A METHODOLOGY FOR DETERMINING CRITICAL SOURCE AREAS OF NITROGEN IN GRAZED HILL PASTURES

Keith Betteridge[1], Siva Ganesh[1], Dongwen Luo[1], Kensuke Kawamura[2], Des Costall[1], Rena Yoshitoshi[2]

[1]*AgResearch Grasslands, PB 11008 Palmerston North, New Zealand 4442*
[2]*Graduate School for International Development and Cooperation, Hiroshima University, 1-5-1 Kagamiyama, Higashi-hiroshima 739-8529, Japan*
keith.betteridge@agresearch.co.nz

Summary
Critical source areas (CSAs) are areas of enriched contaminant sources and hydrological activity that occur in small parts of a catchment or farm, but contribute a disproportionately large amount of contaminants to the environment. Defining areas that represent enriched sources of contaminants is central to isolating and then managing contaminant losses from a CSA. These areas could be camp sites that receive large amounts of nitrogen (N) from urine and N, phosphorous (P) and microbes from faeces. This paper draws on a spatial datasets from a hill country, cattle grazed pasture, in which cows wore a GPS collar to track their movement over six days. Cows were also fitted with a urine sensor so that urine patches could be located in the paddock, and a motion sensor to differentiate between walking, grazing, standing and lying activities. GIS map layers were used to show distribution patterns of cow movement activities, urine deposition, and paddock slope, aspect and elevation. With time spent in a given area of a paddock (greatest when lying down) being correlated with urine distribution ($r = 0.64$), lying time was used as a proxy for urine distribution. Of several modelling approaches evaluated, the Generalised Additive Model (GAM) best predicted where cows would lie, using independent variables slope, elevation, aspect, Northings and Eastings in 25 m^2 grid cells overlying the paddock map. Aggregates of "cells" with high densities of urine patches are considered to be potential CSAs of N. This model will be revised as new data become available and by inclusion of farm features such as trees, shelter belts, water troughs and gateways around which animals typically congregate.

Introduction
With increasing pressure coming on to farmers to minimise environmental pollution from their farming operations, mitigation strategies are required. Nitrogen (N), phosphorous (P) and faecal microbes are pollutants of major concern, where N can be leached as nitrate or emitted as ammonia or nitrous oxide, whereas organic N, inorganic P and faecal microbes move in water, predominantly in overland flow (McDowell et al. 2005; McDowell & Wilcock 2007; McDowell & Srinivasan 2009; McDowell 2012). A critical source area (CSA) is an area of land with a large source of nutrient or faecal contaminants that intersects with a transport mechanism – usually hydrological activity like surface runoff (McDowell & Srinivasan 2009). Where only urine is involved, each urine patch, but typically an aggregation of urine patches such as in a gateway or stock camp, is nutrient rich and can be a CSA of N (CSA_N) with losses emitted as ammonia, nitrous oxide or as nitrate in leachate to groundwater. Toolboxes of potential mitigation strategies exist (Monaghan et al. 2007; Monaghan et al. 2008; Monaghan 2009), but unless these small CSA areas are targeted with the mitigation, the cost of mitigation may to be too high for whole-paddock treatment (McDowell & Srinivasan 2009; Betteridge et al. 2011). Targeting requires knowledge of

