural vehicles have been developed (Adsett et al., 1999) and, for the case of pH, commercialised (Adamchuk et al., 1999). The results have been used to develop field nitrate maps (Sibley et al., 2009) which could be used to define within-field management zones and to calculate variable fertiliser application rates. On-the-go sampling is generally more spatially intensive than manual field sampling, allowing better spatial resolution, although key information on how soil mineral N varies over small spatial scales may not be obtained. This can lead to increased uncertainties of interpolative predictions, especially if the sample volume is small (Schirrmann and Domsch, 2011). Furthermore, increasing the temporal resolution of this approach requires additional economic costs and as both these approaches rely on reactive management, crucial changes in soil mineral N status may be missed.

One approach, which has yet to be explored, is the use of in-situ sensors capable of monitoring soil mineral N in real time. At the time of writing, there are no such quasi-permanent field sensors in use commercially. However, potential for the development and deployment of such sensors exists (Shaw et al., 2014). For example, ion-selective electrodes (ISEs) have many characteristics suitable for soil sensing networks. They are relatively cheap, simple to use, require no mains electrical power supply and the concentration of the target ion can be easily calculated via a pre-calibration. Nitrate (NO_3^-) ISEs have previously been successfully deployed for monitoring streams and agricultural drainage ditches (Le Goff et al., 2002; Le Goff et al., 2003) as well as for on-the-go soil sampling of agricultural soils (Sinfield et al., 2010) and on-farm rapid tests for soil NO_3^- (Shaw et al., 2013). Similarly, ammonium (NH_4^+) ISEs have been used for water monitoring in a variety of situations (Schwarz et al., 2000; Müller et al. 2003). Direct soil measurement, which is essential for the success of in-situ monitoring, has been shown to be possible (Ito et al., 1996; Adamchuk et al., 2005), although improvements in accuracy and robustness of the sensing membrane are

93 required. Increasing use of nano technologies for the construction of electrochemical sensors
94 may result in significant advances in sensor performance (Arrigan, 2004; Atmeh and Alcock-
95 Earley, 2011).

96 Using in-situ sensor networks may enable a step away from predetermined fertiliser N
97 recommendations (Defra, 2010) to a more dynamic system that responds in real-time to
98 changes in growing conditions. It potentially has many benefits compared to on-the-go soil
99 sampling and crop canopy sensing. The data provided by in-situ sensors will be of
100 significantly higher temporal resolution, negating the need for repeated sampling surveys
101 throughout the year, which represent an economic cost to the farmer. Furthermore, this may
102 enable more accurate timing of fertiliser application, reducing the risk of yield penalties
103 caused by N-nutrition deficiencies, and the risk of N transfers to water and air as a result of
104 excessive fertiliser N applications. It is also likely that the data generated at a high temporal
105 resolution by an in-situ sensor networks will increase knowledge of the controls of soil N
106 processes and thus enable development of models which allow for a proactive approach to
107 fertiliser N management. However, there is a trade-off to be made. The increase in temporal
108 resolution gained from in-situ sensor networks may be offset by the costs of achieving
109 sufficient spatial resolution.

110 It is, therefore, important that consideration is made as to how such sensor networks
111 could be optimally designed to enable sufficiently precise estimates of mean field or
112 management zone (MZ) soil N at minimum cost. Two factors complicate this. First, each
113 sensor must attach to a data logger, over a relatively short distance. As data loggers cost
114 more per unit than sensors, and one logger can support several sensors, then feasible
115 networks will comprise sensor clusters, each associated with a logger. As such, sensor
116 networks can be regarded as multi-scale sampling schemes with data loggers (primary units)
117 and sensors (secondary units) randomly placed in an area around each data logger. Second,

soil N is variable, at multiple scales. Efficiently designed sensor networks will have sufficient replication at the most variable scales, achieving this within the constraints of feasible clustered designs. As shown by de Gruijter et al. (2006), the optimum configuration of such a sampling scheme depends on the relative costs of additional primary and secondary units and the within- and between-primary unit variability.

The primary objective of this study was to address the above problem and to investigate how the design of a theoretical network of in-situ soil N sensors could be optimised on a cost-precision basis, to enable monitoring of soil N concentrations in a grassland field. As seen in the discussion above, the feasibility and optimal design of sensor networks depends on the variability of the target properties at different within-field scales. An effective way to collect such information is by spatially nested sampling, which has been used previously to characterise the spatial variation in a range of soil-related variables (Lark, 2011). In nested spatial sampling, sample sites are arranged in a nested hierarchical design which allows the partition of the variance of the measured variable into components associated with a set of pre-determined scales. At the highest level of the hierarchy sample, points are arranged in clusters associated with “mainstations” which may be at randomly-located sites or on nodes of a grid or transect. Within a mainstation, sample points may be divided between two or three stations at level 2 which are separated from each other by some fixed distance. Within each level-2 station, sample points may be ordered at further nested spatial scales.

As such, spatially nested sampling was performed at a range of scales to characterise the within-field spatial variability of amino acids, NH_4^+ and NO_3^-. These results were then used to explore the optimisation of a NH_4^+ and NO_3^- in-situ sensor network design on the basis of both cost and precision. Finally, the potential and challenges of implementing this approach within a PA framework are discussed.

2. Materials and methods

2.1. Field site and soil characteristics

The field used for this study is located within the Henfaes Research Station Abergwyngregyn, Wales, UK (53°14′N 4°01′W). The site has a temperate, oceanic climate, receives an average annual rainfall of 1250 mm and has a mean annual soil temperature at 10 cm depth of 11 °C. The field is roughly rectangular with a perimeter of 559 m and an area of 1.91 ha. It has an average altitude of 12.1 m asl with a slope of 1.5% in a northerly aspect. It is a semi-permanent sheep-grazed grassland, dominated by *Lolium perenne* L. The current ley was established by direct drill in April 2009 using a perennial and hybrid ryegrass mix. The field has been used for both all year round grazing and silage production since 2009, receiving an annual inorganic fertiliser input of between 100 – 130 kg N ha^{-1} in addition to potassium (K), phosphate (P) and sulphur (S) at recommended rates. Lime has also been applied when necessary to restore the pH to a target value of 6.5. In 2014, inorganic fertiliser was applied on 12/5/14 and 11/7/14 at a rate of N:P:K 50:10:10 and 60:4:0 kg ha^{-1}, respectively. The field was grazed until 9/6/14 and the field remained sheep free until the 2/9/14. The soil is a free draining Eutric Cambisol with a sandy clay loam texture and a fine crumb structure.

To assess the chemical characteristics of the soil, replicate samples (n = 4) were collected from 4 randomly located points within the field. For each sample, the vegetation was removed from an area approximating 30 × 30 cm and soil was collected to a depth of 10 cm, representing the Ahp horizon (an Ahp horizon is an Ah horizon which has been subjected to cultivation). This sampling design is not related to the spatially nested design described in section 2.2. The soil was placed in gas-permeable polyethylene bags and transported to the laboratory in a refrigerated box. All of the following procedures were performed on the same

day as field sampling. Soil pH and electrical conductivity were determined in a 1:2.5 (w/v) soil:distilled water suspension using standard electrodes. Moisture content was determined by drying for 24 h at 105 °C. Total C and N were determined with a TruSpec CN analyser (Leco Corp., St Joseph, MI, USA). Dissolved organic carbon (DOC) and dissolved organic nitrogen (DON) were measured in soil extracts (0.5 M K_2SO_4, 1:5 w:v) using an Analytik Jena Multi N/C 2100S (AnalytikJena, Jena, Germany). Chloroform fumigation and incubation (t = 7 days) of 2 g (n = 4) of fresh soil was performed to determine microbial biomass C and N according to Voroney et al. (2008) (K_{EC} = 0.35 K_{EN} = 0.5). Exchangeable cations were extracted using 0.5 M acetic acid (Sparks, 1996) and the filtered extracts analyzed using flame emission spectroscopy (Sherwood 410 flame photometer; Sherwood Scientific, Cambridge, UK). Extractable phosphorus (P) was determined by extraction with 0.5 M acetic acid with subsequent colorimetric analysis using the molybdate blue method of Murphy and Riley (1962). Basal soil respiration was determined in the laboratory at 20 °C using an SR1 automated multichannel soil respirometer (PP Systems Ltd., Hitchin, UK) and steady state CO_2 production rates recorded after 24 h. Potentially mineralisable N was determined using an anaerobic incubation method based on Keeney (1982). Briefly, 5 g field moist soil was place in a 50 ml centrifuge tube, which was then filled to the top with de-ionized H_2O and the tubes sealed. Soils were subsequently incubated in the dark at 40 °C for 7 d. The difference in NH_4^+ content between t = 0 and t = 7 d was attributed to N mineralization.

Above ground biomass was sampled on 26/6/2014. Replicate 1 × 1 m blocks (n = 4) were chosen at random from within the field. The vegetation was cut to ground level, stored in paper bags and subsequently oven-dried at 80 °C to determine dry matter content. A summary of the results is shown in Table 1.

2.2. Sampling design and protocol

193 *Nested sampling for spatial variability:* The aim of the sampling was to characterize
194 the variability of plant-available N forms – amino acid-N, NH_4^+ and NO_3^- – at a range of
195 spatial scales relevant to planning the design of an in-situ sensor network. In particular, it was
196 necessary to examine the relative importance of variance between and within local regions
197 each of which might be represented by a cluster of soil N sensors deployed around a single
198 data logger such that the maximum distance between any two sensors is about 2 m. In a
199 grassland environment it was expected that one of the main sources of variation in soil N
200 would be the uneven and relatively random distribution of urine patches of linear dimensions
201 about 40 cm (Bogaert et al., 2000; Selbie et al., 2015). Variation at larger scales may also be
202 important due to preferential use of certain areas of the field such as tracks, areas of shade
203 and around drinking troughs (Bogaert et al., 2000), which may be reflected in local gradients
204 in soil chemistry. The study field is broadly homogenous in terms of its topography and soil
205 type. Furthermore, a visual inspection of the field revealed no obvious large-scale gradients
206 in vegetation condition which is likely to reflect the broadly homogenous nature of the soil.
207 Previously, the field has received uniform management in terms of its fertiliser and lime
208 inputs and grazing regime. Because of these factors, it was decided to treat the field as
209 singular management unit with a singular mean rather than subdivide the field into separate
210 management zones.
211 Given these considerations, a nested sampling protocol was designed with length
212 scales within each mainstation of 1 cm, 10 cm (intermediate between the fine scale and urine
213 patch scale), 50 cm (urine patch scale) and 2 m (upper bound on the "within-region served by
214 a sensor cluster" scale). To assess spatial variation at larger scales, mainstations were
215 distributed by stratified random sampling with the target field divided into four quarters
216 (strata) of equal area. Four mainstations were established at independently and randomly-
217 selected locations within each quarter (stratum), giving a total of 16 mainstations. The design

of the sampling scheme within each mainstation, was obtained by the optimization procedure of Lark (2011) on the assumption of a fractal or quasi-fractal process in which the variance is proportional to the log of the spatial scale. The objective function was the mean estimation variance of the variance components. With 12 samples per mainstation the total sample size was 192. The sample sites were then selected at each mainstation by randomizing the direction of the vectors between the substations at each level of the design shown in Figure 1, while keeping the lengths of the vectors fixed. For practical purposes, sampling was split over 2 successive days, with 2 strata sampled on day 1 and two on day 2, giving a total of 8 mainstations and 96 samples per day. No duplicate sampling took place as each sample site was visited only once over the 2-day period.

An initial nested sampling campaign was performed over 2 days on the 4th and 5th June, 2014. Following this, all sheep were removed and the field remained ungrazed until 2nd September, 2015. A further nested sampling campaign was performed on the 31st July and 1st August, 2014, 3 weeks after the field received a N fertiliser input of 60 kg N ha^{-1}. These are subsequently referred to as the June nested sampling and the July nested sampling respectively. Sample site locations were set up the day before sampling took place. At each sampling location a soil corer, of diameter 1 cm, was used to sample soil. A 5 cm soil core from between depths of 5 -10 cm was sampled and placed in gas-permeable plastic bags, and stored in a refrigerated box. This depth was chosen as it represents the middle of the rooting zone and would make installation of any in-situ sensor a straight forward process. Following the sampling event, the samples were transferred immediately to the laboratory where they were refrigerated at 4 °C. Extraction of soluble N from soil was performed on the soil cores on the same day as sampling as described below. During the second nested sampling event, duplicate sub-sampling and chemical analysis were performed on 4 out of the 12 samples

from each mainstation in order to make an assessment of the error variance attributable to subsampling and analytical error.

Soil properties are also likely to vary at sub-core (less than 1 cm) scales (Parkin, 1987; Stoyen et al., 2000). As such, a further sampling design and protocol was developed and performed on the 25th June, 2014 to investigate how this micro-heterogeneity affected the spatial variability of N forms at the "aggregate scale" (less than 1 cm). Two sampling locations were chosen at random within each of the 4 strata. At each location, a pair of samples were taken, using the protocol described above, with a distance of 1 cm between each sample. This resulted in a total of 16 core samples. On return to the laboratory the cores where broken apart and 4 "aggregates" of weight 60 – 80 mg were collected (diameters ca. 1-2 mm). These aggregates were then extracted for soluble N and analysed using the protocol described below.

2.3. Extraction and chemical analysis of soil samples

All soil extractions were performed on the same day as sample collection, according to the following protocol. Samples were crumbled by hand, in order to prevent sieving induced N mineralisation (Jones and Willett, 2006; Inselsbacher, 2014). Large stones, roots and vegetation were removed prior to gentle mixing of the sample. To further reduce mineralisation of organic N forms, sub-samples of field-moist soil (2 g) were extracted on ice (175 rev min^{-1}, 15 min) using cooled (5 °C) 0.5 M K_2SO_4 at a soil: extractant ratio of 1:5 (w:v) (Rousk and Jones, 2010). The extracts were centrifuged (4,000 *g*, 15 min), and the resulting supernatant collected and frozen (-18°C) to await chemical analysis. The protocol differed slightly for the soil aggregate samples. Each aggregate, of weight 60 – 80 mg, was placed in a 1.5 ml Eppendorf® micro-centrifuge tube and crumbled gently using a metal spatula. The soil was then extracted in 500 µl of 0.5 M K_2SO_4 as described above. Total free

amino acids (referred to as amino acids) were determined by the *o*-phthaldialdehyde spectrofluorometric method of Jones et al. (2002). NH_4^+ was determined by the salicylate-nitroprusside colorimetric method of Mulvaney (1996) and NO_3^- by the colorimetric Griess reaction of Miranda et al. (2001) using vanadate as the catalyst.