where CSAs are located within a paddock. Two examples of identified CSA_N in hill country have been reported; first, cattle campsites on flat terrain within the hills and sheep campsites at the top of the hill (Betteridge et al. 2010b); second, flat areas on the farm where crops or autumn saved pasture are break-fed such that very high amounts of excreta are returned to small areas of the farm (Betteridge et al. 2011). To enable farmers to treat CSAs with a chosen mitigation strategy they will need to know where these are located. Also, a regulatory body may need independent verification that such areas have been correctly identified and treated. An example of this approach has been reported for sheep grazing a hill country farm in the Lake Taupo region, for which spatio-temporal distribution data for urine, GPS, pasture mass and quality parameters within 25 m^2 grid cells were used to develop a model of the important animal behaviour drivers for where urine would be excreted (Betteridge et al. 2008). For determining CSA_N, a critical assumption is that there is a strong correlation between the time a cow spends lying in an area of the paddock and the proportion of the total daily urination events deposited there. Previous studies and data from this study show the correlation to be between r = 0.57 for time spent in a 100 m^2 grid cell (Betteridge et al. 2010b) and 0.64 for time spent in a 25 m^2 grid cell and the number of urination events occurring in those respective cells. This assumption is important because farmers will not be able to use urine sensors and GPS units on animals to define where they urinate. In this project we have developed a process to predict where cattle will lie in a random paddock for which only slope, aspect, elevation, Northing and Easting values are known for each grid cell of the 5 m * 5 m overlying the paddock. We have not linked faecal distributions to this model although the distribution patterns of urine and faeces are likely to be similar (White et al. 2001).

Trial data

A steep, ~0.5 ha paddock with average slope of the grid cells being 17.8˚ (range 2.5˚ to 40.9˚), at the AgResearch *Ballantrae Research Station* (near Woodville), was stocked with 20 rising 2 yr beef heifers (ave. LWt 264 (SD 70.5) kg) for 5 days. The paddock had a predominantly E to SE aspect and there were no trees to modify cattle behaviour. Pasture quality was not specifically measured but was typical hill country pasture dominated by browntop (*Agrostis capillaris*) that was moderately well utilised in the previous grazing event. Some rank browntop was present on steeper slopes. Each cow was fitted with a GPS collar and urine sensor (Betteridge et al. 2010b) and an IceTag® on the left hind leg for determining Lying, Standing, Walking and Grazing times (Betteridge et al. 2010a). The GPS logged the position every 10 minutes if no movement had been detected or when the animal moved ≥3 m relative to the last logged position. Within a Microsoft® *Excel* spreadsheet all GPS data were converted to 1 second intervals by interpolation between irregularly timed GPS readings. To this were synchronised the IceTag readings at one minute intervals detailing the type of activity and percentage of time within that minute spent doing those activities (Aharoni et al. 2009). The urine sensor used a thermistor suspended under the tail of the cow. The urine sensor recorded time and ambient temperature at 1 second intervals except when urine was excreted when the temperature rose close to body temperature (39˚C). Urination data were synchronised with the other data by time. All data were then translated to values per minute.

Modelling methodology

Using the *R* computing software package (R 2011), a 5 m * 5 m grid cell overlay was developed within which all motion activity and urination event data of all cows were summed. Analyses were based on these cell totals rather than on data for individual cattle. The cells around the paddock boundary were clipped and data apportioned *pro-rata* according to that proportion of the cell area remaining within the paddock after clipping. Activities are plotted frequency densities for visualising where these activities, especially lying and urination, took place. Dung excreted by all cattle during the grazing period was mapped with a hand-held GPS at the completion of grazing. This contrasts with urine distribution which related to urination events recorded only while the urine sensors were working successfully.

Multiple linear regression analysis (MLR)
The main aim was to identify the best predictive regression model for determining ‘Lying time’ based on other measured variables. In the first instance, the MLR was used to quantify effects of measured variables urine, dung, Aspect, Slope and Elevation on lying time (Lying, mins).

Here contour parameters related to each 25 m^2 grid cell and dung and urine were counts within each grid cell. Standardised data are required in all spatial predictive modelling. The Akaike Information Criterion (AIC) was used to determine those variables that would not be used as they contributed least towards the outcome of the final selected model (Akaike 1973). The grid data were also used to generate a correlation matrix amongst all variables.

Other Models assessed
An underlying assumption of the MLR method is that the relationship under study is spatially constant and that the estimated parameters remain constant over space. In heterogeneous environments, such as grazed pastures, especially hill country pastures, variation of parameter values will often change in unison, i.e. they are auto-correlated. Thus, the basic premise of the parameters being ‘stationary’ is violated and the MLR approach is invalid (Wang et al. 2005). Moreover, failure to account for auto-correlation prevents in-depth interpretation of almost all geographical analyses (Jetz et al. 2005) and can lead to incorrect conclusions.

Furthermore, if the main aim is to ‘predict’ the lying time in a grid of a new paddock, then we need a regression model for the response variable ‘Lying’, based on measured variables Aspect, Slope, Elevation, Easting and Northing. Note that for a random paddock on a random farm, only slope, aspect, elevation, Northing and Easting data are readily available for each 5 m * 5 m grid cell overlying the paddock.

Geographically Weighted Regression (GWR, Fotheringham et.al., 2002), k-Nearest Neighbour Regression (kNNR), Random Forest Regression (RFR) and the Generalised Additive Model (GAM) (see Hastie et.al., 2009) are methods that allow for variation in parameters in time and space, thereby overcoming the limitation of MLR analyses of spatially oriented data. Each modelling approach predicts the response value of a given cell by referencing values of surrounding cells in a certain way. kNNR uses the k number of surrounding cells, whereas GAM determines the smoothed response surface based on all surrounding cells. GWR explores spatial non-stationarity of a regression relationship for spatial data by locally fitting a spatially varying coefficient regression model. RFR, on the other hand, builds a large number of regression trees based on bootstrap samples together with a random subset of predictor variables. Tree models are grown without pruning and the final prediction is an ensemble of predictions from all trees.

These methods don’t produce a *model* (that can be easily written up) but rather take an input *training data set* and predict CSAs for the *new paddock data.* Input data must include the Northing and Easting values of each grid cell and <u>all</u> input data must be *standardised* so that the model built on the training data can be used to predict outcomes from the new data where the site will have local slopes, aspects, elevations and location co-ordinates (northing and easting).

R software was utilised to assess several potential models for predicting the location of lying - our proxy for urine hotspots (CSA_N). Models were assessed using Mean-Squared-Error (MSE) and R-squared values in the usual manner (i.e. 're-substitution' approach) as well as via a leave-one-out cross-validation (CV) approach. Furthermore, graphics including grid-plots showing actual and fitted values were used for assessing the goodness-of-fit of the candidate models.

Results and discussion

The paddock's mean slope was 17.8˚ and rose 28 m from the lowest to the highest point (Fig. 1). Although 20 cows were fitted with urine sensors only data from the 8 sensors that provided good data for the 5 days of grazing were used in these analyses. Lying was the most common activity (45% of total time spent doing all activities in a grid cell) and was concentrated on the flatter areas of the paddock, whereas grazing (33%) was much more widely represented across the paddock. Standing time was half that of grazing and walking was only 3% of total timed activities (Fig. 2).The 20 cows produced 247 urination events (Figs. 1, 3), averaging (SD) 9.0 (2.95) events/cow/day. Urine was distributed predominantly on the easy slope areas where cows were shown to have rested. As with GPS units, several urine sensors failed to work for the full trial period. The 1457 dung deposits are also shown in Fig. 3.