2.4. Statistical analysis

Nested Sampling: The aim of the statistical analysis was to compute the variance components attributable to each of the spatial-scales in order to inform the optimisation of the sensor network design. After Box-Cox transformations (see section 2.5 for details of transformations), the n data from the nested sampling may be analysed according to the following statistical model (Webster and Lark, 2013). An $n \times 1$ vector of observations, $\mathbf{y}$, is regarded as a realization of a random variate, $\mathbf{Y}$, where

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}_s\boldsymbol{\eta}_s + \mathbf{U}_m\boldsymbol{\eta}_m + \mathbf{U}_2\boldsymbol{\eta}_2 + \mathbf{U}_{0.5}\boldsymbol{\eta}_{0.5} + \mathbf{U}_{0.1}\boldsymbol{\eta}_{0.1} + \mathbf{U}_{0.01}\boldsymbol{\eta}_{0.01} + \mathbf{U}_r\boldsymbol{\eta}_r. \quad (1)$$

$\mathbf{X}$ is a $n \times p$ design matrix which represents fixed effects in the model (e.g. p levels of a categorical factor, or p continuous covariates) and $\boldsymbol{\beta}$ is a length-p vector of fixed effects coefficients. In this analysis the fixed effects were different means for data collected on two successive days, as it was not possible logistically to sample on one day. Strata were randomly allocated to days (two strata per day) so between-stratum variation is not confounded with any temporal effect. There are 4 strata in the sampling design, and $\mathbf{U}_s$ is a $n \times 4$ design matrix for the strata. If the ith observation is in stratum j then $\mathbf{U}_s[i, j] = 1$ and all other elements in the ith row are zero. The design matrix associates each observation with one of 4 random values in the random variate $\boldsymbol{\eta}_s$. These values are assumed to be independent and identically distributed Gaussian random variables with a mean of zero and a variance σ^2_s which is the between-stratum variance component. Similarly, $\mathbf{U}_m$ is a $n \times 16$ design matrix

for the mainstations, and the variance of $\boldsymbol{\eta}_m$ is the between-mainstation variance component. The terms with subscripts 2, 0.5. 0.1 and 0.01 represent the design matrices and random effects for the components of variation associated with the 2-m, 0.5-m, 0.1-m and 0.01-m scales respectively. If duplicate material from some or all of the soil specimens is analysed then the random effect $\boldsymbol{\eta}_r$ which represents the variation due to subsampling and analytical variation can be estimated, otherwise it is a component of the variation estimated for the finest spatial scale.

Under the linear mixed model Y has covariance matrix **H** where

$$\mathbf{H} = \sigma^2_s\mathbf{U}_s\mathbf{U}_s^T + \sigma^2_m\mathbf{U}_m\mathbf{U}_m^T + \sigma^2_2\mathbf{U}_2\mathbf{U}_2^T + \sigma^2_{0.5}\mathbf{U}_{0.5}\mathbf{U}_{0.5}^T + \sigma^2_{0.1}\mathbf{U}_{0.1}\mathbf{U}_{0.1}^T + \sigma^2_{0.01}\mathbf{U}_{0.01}\mathbf{U}_{0.01}^T + \sigma^2_r\mathbf{U}_r\mathbf{U}_r^T, \quad (2)$$

and the superscript T denotes the transpose of a matrix. The parameters of this matrix are therefore the variance components, and these can be estimated by residual maximum likelihood (REML), see Webster and Lark (2013). Once this has been done then the fixed effects coefficients in the model can be estimated by generalized least squares (see Lark and Cullis, 2004). Note that there is an explicit assumption that the data are a realization of a Gaussian random variable with mean $\mathbf{X}\boldsymbol{\beta}$ dependent on the fixed effects and coefficients.

Because all sampling could not be done in one day the sampling day was randomized within strata, so as not to be confounded with the spatial variance components of interest. For this reason, it is regarded as a fixed effect in the model. The significance of the between-day effect was tested with the Wald statistic as discussed in Lark and Cullis (2004).

The significance of a random effect in the model can be tested by comparing the residual log-likelihood for a model with the term dropped (L^-) with the residual log-likelihood for the full model (all random effects, L). Any variance accounted for by a term which is dropped will contribute to variance at lower levels in the hierarchy (finer spatial

scales) for the dropped model. For this reason the ultimate component of the model (η_r when there are duplicate analyses and $\eta_{0.01}$ otherwise) cannot be dropped. Dropping a term from the model will usually reduce the log-likelihood (and will not increase it). Whether the reduction in likelihood is strong enough evidence that the inclusion of the term in the full model is justified can be assessed by computing Akaike's information criterion (AIC), *A*, for each model:

$$A = -2L + 2P \quad (3)$$

where *P* is the number of parameters in the model. The AIC penalizes model complexity, by selecting the model with smaller AIC, one minimises the expected information loss through the selection decision (Verbeke and Molenberghs, 2000).

Aggregate scale sampling: After Box-Cox transformations (see section 2.5 for details of transformations) the *n* data collected to investigate variation within cores were analysed according to the following statistical model. An $n \times 1$ vector of observations, y, is regarded as a realization of a random variate, Y, where

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}_s\boldsymbol{\eta}_s + \mathbf{U}_p\boldsymbol{\eta}_p + \mathbf{U}_c\boldsymbol{\eta}_c + \mathbf{U}_a\boldsymbol{\eta}_a, \quad (4)$$

As in Equation (1), $\mathbf{X}$ is a design matrix for fixed effects and $\boldsymbol{\beta}$ is a vector of fixed effects coefficients (here just a constant mean). Again, as in Equation (1), Us is a $n \times 4$ design matrix for the strata and $\boldsymbol{\eta}_s$ is assumed to be an independent and identically distributed Gaussian random variate with a mean of zero and a variance σ^2_s. In the same way $\mathbf{U}_p$ and $\boldsymbol{\eta}_p$ are the design matrix and the random variate for the between-pair within-stratum effect, with variance σ^2_p; $\mathbf{U}_c$ and $\boldsymbol{\eta}_c$ are the design matrix and the random variate for the between-core within-pair component, with variance σ^2_c and $\mathbf{U}_a$ and $\boldsymbol{\eta}_a$ are the design matrix and the random variate for the between-aggregate within-core component, with variance σ^2_a. This latter component is effectively the residual as there are no duplicate measurements on any

aggregate. The same method based on the AIC was used to assess the evidence for including each term in the model above the between-aggregate effect.

2.5. Data transformations

The REML estimator for random effects parameters makes an explicit assumption that the random variation in the model is normally distributed. This assumption is of particular importance in the model with nested random effects, because these must be modelled as independent additive components. This is plausible for normal random variables, but not in general otherwise. For this reason, it was necessary to transform the data so that the residuals from any fixed effects could be regarded as normal. To this end, a Box-Cox procedure from the MASS package in R (Venables and Ripley, 2002) was used to apply the Box-Cox transform. Under this transform the variable x is transformed to a normal variable y by finding a maximum likelihood estimate of the parameter λ such that

$$y = \begin{cases} \frac{x^{\lambda}-1}{\lambda}, & \lambda \neq 0 \\ \ln(x), & \lambda = 0 \end{cases}. \quad (6)$$

Note that the log-transformation is a special case of the Box-Cox, with λ=0. In this study, the log-normal transformation was used for cases where the 95% confidence interval of λ included 0. Where this was not the case, the Box-Cox transformation with the maximum likelihood estimate of λ was used.

A disadvantage of transformation is that the results are not on familiar scales, and that the relative and absolute magnitudes of the variances depend on the transformation and the mean. Section 2.6 describes how the variance components estimated on the transformed scales were used to calculate the width of confidence intervals for estimates of field means for forms of N estimated from different sensor arrays by a simple back-transformation of confidence limits on the transformed scale. However, it is not possible to perform a

366 comparable simple back-transformation of variance components for general interpretation of
367 the variability at different scales. This would be true of any Box-Cox transformation.
368 Two sets of results have been presented. The first are the variance components on the
369 transformed scale, the scale of measurement on which each data set can most plausibly be
370 modelled as an additive combination of random components at different scales (Panel A,
371 Figs. 2 & 3; tables 3, 5 & 7). For the different scales, percentages of the total accumulated
372 variances have been calculated, and these values are referred to in the results section to aid
373 comparison between the different N forms and sampling events. Secondly, a non-parametric
374 statistic was computed on the original scales of measurement for each nested analysis. Using
375 the estimated variance components on the transformed scale, 10,000 data values in a nested
376 configuration were simulated with, in the case of the spatially nested sample, pairs of points
377 in contrasting strata, pairs of points at different locations within a stratum and pairs of points
378 within strata separated by 2 m, 0.5 m, 0.1 m and 0.01m respectively. Each simulated value
379 was then back-transformed to the original scale of measurement by inverting the Box-Cox
380 transformation. All these pairs of observations were then examined on the original scale,
381 computing a non-parametric and robust measure of the variability of the differences. This is
382 the median absolute deviation from the median (MAD). If one considers all pair comparisons
383 over 2 m, for example, the median difference is first computed and then the absolute
384 difference between this median difference and each individual pair difference is computed,
385 and the median of all these values is extracted. The MAD is a measure of variability on the
386 original scale of measurement (like the standard deviation). These were extracted for all
387 variables and plotted alongside cumulative plots of the variance components on the
388 transformed scale to aid the interpretation of these latter plots and give an indication of the
389 magnitude of variation that can be expected for measurements on the original scale (panel B,
390 Figs 2 & 3). The same process was followed after the analysis of aggregate-scale variation to

compute the MAD of comparisons between two aggregates within a core, between two cores in a pair, between two sites within a stratum, and between pairs of strata, these are presented in Table 8. MAD values are also explicitly reported in the results section with units of µg N g^{-1}.

2.6. Optimising the design of an in-situ sensor network

The transformed variance components derived from the nested sampling and subsequent statistical analysis were used to examine the theoretical performance of different designs of in-situ soil NH_4^+ and NO_3^- sensor networks. When considering the optimal design, two factors must be considered. Firstly, what is the required level of precision for the estimation of the field mean and how many sensors and data loggers are required to achieve this? Secondly, how can the design be optimised in-terms of achieving a desired level of precision at minimum cost? Alternatively, it may be useful to explore how to design the network to achieve the highest precision possible given a certain budget restriction.

To estimate the level of precision associated with a particular sensor network design, the between-sensor within-logger component of variance, where a cluster of n_e sensors are randomly located within a region of 2 m diameter around each of n_l data logging hubs, which are located by simple random sampling, can be approximated by

$$\sigma^2_{sensor} = \sigma^2_2 + \sigma^2_{0.5} + \sigma^2_{0.1} + \sigma^2_{0.01}, \quad (6)$$

and the between-logger variance by

$$\sigma^2_{logger} = \sigma^2_s + \sigma^2_m. \quad (7)$$

As such, the standard error of the field mean soil N concentration derived from the sensor network can be estimated as follows:

$$\sigma_{mean} = \{(\sigma^2_{logger} / n_l) + (\sigma^2_{sensor} / n_l n_e)\}^{½} \quad . \quad (8)$$

This allows the 95% confidence interval of the field mean estimations to be calculated, given the variance components calculated from the nested sampling, for particular combinations

and numbers of data loggers (n_l) and sensors (n_e). These calculations were performed on the transformed scale prior to back-transformation of the 95% confidence interval to the original scale of measurement.

In order to demonstrate how the design may be optimised on a cost basis it was necessary to decide on a unit cost for a data logger and a sensor. Given the potential of electrochemical sensors for in-situ monitoring, it was decided that the unit cost for the sensor would be £200, based on the cost of a commercially available NH_4^+ or NO_3^- ISE (ELIT 8021, ELIT 003, Nico2000, Harrow, UK) and £2000 for the data logger, based on the cost of a commonly used data logger (DL2e DeltaT, Cambridge, UK). Whilst these costs are somewhat arbitrary, it does allow useful comparison between designs to be made. It would also be possible to change these unit costs to explore how using different sensors and loggers may affect the optimisation of the network.

The 95% confidence intervals were computed for sensor network designs that consisted of 1 to 10 data loggers with 2 to 15 sensors distributed equally among the loggers, 15 being the maximum number of sensor ports on the data logger (DL2e DeltaT, Cambridge, UK). This allowed construction of graphs (Fig. 4) which illustrate the total cost for each design plotted against the resulting 95% confidence interval of the estimated field mean.

3. Results

3.1. June nested sampling - evaluating the spatial variation of soluble N in soil prior to application of N fertiliser

The mean concentrations of amino acid, NH_4^+ and NO_3^- were found to be fairly similar with values of 1.44, 1.87 and 1.71 respectively (Table 2). All of the N-forms had a positively skewed distribution. This was especially the case for NH_4^+, which had a skewness value of 12.82 and a maximum value of 80.49 µg N g^{-1} (histograms of distributions can be

441 viewed in the supplementary information, Fig. S1). The Box-Cox parameter, λ, for each
442 variable had a 95% confidence interval which excluded zero, so the maximum likelihood
443 estimate of λ (Table 2) was used to transform each variable. Plots of the profile likelihood
444 for λ, with the 95% confidence interval can be viewed in the supplementary information, Fig.
445 S2.