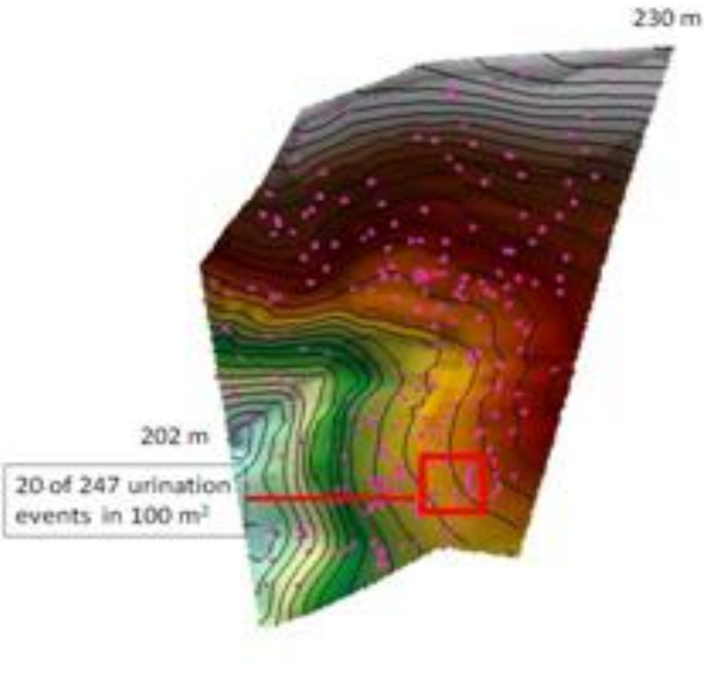

Figure 1. Contour map showing range in elevation (m asl) with 1 m contour lines. Pink spots represent location of urination sites determined by urine sensor and GPS, with most being on the flatter (most widely space contour lines) areas of this paddock.

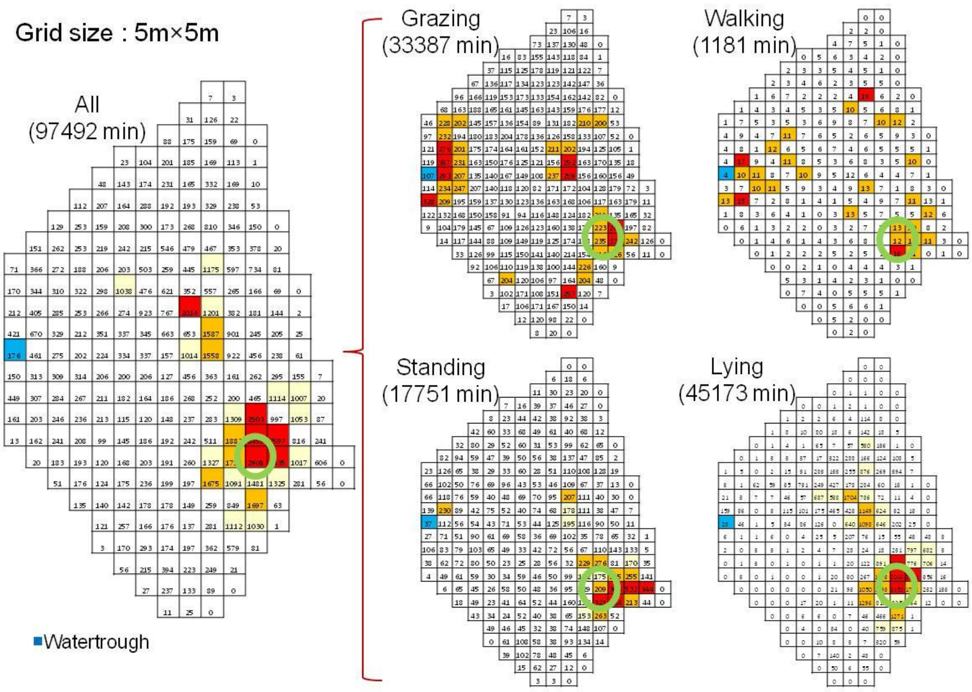

Figure 2. Sum of time spent by all cows within grid cells (left side)[lemon 1000-1600, orange 1600- 2000, red >2000 mins.]. Plots on the right indicate the areas where greatest activities took place in each paddock with lemon increasing through to red indicates increasingly larger values of time spent performing activities. The circle is a reference marker for the reader that circles the same plot in each diagram.

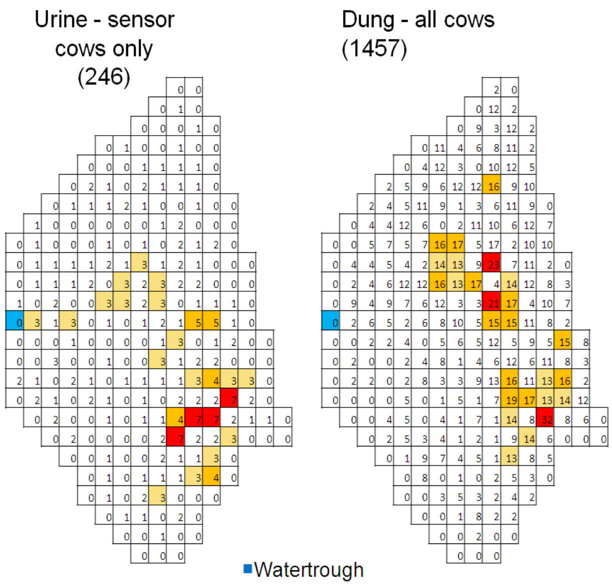

Figure 3. Frequency distribution pattern of urine determined from working urine sensors and all dung deposits of all cows measured manually with a GPS at the conclusion of the trial. Colour codes are as for Fig. 2.

MLR analyses

Multiple linear regression analysis of trial data showed lying time was moderately related to urine events, less strongly related to faecal output and negatively impacted by slope. Whereas dung was included in this model it was not used in subsequent models since dung is a variable that we have rarely mapped.

$$\text{Lying} = 0.53\text{Urine} + 0.28\text{Dung} - 0.18\text{Slope} - 0.001 \; (R^2 = 0.529, P < 0.01)$$

Table 1. Correlation coefficients amongst measured variables, with those of greater importance being highlighted in bold.

	Grazing	Walking	Standing	Lying	Urine	Dung	Aspect	Slope	Elevation
Grazing	1	0.67	0.60	0.41	0.39	0.35	-0.17	0.15	0.04
Walking	0.67	1	0.48	0.35	0.27	0.34	-0.11	-0.07	0.12
Standing	0.60	0.48	1	0.60	0.44	0.36	-0.02	-0.21	-0.07
Lying	0.41	0.35	0.60	1	**0.64**	**0.49**	0.11	**-0.34**	-0.07
urine	0.39	0.27	0.44	0.64	1	0.31	0.09	-0.16	-0.09
dung	0.35	0.34	0.36	0.49	0.31	1	0.01	-0.25	0.16
Aspect	-0.17	-0.11	-0.02	0.11	0.09	0.01	1	-0.31	-0.48
Slope	0.15	-0.07	-0.21	-0.34	-0.16	-0.25	-0.31	1	0.09
Elevation	0.04	0.12	-0.07	-0.07	-0.09	0.16	-0.48	0.09	1

Correlation coefficients amongst the variables (Table 1) show urine deposition is most strongly correlated with lying time. The correlations amongst activity states (lying, standing, walking, grazing) occurs because within any one 25 m^2 grid cell several animals will have visited multiple times, often performing different activities. Because lying and standing are both states of resting, these were combined to reveal a stronger correlation with urination than between urination and lying or standing, alone. Notable, but weak correlations are also seen between urination events with standing and grazing, confirming our observation that urine is also deposited in regions of the paddock other than where they lie down. The negative correlation between lying and slope shows that cows mainly rested on the flatter areas (0 to 12° - data not presented) of this steep paddock.

A comparison of MLR, GWR, GAM, kNNR and RFR models

Based on the *Training* mean square error (MSE = sum$(y\text{-}f)^2$/n, y – actual, f – fitted, n – sample size) values (in Table 2), lowest was with the RFR model followed by GAM, kNNR and GWR, the highest being MLR. However, the CV MSE values indicate that kNNR followed by GAM are the best models. The R^2 values (= 1 – {sum$(y\text{-}f)^2$/sum$(y\text{-}\bar{y})^2$}, $\bar{y}$ – mean(y)) indicate that RFR explains about 90% and 50% of the variation in lying times under ‘re-substitution’ and CV model assessment scenarios respectively, followed closely by the GAM and kNNR models with values 83% & 52% and 82% & 57% respectively.