446 The different forms of N showed slightly different scale-dependencies, although in
447 general, short-range variance dominated (Fig. 2 and Table 3). On the original units, the MAD
448 for comparisons where all sources of variation contributed were largest for NO_3^- (0.92 μg N
449 g^{-1}), followed by NH_4^+ (0.75 μg N g^{-1}), and amino acids with (0.51 μg N g^{-1}). For amino
450 acids, the 1-cm scale had the largest variance component, constituting 59% of the total
451 accumulated variance. The 10-cm and the between-mainstations within-strata term were also
452 considered important spatial components (as judged by AIC; see Table 3 and Table S1). For
453 NH_4^+, the 1-cm scale had the largest variance component, constituting 63% of the total
454 accumulated variance. However, for spatial scales greater than 1 cm, only the between-
455 mainstations within-strata term was considered important. For NO_3^-, the relative importance
456 of variation at scales coarser than 1cm was larger than for other forms of N (Fig 2) with the
457 10-cm scale having the largest variance component, constituting 28% of the total
458 accumulated variance on the transformed scale. A comparable pattern was seen with the
459 MAD values. Furthermore, all the spatial scales, with the exception of the 2-m scale,
460 exhibited variance that was considered important. Short-range scale variation still dominated
461 though, with 70% of the variance occurring at spatial scales up to 50 cm. It should be noted
462 that the 1-cm scale component will also include any measurement error.
463
464 *3.2. July nested sampling - evaluating the spatial variation of soluble N in soil after*
465 *application of N fertiliser*

The mean concentrations of amino acid, NH_4^+ and NO_3^- were found to be fairly similar with values of 1.25, 1.96 and 1.36 respectively (Table 2). All of the N-forms had slight positively skewed distributions. NH_4^+ displayed the largest positive skew, with a skewness value of 3.28 and a maximum value of 9.88 μg N g^{-1} (histograms of distributions can be viewed in the supplementary information, Fig. S3). The Box-Cox parameter, λ, for each variable had 95% confidence interval which excluded zero, so the maximum likelihood estimate of λ (Table 4) was used to transform each variable. Plots of the profile likelihood for λ, with the 95% confidence interval can be viewed in the supplementary information, Fig. S4.

The different forms of N showed slightly different scale-dependencies, although in general short-range variance dominated (Fig. 3 and Table 5). On the original units, the MAD for comparisons where all sources of variation contributed were largest for NH_4^+ (0.88 μg N g^{-1}), followed by NO_3^- (0.64 μg N g^{-1}) and amino acids with (0.41 μg N g^{-1}). For amino acids, the between mainstations within-strata had the largest variance component, constituting 36% of the total accumulated variance, although 58% of the total accumulated variance occurred at scales up to 10 cm. The 1-cm, 10-cm and the between-mainstations within-strata term were considered important spatial components (as judged by AIC; see Table 5 and Table S2). For NH_4^+, the 1-cm scale had the largest variance component, constituting 55% of the total accumulated variance. Spatial scales greater than 10 cm accounted for only 13% of the total accumulated variance. Only the 1-cm and the between-mainstations within-strata terms were considered important spatial components. For NO_3^-, the between-mainstations within strata scale was the largest variance component, constituting 39% of the total accumulated variance. The 1-cm, 10-cm and the between-mainstations within-strata term were considered important spatial components. Short-range scale variation still dominated though, with 61% of the variance occurring at spatial scales up to 50 cm.

Duplicate measurements on 4 samples from each mainstation allowed the 1-cm spatial variance component to be resolved from the subsampling and measurement error. As this residual term formed the ultimate term in the model, it allowed an assessment of the importance of the 1-cm spatial component. For all of the N forms, the 1-cm scale was considered an important spatial component, and was larger than the residual variance. However, the residual variance, which was similar for all N forms, constitutes a substantial component of the accumulated variance and was, for all N forms, larger than the variance at 50 cm and 2 m.

3.3. Aggregate-scale variability of soluble N in soil

In the case of the aggregate-scale data, the 95% confidence interval for the Box-Cox parameter, λ, for each variable included zero (see supplementary material Fig. S6), so the log-transformation was applied.

In all cases, the largest variance component was found to be the between-aggregate within-core scale (table 7 & 8). For NH_4^+ and NO_3^-, 91% and 80% respectively of the total accumulated variance occurred at this scale, which was an order of magnitude higher than the variance at the between core scale. The variance at the aggregate scale for amino acid-N was lower at 66%. It should be noted that any analytical error that occurred will also appear in this variance component. The between-core component, which represents the 1-cm spatial scale, was considered important (as judged by AIC; see Table 7 and Table S3) for amino acids and NO_3^-, but not NH_4^+. Neither the between-pair component, which is similar to the between-mainstations scale, nor the between-strata component, were considered to be important spatial components. However, the stratum and mainstation scale in this analysis were based on limited replication. The focus of this particular sampling exercise was on the aggregate and core scale, so general conclusions from these results about the importance of coarser-

516 scale variation are not made.

517

518 *3.5. Optimisation of a within-field sensor network for monitoring soluble N in soil*

519 The optimisation of the design of an in-situ network of NO_3^- and NH_4^+ sensors can be
520 explored using the graphs in Fig. 4. The graphs show how increasing both the number of
521 sensors per data logger, and increasing the number of data loggers, reduces the width of the
522 95% confidence interval of the estimated field mean derived from the sensor network. There
523 are differences in the results between NO_3^- and NH_4^+ and between sampling events. For
524 example, to achieve a 95% confidence interval width of no more than 1 µg N g^{-1} for a NO_3^-
525 sensor network, given the spatial variation observed in the June sampling event, would
526 require 3 data loggers each with 11 sensors at a cost of £8200. For the July sampling, 2 data
527 loggers each with 7 sensors would be sufficient, at a lower cost of £5400. For a NH_4^+ sensor
528 network, given the spatial variation observed in the June sampling event, 2 data loggers each
529 with 6 sensors, at a cost of £5200 would be required to achieve a 95% confidence interval
530 width of no more than 1 µg N g^{-1}. For the July sampling, 2 data loggers each with 8 sensors,
531 at a slightly higher cost of £5600, would be required. Reducing the width of the 95%
532 confidence interval substantially below 1 µg N g^{-1} dry soil would result in a large cost
533 increase, with small marginal improvement on increasing the size of the network. For a NO_3^-
534 sensor network, given the spatial variation observed in the June sampling event, reducing the
535 width of the confidence interval to less than 0.5 µg N g^{-1} would require 10 loggers each with
536 12 sensors, at a cost of £22400.

537 An alternative approach is to optimise the sensor network design within the
538 constraints of a fixed budget. A budget of £5000 for a NO_3^- sensor network could provide a
539 single data logger with 15 sensors or 2 data loggers each with 5 sensors. This could be used
540 to provide a single logger with 15 sensors on each, or two loggers with 5 sensors on each.

The width of the confidence interval for these two options is 2.12 and 1.70 µg N g^{-1} dry soil respectively, so the second option is the rational choice.

4. Discussion

4.1. Spatial variation of soluble N at within-field scales

The dominance of short range variation (i.e. less than 2 m) for all the N forms may be attributed to the relatively random and uneven deposition of N from sheep excreta within the context of a broadly homogenous field. The proportion of the total accumulated variance attributed to the 1-cm scale was much larger for amino acid-N and NH_4^+-N than NO_3-N which may be related to their relative diffusion coefficients, interactions with the solid phase (Owen and Jones, 2001) and the rapid rate of amino acid turnover and mineralisation in this soil (Jones et al., 2004; Wilkinson et al., 2014).. Similar small-scale variation of NO_3^- in grazed pastures has been identified in previous studies, with semi-variograms exhibiting the range of spatial dependency of less than 5 m (White et al., 1987; Broeke et al., 1996; Wade et al., 1996; Bogaert et al., 2000), and a nugget variance of 60% (Bogaert et al., 2000). These results contrast with similar studies performed on arable soils, which were characterised by ranges of spatial dependencies for NO_3^- of greater than 39 m (Van Meirvenne et al., 2003; Haberle et al., 2004). The observed variation at larger spatial-scales, especially NO_3^-, could be due to the habit of sheep to frequent certain areas of the pasture such as paths, a drinking trough and areas of shade (Bogaert et al., 2000).

It is unlikely that the observed variation at the "aggregate" scale is driven by the deposition of sheep excreta. Previous studies of spatial variation in soil N, in the context of within-field scales, have not investigated variation over such small scales. This small-scale variation is likely due to the inherent micro-heterogeneity of soil properties, for example, the abundance of plant roots and mycorrhizal hyphae (Stoyan et al., 2000), availability of labile

566 organic matter (Parkin, 1987; Wachinger et al., 2000), earthworm channels and the
567 composition and abundance of the microbial community (Grundmann and Debouzie, 2000;
568 Nunan et al., 2002), which in turn will affect biogeochemical processes controlling soil N
569 concentrations.

570 There was also some suggestion of a spatio-temporal interaction as evidenced by
571 small differences in the spatial dependencies of the N forms between the June and July nested
572 sampling events. In the case of NO_3^-, the total accumulated variance was lower, with more of
573 the observed variance attributed to scales greater than 2 m for the July sampling. This change
574 may be attributed to the removal of sheep and the associated local inputs of N, combined with
575 N fertilisation of the field (60 kg N ha^{-1}) that occurred 3 weeks prior to the second nested
576 sampling event.

577

578 *4.2. Optimisation of designing a within-field soil N sensor network*

579 This study clearly demonstrates how nested sampling combined with geostatistical
580 analysis can be used to optimise the design of an in situ sensor network. Furthermore, given
581 knowledge of logger and sensor costs it is possible to rationalise planning decisions on a cost-
582 precision basis. Essentially, the shape of the curves within the optimisation graphs (Fig. 5)
583 reflects the observed spatial variation. The largest observed accumulated variance was for
584 NO_3^- from the June nested sampling. Consequently, a NO_3^- sensor network based on this
585 spatial variation would require a greater number of data loggers and sensors (with a resultant
586 cost increase), to achieve a desired level of precision when compared to a NO_3^- sensor
587 network based on the July nested sampling results and a NH_4^+ sensor network based on either
588 the June or July nested sampling results. The distribution of variance across the spatial scales
589 also affects the most efficient use of a specific budget, in terms of the choice of number of
590 data loggers and sensors used and the resulting width of the 95% confidence interval. For

example, a budget of £7000 could be used to purchase either 2 data loggers with 15 sensors on each or 3 data loggers with 5 sensors on each. Given the observed variation from the June nested sampling, the former of the above choices gives the most efficient design for NH_4^+ and the later for NO_3^-. This reflects the fact that, compared to NO_3^-, a greater proportion of the total accumulated variance for NH_4^+, is attributed to scales less than 2 m.

It is important to note that the data used for these calculations were derived from the nested sampling which used a soil corer of 1 cm diameter. As such, these calculations are based on the assumption that any given sensor used for the in-situ network would have a similar sized zone of influence. Results derived from the aggregate-scale sampling exhibited large variation at the sub 1-cm scale, which for NH_4^+ and NO_3^- was an order of magnitude larger than the 1-cm scale. This will have significant implications when using sensors, for which the measurement represents an integration over a sensor-soil contact area of diameter less than 1 cm, as they will be affected by the observed "aggregate" scale variation. If this variation is not considered when designing the sensor network, it is likely that the precision of the estimated field mean would be overestimated. To compensate for this, more local replication at the sub 1-cm scale and hence an increase in the size of the sensor network would be required for an acceptable level of precision to be achieved, resulting in increased costs. To explore optimisation of a network of such sensors, further sampling at the sub 1-cm scale would be required. Ideally this would involve a similar level of replication, across all scales, to that which was used in the July nested sampling campaign. This evidence may also be quite instructive for optimising sensor design, as sensors with larger sampling areas will encompass more of this small-scale variation.

Within this optimisation, no consideration has been made to the observed depth effects. The resulting estimate of the field mean derived from the sensor network would therefore, only be applicable to the 5-10 cm depth. Rooting depth, and therefore, nutrient

uptake, in fields adjacent to the study site has previously been observed to a depth of 30 cm (Jones et al., 2004) and a decrease in plant-available with depth has been observed in the study field (see supplementary information, Fig. S4). As such, any quantification of plant-available N derived from the sensor network should be adjusted for observed depth effects. In the case of cereals, which may root to depths in excess of 1.5 m, both topsoil and subsoil sensors will probably be required to avoid bias and gain a representative pattern of soluble N within the field. Logistically, however, the deployment of sensors in subsoils represents a significant challenge.

4.3. Potential for use of in-situ sensor networks within precision agriculture

Adoption of in-situ networks not only relies on the development of suitable sensors but also on evidence of the cost-benefit. Given the results of the sensor network optimisation, it is probable that uptake of this approach will incur significant costs; especially if a high level of precision is required. As such, it is likely that this approach will be limited initially to high value horticultural crops. As the cost of sensor and data logging technology continues to fall, adoption by arable and pastoral agriculture may increase. One key factor that will affect the cost of a sensor network is the required precision of the resulting estimated mean and this in turn may depend on how the generated data is used to improve fertiliser management. The creation of a decision support tool that will bring about improved NUE, and hence reduce input costs and/or increase profits, requires significant future research and is not an insignificant challenge.

It is important to consider how the approach used in this study could be applied to field exhibiting significant random and non-random (i.e. a gradient) large-scale variation. It is possible that the field could be split into management zones each with their own sensor network. These management zones could be delimitated on the basis of *a priori* knowledge of

641 variables that may affect or indicate soil N status such as topography (Kravchenko and
642 Bullock, 2000), soil type (Moral et al., 2011), yield variability (Diker et al., 2004) and
643 farmers knowledge (Fleming et al., 2000). Alternatively, proximal or remote sensing, such as
644 electromagnetic induction, may allow rapid and cost effective identification of large scale
645 heterogeneity of soil physical properties (Hedley et al., 2004; King et al., 2005). However,
646 the extent to which these variables correlate to soil N concentration is likely to be site specific
647 and so may require some ground truthing. A further broad question which needs to be
648 addressed with respect to management zones, is at what point does the magnitude and the
649 spatial-scale of soil N variation become sufficiently large enough to justify site-specific
650 agriculture?