Considering the fact that the CV based assessment of the models is more reliable than those based on the 're-substitution' approach, we recommend kNNR and GAM models for predicting lying times (in a 5m x 5m grid) using the associated grid values of Aspect, Slope, Elevation, Eastings and Northings.

Table 2. Mean-Squared-Error (MSE) and R-squared estimated values via 're-substitution' (Training) and 'cross validation' (CV). "*" indicates that prediction facility for GWR model is unavailable in R software.

Model	Training Fitted MSE	Training Fitted R-square	CV Fitted MSE	CV Fitted R-square
MLR	132671	0.1639	137939	0.1307
GWR	35390	0.7770	*	*
RFR	16311	0.8972	79031	0.5019
kNNR	27853	0.8245	68774	0.5666
GAM	27523	0.8265	76461	0.5181

The observed vs fitted graphs in Figures 4 and 5 clearly show that GAM followed by kNNR (and GWR with respect to 'training') are the best models for predicting lying times compared to actual lying times ((a) in Figs 4, 5).

This conclusion is based purely on data collected from a single paddock. Ideally, the fitted model needs to be tested on independent paddock(s) with known values of predictor as well as response variables.

The model(s) will be further developed using *new* paddocks for which we have Contour, Easting, Northing and animal activity data. These paddocks will need to vary in size, Aspect, Elevation and Slope ensuring that there are different ratios of hill and flat areas amongst the paddocks. This will ensure a more robust model of where cattle nutrient hotspots will likely be found in a randomly selected paddock for which the farm manager wishes to apply a N loss mitigation strategy.

A truly robust model will also needs to recognise physical features (trees, hedges, water troughs and gateways) within a paddock that are likely to entice animals to excrete disproportionate amounts of urine within close proximity. Such features will need to be added manually into a GIS map layer using local knowledge or an aerial photograph. Furthermore, these nutrient hotspots of organic N, inorganic P and faecal microbes will need to be linked to a hydrology model to determine transport.

Figure 4. Grid plots for frequency of *lying time* showing actual (a) and fitted values based on five models (b-f) for the 'training' case.

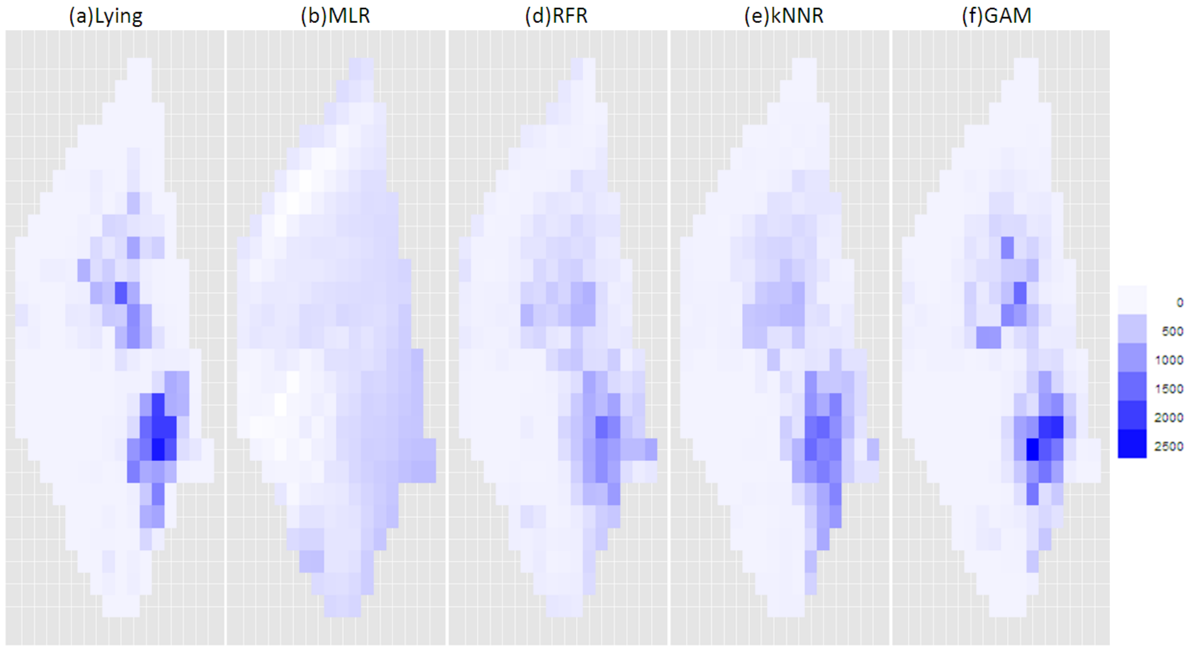

Figure 5. Grid plots for 'Lying time' showing actual (a) and predicted values (b-f) for cross-validation case. R is unable to predict Lying activity using (c)GWR.

Conclusions

Cattle fitted with a GPS and motion sensor to determine where they lie down, as a proxy for where they excrete a large proportion of their daily urine, can be used to develop models for predicting campsite locations. Such sites, once located, could be targeted by farmers to mitigate these disproportionately high N losses, relative the losses from the remaining areas of the paddock. We recommend acquiring Slope, Elevation, Aspect, Easting and Northing data for a paddock, based on a 5 m x 5 m grid. The Generalised Additive Model may be the best for this purpose, but this can be confirmed only once many more datasets are assembled to develop a robust model.

Acknowledgements

We thank Sophie Balladur, French Intern student, Institut National Agronomique de Paris-Grignon/AgroParistech, France, for technical assistance and John Napier and Brad Gordon, AgResearch *Ballantrae* Research Station, for their assistance with animals and pasture management. The trial was Funded by the Foundation for Science Research and Technology (C10X315, Nitrogen and Taupo; Clean Water Productive Land C10X006) and the Ministry for Science and Innovation Japan – New Zealand Science Collaboration Programme (C10X1011, Managing Farm Spatial variability).

References

Aharoni, Y.; Henkin, Z.; Ezra, A.; Dolev, A.; Shabtay, A.; Orlov, A.; Yehuda, Y.; Brosh, A. 2009: Grazing behavior and energy costs of activity: a comparison between two types of cattle. *J. Anim Sci doi:10.2527/jas.2008-1505*:

Akaike, H. 1973: Information theory and an extension of the maximum likelihood principle. Pp. 267-281 *in*:2nd Symposium on Information Theory, Petrov, B.; F.Csake *eds.* Akademiai Kiado, Budapest,

Betteridge, K.; Costall, D.; Balladur, S.; Upsdell, M.; Umemura, K. 2010a: Urine distribution and grazing behaviour of female sheep and cattle grazing a steep New Zealand hill pasture *Animal Production Science 50*: 624–629.

Betteridge, K.; Hoogendoorn, C.; Costall, D.; Carter, M.; Griffiths, W. 2010b: Sensors for detecting and logging spatial distribution of urine from grazing cows and ewes. *Computers and Electronics in Agriculture 73*: 66-73.

Betteridge, K.; Kawamura, K.; Costall, D.; Carter, M.; Hoogendoorn, C.; Griffiths, W.; Inoue, Y. 2008: Tools to determine impact of animal behaviour on nitrogen leaching and nitrous oxide emissions. Pp. 286-298 *in*:Carbon and Nutrient Management in Agriculture, Massey University, Palmerston North, New Zealand, Fertiliser and Lime Research Centre.