651 The success of the approach used here to optimise a sensor network requires temporal
652 stability of spatial variation (Sylvester-Bradley et al., 1999; Shi et al., 2002). Given
653 significant spatio-temporal interaction, the results from any sensor network could no longer
654 be considered accurate or precise. In this study there was evidence of a slight spatio-temporal
655 interaction which was related to the removal of sheep from the field and the application of N
656 fertiliser. An alternative approach to that advocated here, would be the implementation of a
657 grid network, with sensor arrays at each node to account for small-scale soil variation. This
658 would enable temporal, large-scale spatial variation and their interaction to be monitored.
659 Kriging techniques could then be used to produce dynamic maps of soil N concentrations
660 which could be used to inform variable-rate fertiliser management. However, this approach is
661 likely to require significantly more sensing units and data loggers with a resulting cost
662 increase.

663

664 **5. Conclusions**

This study demonstrates how a network of in-situ soil N sensors could be efficiently designed and optimised on the basis of cost and precision. To achieve this, the spatial variation of plant available N – amino acids, NH_4^+ and NO_3^- – within the soil of a grazed grassland field was investigated using a nested sampling approach and geo-statistical analysis. Variation of all N forms at small scales (less than 2 m) was shown to be dominant, with further large variance evident at scales less than 1 cm. The observed variation was considered in line with previous work and was attributed to the random input of N to the soil via sheep excreta and the inherent heterogeneity of soil at the aggregate scale. Based on the observed spatial variance observed in the June nested sampling, and given a budget of £5000, the NO_3^- field mean could be estimated with a 95% confidence interval width of 1.70 µg N g^{-1} using 2 randomly positioned data loggers each with 5 sensors. Achieving a 95% confidence interval width substantially lower than 1.70 µg N g^{-1} would require significant extra cost. Adoption of in-situ sensor networks will depend upon the development of suitable low-cost sensors, demonstration of the cost-benefit and the construction of a decision support system that utilises the generated data to improve the NUE of fertiliser N management.

Acknowledgements

The authors would like to recognise the funding for this work provided by the UK Agriculture and Horticulture Development Board. RML's contribution appears with the permission of the Director of the British Geological Survey (NERC). RS would also like to thank Prof. A. J. Miller (John Innes Centre, Norwich) for his excellent PhD supervision, Llinos Hughes and Mark Hughes for all their help with field work at the Henfaes research station. Finally, the authors would like to acknowledge the reviewers' excellent contributions to this paper.

690 **References**

691 Adamchuk, V.I., Lund, E.D., Sethuramasamyraja, B., Morgan, M.T., Dobermann, A., Marx,
692 D.B., 2005. Direct measurement of soil chemical properties on-the-go using ion-
693 selective electrodes. Comput. Electron. Agric. 48: 272-294.

694 Adamchuk, V.I., Morgan, M.T., Ess, D.R., 1999. An automated sampling system for
695 measuring soil pH. Trans. ASAE. 42: 885-891.

696 Adsett, J.F., Thottan, J.A., Sibley, K.J., 1999. Development of an automated on-the-go soil
697 nitrate monitoring system. Appl. Eng. Agric. 15: 351-356.

698 Arrigan, D.W.M., 2004. Nanoelectrodes, nanoelectrode arrays and their applications.
699 Analyst. 129: 1157-1165.

700 Atmeh, M., Alcock-Earley, B.E., 2011. A conducting polymer/Ag nanoparticle composite as
701 a nitrate sensor. J. Appl. Electrochem. 41: 1341-1347.

702 Bogaert, N., Salomez, J., Vermoesen, A., Hofman, G., Van Cleemput, O., Van Meirvenne,
703 M., 2000. Within-field variability of mineral nitrogen in grassland. Biol. Fertility Soils.
704 32: 186-193.

705 Broeke, M., Groot, d.W., Dijkstra, J., 1996. Impact of excreted nitrogen by grazing cattle on
706 nitrate leaching. Soil Use Manage. 12: 190-198.

707 Cassman, K.G., Dobermann, A., Walters, D.T., 2002. Agroecosystems, nitrogen-use
708 efficiency, and nitrogen management. Ambio. 31: 132-140.

709 de Gruijter, J., Brus, D.J., Bierkens, M.F.P., Knotters, M., 2006. Sampling for natural
710 resource monitoring. Springer Science & Business Media. Springer-Verlag, Berlin.

711 Deen, B., Janovicek, K., Bruulsema, T., Lauzon, J., 2014. Predicting year-year, field level
712 variation in maize nitrogen fertilizer requirement. Cordovil, C. M. d. S., Ed.;
713 Proceedings of the 18th Nitrogen Workshop - The nitrogen challenge: building a
714 blueprint for nitrogen use efficiency and food security. Lisbon, Portugal. pp 65-67.

715 Defra, 2010. Fertiliser Manual (RB209). 8th Edition. The Stationary Office (TSO). Norwich,
716 UK.
717 Diacono, M., Rubino, P., Montemurro, F., 2013. Precision nitrogen management of wheat. A
718 review. Agron. Sustain. Dev. 33: 219-241.
719 Diker, K., Heermann, D., Brodahl, M., 2004. Frequency analysis of yield for delineating
720 yield response zones. Precis. Agric. 5: 435-444.
721 Dobermann, A., Blackmore, S., Cook, S.E., Adamchuk, V.I., 2004. Precision Farming:
722 Challenges and Future Directions. Fisher, T., Turner, N., Angus, J., McIntyre, L. and
723 Rob, M., Eds.; New directions for a diverse planet. Proceedings of the 4th International
724 Crop Science Congress, Brisbane, Australia, 26 Sep – 1 Oct 2004. The Regional
725 Institute, Gosford, Australia. pp 217-237.
726 Fleming, K., Westfall, D., Wiens, D., Brodahl, M., 2000. Evaluating farmer defined
727 management zone maps for variable rate fertilizer application. Precis. Agric. 2: 201-215.
728 Grundmann, G., Debouzie, D., 2000. Geostatistical analysis of the distribution of NH_4^+ and
729 NO_2^--oxidizing bacteria and serotypes at the millimeter scale along a soil transect.
730 FEMS Microbiol. Ecol. 34: 57-62.
731 Haberle, J., Kroulik, M., Svoboda, P., Lipavsky, J., Krejcova, J., Cerhanova, D., 2004. The
732 spatial variability of mineral nitrogen content in topsoil and subsoil. Plant Soil Environ.
733 50: 425-433.
734 Hedley, C., Yule, I., Eastwood, C., Shepherd, T., Arnold, G., 2004. Rapid identification of
735 soil textural and management zones using electromagnetic induction sensing of soils.
736 Soil Res. 42: 389-400.
737 Inselsbacher, E., 2014. Recovery of individual soil nitrogen forms after sieving and
738 extraction. Soil Biol. Biochem. 71: 76-86.

739 Ito, S., Baba, K., Asano, Y., Takesako, H., Wada, H., 1996. Development of a nitrate ion-
740 selective electrode based on an urushi matrix membrane and its application to the direct
741 measurement of nitrate-nitrogen in upland soils. Talanta. 43: 1869-1881.
742 Jones, D. L.; Owen, A. G.; Farrar, J. F. 2002. Simple method to enable the high resolution
743 determination of total free amino acids in soil solutions and soil extracts. Soil Biol.
744 Biochem. 34: 1893-1902.
745 Jones, D.L., Shannon, D., Murphy, D., Farrar, J., 2004. Role of dissolved organic nitrogen
746 (DON) in soil N cycling in grassland soils. Soil Biol. Biochem. 36: 749-756.
747 Jones, D.L., Willett, V.B., 2006. Experimental evaluation of methods to quantify dissolved
748 organic nitrogen (DON) and dissolved organic carbon (DOC) in soil. Soil Biol.
749 Biochem. 38: 991-999.
750 Keeney, D.R., 1982. Nitrogen - availability indices. Page, A. L., Ed. Methods of Soil
751 Analysis. Part 2, 2nd ed. Chemical and Microbiological Properties. SSSA and ASA.
752 Madison, WI, USA, pp 711-733.
753 Kim, H., Sudduth, K.A., Hummel, J.W., 2009. Soil macronutrient sensing for precision
754 agriculture. J. Environ. Monit. 11: 1810-1824.
755 King, J., Dampney, P., Lark, R., Wheeler, H., Bradley, R., Mayr, T., 2005. Mapping potential
756 crop management zones within fields: use of yield-map series and patterns of soil
757 physical properties identified by electromagnetic induction sensing. Precis. Agric. 6:
758 167-181.
759 Kravchenko, A.N., Bullock, D.G., 2000. Correlation of corn and soybean grain yield with
760 topography and soil properties. Agron. J. 92: 75-83.
761 Lark, R.M., 2011. Spatially nested sampling schemes for spatial variance components: Scope
762 for their optimization. Comput. Geosci. 37: 1633-1641.

763 Lark, R.M., Cullis, B., 2004. Model-based analysis using REML for inference from
764 systematically sampled data on soil. Eur. J. Soil Sci. 55: 799-813.
765 Lark, R.M., Wheeler, H. 2003. Experimental and analytical methods for studying within-field
766 variation of crop responses to inputs. In: J.V. Stafford, A., Werner (Eds). Precision
767 Agriculture. Proceedings of the 4th European Conference on Precision Agriculture.
768 Wageningen, the Netherlands. pp 341-346.
769 Le Goff, T., Braven, J., Ebdon, L., Chilcott, N., Scholefield, D., Wood, J., 2002. An accurate
770 and stable nitrate-selective electrode for the in situ determination of nitrate in
771 agricultural drainage waters. Analyst. 127: 507-511.
772 Le Goff, T., Braven, J., Ebdon, L., Scholefield, D., 2003. Automatic continuous river
773 monitoring of nitrate using a novel ion-selective electrode. J. Environ. Monit. 5: 353-
774 358.
775 McBratney, A., Whelan, B., Ancev, T., Bouma, J. 2005. Future directions of precision
776 agriculture. Precis. Agric. 6: 7-23.
777 Miranda, K. M.; Espey, M. G.; Wink, D. A. 2001. A rapid, simple spectrophotometric
778 method for simultaneous detection of nitrate and nitrite. Nitric Oxide-Biol. Ch. 5: 62-71.
779 Moral, F.J., Terrón, J.M., Rebollo, F.J., 2011. Site-specific management zones based on the
780 Rasch model and geostatistical techniques. Comput. Electron. Agric. 75: 223-230.
781 Müller, B., Reinhardt, M., Gächter, R. 2003. High temporal resolution monitoring of
782 inorganic nitrogen load in drainage waters. J. Environ. Monit. 5: 808-812.
783 Mulvaney, R. L. 1996. Nitrogen - Inorganic Forms. Sparks, D. L., Ed. Methods of Soil
784 Analysis. Part 3. Chemical Methods. Soil Science Society of America. pp 1123-1184.
785 Murphy, J., Riley, J., 1962. A modified single solution method for the determination of
786 phosphate in natural waters. Anal. Chim. Acta. 27: 31-36.

787 Nunan, N., Wu, K., Young, I., Crawford, J., Ritz, K., 2002. In situ spatial patterns of soil
788 bacterial populations, mapped at multiple scales, in an arable soil. Microb. Ecol. 44: 296-
789 305.
790 Owen, A., Jones, D., 2001. Competition for amino acids between wheat roots and rhizosphere
791 microorganisms and the role of amino acids in plant N acquisition. Soil Biol. Biochem.
792 33: 651-657.
793 Parkin, T.B., 1987. Soil microsites as asource of denitrification variability. Soil Sci. Soc. Am.
794 J. 51: 1194-1199.
795 Pierce, F.J., Nowak, P., 1999. Aspects of precision agriculture. Adv. Agron. 67: 1-85.
796 Raun, W.R., Johnson, G.V., 1999. Improving nitrogen use efficiency for cereal production.
797 Agron. J. 91: 357-363.
798 Robertson, G.P., Vitousek, P.M., 2009. Nitrogen in agriculture: balancing the cost of an
799 essential resource. Annu. Rev. Env. Resour. 34: 97-125.
800 Rousk, J., Jones, D.L., 2010. Loss of low molecular weight dissolved organic carbon (DOC)
801 and nitrogen (DON) in H_2O and 0.5 M K_2SO_4 soil extracts. Soil Biol. Biochem. 42:
802 2331-2335.
803 Schirrmann, M., Domsch, H., 2011. Sampling procedure simulating on-the-go sensing for
804 soil nutrients. J. Plant Nutr. Soil. Sc. 174: 333-343.
805 Schwarz, J., Kaden, H., Pausch, G. 2000. Development of miniaturized potentiometric nitrate
806 and ammonium selective electrodes for applications in water monitoring. Fresen. J. Anal.
807 Chem. 367: 396-398.
808 Selbie, D.R., Buckthought, L.E., Shepherd, M.A., 2015. The challenge of the urine patch for
809 managing nitrogen in grazed pasture systems. Adv. Agron. 129: 229-292.
810 Shahandeh, H., Wright, A.L., Hons, F.M., Lascano, R.J., 2005. Spatial and temporal variation
811 of soil nitrogen parameters related to soil texture and corn yield. Agron. J. 97: 772-782.

812 Shanahan, J.F., Kitchen, N.R., Raun, W.R., Schepers, J.S., 2008. Responsive in-season
813 nitrogen management for cereals. Comput. Electron. Agric. 61: 51-62.
814 Shaw, R., Williams, A.P., Miller, A., Jones, D.L., 2014. Developing an in situ sensor for real
815 time monitoring of soil nitrate concentration. Hopkins, A., Collins, R. P., Fraser, M. D.,
816 King, V. R., Lloyd, D. C., Moorby, J. M. and Robson, P. R. H., Eds.; EGF at 50: The
817 future of European grasslands. Proceedings of the 25th General Meeting of the European
818 Grassland Federation, Aberystwyth, Wales, 7-11 September 2014. IBERS, Aberystwyth
819 University, UK. pp 273-275.
820 Shaw, R., Williams, A.P., Miller, A., Jones, D.L., 2013. Assessing the potential for ion
821 selective electrodes and dual wavelength UV spectroscopy as a rapid on-farm
822 measurement of soil nitrate concentration. Agriculture. 3: 327-341.
823 Shi, Z., Wang, K., Bailey, J., Jordan, C., Higgins, A., 2002. Temporal changes in the spatial
824 distributions of some soil properties on a temperate grassland site. Soil Use Manage. 18:
825 353-362.
826 Sibley, K.J., Astatkie, T., Brewster, G., Struik, P.C., Adsett, J.F., Pruski, K., 2009. Field-scale
827 validation of an automated soil nitrate extraction and measurement system. Precis. Agric.
828 10: 162-174.
829 Sinfield, J.V., Fagerman, D., Colic, O., 2010. Evaluation of sensing technologies for on-the-
830 go detection of macro-nutrients in cultivated soils. Comput. Electron. Agric. 70: 1-18.
831 Sparks, D.L., Ed.; 1996. Methods of Soil Anaylsis Part 3 - Chemical Methods. SSSA book
832 series No. 5. American Society of Agronomy, Madison, WI.
833 Stoyan, H., De-Polli, H., Böhm, S., Robertson, G.P., Paul, E.A., 2000. Spatial heterogeneity
834 of soil respiration and related properties at the plant scale. Plant Soil. 222: 203-214.

Sutton, M.A., Howard, C.M., Erisman, J.W., Billen, G., Bleeker, A., Grennfelt, P., van Grinsven, H., Grizzetti, B., Eds.; 2011. The European nitrogen assessment. Sources, effects and policy perspectives. Cambridge University Press, Cambridge, UK.

Sylvester-Bradley, R., Lord, E., Sparks, D.L., Scott, R.K., Wiltshire, J., Orson, J., 1999. An analysis of the potential of precision farming in Northern Europe. Soil Use Manage. 15: 1-8.

Tilman, D., Balzer, C., Hill, J., Befort, B.L., 2011. Global food demand and the sustainable intensification of agriculture. Proc. Natl. Acad. Sci. U. S. A. 108: 20260-20264.

Venables, W. N. & Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth Edition. Springer, New York

Verbeke, G., Molenberghs, G., 2000. Linear Mixed Models for Longitudinal Data. Springer, New York, USA. pp. 19-29.

van Meirvenne, M., Maes, K., Hofman, G. 2003. Three-dimensional variability of soil nitrate-nitrogen in an agricultural field. Biol. Fertility Soils. 37: 147-153.

Voroney, R.P., Brookes, P.C., Beyaert, R.P., 2008. Soil microbial biomass C, N, P and S . Carter, M.R., Gregorich, E.G., Eds. Soil sampling and methods of analysis, 2nd edn. CRC Press. FL, USA, pp 637-651.

Wachinger, G., Fiedler, S., Zepp, K., Gattinger, A., Sommer, M., Roth, K., 2000. Variability of soil methane production on the micro-scale: spatial association with hot spots of organic material and Archaeal populations. Soil Biol. Biochem. 32: 1121-1130.

Wade, S.D., Foster, I.D., Baban, S.M., 1996. The spatial variability of soil nitrates in arable and pasture landscapes: Implications for the development of geographical information system models of nitrate leaching. Soil Use Manage. 12: 95-101.

Webster, R., Lark, R.M., 2013. Field sampling for environmental science and management. Routledge, Abingdon, UK.

White, R., Haigh, R.A., Macduff, J., 1987. Frequency distributions and spatially dependent variability of ammonium and nitrate concentrations in soil under grazed and ungrazed grassland. Fert. Res. 11: 193-208.

Wilkinson, A., Hill, P.W., Farrar, J.F., Jones, D.L., Bardgett, R.D., 2014. Rapid microbial uptake and mineralization of amino acids and peptides along a grassland productivity gradient. Soil Biol. Biochem. 72: 75-83.

866 **Figure legends**

867 **Figure 1.** The optimised sampling design of a mainstation used to perform spatial nested
868 sampling in a 1.9 ha grassland field. Distances between sampling points were fixed but
869 angles were randomized, with the exception of the 2 m vectors.

870 **Figure 2.** Accumulated variance components of the Box-Cox transformed data (Panel A)
871 from the finest to coarsest spatial scale, derived from the June nested sampling results
872 (before fertiliser addition). Panel B shows median absolute deviation from the median
873 (MAD) of differences over the nested spatial intervals on the original scale of
874 measurement. Source is the spatial-component in meters, with M and S representing the
875 between-mainstation and between-strata components respectively.

876 **Figure 3.** Accumulated variance components of the Box-Cox transformed data (Panel A)
877 from the finest to coarsest scale, derived from the July nested sampling results (after
878 fertiliser addition). Panel B shows median absolute deviation from the median (MAD)
879 of differences over the nested spatial intervals on the original scale of measurement.
880 Source is the spatial-component in meters, with M and S representing the between-
881 mainstation and between-strata components respectively.

882 **Figure 4.** Width of the 95% confidence interval for alternative sensor network designs of
883 different cost computed to facilitate monitoring of soil N in a 1.9 ha grassland field.
884 Values are computed from variance components from nested sampling of nitrate in (a)
885 June (before fertiliser addition) and (b) July (after fertiliser addition) and of ammonium
886 in (c) June (before fertiliser addition) and (d) July (after fertiliser addition) and on the
887 basis of unit costs for a sensor and a data logger of £200 and £2000 respectively. Note
888 that the arrays comprise 1-10 loggers and a maximum of 15 sensors per logger. To
889 allow a common range of values on the ordinates of these graphs, and to facilitate
890 interpretation, arrays with fewer than five sensors in total have been excluded from

891 Figure 4(a) and arrays with fewer than three sensors have been excluded from Figures
892 4(b-d).

Table 1

Background properties of the agricultural grassland Eutric Cambisol used in the study. Values represent means ± SEM (n = 4). All soil values are expressed on a dry weight soil basis.

Site property	Mean ± SEM
pH	6.57 ± 0.05
EC (µS cm^{-1})	26.5 ± 1.0
Basal soil respiration (mg CO_2 kg^{-1} h^{-1})	12.61 ± 1.04
Total soil C (g C kg^{-1})	25.35 ± 1.47
Total soil N (g N kg^{-1})	2.95 ± 0.06
Soil C:N	8.62 ± 0.64
DOC (mg C kg^{-1})	70.08 ± 2.57
DON (mg N kg^{-1})	10.48 ± 1.07
Mineralisable N (mg N kg^{-1} d^{-1})	3.92 ± 0.54
Microbial C (g C kg^{-1})	1.03 ± 0.10
Microbial N (g N kg^{-1})	0.16 ± 0.01
Exchangeable Ca (mg Ca kg^{-1})	501 ± 122
Exchangeable K (mg K kg^{-1})	46.05 ± 12.61
Exchangeable Na (mg Na kg^{-1})	25.43 ± 5.13
Available P (mg P kg^{-1})	7.38 ± 2.02
Above ground biomass (t DM ha^{-1})	1.56 ± 0.14

Table 2

Summary statistics describing the spatial variability of soluble N (μg N g^{-1}) derived from the nested sampling of a grassland soil prior to the application of N fertiliser. Alongside the raw data, an estimate of the Box-Cox transformation parameter (λ) is also provided.

Variable	Mean	Median	Minimum	Maximum	Skewness	λ
Nitrate	1.71	1.10	0.29	22.51	5.41	-0.426
Ammonium	1.87	1.27	0.29	80.49	12.82	-0.541
Amino acid	1.44	1.39	0.65	5.20	3.37	-0.492

Table 3

Variance components for the (Box-Cox transformed) variables and associated Wald tests describing the spatial variability of soluble N derived from the nested sampling of a grassland soil prior to the application of N fertiliser. The Wald statistic and associated p-value describe differences between the two sampling days. Those variance components marked with an asterisk are ones which caused an increase in AIC if they were dropped from the model (finest scale cannot be dropped).

Variable	Variance component						Wald statistic	*p*-value
	σ^2_s	σ^2_m	σ^2_2	$\sigma^2_{0.5}$	$\sigma^2_{0.1}$	$\sigma^2_{0.01}$		
Nitrate	0.0629*	0.0362*	0.0	0.0795*	0.0937*	0.0628	0.001	0.974
Ammonium	0.0087	0.0121*	0.0078	0.00008	0.0153	0.0751	6.8	0.009
Amino acid	0.0058	0.0035*	0.0	0.0	0.0124*	0.0307	1.89	0.17

Table 4

Summary statistics describing the spatial variability of soluble N (μg N g^{-1}) derived from the nested sampling of a grassland soil after the application of N fertiliser. Alongside the raw data, an estimate of the Box-Cox transformation parameter (λ) is also provided.

Variable	Mean	Median	Minimum	Maximum	Skewness	λ
Nitrate	1.36	1.25	0.26	3.45	0.89	0.302
Ammonium	1.96	1.71	0.26	9.88	3.28	-0.424
Amino acid	1.25	1.18	0.56	4.40	2.58	-0.481

Table 5

Variance components for the (Box-Cox transformed) variables and associated Wald tests describing the spatial variability of soluble N derived from the nested sampling of a grassland soil after the application of N fertiliser. The Wald statistic and associated *p*-value describe differences between the two sampling days. Those variance components marked with an asterisk are ones which caused an increase in AIC if they were dropped from the model (finest scale cannot be dropped).

Variable	Variance component							Wald statistic	*p*-value
	σ^2_s	σ^2_m	σ^2_2	$\sigma^2_{0.5}$	$\sigma^2_{0.1}$	$\sigma^2_{0.01}$	σ^2_ε		
Nitrate	0.0	0.0638*	0.0	0.0052	0.049*	0.031*	0.0131	7.89	0.005
Ammonium	0.0039	0.0069*	0.0	0.0	0.015	0.045*	0.0109	15.43	8.60×10^{-15}
Amino acid	0.002	0.0241*	0.0025	0.0	0.0086*	0.0199*	0.0103	0.708	0.4

Table 6

Summary statistics describing the aggregate-scale variability of soluble N (μg N g^{-1}) within a grassland soil.

Variable	Mean	Median	Minimum	Maximum	Skewness
Nitrate	1.20	1.04	0.19	3.13	0.80
Ammonium	2.00	1.78	0.30	5.85	1.24
Amino acid	1.56	1.50	0.77	2.69	0.49

Table 7

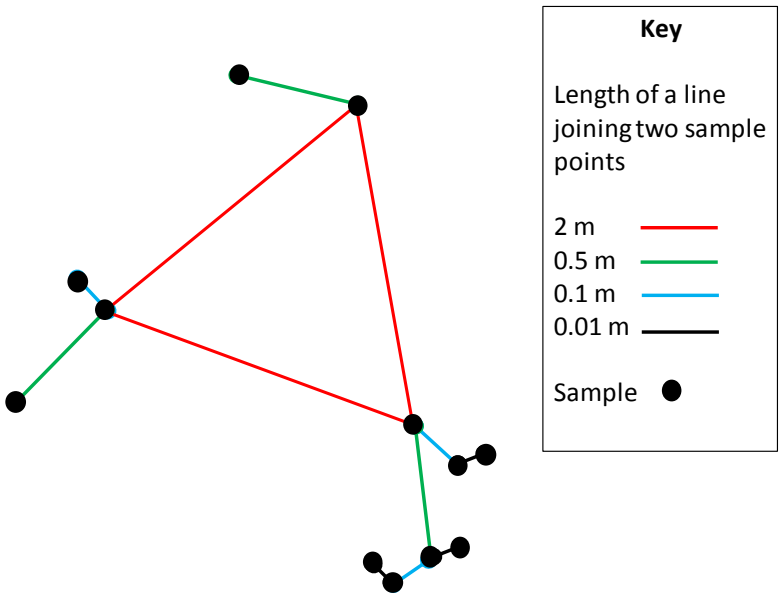
Variance components for the (log-transformed) variables describing the aggregate-scale spatial variability of soluble N in a grassland soil. Those variance components marked with an asterisk are ones which caused an increase in AIC if they were dropped from the model (finest scale cannot be dropped).