Betteridge, K.; Li, F.; Costall, D.; Roberts, A.; Catto, W.; Richardson, A.; Gates, J. 2011: Targeting DCD at critical source areas as a nitrogen mitigation strategy for hill country farmers. Pp. *in*:Adding to the knowledge base for the nutrient manager, Currie, L D and Christensen, C. L. *ed.* Massey University, Palmerston North, Fertilizer and Lime Research Centre.

Fotheringham, A.S., Brunsdon, C., and Charlton, M.E., 2002, Geographically Weighted Regression: The Analysis of Spatially Varying Relationships, Chichester: Wiley.

Hastie, T., Tibshirani, R. and Friedman, J., 2009, The Elements of Statistical Learning: Data Mining, Inference and Prediction", Springer, 2nd Ed. Available at http://www-stat.stanford.edu/~tibs/ElemStatLearn/

Jetz, W.; Rahbek, C.; Lichstein, J. 2005: Local and global approaches to spatial data analysis in ecology. *Global ecology and biogeography* 97-98.

McDowell, R. W. 2012: Challenges and opportunities to decrease phosphorus losses. Pp. *in*:Advanced Nutrient Management: Gains from the Past - Goals for the Future, Currie, L. D. and Christensen, C. L. *ed.* Massey University, Palmerston North, New Zealand, Fertilizer and Lime Research Centre.

McDowell, R. W.; Monaghan, R. M.; Wheeler, D. 2005: Modelling phosphorus losses from pastoral farming systems in New Zealand. *New Zealand Journal of Agricultural Research 48*: 131-141.

McDowell, R. W.; Srinivasan, M. S. 2009: Identifying critical source areas for water quality: 2. Validating the approach for phosphorus and sediment losses in grazed headwater catchments. *Journal of Hydrology 379*: 68-80.

McDowell, R. W.; Wilcock, R. J. 2007: Sources of sediment and phosphorus in stream flow of a highly productive dairy farmed catchment. *Journal of Environmental Quality 36*: 540-548.

Monaghan, R. M. 2009: The BMP Toolbox - selecting the right Best Management Practice for mitigating farming impacts on water quality. Pp. 328-335 *in*:Proceedings of the Fertiliser and Lime Research Centre, Currie, L. D. and Lindsay, C. L. *ed.* Massey University, Palmerston North, New Zealand., Fertiliser and Lime Research Centre.

Monaghan, R. M.; de Klein, C. A. M.; Muirhead, R. W. 2008: Prioritisation of farm scale remediation efforts for reducing losses of nutrients and faecal indicator organisms to waterways: A case study of New Zealand dairy farming. *Journal of Environmental Management 87*: 609-622.

Monaghan, R. M.; Hedley, M. J.; Di, H. J.; McDowell, R. W.; Cameron, K. C.; Ledgard, S. F. 2007: Nutrient management in New Zealand pastures - Recent developments and future issues. *New Zealand Journal of Agricultural Research 50*: 181-201.

R 2011: R computing language. *In*: *ed.* The R Foundation for Statisitcal Computing. R Copyright © 2011 Pp. [http://www.r-project.org/]

Wang, Q.; Li, Y.; Klassen, W. 2005: Influence of summer cover crops on conservation of soil water and nutrients in a subtropical area. *Journal of Soil and Water Conservation 60*: 58-63.

White, S. L.; Sheffield, R. E.; Washburn, S. P.; King, L. D.; Green, J. T. J. 2001: Spatial and Time Distribution of Dairy Cattle Excreta in an Intensive Pasture System *In*: *ed.* Journal of Environmental Quality. White SL, Sheffield RE, Washburn SP, King LD, Green JT Jr. (2001) Spatial and Time Distribution of Dairy Cattle Excreta in an Intensive Pasture System Journal of Environmental Quality 30, 2180-2187. Pp. 2180-2187.