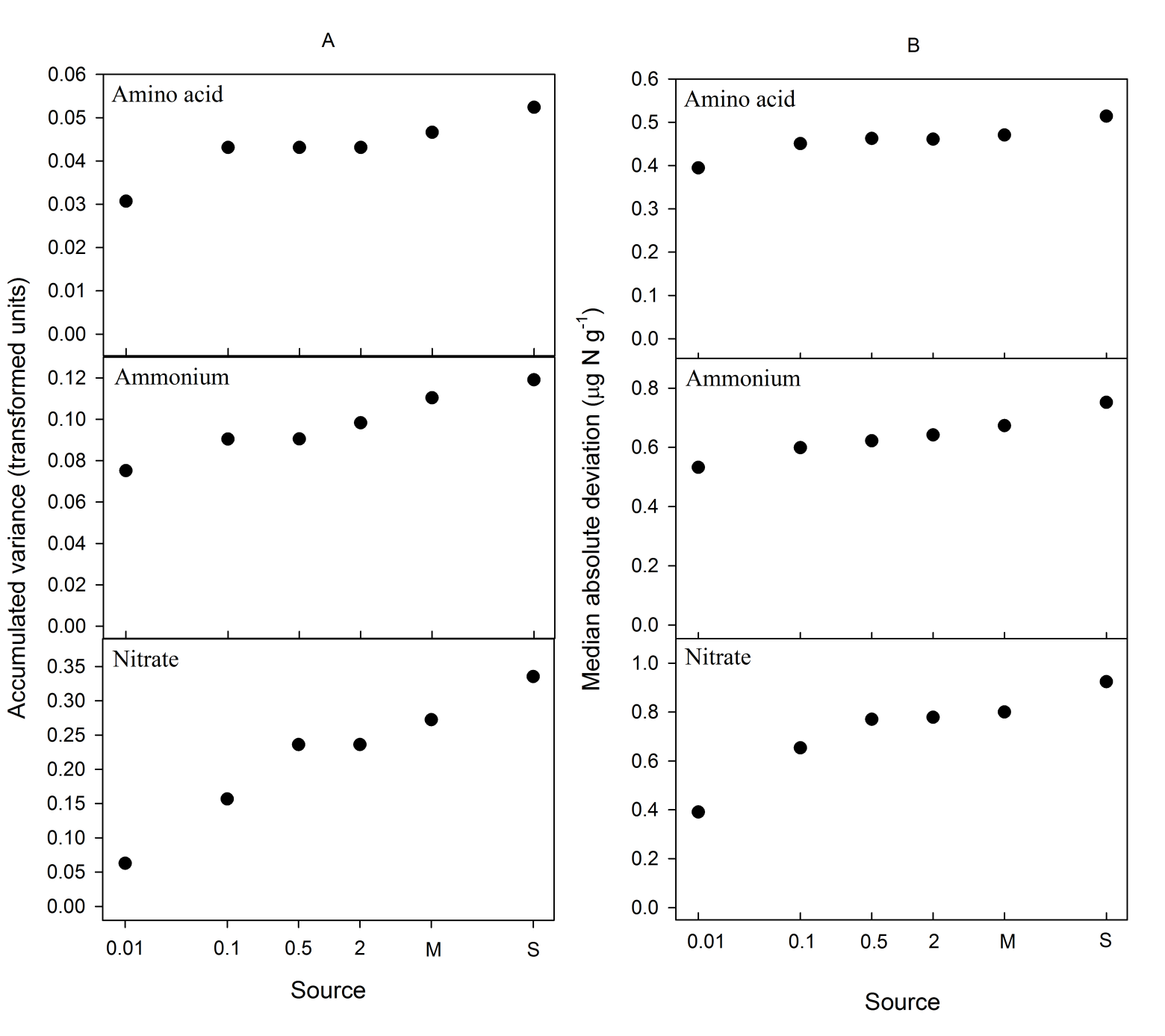
Variable	Variance component			
	σ^2_s	σ^2_p	σ^2_c	σ^2_a
Nitrate	0.0	0.0	0.072*	0.295
Ammonium	0.0	0.0293	0.003	0.3074
Amino acid	0.0	0.0031	0.0132*	0.0321

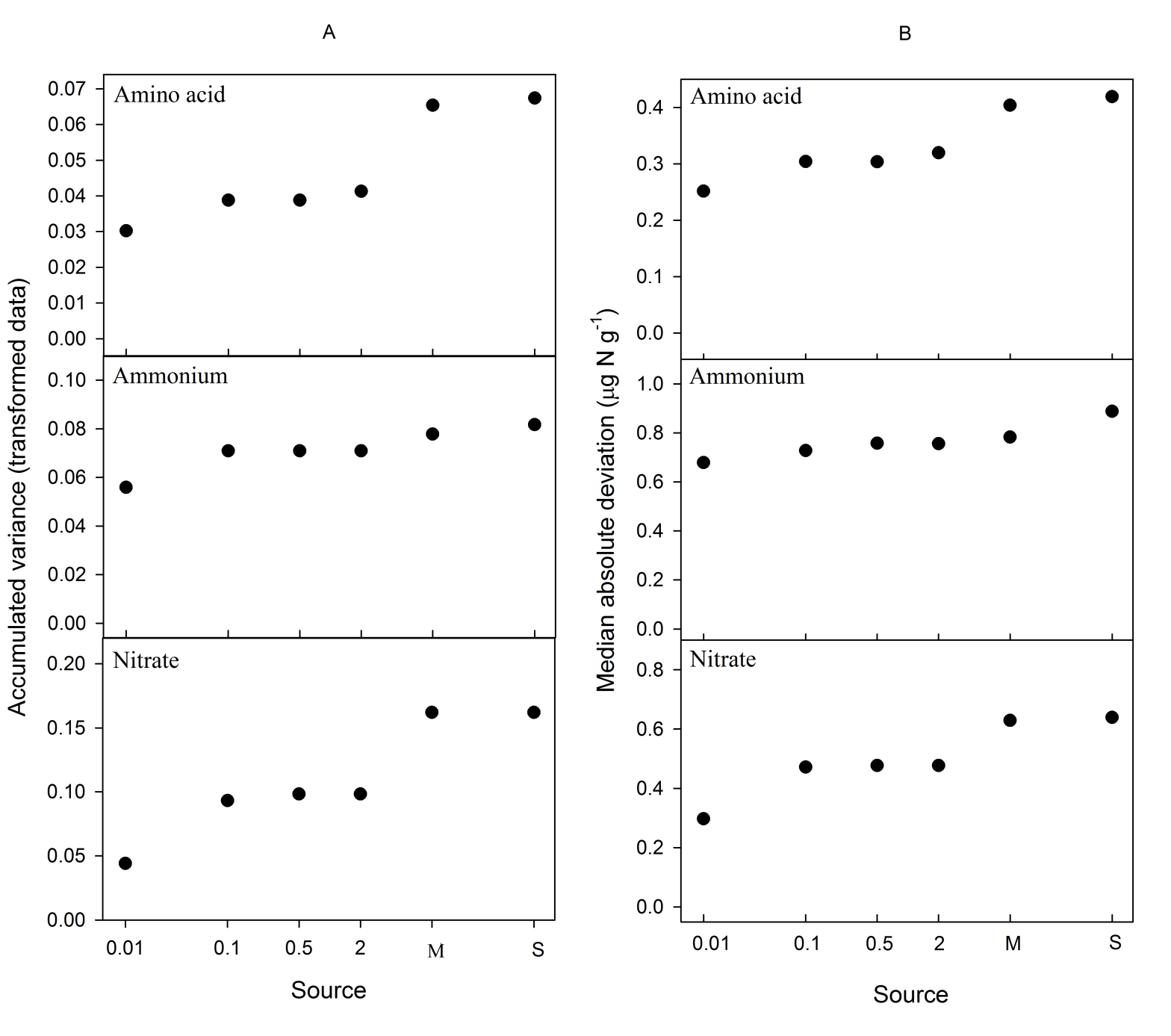
Table 8

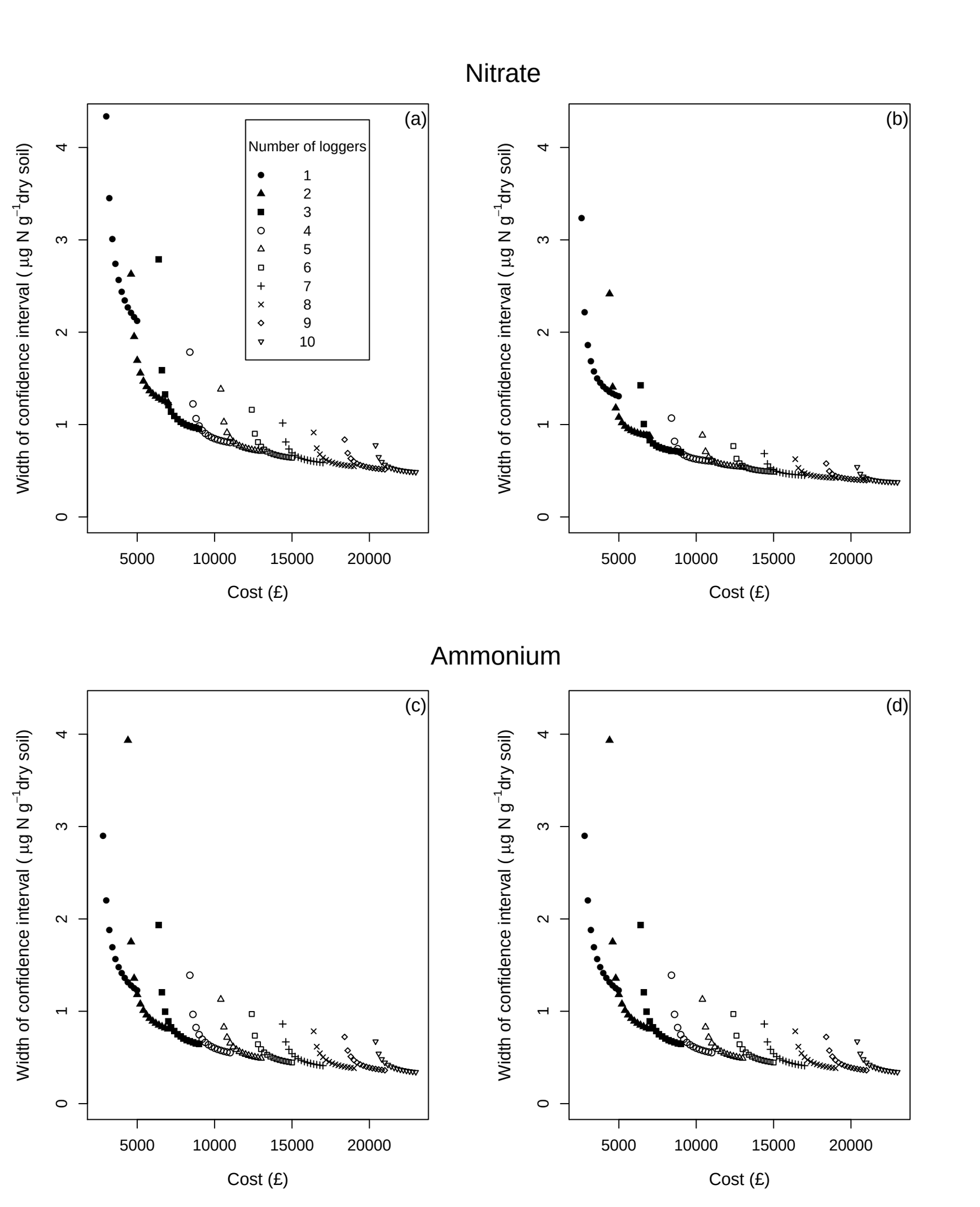
Median absolute deviations from the median (MAD) describing the aggregate-scale spatial variability of soluble N (μg N g^{-1}) in a grassland soil. Comparisons are nested, so the stratum-scale MAD includes the pair, core and aggregate-scale.

Variable	MAD (μg N g^{-1})			
	Aggregate	Core	Pair	Stratum
Nitrate	0.75	0.85	0.85	0.85
Ammonium	1.27	1.30	1.34	1.35
Amino acid	0.39	0.45	0.46	0.46


Key
Length of a line joining two sample points
2 m
0.5 m
0.1 m
0.01 m
Sample

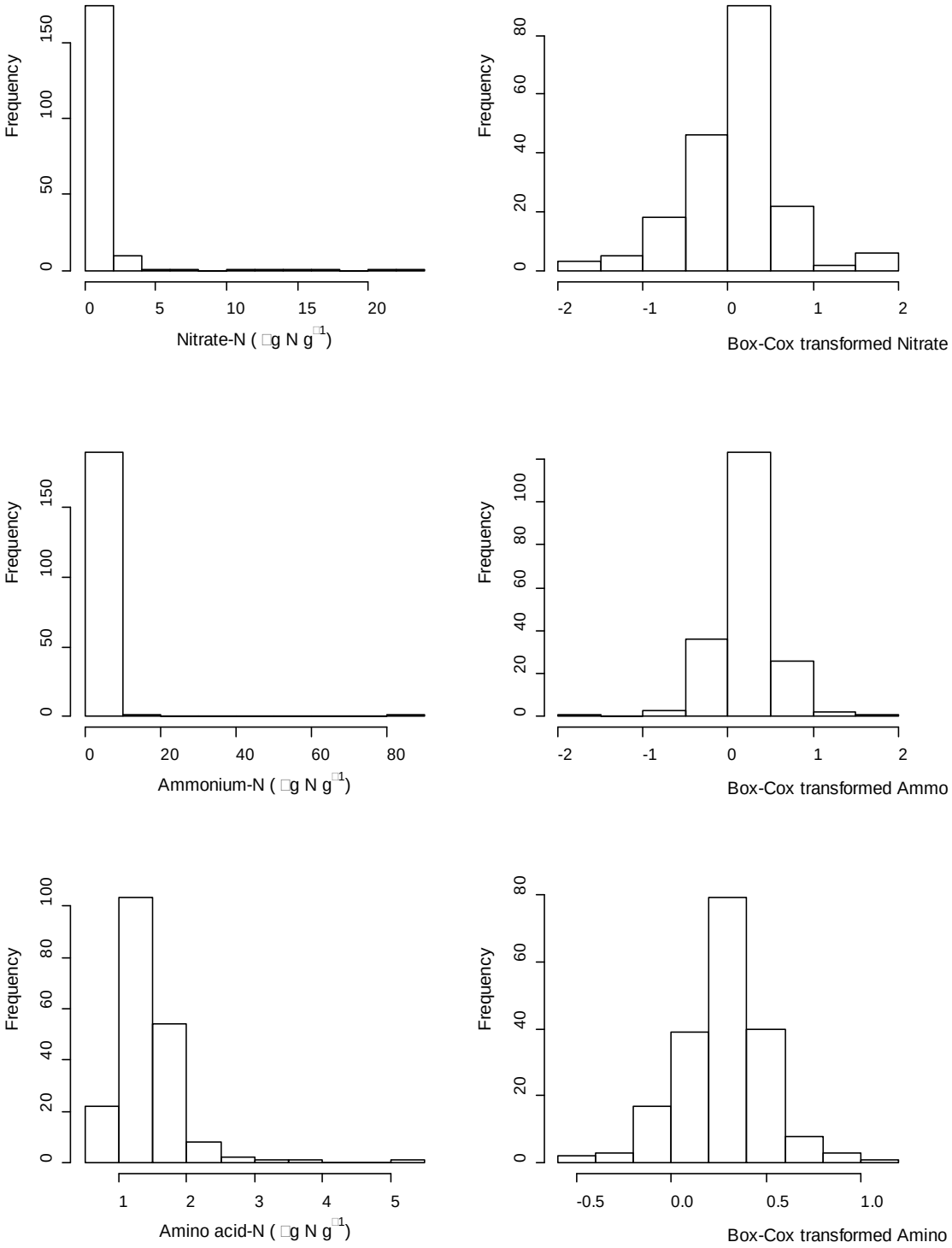

A
B
Amino acid
Ammonium
Nitrate
Accumulated variance (transformed units)
Median absolute deviation (μg N g-1)
0.01
0.1
0.5
2
M
S
Source


A
B
Amino acid
Ammonium
Nitrate
Accumulated variance (transformed data)
Median absolute deviation (μg N g-1)
0.01
0.1
0.5
2
M
S
Source


Nitrate
(a)
(b)
Number of loggers
1
2
3
4
5
6
7
8
9
10
Width of confidence interval (µg N g^{-1}dry soil)
Cost (£)
0
1
2
3
4
5000
10000
15000
20000
Ammonium
(c)
(d)

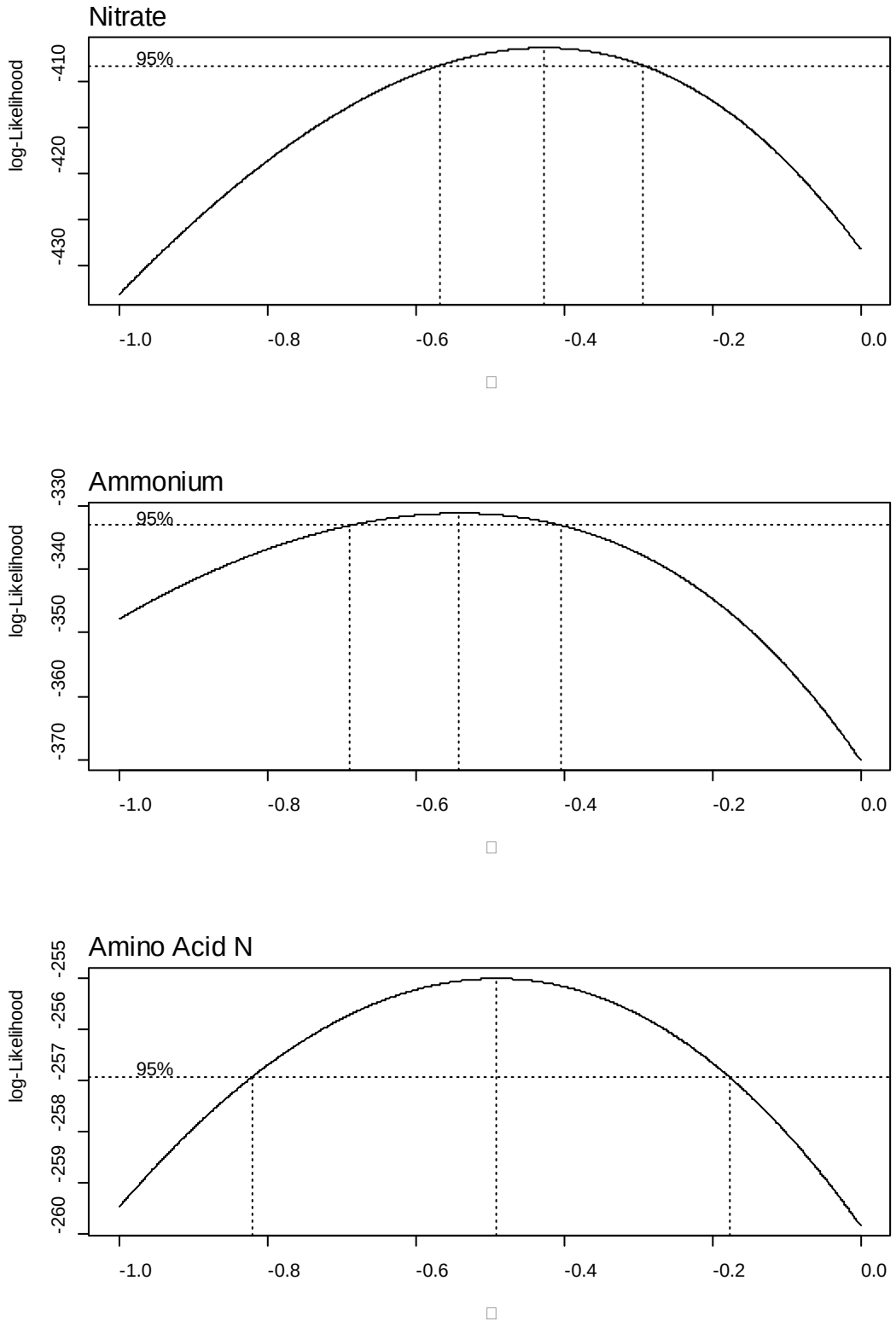
Supplementary information for Shaw et al. 2016, “Characterising the within-field scale spatial variation of nitrogen in a grassland soil to inform the efficient design of in-situ nitrogen sensor networks for precision agriculture”.

Figure S1. Histograms of the absolute (µg N g^{-1}) and Box-Cox transformed units of soil nitrate, ammonium and amino acid concentrations from the June nested sampling prior to fertiliser addition sampling (*n* = 192 for each N form).



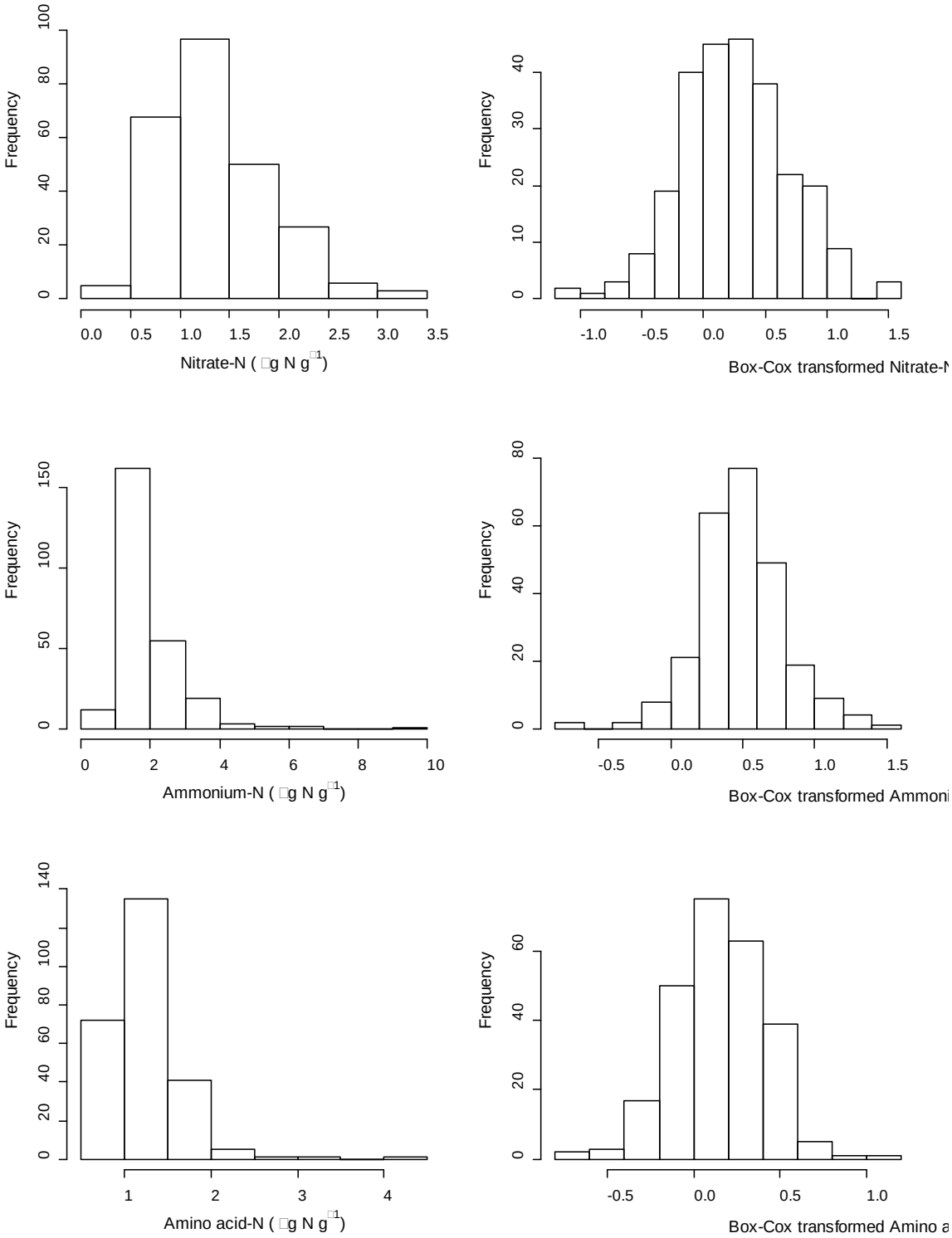
Supplementary information for Shaw et al. 2016, “Characterising the within-field scale spatial variation of nitrogen in a grassland soil to inform the efficient design of in-situ nitrogen sensor networks for precision agriculture”.

Figure S2. Profile likelihood plot for the λ parameter of the Box-Cox transformation for soil nitrate, ammonium and amino acid concentrations from the June nested sampling prior to fertiliser addition.



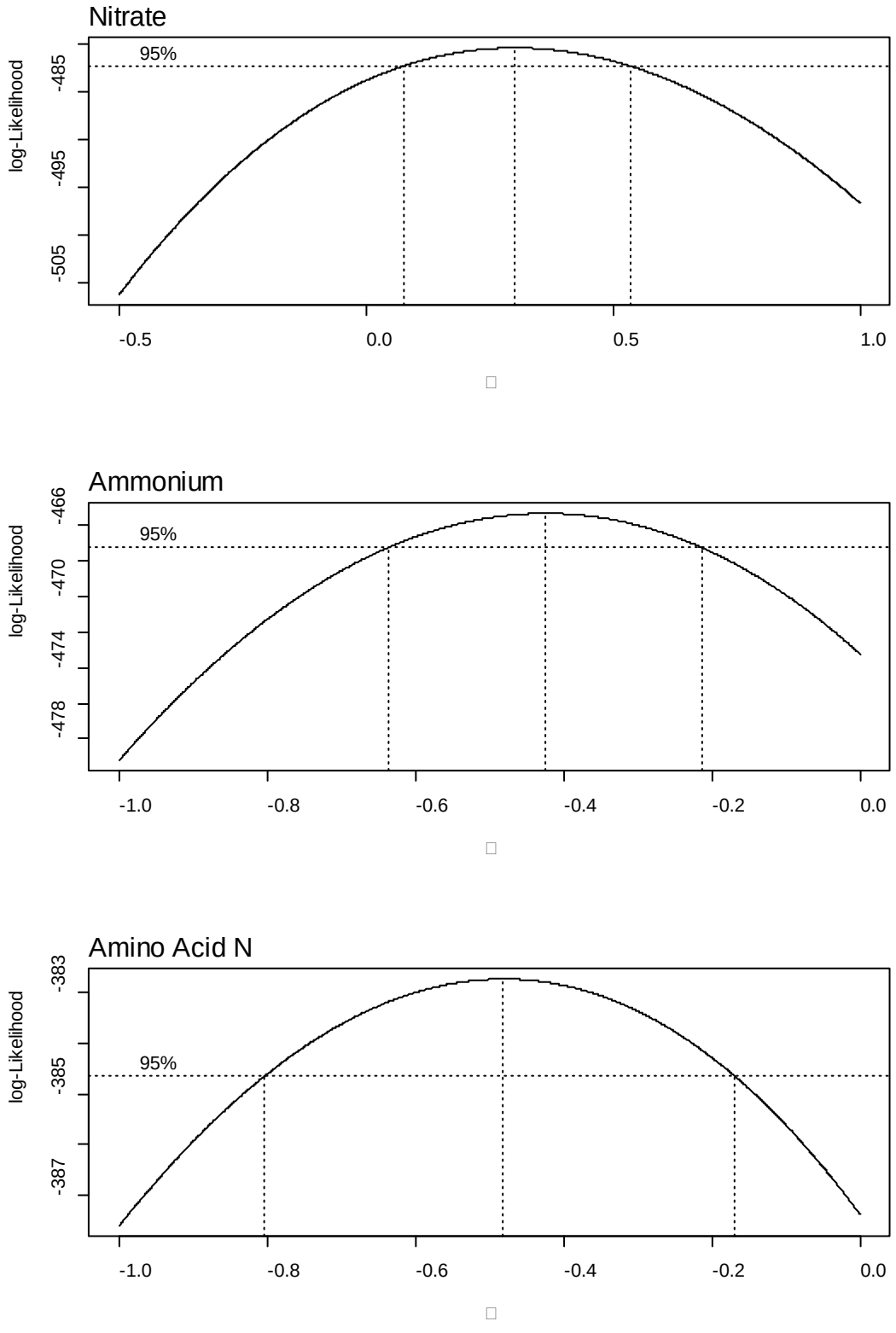
Supplementary information for Shaw et al. 2016, “Characterising the within-field scale spatial variation of nitrogen in a grassland soil to inform the efficient design of in-situ nitrogen sensor networks for precision agriculture”.

Figure S3. Histograms of the absolute (µg N g^{-1}) and Box-Cox transformed units of soil nitrate, ammonium and amino acid concentrations from the July nested sampling following fertiliser addition sampling (*n* = 192 for each N form).



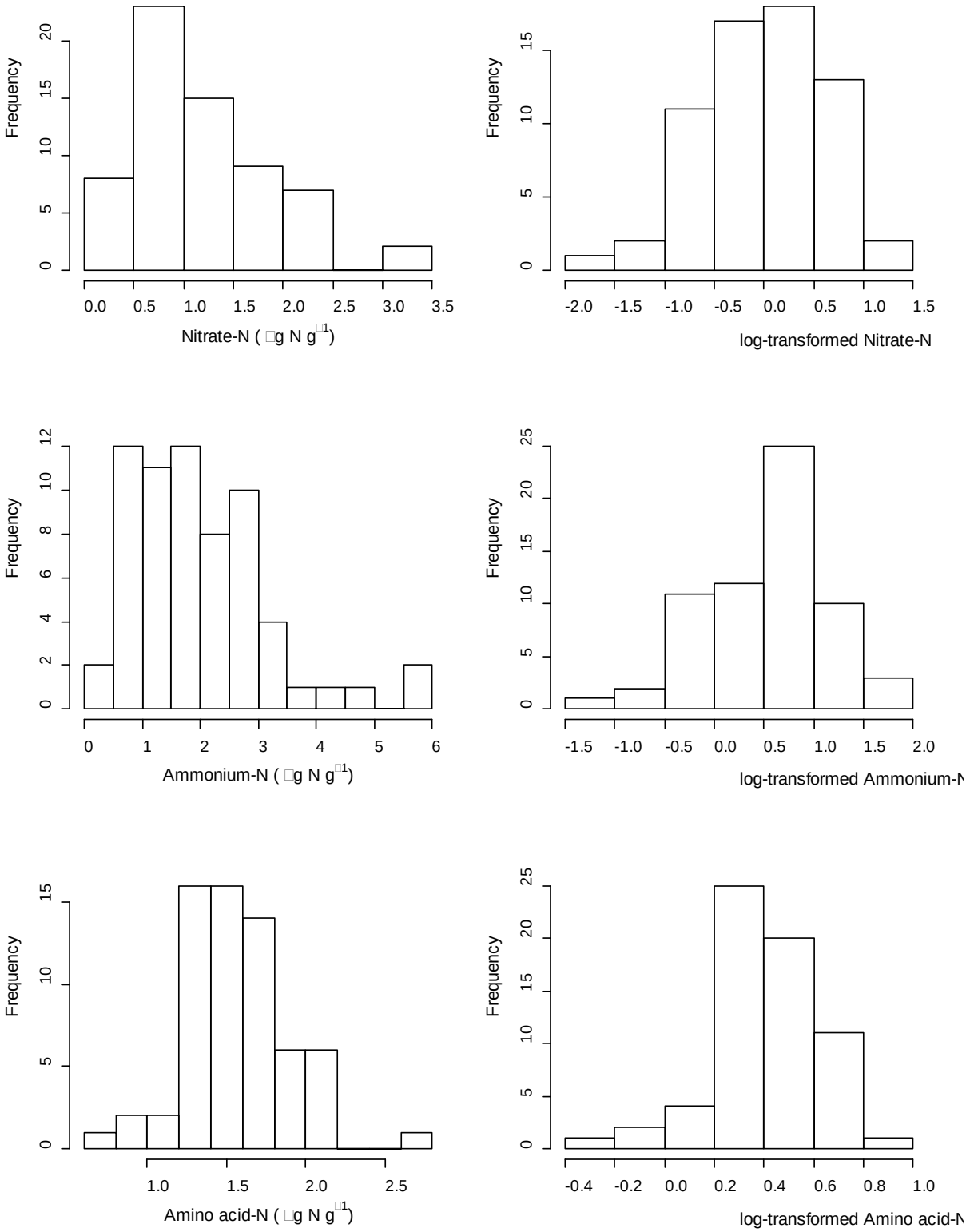
Supplementary information for Shaw et al. 2016, “Characterising the within-field scale spatial variation of nitrogen in a grassland soil to inform the efficient design of in-situ nitrogen sensor networks for precision agriculture”.

Figure S4. Profile likelihood plot for the λ parameter of the Box-Cox transformation for soil nitrate, ammonium and amino acid concentrations from the July nested sampling prior to fertiliser addition.



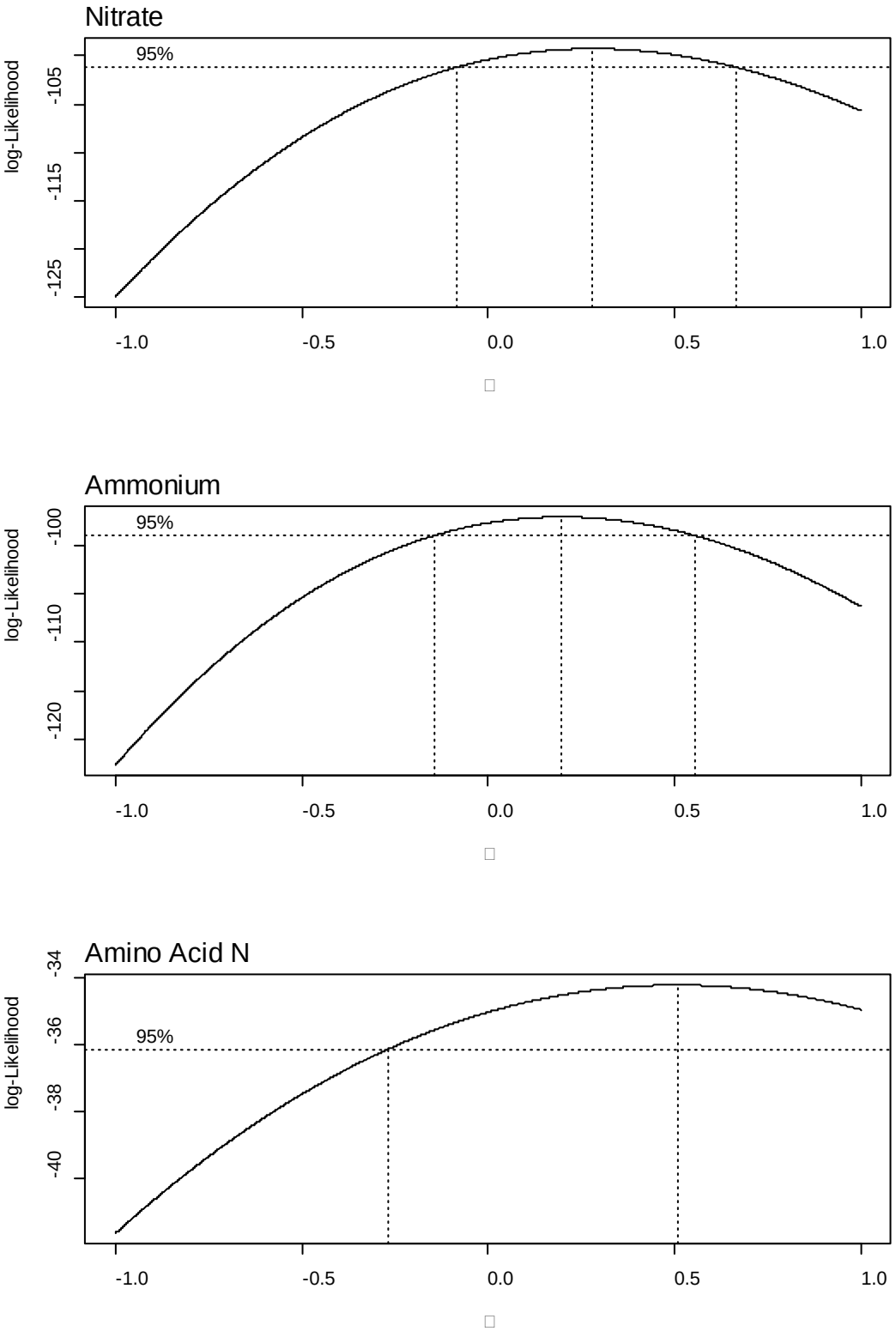
Supplementary information for Shaw et al. 2016, “Characterising the within-field scale spatial variation of nitrogen in a grassland soil to inform the efficient design of in-situ nitrogen sensor networks for precision agriculture”.

Figure S5. Histograms of the absolute (µg N g^{-1}) and Box-Cox transformed units of soil nitrate, ammonium and amino acid concentrations from the aggregate-scale sampling (*n* = 192 for each N form).



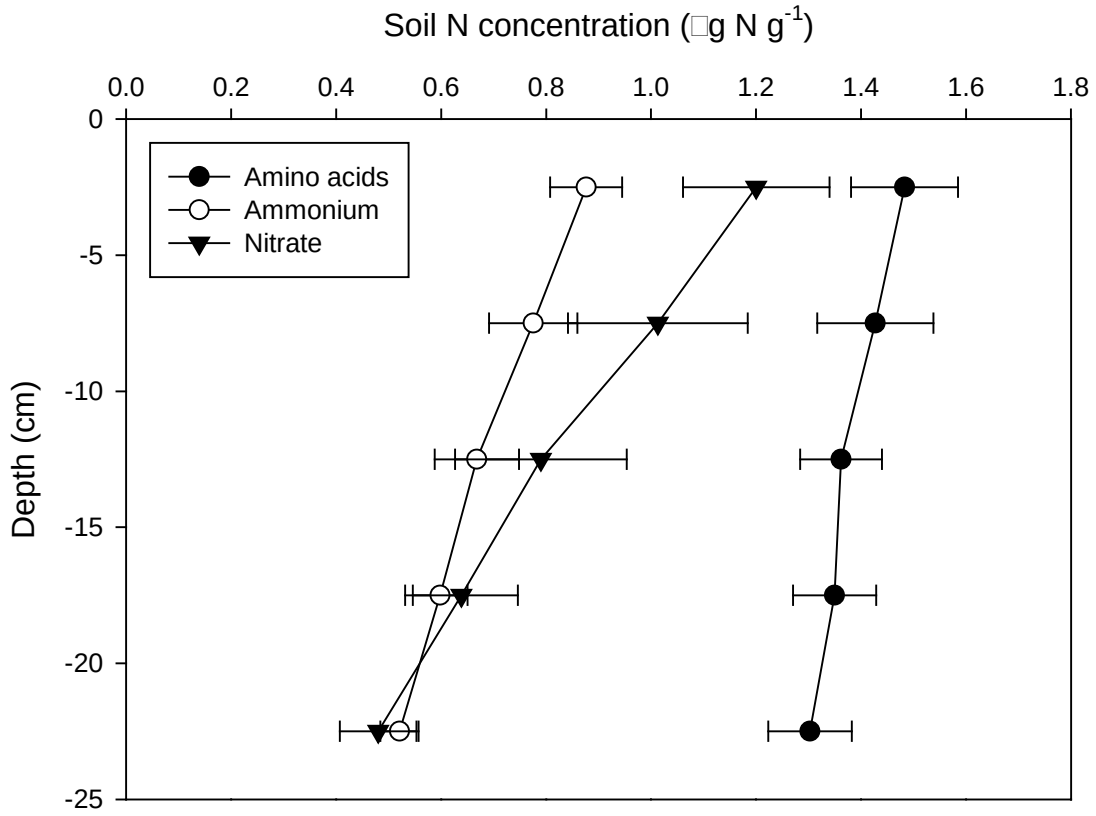
Supplementary information for Shaw et al. 2016, “Characterising the within-field scale spatial variation of nitrogen in a grassland soil to inform the efficient design of in-situ nitrogen sensor networks for precision agriculture”.

Figure S6. Profile likelihood plot for the λ parameter of the Box-Cox transformation for soil nitrate, ammonium and amino acid concentrations from the aggregate-scale sampling.



Supplementary information for Shaw et al. 2016, "Characterising the within-field scale spatial variation of nitrogen in a grassland soil to inform the efficient design of in-situ nitrogen sensor networks for precision agriculture".

Figure S7. Variability of soil nitrate, ammonium and amino acid with soil depth. Data points represent means ± SEM (n = 12) of soil N concentrations (μg N g^{-1}) for each 5 cm depth increment.



Supplementary information for Shaw et al. 2016, "Characterising the within-field scale spatial variation of nitrogen in a grassland soil to inform the efficient design of in-situ nitrogen sensor networks for precision agriculture".

Table S1. Akaike's information criterion (AIC) values for the full model used to describe the spatial variation of soil N forms in a grassland soil derived from the June nested sampling event. The resulting AIC values when each variance component is dropped in turn from the full model are also shown. Where the variance for variance components is zero, no AIC value is reported. Variance components refer to distances in meters, with m and s representing the between-mainstation and between-strata components respectively.

Variable	AIC value for full model	AIC value if variance component is dropped from model				
		σ^2_s	σ^2_m	σ^2_2	$\sigma^2_{0.5}$	$\sigma^2_{0.1}$
Nitrate	–88.35	–86.78	–87.19	—	–85.85	–72.06
Ammonium	–217.59	–218.44	–216.81	–219.12	–219.59	–218.91
Amino Acid	–369.1	–370.01	–368.81	—	—	–368.21

Table S2. Akaike’s information criterion (AIC) values for the full model used to describe the spatial variation of soil N forms in a grassland soil derived from the July nested sampling event. The resulting AIC values when each variance component is dropped in turn from the full model are also shown. Where the variance for variance components is zero, no AIC value is reported. Variance components refer to distances in meters, with m and s representing the between-mainstation and between-strata components respectively.

Variable	AIC value for full model	AIC value if spatial component is dropped from model					
		σ^2_s	σ^2_m	σ^2_2	$\sigma^2_{0.5}$	$\sigma^2_{0.1}$	$\sigma^2_{0.01}$
Nitrate	–412.87	—	–395.20	—	–414.69	–400.52	–376.95
Ammonium	–461.87	–463.08	–460.95	—	—	–462.02	–408.89
Amino Acid	–560.68	–562.60	–547.14	–562.11	—	–560.48	–533.60

Table S3. Akaike’s information criterion (AIC) values for the full model used to describe the aggregate-scale spatial variation of soil N forms in a grassland soil. The resulting AIC values when each variance component is dropped in turn from the full model are also shown. Where the variance for variance components is zero, no AIC value is reported. The variance components s, p and c representing the between-strata, the between-pair and between core components respectively.

Variable	Full model	Term dropped from model		
		σ^2_s	σ^2_p	σ^2_c
Nitrate	10.71	—	—	11.8
Ammonium	7.27	—	5.82	5.28
Amino Acid	–123.0	—	–124.9	–121.0